



**HAL**  
open science

## Statistical analysis of growth-fragmentation models

Adelaïde Olivier

► **To cite this version:**

Adelaïde Olivier. Statistical analysis of growth-fragmentation models. Statistics [math.ST]. Université Paris Dauphine - Paris IX, 2015. English. NNT : 2015PA090047 . tel-01235239

**HAL Id: tel-01235239**

**<https://hal.science/tel-01235239>**

Submitted on 29 Nov 2015

**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.

ÉCOLE DOCTORALE DE DAUPHINE

## THÈSE DE DOCTORAT

présentée par

**Adélaïde OLIVIER**

pour l'obtention du grade de

DOCTEUR EN SCIENCES de l'UNIVERSITÉ PARIS-DAUPHINE

---

### Analyse statistique des modèles de croissance-fragmentation

---

Soutenue le 27 Novembre 2015, devant le jury composé de Mmes et MM.

#### Directeurs de thèse

Marie DOUMIC

Ingénieur en Chef des Ponts et Forêts, INRIA

Marc HOFFMANN

Professeur, Université Paris-Dauphine

#### Rapporteurs

Eva LÖCHERBACH

Professeur, Université de Cergy-Pontoise

Patricia REYNAUD-BOURET

Directeur de Recherche, Université Nice Sophia Antipolis

#### Examineurs

Christophe GIRAUD

Professeur, Université Paris-Sud

Stéphane MISCHLER

Professeur, Université Paris-Dauphine

Benoît PERTHAME

Professeur, Université Pierre et Marie Curie

Alexandre TSYBAKOV

Professeur, Université Pierre et Marie Curie

Cette thèse a été financée par le CENTRE DE LA RECHERCHE EN ÉCONOMIE ET EN STATISTIQUE.

Elle a été préparée conjointement au Laboratoire de Statistique du CREST

CREST - ENSAE ParisTech  
3 avenue Pierre Larousse  
92 245 Malakoff Cedex, France

et au CEREMADE de l'Université Paris-Dauphine

CEREMADE, CNRS-UMR 7534  
Université Paris-Dauphine  
Place du maréchal de Lattre de Tassigny  
75 775 Paris Cedex 16, France.

---

## Résumé

---

**E**TUDIER la croissance d'une population, dite structurée, de cellules qui se divisent selon un taux de division inconnu, fonction d'un trait biologique, est le cœur de ce travail. Le trait biologique régissant la division est appelé variable structurante, et peut représenter l'âge de la cellule, la taille, la teneur d'un certain type de protéine, le taux de croissance, *etc.* Le champ mathématique afférent se situe à l'interface de la statistique des processus, de l'estimation fonctionnelle ainsi que de l'analyse.

D'un point de vue stochastique, l'objet d'étude est un processus de croissance-fragmentation : un système de particules où chaque particule présente une caractéristique individuelle sur laquelle notre attention se porte plus particulièrement – l'âge et la taille constituant les deux exemples paradigmatiques considérés dans ce travail. Deux phénomènes régissent l'évolution du système. D'une part, au cours de leur vie, les particules évoluent de manière déterministe connue : les particules vieillissent, croissent. D'autre part, les particules se divisent au bout d'un temps aléatoire : une particule d'âge  $a$  ou de taille  $x$  se divise en deux nouvelles particules (d'âge 0, de taille initiale  $x/2$ ) selon un taux instantané, noté  $B$ , dépendant de l'âge ou de la taille de la particule. Le taux de division  $B$  est un objet central d'intérêt.

Différents cadres inférentiels sont alors envisageables pour traiter le problème de l'**estimation non-paramétrique du taux de division**, le premier objectif de ce travail :

1) Observation continue *jusqu'à une date  $T$  donnée*. Des difficultés sont intrinsèquement liées au cadre du temps continu, dont un phénomène de biais de sélection. Dans le cadre du modèle structuré en âge, un estimateur du taux de division dépendant de l'âge,  $a \rightsquigarrow B(a)$ , est construit dans le Chapitre I convergeant à vitesse  $\exp(\lambda_B \frac{s}{2s+1} T)$ , avérée optimale, où  $s$  est la régularité supposée connue de  $B$  et  $\lambda_B$  le paramètre de Malthus.

2) Observation *jusqu'à une génération donnée* dans l'arbre généalogique de la population. Ce schéma d'observation permet de s'affranchir du biais de sélection. Dans le modèle en taille, à l'aide de techniques d'estimation par seuillage dans une base d'ondelettes, un estimateur du taux de division dépendant de la taille,  $x \rightsquigarrow B(x)$ , est alors construit : il est adaptatif et converge aux vitesses minimax non-paramétriques classiques. La construction de cet estimateur découle de l'étude générale que nous conduisons dans le Chapitre II sur les chaînes de Markov bifurquantes, autrement dit adaptées à la structure d'arbre binaire.

Le second objectif est l'**étude de la transmission d'un trait biologique** d'une cellule à une autre. Le Chapitre **II** y contribue dans un cadre que nous souhaitons le plus général possible, celui des chaînes de Markov bifurquantes, et le Chapitre **III** y répond au travers d'une structure auto-régressive, qualifiée de *bifurquante* pour souligner la structure d'arbre de notre problème. Cette modélisation se prête bien au trait biologique qu'est le taux de croissance : deux cellules sœurs possèdent un taux de croissance dépendant du taux de croissance de leur parent. Les deux fonctions d'auto-régression, qui autorisent l'asymétrie dans la transmission, sont à présent les objets d'intérêt. Leur estimation est conduite sans supposer de forme *a priori*, ni linéaire ni autre. L'étude des estimateurs que nous conduisons s'étend depuis le contrôle du risque quadratique jusqu'à un principe de déviations modérées. Une des questions clés concerne la symétrie de la transmission du trait, ce qui revient, avec la modélisation choisie, à se demander si les deux fonctions d'auto-régression sont identiques, et nous construisons un test à cet effet dans notre cadre non-paramétrique.

Le troisième objectif est l'**étude comparative de populations à travers leur paramètre de Malthus** régissant la croissance exponentielle du nombre de leurs individus. La question que nous étudions, directement issue d'un problème biologique, est la suivante : comment la variabilité entre cellules influence-t-elle la croissance globale de la population ? Le Chapitre **IV** envisage un enrichissement des modèles structurés en âge ou taille précédents, chaque cellule présentant désormais un taux de vieillissement ou un taux de croissance qui lui est propre. Des techniques d'analyse des équations aux dérivées partielles et des techniques stochastiques sont conjointement mises en œuvre et nous montrons que, selon les propriétés du taux de division, la variabilité individuelle dans le taux de vieillissement ou de croissance peut aussi bien *avantager* que pénaliser la croissance globale d'une population cellulaire.

**Mots-clés : Statistique des processus aléatoires.** Processus de Markov déterministe par morceaux, processus de branchement, chaîne de Markov bifurquante, processus auto-régressif bifurquant, arbre de Galton-Watson. **Statistique non-paramétrique.** Vitesse de convergence minimax, adaptativité, estimateur à noyau, estimateur à ondelettes avec seuillage. **Probabilités.** Many-to-one formulae, inégalité de déviations, principe de déviations modérées. **Équations aux dérivées partielles.** Équation de croissance-fragmentation, équations structurées, équation structurée en âge, équation structurée en taille, problème aux valeurs propres, paramètre de Malthus.

---

## Abstract

---

**T**HIS dissertation is concerned with growth-fragmentation models, implemented for investigating the growth of a population of cells. Cell division occurs at an unknown rate, depending on a biological feature called structuring variable, which can be age, size, growth rate, some proteic content, *etc.* The mathematical framework includes statistics of processes, nonparametric estimations and analysis.

From a stochastic point of view, we are dealing with the evolution of a system of particles. We are especially interested in an individual feature of the particles – age and size being the two paradigmatic examples. The evolution of the system is then driven by two phenomena. First, particles evolve on a deterministic basis: they age, they grow. Secondly, particles split randomly: a particle of age  $a$  or size  $x$  splits into two particles (of age 0, of size at birth  $x/2$ ), at a rate  $B$  depending on the age or on the size of the splitting particle. The division rate  $B$  is an unknown function, on which we focus our attention.

The first aim of this work is to **get a nonparametric estimate of the division rate**. Various observation schemes may be considered:

1) Continuous time observation *up to time*  $T$ . This scheme, which differs radically from those previously used, entails specific difficulties – mainly a bias selection. In Chapter I we build, in the age-structured model framework, an estimator of the *age-dependent* division rate,  $a \rightsquigarrow B(a)$ ; moreover, the rate of convergence  $\exp(\lambda_B \frac{s}{2s+1} T)$  is reached and proved to be optimal –  $s$  denoting the smoothness of  $B$  and  $\lambda_B$  the Malthus parameter.

2) Observation *up to a given generation* in the genealogical tree of the population. In contrast with the previous scheme, this one does not involve any bias selection. In the size-structured model framework, a wavelet thresholding estimator of the *size-dependent* division rate,  $x \rightsquigarrow B(x)$ , is defined, which is adaptative and converges at classical minimax rates. A fundamental tool for building such an estimator is the general study we carry out in Chapter II concerning bifurcating Markov chains, a generalization of Markov chains adapted to a binary tree structure.

The second objective is to **study the transmission of a biological feature** from one cell to another. In Chapter II, we aim at giving answers in a framework we wish as general as possible – this chapter is devoted to bifurcating Markov chains. Chapter III sheds some light on this problem by studying a so-called bifurcating autoregressive process – that is to say, adapted to the tree struc-

ture. This model turns out to be particularly convenient when the considered feature of cells is the growth rate: two daughter cells have a growth rate depending on the growth rate of their parents. The two autoregressive functions, which allow asymmetry in the transmission of the individual feature, are now the parameters of interest. We estimate them without any a priori assumption – as linearity, or other. Our study of the estimators ranges from the control of the quadratic loss to a moderate deviations principle. Whether symmetry occurs or not in the transmission of the growth rate is a key issue – a question which is equivalent, in the chosen bifurcating autoregressive model, to know whether the two autoregressive functions are equal or not. We build a test to answer that question.

The third objective is to **compare different populations of cells through their Malthus parameter**, which governs the global growth. The question we are interested in directly comes from biology: how does the variability among cells affect the global growth of the population? In Chapter IV we enrich the previous age-structured and size-structured models – each cell having now its own aging or growth rate. Techniques to analyse partial differential equations and stochastic techniques are used altogether. We show that individual variability in the aging or growth rate can either slow down or, surprisingly, *speed up* the global growth, depending on the properties of the division rate.

**Keywords: Statistics of stochastic processes.** Piecewise deterministic Markov process, branching process, bifurcating Markov chains, bifurcating autoregressive processes, Galton-Watson trees. **Non-parametric statistics.** Minimax rate of convergence, kernel estimator, wavelet thresholding estimator. **Probability.** Many-to-one formulae, deviations inequality, moderate deviations principle. **Partial differential equation.** Growth-fragmentation equations, structured equations, age-structured equation, size-structured equation, eigenproblem, Malthus parameter.

---

## Remerciements

---

A tous ceux qui ont encouragé ce travail – professeurs, parents, amis – merci.

Je souhaite ouvrir ces remerciements en adressant à mes directeurs, Marie et Marc, mon profond témoignage de reconnaissance. Marc, je vous suis redevable avant toute chose de ma vocation statistique : merci de l'avoir fait naître et encouragée lors de mes études à l'ENSAE, merci de l'avoir éprouvée et affermie lors des trois années écoulées. Marie, merci de m'avoir ouvert la porte du monde des EDP, merci de m'avoir fait partager cet inextinguible enthousiasme qui fait mon admiration. À vous deux, merci pour votre constante bienveillance ; merci de m'avoir fait grandir mathématiquement, scientifiquement.

Je remercie Lydia ROBERT pour nos échanges sur la bactérie *E. coli* et les nombreux problèmes mathématiques qui en sont ressortis. Ma collaboration avec Valère BITSEKI fût très enrichissante. Merci de sa très grande patience lors de nos séances de travail, et d'avoir su tourner nos différends mathématiques en dérision. Puisse notre fructueuse collaboration se poursuivre.

Je sais gré à Eva LÖCHERBACH et Patricia REYNAUD-BOURET d'avoir accepté avec enthousiasme d'assurer la charge de rapporter ma thèse. Je suis touchée de l'intérêt que vous avez porté à mes travaux. À l'une comme à l'autre, merci pour vos lectures attentives.

Je remercie vivement Christophe GIRAUD, Stéphane MISCHLER, Benoît PERTHAME et Sacha TSYBAKOV de prendre part au jury de ma thèse.

Je remercie mes professeurs de statistique à l'ENSAE : par ordre chronologique, Emmanuelle GAUTHERAT, en introduction à la statistique ; Nicolas CHOPIN, en statistique paramétrique ; Marc HOFFMANN, en tests statistiques ; Sacha TSYBAKOV, en statistique non-paramétrique ; Patrice BERTAIL, en statistique semi-paramétrique ; Arnak DALALYAN, en apprentissage ; Judith ROUSSEAU, en statistique bayésienne ; Christian ROBERT, en statistique computationnelle. Je remercie également l'ensemble de mes professeurs du M2 Modélisation aléatoire de Paris 7.

Le laboratoire de statistique du CREST fût un environnement scientifique propice à cette thèse. J'ai pu y suivre des cours de recherche ainsi que des exposés hebdomadaires variés et enrichissants. Je remercie tous ses membres pour leur bienveillance à mon égard. Merci à mes collègues de



bureau d'avoir accepté mes incursions jusqu'au laboratoire du CEREMADE. Merci également aux doctorants du CEREMADE pour leur accueil ; je suis très heureuse d'avoir participé à l'organisation du séminaire des jeunes chercheurs.

Ce fût un plaisir d'enseigner les statistiques à l'ENSAE pendant trois années. Merci à ceux qui ont veillé pendant cette période à la bonne organisation des TD. Je remercie également toutes les personnes qui ont accompagné mes différentes démarches, au CREST comme à Dauphine : départs en conférence, résolution de soucis informatiques et, plus récemment, dépôt de ce manuscrit.

Je tiens à nommer ici : Céline DUVAL et Adrian IUGA, qui m'ont précédée, pour leurs conseils ; Olivier COLLIER, pour notre franche entente ; Pierre GRUET pour le partage de nos joies ou peines mathématiques – les déjeuners qui ponctuèrent cette thèse occuperont leur juste place dans mon souvenir.

Je souhaite également nommer les doctorants rencontrés en conférence : en particulier, par ordre chronologique, Samvel GASPARYAN, au Mans ; Carole BINARD, à Saint-Flour ; Tengyuan LIANG, à Oberwolfach.

Un chaleureux merci à Marie-Carmen GIRALT pour trois années passionnantes de théâtre en espagnol : ce temps d'évasion me fût grandement bénéfique lors de mes études.

J'adresse une pensée à Mme DE WARREN auprès de qui j'ai passé ces dernières années grâce au concours d'Ensemble 2 Générations à qui je dois beaucoup.

Une pensée également à mes amies thouarsaises, aux brillants parcours loin des sphères mathématiques. Puisse notre correspondance perdurer, elle est précieuse.

Olivier, qui avez suivi les développements de cette thèse depuis ses débuts jusqu'à sa finalisation, merci de m'avoir entourée de constantes attentions comme vous l'avez fait, merci d'avoir illuminé ces trois années.

Mes plus affectueuses pensées vont à mes proches. Pour les soins qu'ils me prodiguent depuis toujours, à mes parents, je rends grâce.

*A.O.*

---

# Sommaire

---

|   |           |
|---|-----------|
| Résumé . . . . .  | i         |
| Abstract . . . . .  | iii       |
| Remerciements . . . . .   | v         |
| <b>Chapitre Introductif</b>   | <b>1</b>  |
| 1 OBJET D'ÉTUDE . . . . .   | 1         |
| 2 CADRE MATHÉMATIQUE ET STATISTIQUE . . . . .   | 3         |
| 2.1 Description des modèles de croissance-fragmentation . . . . .                                     | 3         |
| 2.2 Comportements en temps long . . . . .   | 6         |
| 2.3 Estimation du taux de division : quel cadre d'inférence ? . . . . .                               | 9         |
| 2.4 Estimation du taux de division : état de l'art . . . . .  | 11        |
| 3 PREMIÈRE PARTIE : ESTIMATION DU TAUX DE DIVISION EN ÂGE . . . . .                                   | 14        |
| 3.1 Cadre et difficultés inhérentes au temps continu . . . . .  | 14        |
| 3.2 Synthèse des résultats nouveaux . . . . .   | 15        |
| 4 DEUXIÈME PARTIE : ÉTUDE DE LA TRANSMISSION D'UN TRAIT . . . . .                                     | 19        |
| 4.1 Cadre général et exemples paradigmatiques . . . . .   | 19        |
| 4.2 Synthèse des résultats nouveaux . . . . .   | 22        |
| 5 TROISIÈME PARTIE : ÉTUDE COMPARATIVE DE POPULATIONS . . . . .                                       | 26        |
| 5.1 Question biologique . . . . .   | 27        |
| 5.2 Résultats théoriques sur le modèle jouet en âge avec variabilité . . . . .                        | 28        |
| 5.3 Résultats numériques sur le modèle en taille et incrément de taille avec<br>variabilité . . . . . | 30        |
| 6 PERSPECTIVES . . . . .  | 33        |
| 6.1 Estimation du paramètre de Malthus à partir d'une branche . . . . .                               | 33        |
| 6.2 Problème inverse dans le modèle taille-incrément . . . . .  | 34        |
| 7 COMPOSITION DE LA THÈSE . . . . .   | 35        |
| RÉFÉRENCES . . . . .  | 36        |
| <b>I Processus d'âge branchant : estimation non-paramétrique du taux de division</b>                  | <b>40</b> |
| 1 INTRODUCTION . . . . .  | 43        |
| 1.1 Motivation . . . . .  | 43        |
| 1.2 Objectives and results . . . . .  | 44        |

|     |   |    |
|-----|---|----|
| 2   | RATE OF CONVERGENCE FOR BIASED EMPIRICAL MEASURES . . . . .         | 46 |
| 2.1 | Continuous time rooted trees . . . . .                              | 46 |
| 2.2 | The limiting objects . . . . .                                      | 48 |
| 2.3 | Convergence results for biased empirical measures . . . . .         | 49 |
| 3   | STATISTICAL ESTIMATION . . . . .                                    | 51 |
| 3.1 | Construction of an estimation procedure . . . . .                   | 51 |
| 3.2 | Convergence results for $\widehat{B}_T(x)$ . . . . .                | 52 |
| 3.3 | Discussion of the results . . . . .                                 | 53 |
| 4   | NUMERICAL IMPLEMENTATION . . . . .                                  | 56 |
| 5   | PROOFS . . . . .  | 58 |
| 5.1 | Preliminaries . . . . .   | 58 |
| 5.2 | Proof of Theorems 3 and 4 . . . . .                                 | 60 |
| 5.3 | Proof of Proposition 5 . . . . .                                    | 64 |
| 5.4 | Proof of Theorem 7 . . . . .  | 64 |
| 5.5 | Proof of Theorem 8 . . . . .  | 73 |
| 5.6 | Proof of Proposition 9 . . . . .                                    | 75 |
| 6   | APPENDIX . . . . .  | 76 |
| 6.1 | Heuristics for the convergences to the limits (8) and (9) . . . . . | 76 |
| 6.2 | Proof of Proposition 10 . . . . .                                   | 78 |
| 6.3 | Proof of (26) and (27) of Proposition 11 . . . . .                  | 80 |
| 6.4 | Proof of Lemma 16 . . . . .   | 81 |
|     | REFERENCES . . . . .  | 82 |

## II Chaîne de Markov bifurquante : estimation adaptative des transitions et mesure invariante 84

|     |   |     |
|-----|---|-----|
| 1   | INTRODUCTION . . . . .  | 87  |
| 1.1 | Bifurcating Markov chains . . . . .   | 87  |
| 1.2 | Objectives . . . . .  | 88  |
| 1.3 | Main results and organisation of the chapter . . . . .                                    | 89  |
| 2   | DEVIATIONS INEQUALITIES FOR EMPIRICAL MEANS . . . . .                                     | 91  |
| 3   | STATISTICAL ESTIMATION . . . . .  | 93  |
| 3.1 | Atomic decompositions and wavelets . . . . .  | 94  |
| 3.2 | Estimation of the invariant density $\nu$ . . . . .                                       | 95  |
| 3.3 | Estimation of the density of the mean transition $\mathcal{Q}$ . . . . .                  | 96  |
| 3.4 | Estimation of the density of the $\mathbb{T}$ -transition $\mathcal{P}$ . . . . .         | 97  |
| 4   | APPLICATIONS . . . . .  | 98  |
| 4.1 | Estimation of the size-dependent splitting rate in a growth-fragmentation model . . . . . | 98  |
| 4.2 | Bifurcating autoregressive process . . . . .  | 100 |
| 4.3 | Numerical illustration . . . . .  | 102 |
| 5   | PROOFS . . . . .  | 105 |
| 5.1 | Proof of Theorem 4 (i) . . . . .  | 105 |
| 5.2 | Proof of Theorem 4 (ii) . . . . .   | 107 |
| 5.3 | Proof of Theorem 5 (i) . . . . .  | 109 |

|      |  |     |
|------|--|-----|
| 5.4  | Proof of Theorem 5 (ii)                | 111 |
| 5.5  | Proof of Theorem 8                     | 111 |
| 5.6  | Preparation for the proof of Theorem 9 | 112 |
| 5.7  | Proof of Theorem 9, upper bound        | 113 |
| 5.8  | Proof of Theorem 9, lower bound        | 114 |
| 5.9  | Proof of Theorem 10                    | 116 |
| 5.10 | Proof of Theorem 12                    | 117 |
| 6    | APPENDIX                               | 118 |
| 6.1  | Proof of Lemma 11                      | 118 |
| 6.2  | Proof of Lemma 14                      | 119 |
| 6.3  | Proof of Lemma 15                      | 120 |
|      | REFERENCES                             | 121 |

### III Processus bifurquant auto-régressif non linéaire : estimation des fonctions auto-régressives 123

|     |   |     |
|-----|---|-----|
| 1   | INTRODUCTION  | 126 |
| 1.1 | A generalisation of the bifurcating autoregressive model  | 126 |
| 1.2 | Estimation of the autoregressive functions and objectives | 127 |
| 2   | NON-ASYMPTOTIC BEHAVIOUR                                  | 128 |
| 2.1 | Tagged-branch chain                                       | 128 |
| 2.2 | Model constraints   | 130 |
| 2.3 | Main results  | 131 |
| 3   | ASYMPTOTIC BEHAVIOUR                                      | 133 |
| 3.1 | Main results  | 133 |
| 3.2 | Construction of an asymmetry test                         | 135 |
| 4   | DISCUSSION  | 137 |
| 5   | NUMERICAL IMPLEMENTATION                                  | 138 |
| 6   | PROOFS  | 142 |
| 6.1 | Preliminary   | 143 |
| 6.2 | Estimation of the density of the invariant measure        | 144 |
| 6.3 | Proof of Theorem 5  | 147 |
| 6.4 | Proof of Theorem 6  | 149 |
| 6.5 | Proof of Proposition 7                                    | 151 |
| 6.6 | Proofs of Theorems 8 and 9                                | 151 |
| 6.7 | Proof of Theorem 13                                       | 155 |
| 6.8 | Proof of Theorem 15                                       | 159 |
| 7   | APPENDIX  | 161 |
| 7.1 | Proof of Lemma 16   | 161 |
| 7.2 | Proof of Lemma 19   | 162 |
| 7.3 | Proof of Lemma 21   | 162 |
| 7.4 | Proof of Lemma 22   | 163 |
|     | REFERENCES  | 164 |

|           |   |            |
|-----------|---|------------|
| <b>IV</b> | <b>Influence de la variabilité individuelle sur le paramètre de Malthus</b> | <b>167</b> |
| 1         | INTRODUCTION . . . . .  | 170        |
| 1.1       | Introducing individual variability . . . . .                                | 170        |
| 1.2       | Estimating the Malthus parameter . . . . .                                  | 174        |
| 1.3       | Main results and outline . . . . .  | 176        |
| 2         | THE AGE-STRUCTURED MODEL WITH VARIABILITY . . . . .                         | 178        |
| 2.1       | Description of the model . . . . .  | 178        |
| 2.2       | The Malthus parameter . . . . .   | 181        |
| 2.3       | Influence of variability on the Malthus parameter . . . . .                 | 182        |
| 2.4       | Estimation of the Malthus parameter from one branch . . . . .               | 187        |
| 3         | THE SIZE-INCREMENT-STRUCTURED MODEL WITH VARIABILITY . . . . .              | 188        |
| 3.1       | Description of the model . . . . .  | 188        |
| 3.2       | The Malthus parameter . . . . .   | 189        |
| 3.3       | Influence of variability on the Malthus parameter . . . . .                 | 190        |
| 3.4       | Estimation of the Malthus parameter from one branch . . . . .               | 200        |
| 4         | TWO OTHER MODELS OF INTEREST . . . . .                                      | 203        |
| 4.1       | Size-structured model with variability . . . . .                            | 203        |
| 4.2       | Model introduced by Amir . . . . .  | 205        |
| 5         | PERSPECTIVES . . . . .  | 208        |
| 6         | PROOFS . . . . .  | 209        |
| 6.1       | Proof of Theorem 3 . . . . .  | 209        |
| 6.2       | Proof of Theorem 4 . . . . .  | 213        |
| 6.3       | Proof of Theorem 7 . . . . .  | 219        |
| 7         | APPENDIX . . . . .  | 222        |
| 7.1       | Proof of Lemma 9 . . . . .  | 222        |
| 7.2       | Proof of Lemma 11 . . . . .   | 223        |
| 7.3       | Complements on the model introduced by Amir . . . . .                       | 224        |
| 8         | SUPPLEMENTARY FIGURES AND TABLES . . . . .                                  | 225        |
|           | REFERENCES . . . . .  | 233        |

# CHAPITRE INTRODUCTIF

---

## Les modèles de croissance-fragmentation

---

### 1 OBJET D'ÉTUDE

Mes travaux théoriques sont pensés en lien étroit avec un champ d'application : l'étude de la croissance d'une population de cellules. À un niveau microscopique, deux phénomènes régissent cette croissance. Le premier est la croissance individuelle des cellules par ingestion d'un nutriment présent dans leur milieu, selon une dynamique déterministe connue, supposée commune à toutes les cellules au moins dans un premier temps. Le second phénomène est la division de chaque cellule, au bout d'un certain temps, en deux nouvelles cellules, disons de tailles égales. La multiplication des cellules s'opère ainsi et se trouve illustrée en Figure 1, représentant l'image d'une population de bactéries à deux instants successifs. Il s'agit alors de comprendre, de modéliser, le temps inconnu au bout duquel se divise chaque cellule. À cet effet une caractéristique individuelle, la taille par exemple, est distinguée comme variable dite *structurante*. Le postulat suivant est alors en vigueur : une cellule se divise *selon* sa taille. Autrement dit, le temps de division d'une cellule est vu comme un temps aléatoire, dont la loi dépend de la variable structurante qui sous-tend la dynamique. Les traits individuels privilégiés dans ce travail sont la taille et l'âge, mais d'autres caractéristiques peuvent venir additionnellement structurer le modèle : incrément de taille, taux de vieillissement ou taux de croissance par la suite, ou encore teneur en un certain type de protéine, *etc*. La division est appréhendée comme un phénomène aléatoire, par opposition à la croissance qui s'effectue selon une trajectoire fixée.

Le mécanisme de division, lui seul à l'origine de la multiplication des cellules, se trouve résumé par un taux de division, fonction de la variable structurante considérée, déterminant la loi des temps de vie des cellules. Ce taux de division caractérise entièrement le modèle, si l'on suppose la trajectoire de croissance individuelle connue, et il se trouve à ce titre être l'objet le plus naturel d'étude. Néanmoins la connaissance d'un certain nombre d'autres quantités, fonctions du taux de division lui-même certes, peut être recherchée de façon directe. La vitesse de croissance de la population est par exemple une quantité clé, et est *a priori* plus facile d'accès que le taux de division lui-même.

Dans ce contexte, les axes d'étude que nous privilégions sont les suivants :

- (A) Comprendre le mécanisme de division à travers la reconstruction du taux de division en tant que fonction de la variable structurante. [Chapitres I et II]
- (B) Étudier la transmission du trait individuel d'une cellule à ses descendants, étudier le trait d'une cellule typique. [Chapitres II et III]
- (C) Comparer différentes populations à travers leur vitesse de croissance. [Chapitre IV]

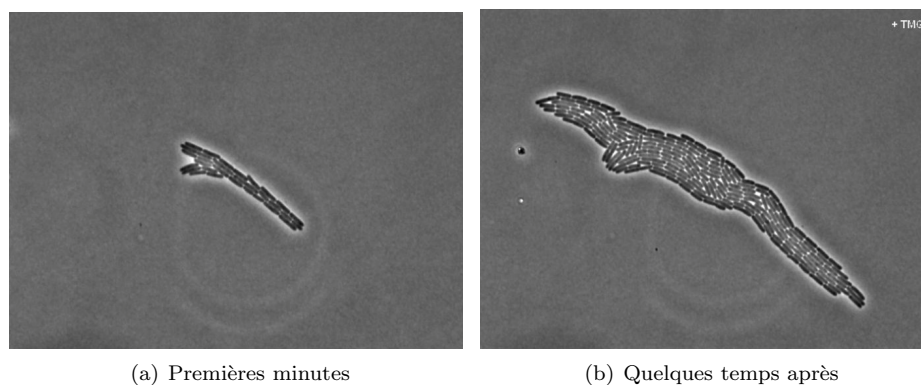


FIGURE 1 – Croissance d'une population de bactéries *E. coli*. La bactérie *E. coli* croit de façon exponentielle et se divise en son milieu en deux cellules de tailles égales. Le temps de doublement de la population, en milieu riche, est de 20 minutes, tandis qu'en milieu pauvre, il est d'une heure environ. Source : images fournies par Lydia Robert (INRA).

L'étude des axes (A) et (B) s'appuiera sur une modélisation microscopique stochastique, tandis que nous mobiliserons pour l'axe (C) à la fois outils déterministes et outils stochastiques. L'axe (A) est développé dans les Chapitres I et II dans des cadres inférentiels distincts, en temps continu ou discret, et pour des modèles distincts, avec l'âge ou la taille comme variable structurante. Le Chapitre II comprend également une étude plus générale se rapportant à l'axe (B). Le Chapitre III est un prolongement du chapitre précédent et considère l'axe (B) au travers d'une modélisation plus spécifique. Enfin, le Chapitre IV est dédié à l'axe (C).

Travaux déterministes et stochastiques consacrés aux modèles de croissance-fragmentation coexistent de longue date – citons les ouvrages de B. Perthame [Per07] et J. Bertoin [Ber06] qui adoptent respectivement ces deux angles d'étude. Ces modèles pouvant servir à décrire la division cellulaire [BA67], nous proposons de garder cette image des populations de cellules en tête, ainsi que le vocabulaire associé, tout au long de cet écrit. Nous soulignons cependant que le champ des applications est bien plus large et englobe le protocole TCP/IP utilisé pour internet, la polymérisation de protéines, la décharge de neurones en réseaux. Pour l'approche déterministe, citons [BMR, CLO+09, PPS10], tandis que [ABG+14] recense nombre des travaux stochastiques relatifs à ces sujets.

## 2 CADRE MATHÉMATIQUE ET STATISTIQUE

### 2.1 DESCRIPTION DES MODÈLES DE CROISSANCE-FRAGMENTATION

Les deux exemples paradigmatiques, étudiés dans ce travail, sont le modèle structuré en âge et le modèle structuré en taille. Aussi présentons-nous ici conjointement ces deux modèles. Concernant l'approche stochastique, les deux descriptions que nous donnons, en temps continu et généalogique, structurent fondamentalement la façon d'aborder les questions liées aux modèles de croissance-fragmentation que nous nous posons, et conduisent à l'utilisation de techniques probabilistes très différentes. L'approche déterministe vient également élargir le spectre des outils d'investigation.

#### *Approche stochastique : description en temps continu*

Définissons le processus recensant à l'instant  $t$  la valeur du trait biologique, l'âge ou la taille, de toutes les cellules vivantes (par ordre décroissant par exemple),

$$X(t) = (X_1(t), X_2(t), \dots), \quad t \geq 0,$$

à valeurs dans  $\cup_{k=0}^{\infty} [0, \infty)^k$ . Ce processus entre dans la classe des processus de Markov déterministes par morceaux, PDMP selon l'acronyme anglais, mais également dans la classe des processus de branchement. Lorsque la variable d'intérêt est l'âge, le processus  $X$  est plus connu sous le nom de processus de Bellman-Harris, [Har63].

Concernant les processus de Markov déterministes par morceaux spécifiquement (sans structure branchante), nous indiquons la revue détaillée des travaux disponibles réalisée par R. Azaïs *et al.* dans [ABG+14]. D'un point de vue statistique, R. Azaïs s'intéresse à l'estimation du noyau de transition dans [Aza14a] et à celle du taux de saut dans [Aza14b]. L'estimation du taux de saut a été également abordée par N. Krell dans [Kre14].

E. Löcherbach dans [Löc02a, Löc02b] étudie des processus markovien branchants généraux, sous les angles probabiliste et statistique. Pour les travaux statistiques dans un cadre non-paramétrique, qui sera également le nôtre, citons ceux de R. Höpfner, M. Hoffmann et E. Löcherbach [HHL02], traitant de l'estimation du taux de branchement d'un processus de diffusion branchant. Citons aussi ceux de M. Hoffmann et N. Krell [HK11] étudiant une chaîne de fragmentation.

#### *Approche stochastique : description généalogique*

Le mécanisme de division des cellules en deux nouvelles cellules amène à considérer une structure d'arbre. Plus précisément, nous introduisons l'arbre binaire

$$\mathbb{T} = \bigcup_{n=0}^{\infty} \{0, 1\}^n,$$

vu désormais comme l'arbre généalogique d'une population (infinie) de cellules. Chaque nœud de l'arbre  $u \in \mathbb{T}$  est identifié à une cellule et se voit attribuer la marque  $(\zeta_u, b_u)$  dans le modèle structuré en âge, la marque  $(\zeta_u, b_u, \xi_u)$  dans le modèle structuré en taille, où  $\zeta_u$  désigne la durée de vie de la cellule  $u$ ,  $b_u$  sa date de naissance et  $\xi_u$  sa taille à la naissance.



**Modèle structuré en âge.** Précisons que nous parlons ici de l'âge selon le sens courant, comme le temps écoulé depuis la naissance. Le premier phénomène à l'œuvre est le vieillissement des cellules, l'âge de la cellule  $u$  à l'instant  $t \in [b_u, b_u + \zeta_u)$  s'écrivant

$$\zeta_u^t = t - b_u.$$

Le second phénomène est la division des cellules en deux nouvelles cellules d'âge nul. La division d'une cellule se produit à un taux dépendant de son âge. En prenant une fonction  $B : [0, \infty) \rightarrow [0, \infty)$  telle que  $\int_0^\infty B(a)da = \infty$  pour taux de division, une cellule d'âge  $a$  se divise avec probabilité  $B(a)dt$  entre les instants  $t$  et  $t + dt$ , ce qui se traduit par

$$\mathbb{P}(\zeta_u \in [t, t + dt) | \zeta_u \geq t) = B(t)dt, \quad (1)$$

où  $t$  représente précisément l'âge d'une cellule au bout d'un temps de vie  $t$ . Les durées de vie  $(\zeta_u, u \in \mathbb{T})$  forment une collection de variables indépendantes et identiquement distribuées de densité  $f_B(a) = B(a) \exp(-\int_0^a B(s)ds)$ ,  $a \geq 0$ .

**Modèle structuré en taille.** Nous suivons la description donnée par M. Doumic *et al.* dans [DHKR15]. Le premier phénomène à l'œuvre est la croissance des cellules. Considérons une croissance exponentielle, à un taux de croissance commun à toutes les cellules, au moins pour une première modélisation, et supposé connu. Nous supposons donc que l'évolution de la taille de la cellule  $u$  pendant sa durée de vie se fait exponentiellement à taux  $\tau > 0$  depuis sa taille à la naissance  $\xi_u$ . La taille de cette cellule en  $t \in [b_u, b_u + \zeta_u)$  est alors donnée par

$$\xi_u^t = \xi_u \exp(\tau(t - b_u)). \quad (2)$$

Le second phénomène est la division des cellules en deux nouvelles cellules de tailles égales. Une cellule  $u$  donne naissance aux cellules  $u_0 = (u, 0)$  et  $u_1 = (u, 1)$  de tailles à la naissance

$$\xi_{u_0} = \xi_{u_1} = \frac{1}{2} \xi_u \exp(\tau \zeta_u), \quad (3)$$

où  $\xi_u \exp(\tau \zeta_u)$  représente la taille à la division de la cellule  $u$ . La division d'une cellule se produit à un taux dépendant de sa taille. En prenant une fonction  $B : (0, \infty) \rightarrow [0, \infty)$  telle que  $\int_0^\infty x^{-1} B(x)dx = \infty$  pour taux de division, une cellule de taille  $x$  se divise avec probabilité  $B(x)dt$  entre les instants  $t$  et  $t + dt$ , ce qui se traduit par

$$\mathbb{P}(\zeta_u \in [t, t + dt) | \zeta_u \geq t, \xi_u = x_b) = B(x_b \exp(\tau t))dt, \quad (4)$$

où  $x_b \exp(\tau t)$  représente la taille d'une cellule de taille initiale  $x_b$  au bout d'un temps de vie  $t$ .

La suite  $(\xi_u, u \in \mathbb{T})$  forme une chaîne de Markov sur un arbre, de transition

$$\mathbb{P}((\xi_{u_0}, \xi_{u_1}) \in dx_0 dx_1 | \xi_u = x). \quad (5)$$

Nous appellerons un tel objet *chaîne de Markov bifurquante*, à la suite de J. Guyon [Guy07], et sa transition sera nommée plus spécifiquement  $\mathbb{T}$ -transition pour souligner la structure d'arbre. Le

qualificatif *bifurquant* est préféré à *branchant* puisque nous autorisons ici les traits de deux cellules sœurs à être dépendants, conditionnellement au trait de leur parent. Une définition générale est donnée en Section 4.

Soulignons pour finir que les deux spécifications précédentes sont conformes au développement de la bactérie *E. coli*. En effet, il a été établi expérimentalement que, pour cette bactérie, la croissance se fait exponentiellement (voir [RHK+14], Figure 2). Par ailleurs l'asymétrie dans la division chez *E. coli* est très légère (voir [MHT66] ou [SRBA14], Figure 8) et c'est pourquoi nous la négligeons.

### Approche déterministe : équation de croissance-fragmentation

À la vision microscopique précédente, donnant naturellement lieu à une analyse en termes stochastiques, correspond une vision macroscopique, déterministe quant à elle. Macroscopiquement, l'on s'intéresse à l'évolution temporelle de la quantité globale de cellules d'un âge ou d'une taille donnée, et cette évolution se traduit par une équation aux dérivées partielles.

**Modèle structuré en âge.** Nous considérons la quantité à l'instant  $t$  de cellules d'âge  $a$ , notée  $n(t, a)$ , satisfaisant l'équation d'évolution

$$\begin{cases} \frac{\partial}{\partial t} n(t, a) + \frac{\partial}{\partial a} n(t, a) + B(a)n(t, a) = 0, \\ n(t, a = 0) = 2 \int_0^\infty B(s)n(t, s)ds, \quad n(t = 0, a) = n^{\text{in}}(a), \end{cases} \quad (6)$$

où  $n^{\text{in}}$  est une condition initiale. Un terme de transport rend compte du vieillissement des cellules, tandis qu'un terme de fragmentation témoigne de la division des cellules à taux  $B$ , dépendant de l'âge  $a$ . La condition au bord assure que la quantité de cellules d'âge nul qui naissent est le double de la quantité de cellules qui se divisent.

Le modèle décrit par (6) est certes relativement limité pour décrire la croissance d'une population de cellules. Il n'en reste pas moins un modèle archétypique, qu'il est important de comprendre, et qu'il est possible d'enrichir. Par exemple, les différentes phases du cycle de la division cellulaire peuvent être modélisées par des équations de type (6) assemblées pour constituer un système différentiel. De nombreux autres enrichissements de ce modèle sont décrits par B. Perthame [Per07] (Section 3.9).

**Modèle structuré en taille.** Pour  $n(t, x)$ , la quantité à l'instant  $t$  de cellules de taille  $x$ , l'équation de croissance-fragmentation

$$\begin{cases} \frac{\partial}{\partial t} n(t, x) + \frac{\partial}{\partial x} (\tau x n(t, x)) + B(x)n(t, x) = 4B(2x)n(t, 2x), \\ n(t = 0, x) = n^{\text{in}}(x) \end{cases} \quad (7)$$

est valable. Le mécanisme capturé par cette équation d'évolution peut être vu comme une équation de bilan de masse (voir J. A. J. Metz et O. Dieckmann [MD86] ou H. T. Banks *et al.* [BST+11]) :

la quantité de cellules de taille  $x$  à l'instant  $t$  est alimentée par un terme de transport  $\tau x$  qui rend compte de la croissance par ingestion d'un nutriment dans le milieu expérimental, et chaque cellule se divise en deux enfants de tailles identiques à la naissance, selon un taux de division  $B$ , dépend de sa taille. Le facteur 4 provient de la multiplication de deux facteurs 2, une cellule de taille  $2x$  donnant naissance à 2 cellules de taille  $x$ .

De nombreux travaux mathématiques existent sur les équations d'évolution structurées en âge ou en taille. Citons de nouveau l'ouvrage de B. Perthame [Per07] pour une vue d'ensemble du sujet. Le modèle structuré à la fois en âge et en taille est étudié par M. Doumic [Dou07], et nous renvoyons également aux références afférentes.

### *Lien entre les approches stochastique et déterministe*

Les deux approches précédentes coïncident selon le sens suivant. Notons  $\Upsilon_u^t$  la valeur à l'instant  $t$  d'un trait biologique quelconque associé à la cellule  $u \in \mathbb{T}$  et  $\delta_x$  la masse de Dirac au point  $x$ . La mesure empirique moyenne

$$n(t, dy) = \mathbb{E} \left[ \sum_{i=1}^{\infty} \mathbf{1}_{\{X_i(t) > 0\}} \delta_{X_i(t)}(dy) \right] = \mathbb{E} \left[ \sum_{u \in \mathbb{T}} \mathbf{1}_{\{b_u \leq t < b_u + \zeta_u\}} \delta_{\Upsilon_u^t}(dy) \right]$$

résout dans un sens faible l'équation d'évolution (6) quand  $\Upsilon_u^t = \zeta_u^t$  et l'équation d'évolution (7) quand  $\Upsilon_u^t = \xi_u^t$ . Pour une démonstration (via l'utilisation de many-to-one formulae), nous citons M. Doumic *et al.* [DHKR15] concernant le modèle structuré en taille et renvoyons au Chapitre IV pour le modèle structuré en âge. Citons par ailleurs les travaux de V. C. Tran [Tra08], et ceux de B. Cloez [Clo11] par exemple, qui incluent l'étude de limites en grande population dans un cadre général, et établissent par là un lien de même type.

## 2.2 COMPORTEMENTS EN TEMPS LONG

Nous notons désormais  $\Upsilon_u$  le ou les traits biologiques caractérisant la cellule  $u$ , à la naissance ou à la division, selon la convention choisie ou la pertinence. Cette caractéristique peut représenter l'âge à la division ( $\Upsilon_u = \zeta_u$ ), la taille à la naissance ( $\Upsilon_u = \xi_u$ ), ou autre.

### *Fragment marqué discret ou continu*

L'étude de  $(X(t), t \geq 0)$  comme celle de  $(\Upsilon_u, u \in \mathbb{T})$  se fondent sur un objet central nommé *fragment marqué* (voir l'ouvrage de J. Bertoin [Ber06] par exemple).

**Fragment marqué discret.** Un fragment marqué en temps discret est construit de la façon suivante. Pour  $(\epsilon_1, \epsilon_2, \dots)$ , une suite de variables aléatoires de Bernoulli de paramètre  $1/2$ , nous posons  $\chi_0 = \Upsilon_\emptyset$  où  $\emptyset$  représente la cellule initiale et pour  $m \geq 1$ ,

$$\chi_m = \Upsilon_{\emptyset_{\epsilon_1 \dots \epsilon_m}}.$$

La suite  $(\chi_m, m \geq 0)$  forme une chaîne de Markov et revient à suivre les caractéristiques des cellules appartenant à une branche choisie aléatoirement dans l'arbre binaire  $\mathbb{T}$  (le choix d'une cellule se

faisant uniformément à chaque branchement).

**Fragment marqué continu, modèle en âge.** Le fragment marqué discret des âges à la division peut être suivi en temps continu et la version continue est un processus  $(\tilde{\chi}_t, t \geq 0)$  de générateur infinitésimal

$$\mathcal{A}_B g(a) = g'(a) + B(a)(g(0) - g(a)),$$

pour  $g$  fonction test suffisamment lisse. Un phénomène de biais de sélection, spécifique au cadre du temps continu, que nous détaillons par la suite, requiert l'introduction d'un fragment marqué dit *biaisé* (voir [BDMT11, Clo11] dans un cadre général). Il s'agit du processus  $(\tilde{\chi}_t, t \geq 0)$  dont la dynamique est décrite par le générateur

$$\mathcal{A}_{H_B} g(x) = g'(a) + H_B(a)(g(0) - g(a)),$$

où  $H_B : (0, \infty] \rightarrow [0, \infty)$  est un taux de division, autre que  $B$ , dit *biaisé*. Le fragment marqué biaisé  $\tilde{\chi}$  se distingue donc du fragment marqué précédent,  $\tilde{\chi}$ , par son taux de saut.

**Fragment marqué continu, modèle en taille.** Suivre en temps continu le fragment marqué discret des tailles à la naissance définit un processus  $(\tilde{\chi}_t, t \geq 0)$  de générateur infinitésimal

$$\mathcal{A}_B g(x) = \tau x g'(x) + B(x)(g(\frac{x}{2}) - g(x)),$$

pour  $g$  une fonction test suffisamment lisse. La dynamique du fragment marqué *biaisé*  $(\tilde{\chi}_t, t \geq 0)$  est quant à elle décrite par

$$\mathcal{A}_{H_B} g(x) = \tau x g'(x) + H_B(x)(g(\frac{x}{2}) - g(x)),$$

où  $H_B : (0, \infty] \rightarrow [0, \infty)$  est, de nouveau, un taux de division, autre que  $B$ , dit *biaisé*.

À la chaîne de Markov  $(\chi_m, m \geq 0)$  et au processus de Markov  $(\tilde{\chi}_t, t \geq 0)$ , nous associons, dès qu'elles existent, leurs mesures invariantes respectives,  $\nu_B$  et  $\tilde{\nu}_B$ .

### Approximations connues

Les résultats connus se déclinent selon les trois descriptions précédentes.

- 1) La moyenne empirique des caractéristiques des cellules appartenant aux  $n$  premières générations (ensemble noté  $\mathbb{T}_n$ ) est approchée comme cela :

$$\sum_{u \in \mathbb{T}_n} g(\Upsilon_u) \approx |\mathbb{T}_n| \nu_B(g),$$

en probabilité, pour  $n$  grand, où la notation  $\nu_B(g)$  est utilisée pour  $\int g d\nu_B$ , et  $\nu_B$  désigne la mesure invariante associée au fragment marqué discret  $\chi$  que nous venons de définir. Cette approximation est complétée par l'étude des fluctuations : un théorème de normalité asymptotique accompagne cette loi des grands nombres dans l'article de J. Guyon [Guy07],

et les fluctuations sont d'ordre  $|\mathbb{T}_n|^{1/2}$ . Citons J.-F. Delmas et L. Marsalle [DM10] pour la généralisation à un arbre de Galton-Watson. Un contrôle en norme quadratique moyenne est en outre établi par Doumic *et al.* [DHKR15], des inégalités de déviations sont démontrées dans les travaux de S. V. Bitseki Penda *et al.* [BDG14] (ainsi que dans le Chapitre II).

- 2) La moyenne empirique des caractéristiques à l'instant  $t$  des cellules vivantes se comporte quant à elle de la façon suivante,

$$\sum_{\substack{u \in \mathbb{T}, \\ b_u \leq t < b_u + d_u}} g(\Upsilon_u^t) \approx e^{\lambda_B t} \tilde{\nu}_B(w_B g) Z_B$$

en probabilité, pour  $t$  grand, où  $\lambda_B$  est le paramètre de Malthus (défini en 3) ci-dessous)  $\tilde{\nu}_B$  est la mesure invariante associée au fragment marqué biaisé  $(\tilde{\chi}_t, t \geq 0)$  défini ci-dessus,  $w_B$  une fonction de poids ne dépendant que de  $B$  (et dont l'expression, explicite ou non, varie selon le modèle choisi pour cadre) et  $Z_B$  une variable aléatoire non-dégénérée. Un résultat similaire peut être écrit pour la moyenne empirique des tailles des cellules se divisant avant  $t$  (telles que  $b_u + \zeta_u < t$ ). Les deux références clés à ce sujet sont les travaux de V. Bansaye, J.-F. Delmas, L. Marsalle et V. C. Tran d'une part, B. Cloez d'autre part [BDMT11, Clo11]. Pour le cas spécifique de l'âge ( $\Upsilon_u = \zeta_u$ ), nous proposons une étude concernant la vitesse de convergence liée à cette approximation dans le Chapitre I.

- 3) Le comportement en temps long de la solution temporelle de l'équation d'évolution est donné par

$$n(t, y) \approx e^{\lambda_B t} N_B(y),$$

pour une certaine norme pondérée, quand  $t$  est grand, où  $\lambda_B$  apparaît comme la valeur propre d'un problème aux valeurs propres associé à (6) ou à (7) selon le modèle considéré, dont  $N_B$  est une solution. Pour le modèle structuré en âge,

$$N_B'(a) + B(a)N_B(a) = -\lambda_B N_B(a), \quad N_B(a=0) = 2 \int_0^\infty B(s)N_B(s)ds, \quad (8)$$

tandis que pour le modèle structuré en taille,

$$(\tau x N_B(x))' + B(x)N_B(x) - 4B(2x)N_B(2x) = -\lambda_B N_B(x), \quad (9)$$

avec  $\int_0^\infty N_B(x)dx = 1$ . La quantité  $\lambda_B$  est nommée *paramètre de Malthus*, elle régit la croissance exponentielle de la population en temps grand, et  $N_B$  est nommé *profil stationnaire*. Parmi les nombreuses références à ce sujet, citons S. Mischler et J. Scher [MS13], K. Pakdaman, B. Perthame et D. Salort [PPS14] pour les articles les plus récents traitant de la vitesse de convergence (obtenue par des techniques de trou spectral) à laquelle cette limite est atteinte.

À la lumière de ces trois résultats, réinterprétons tour à tour et rendons plus précis les axes d'étude formulés en incipit :

(A) De chacun des résultats précédents peut découler une méthode de reconstruction du taux de division  $x \rightsquigarrow B(x)$ , selon la possibilité d'estimer  $\nu_B, \tilde{\nu}_B$  ou  $(\lambda_B, N_B)$  puis d'inverser les problèmes  $B \rightsquigarrow \nu_B, B \rightsquigarrow \tilde{\nu}_B$ , abstraction faite des difficultés que peuvent générer la fonction de poids  $w_B$  et l'aléa représenté par  $Z_B$ , ou encore  $B \rightsquigarrow (\lambda_B, N_B)$ . L'exploration de cet axe dépend donc radicalement du cadre inférentiel que l'on se donne.

(B) L'étude de la transmission du trait biologique d'une cellule à une autre s'accorde avec la description stochastique en temps généalogique. Elle n'est pertinente que pour le modèle structuré en taille (les âges à la division étant indépendants). L'étude du trait d'une cellule typique, ou d'une lignée de cellules typique, peut se faire en particulier via l'étude du fragment marqué discret (à travers sa transition, sa mesure invariante) ainsi que via l'étude de la  $\mathbb{T}$ -transition (5).

(C) Le paramètre clé régissant la croissance de la population est le paramètre de Malthus  $\lambda_B$ , qui fournit un critère simple de comparaison des populations.

### 2.3 ESTIMATION DU TAUX DE DIVISION : QUEL CADRE D'INFÉRENCE ?

Avant d'aborder la problématique (A), définissons les observables à notre disposition. Différents scénarios sont envisageables. De façon imagée, une photographie de la population à un instant  $T > 0$  peut être prise (voir Figure 1(a) ou 1(b)), ou un film de la population être enregistré entre deux instants, disons 0 et  $T$ .

#### *Schémas d'observation en temps continu*

**Observation au temps  $T$ .** Introduisons l'ensemble des cellules vivantes à un instant  $t \geq 0$  fixé,

$$\partial\mathcal{T}_t = \{u \in \mathbb{T}, b_u \leq t < b_u + \zeta_u\},$$

où, nous le rappelons,  $b_u$  est la date de naissance de  $u$  et  $\zeta_u$  sa durée de vie. Peut nous être fournie l'observation de  $X(T)$ , ou encore, de manière équivalente, l'observation de

$$(\Upsilon_u^T, u \in \partial\mathcal{T}_T),$$

$\Upsilon_u^T$  désignant la valeur à l'instant  $T$  du trait de la cellule  $u$ . Dit autrement, les valeurs courantes du ou des traits individuels choisis comme variables structurantes, des cellules vivantes à l'instant  $T$  sont observées.

**Observation jusqu'au temps  $T$ .** De façon analogue, introduisons l'ensemble des cellules qui se divisent avant un instant  $t \geq 0$  fixé,

$$\mathring{\mathcal{T}}_t = \{u \in \mathbb{T}, b_u + \zeta_u < t\}.$$

L'observation de  $(X(t), 0 \leq t \leq T)$  est accessible, ou encore l'observation de

$$(\Upsilon_u \in \mathring{\mathcal{T}}_T).$$

Les caractéristiques individuelles (à la naissance ou à la division) de toutes les cellules de la population avant l'instant  $T$  sont donc observées.

La Figure 2(a) ci-dessous représente ces deux types d'information : les ensembles  $\partial\mathcal{T}_T$  et  $\mathring{\mathcal{T}}_T$  sont représentés respectivement en rouge et bleu.

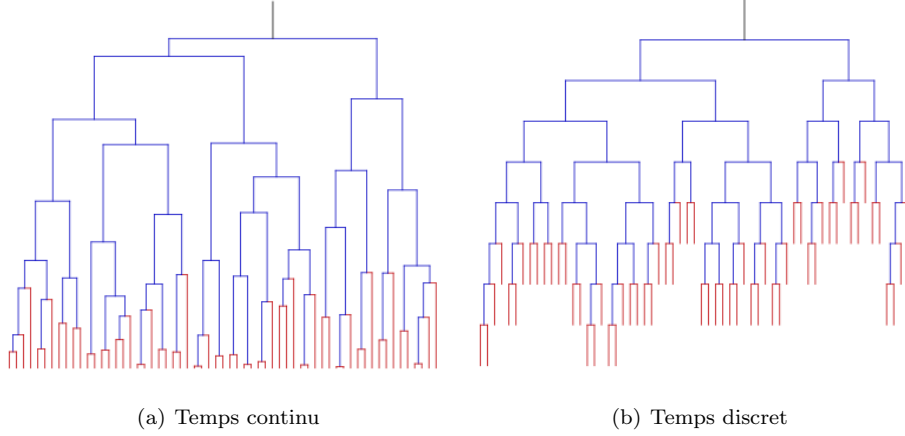


FIGURE 2 – Réalisation d'un processus de Markov branchant déterministe par morceaux jusqu'à une date  $T$  fixée. À gauche : la longueur de chaque segment est proportionnelle à la durée de vie d'une particule. À droite : représentation généalogique de la population, pour la même réalisation du processus.

Il est raisonnable d'espérer pouvoir reconstruire le taux de division  $x \rightsquigarrow B(x)$  à partir de l'un ou l'autre de ces schémas d'observation, en prenant pour base le résultat **2)** de la partie précédente (Section 2.2). Cependant le problème demeure complexe (voir la Section 3.1 détaillant les difficultés inhérentes au cadre du temps continu), ce pourquoi des schémas d'observation approchants ont été envisagés jusqu'à présent.

### Schémas d'observation approchants

**Observation en temps grand sous le profil stationnaire.** Après un certain temps d'évolution, suffisamment grand, les caractéristiques de cellules prises au hasard dans la population peuvent être observées. Le cadre statistique correspondant est alors celui d'un échantillon de  $n$  variables aléatoires  $(X_1, \dots, X_n)$  indépendantes et identiquement distribuées de densité  $N_B$ , le profil stationnaire défini en **3)** de la Section 2.2.

Ceci n'est qu'un cadre approchant l'observation à un temps  $T$  fixé, dans la mesure où toutes les corrélations existant entre les tailles des cellules sont ici négligées.

**Observation généalogique (jusqu'à la génération  $n$ ).** La technologie actuelle permet l'accès à beaucoup plus d'informations sur une population cellulaire que par le passé. Il est désormais possible de suivre l'évolution d'une colonie de cellules le long d'un arbre binaire en identifiant pour chacune, en particulier, sa taille à la naissance, ainsi que sa taille à la division et son temps de vie.

Notons  $\mathbb{T}_n = \{u \in \mathbb{T}, |u| \leq n\}$  où  $|u|$  désigne la génération à laquelle appartient  $u$  (avec la convention  $|\emptyset| = 0$ ). Les observés statistiques se composent des caractéristiques individuelles des cellules (à la naissance ou à la division) jusqu'à une profondeur  $n$  donnée dans l'arbre généalogique

de la population,

$$(\Upsilon_u, u \in \mathbb{T}_n).$$

Ce schéma d'observation approche le schéma d'observation entre la date 0 et la date  $T$ . Il s'en éloigne dans la mesure où les cellules nées entre 0 et  $T$  ne forment pas un sous-arbre plein jusqu'à une génération donnée mais composent au contraire un sous-ensemble aléatoire de  $\mathbb{T}$ , comme cela est illustré par la Figure 2(b). (D'un point de vue expérimental, les caractéristiques des cellules du sous-arbre  $\mathbb{T}_n$  sont extraites à partir d'un arbre observé du type de celui de la Figure 2(b).)

## 2.4 ESTIMATION DU TAUX DE DIVISION : ÉTAT DE L'ART

Nous rappelons que l'inconnue des modèles établis est le taux de division  $B$ , fonction de l'âge ou de la taille, et le problème de son estimation est de nature non-paramétrique.

### *Résultats connus sur le modèle en taille*

Jusqu'ici la reconstruction de  $B$  s'est faite sous les deux schémas d'observation que nous avons qualifiés d'approchants [DHR12, DHKR15].

**Observation en temps grand sous le profil stationnaire.** L'équation (7) est bien connue en analyse et d'un point de vue numérique, [PZ07, DPZ09, DT13, BDE14]. Ces travaux s'attachent à reconstruire globalement la fonction  $(B(x), x \in \mathcal{D})$ , en norme  $L^2(\mathcal{D})$  par exemple, où  $\mathcal{D}$  est un compact de  $(0, \infty)$ , lorsque seule une solution perturbée de la solution  $n(t, x)$  est à disposition. La stratégie consiste à approcher en temps grand la solution de (7) par son analogue stationnaire  $e^{\lambda_B t} N_B(x)$  (approximation 3), Section 2.2). À partir d'observations perturbées de  $\lambda_B$  et  $N_B$ , en inversant l'équation (9), une solution approchée de  $B(x)$  est obtenue, accompagnée d'estimations déterministes d'erreur.

En 2011, M. Doumic, M. Hoffmann, P. Reynaud-Bouret et V. Rivoirard [DHR12] ont posé les jalons d'un programme d'estimation statistique pour les modèles structurés : à un instant  $T$  grand, sont observées les tailles de  $n$  cellules parmi les cellules vivantes. Les corrélations entre cellules sont négligées et dans le même esprit que les travaux sus-cités, le problème approché (ou stationnaire) est envisagé, le temps d'observation  $T$  étant supposé assez grand pour que le régime stationnaire de la population soit une approximation valide.

Dans un cadre statistique rigoureusement établi, une méthode d'estimation par noyau, accompagnée de l'utilisation de techniques récentes de Goldenschluger-Lepski pour le choix de la fenêtre de lissage du noyau, permet de construire un estimateur  $x \in \mathcal{D} \rightsquigarrow \widehat{B}_n(x)$  satisfaisant l'estimation d'erreur

$$\mathbb{E}[\|\widehat{B}_n - B\|_{L^2(\mathcal{D})}^2]^{1/2} \lesssim n^{-s/(2s+3)}, \quad (10)$$

uniformément en  $B$  dans une certaine classe bien choisie, où  $s$  est la régularité Sobolev de  $B$ .

Précisons ici que la méthode d'estimation mise en place ne nécessite pas la connaissance *a priori* de cette régularité  $s$  et la méthode est qualifiée à ce titre d'*adaptive*. L'exposant  $2s/(2s+3)$  obtenu ici, en accord avec les prédictions déterministes de [DPZ09], indique un problème inverse d'ordre 1.



**Observation généalogique (jusqu'à la génération  $n$ ).** Adoptons à présent le point de vue généalogique et imaginons que les tailles à la naissance de toutes les cellules jusqu'à une génération fixée  $n$  sont observées. Dans ce cadre, l'estimation du taux de division  $B$ , fonction de la variable taille, a été conduite en 2012 par M. Doumic, M. Hoffmann, N. Krell et L. Robert [DHKR15]. La procédure d'estimation développée se fonde sur la relation

$$B(x) = \frac{\tau x}{2} \frac{\nu_B(x/2)}{\int_{x/2}^x \nu_B(y) dy}, \quad x > 0, \quad (11)$$

valide à condition que le fragment marqué discret déjà introduit (Section 2.2) admette une mesure invariante de densité  $\nu_B$  (par rapport à la mesure de Lebesgue). Un estimateur à noyau  $\hat{\nu}_n$  de  $\nu_B$  est construit à l'aide des observations  $(\xi_u, u \in \mathbb{T}_n)$  (pensons à l'approximation 1) de la Section 2.2), permettant la construction d'un estimateur  $\hat{B}_n$  de  $B$  et le contrôle de l'erreur

$$\mathbb{E}[\|\hat{B}_n - B\|_{L^2(\mathcal{D})}^2] \lesssim |\mathbb{T}_n|^{-2s/(2s+1)}. \quad (12)$$

Ce contrôle est valide uniformément en  $B$  dans une certaine classe bien choisie, où  $s$  est la régularité hölderienne de  $B$ .

La vitesse de convergence de l'estimateur fait intervenir un exposant  $2s/(2s+1)$  et améliore la vitesse de (10) en celle d'un problème direct, dit aussi bien posé. (Le nombre d'observations est  $n$  dans le premier cas et  $|\mathbb{T}_n| = 2^{n+1} - 1$  dans le second.)

Précisons que l'estimateur à noyau construit n'est pas adaptatif : le choix de la fenêtre de lissage optimale dépend de la connaissance *a priori* de la régularité de  $B$ , inconnue en pratique. Nous annonçons dès à présent que le Chapitre II viendra, en particulier, affiner les résultats de [DHKR15] dans deux directions principales : une procédure adaptative d'estimation sera proposée et l'optimalité au sens minimax de la vitesse de convergence sera démontrée.

### Remarques sur le modèle en âge

Les deux schémas d'observation dits approchants appliqués au modèle structuré en âge n'ont qu'un intérêt très limité.

Considérons d'abord l'observation en temps grand sous le profil stationnaire. Dans ce cadre, sont observés les âges de  $n$  cellules indépendantes de densité  $N_B$ , où  $N_B$  est le profil stationnaire solution de (8). Cette équation fournit un lien explicite entre le taux de division  $B$  d'intérêt et  $N_B$ . En effet, elle se réécrit

$$B(a) = -\frac{N'_B(a)}{N_B(a)} - \lambda_B.$$

Par conséquent, en imaginant que  $\lambda_B$  est connu (ou issu d'une estimation préliminaire), l'estimation de  $N'_B$  et  $N_B$  rend directement possible l'estimation de  $B$ . Notons qu'une vitesse de problème inverse est attendue, du fait de la présence de la dérivée de la densité  $N_B$ .

En ce qui concerne l'observation généalogique, l'échantillon  $(\zeta_u, u \in \mathbb{T}_n)$  des durées de vie des cellules jusqu'à la génération  $n$ , est composé de variables indépendantes et identiquement distribuées selon la densité  $f_B$ . La structure d'arbre binaire n'intervient donc en rien, et l'estimation de  $B$  est bien connue puisque pour tout  $a \geq 0$ ,  $B(a) = f_B(a)/(1 - \int_0^a f_B(s) ds)$ .

*Vue d'ensemble sur les modèles en âge et taille*

Le cadre inférentiel étant posé, l'état de l'art étant circonscrit, nous déclinons l'axe d'étude **(A)** selon les questions suivantes :

**(A.i)** À partir d'une observation *généalogique*, quelle procédure mettre en place pour que la reconstruction du taux de division *en taille* soit *adaptative et optimale* ?

**(A.ii)** À partir d'une observation *en temps continu*, comment reconstruire le taux de division, *en âge ou taille* ?

**(A.iii)** À partir d'une observation *en temps continu*, quelle procédure mettre en place pour que la reconstruction du taux de division, *en âge ou taille*, soit *adaptative et optimale* ?

Les résultats concernant le modèle structuré en âge consignés dans la première colonne de la Table 1 sont ceux obtenus dans le Chapitre I et nous les présentons plus en détail dans la partie suivante (Section 3). Ils constituent le premier pas significatif vers la compréhension de l'inférence dans un cadre en temps continu. Ce modèle en âge, dans sa simplicité apparente, renferme en réalité les difficultés essentielles qu'il est important de comprendre avant de traiter le modèle, plus adéquat pour la modélisation de la division d'*E. coli* (voir [RHK+14]) mais plus complexe, structuré en taille, dans un cadre continu (nous étayons cette allégation par la suite).

| Observation  | Modèle en âge  | Modèle en taille  |
|--|--|---|
| <b>Sous la densité stationnaire</b><br>( $n$ v.a. i.i.d) | $n^{-s/(2s+3)}$  | $n^{-s/(2s+3)}$<br>+ adaptation ([DHRR12])  |
| <b>En <math>T</math></b>                                 | $(e^{\lambda_B T})^{-s/(2s+3)}$<br>(Chapitre I)              | ?   |
| <b>Jusqu'à la génération <math>n</math></b>              | $ \mathbb{T}_n ^{-s/(2s+1)}$                                 | $ \mathbb{T}_n ^{-s/(2s+1)}$ ([DHKR15])<br>+ adaptation + optimalité<br>(Chapitre II) |
| <b>Entre 0 et <math>T</math></b>                         | $(e^{\lambda_B T})^{-s/(2s+1)}$<br>+ optimalité (Chapitre I) | ?   |

TABLE 1 – Synthèse des principaux résultats existants concernant l'estimation du taux de division  $B$  en tant que fonction de l'âge ou de la taille. Vitesses de convergence des estimateurs construits en fonction de la régularité  $s$  de  $B$  – nous précisons quand la méthode d'estimation est adaptative et si la vitesse de convergence est avérée optimale. Modèle structuré en âge :  $\lambda_B$  désigne le paramètre de Malthus.

La Table 1 synthétise dans sa seconde colonne les résultats existants que nous venons de décrire concernant le modèle structuré en taille, ainsi que les résultats nouveaux apportés par cette thèse. Le problème de la reconstruction du taux de division  $B$ , fonction de la variable *taille*, pour les deux schémas d'observation en temps continu reste à résoudre pour ce modèle.

Notons d'ores et déjà que les schémas les moins informatifs, à un instant  $T$ , avec ou sans approximation stationnaire, conduisent à des vitesses de convergence d'un problème inverse d'ordre 1. Les schémas d'observation les plus complets, entre les temps 0 et  $T$  ou entre les générations 0 et  $n$ , conduisent quant à eux à des vitesses d'un problème direct, dont l'optimalité au sens minimax est avérée, alors qu'elle reste à établir dans les cas précédents.

### 3 PREMIÈRE PARTIE : ESTIMATION DU TAUX DE DIVISION EN ÂGE

Nous traitons ici de la question (A.ii), à laquelle le Chapitre I est consacrée, pour le modèle structuré en *âge*. Nous nous concentrons sur le modèle structuré en âge en particulier, mais le but est de comprendre les difficultés générales liées au cadre inférentiel en temps continu, et d'y apporter les premières réponses.

#### 3.1 CADRE ET DIFFICULTÉS INHÉRENTES AU TEMPS CONTINU

Selon la terminologie utilisée jusqu'à présent dans ce chapitre introductif, l'objet d'étude du Chapitre I est un modèle structuré en âge : chaque cellule est caractérisée par son âge et se divise à un taux dépendant de son âge courant. Le processus qui en résulte est en réalité plus connu sous le nom de processus en âge branchant, processus introduit et étudié par Harris [Har63], aussi dénommé processus de Bellman-Harris. L'objectif est l'estimation non-paramétrique du taux de branchement, ou taux de division, en tant que fonction de l'âge, lorsque les observables sont collectées entre la date 0 et une date  $T$  fixée.

D'un point de vue stochastique microscopique, nous modélisons un système très simple de particules évoluant indépendamment les unes des autres. Une particule  $u$  vit pendant une durée  $\zeta_u$  aléatoire. Après la durée  $\zeta_u$ , la particule  $u$  est remplacée par  $\nu_u$  particules filles. Le nombre de particules filles  $\nu_u$ , indépendant de la durée de vie  $\zeta_u$ , est distribué selon la loi de probabilité  $(p_k, k \geq 2)$  telle que  $\sum_{k \geq 2} kp_k = m < \infty$ . Le cadre est donc légèrement plus général que celui de la division cellulaire puisque que la naissance de plus de deux enfants est autorisée à chaque instant de branchement. Les particules se divisent selon le taux de division  $B$ , à savoir

$$\mathbb{P}(\zeta_u \in [a, a + da) \mid \zeta_u \geq a) = B(a)da$$

avec  $B : [0, \infty) \rightarrow [0, \infty)$ ,  $\int_0^\infty B(a)da = \infty$ . Autrement dit, les durées de vie  $\zeta_u$  des particules de la population sont indépendantes et identiquement distribuées selon la densité

$$f_B(a) = B(a) \exp\left(-\int_0^a B(s)ds\right).$$

Le lien entre la densité des durées de vie  $f_B$  et le taux de division  $B$  s'écrit de façon équivalente

$$B(a) = \frac{f_B(a)}{1 - \int_0^a f_B(s) ds}, \quad (13)$$

et c'est à partir de cette formule que nous établissons une stratégie de reconstruction de  $B$ , étant donné les observables décrites ci-dessous.

Nous observons l'évolution du système de particules entre la date 0 et la date  $T$ . Plus précisément, nous observons  $(X(t), t \in [0, T])$  où  $X(t)$  recense les âges de toutes les cellules vivantes à l'instant  $t$ , ce qui revient à l'observation des durées de vie (éventuellement tronquées) suivantes, accompagnées du nombre d'enfants :

$$((\zeta_u, \nu_u), u \in \mathring{\mathcal{T}}_T) \quad \text{et} \quad (\zeta_u^T, u \in \partial\mathcal{T}_T)$$

où nous désignons par  $\mathring{\mathcal{T}}_T$  l'ensemble des cellules qui se divisent avant la date  $T$ , et par  $\partial\mathcal{T}_T$  l'ensemble des cellules vivantes à la date  $T$  (représentés respectivement en bleu et rouge sur la Figure 2(b)). Notons que deux types d'observations coexistent : durées de vie complètes  $\zeta_u$  et durées de vie censurées,  $\zeta_u^T = T - b_u$ , où  $b_u$  représente la date de naissance de  $u$ . Les observables sont donc constituées des longueurs des branches de l'arbre en temps continu typique de la Figure 2(a) (durées de vie complètes et censurées étant figurées respectivement en bleu et rouge).

Dans le cadre de ce modèle simple, les difficultés sont engendrées uniquement par le mode d'observation adopté :

- 1) Le nombre d'observations est aléatoire et sa loi dépend de  $B$  inconnu.
- 2) Il existe par ailleurs une dépendance complexe entre les durées de vie observées.
- 3) L'observation des durées de vie les plus courtes est favorisée par ce mode d'observation et c'est ce que nous nommons *biais de sélection*.

Voyons comment ces trois difficultés de natures différentes se traduisent dans les résultats que nous obtenons et dans les techniques de preuve que nous utilisons.

## 3.2 SYNTHÈSE DES RÉSULTATS NOUVEAUX

### *Comportement asymptotique des moyennes empiriques*

L'asymptotique considérée par la suite est celle où le temps  $T$  d'observation du processus tend vers l'infini.

**Lois des grands nombres existantes.** Une étude des moyennes empiriques en temps continu pour des processus de Markov branchants généraux a été réalisée par V. Bansaye, J.-F. Delmas, L. Marsalle et V. C. Tran [BDMT11], dans le cas d'un taux de branchement constant, puis complétée

par B. Cloez [Clo11], pour des taux de branchement quelconques. Deux résultats correspondent aux deux types d'observables à disposition. Dans le cas particulier du processus en âge que nous étudions, d'une part, pour les particules vivantes à l'instant  $T$ ,

$$\mathcal{E}^T(\partial\mathcal{T}_T, g) = \frac{1}{|\partial\mathcal{T}_T|} \sum_{u \in \partial\mathcal{T}_T} g(\zeta_u^T) \xrightarrow{\mathbb{P}} \partial\mathcal{E}_B(g) = \int_0^\infty g(a) \frac{\lambda_B m}{m-1} e^{-\lambda_B a} e^{-\int_0^a B(s) ds} da, \quad (14)$$

et d'autre part, pour les particules se divisant avant  $T$ ,

$$\mathcal{E}^T(\mathring{\mathcal{T}}_T, g) = \frac{1}{|\mathring{\mathcal{T}}_T|} \sum_{u \in \mathring{\mathcal{T}}_T} g(\zeta_u) \xrightarrow{\mathbb{P}} \mathring{\mathcal{E}}_B(g) = \int_0^\infty g(a) m e^{-\lambda_B a} f_B(a) da, \quad (15)$$

où  $\lambda_B$  désigne le paramètre de Malthus, unique solution de

$$\int_0^\infty m e^{-\lambda_B a} f_B(a) da = 1.$$

Le paramètre de Malthus est central dans cette étude ; il régit la croissance (exponentielle) de la population et satisfait

$$\mathbb{E}[|\partial\mathcal{T}_T|] \sim \mathbb{E}[|\mathring{\mathcal{T}}_T|] \sim e^{\lambda_B T}, \quad (16)$$

quand  $T$  est grand, à constante près ne dépendant que de  $B$ .

Comparons d'abord (14) et (15). La représentation (13) de  $B$  nous laisse voir que sa reconstruction passe par celle de la densité  $f_B$ , et là se trouve l'enjeu – le dénominateur de (13), comme intégrale d'une densité, pouvant s'estimer à vitesse paramétrique. Le comportement asymptotique de la moyenne empirique sur  $\mathring{\mathcal{T}}_T$  laisse précisément apparaître la densité  $f_B$ , tandis que le comportement asymptotique de la moyenne empirique sur  $\partial\mathcal{T}_T$  ne fait apparaître  $B$  qu'à travers son intégrale (abstraction faite pour l'heure des fonctions de poids inconnues intervenant). L'utilisation de (15) conduira donc *a priori* à un problème direct, dit aussi bien posé, tandis que (14) conduira *a priori* à un problème inverse d'ordre 1, dit aussi mal posé.

Concernant (15) en particulier, asymptotiquement, tout se passe comme si nous observions un échantillon de densité non pas  $f_B$  mais de densité  $x \rightsquigarrow m e^{\lambda_B x} f_B(x)$ , du fait du biais de sélection. C'est de cette façon qu'est modélisé un biais de sélection dans S. Efromovitch [Efr13] dans le cas d'observations indépendantes et identiquement distribuées, pour une fonction de biais connue. Outre le fait que l'analogie n'est valable qu'asymptotiquement, nous rappelons dans notre cas la difficulté **2**) et nous soulignons que la fonction de biais dépend de  $B$  lui-même.

**Identification d'une vitesse de convergence.** Les lois des grands nombres (14) et (15) peuvent être affinées, et nous établissons que :

**Résultat 1.** *Sous des conditions appropriées d'ergodicité du fragment marqué biaisé, les suites*

$$e^{\lambda_B T/2} (\mathcal{E}^T(\partial\mathcal{T}_T, g) - \partial\mathcal{E}_B(g)) \quad \text{et} \quad e^{\lambda_B T/2} (\mathcal{E}^T(\mathring{\mathcal{T}}_T, g) - \mathring{\mathcal{E}}_B(g))$$

*sont tendues, quand  $T \rightarrow \infty$ , uniformément sur une classe de taux de division  $B$  minorés et majorés et de fonctions tests  $g$  bornées.*

Compte tenu de (16), la vitesse identifiée, en  $e^{\lambda_B T/2}$ , est bien la vitesse attendue, de type paramétrique. Seul un critère de tension est obtenu en raison des normalisations aléatoires par  $|\partial\mathcal{J}_T|$  et  $|\mathring{\mathcal{J}}_T|$ . L'outil principal auquel nous faisons appel est constitué des *many-to-one formulae*, développées dans [BDMT11, Clo11]. Les many-to-one formulae permettent de réduire, en espérance, l'étude d'une somme sur les ensembles  $\partial\mathcal{J}_T$  ou  $\mathring{\mathcal{J}}_T$  à l'étude d'un processus de Markov unidimensionnel. Plus précisément,

$$\mathbb{E}\left[\sum_{u \in \partial\mathcal{J}_T} g(\zeta_u^T)\right] = m^{-1} e^{\lambda_B T} \mathbb{E}[g(\tilde{\chi}_T) B(\tilde{\chi}_T)^{-1} H_B(\tilde{\chi}_T)]$$

tandis que

$$\mathbb{E}\left[\sum_{u \in \mathring{\mathcal{J}}_T} g(\zeta_u)\right] = m^{-1} \int_0^T e^{\lambda_B s} \mathbb{E}[g(\tilde{\chi}_s) H_B(\tilde{\chi}_s)] ds$$

avec  $(\tilde{\chi}_t, t \geq 0)$  un processus d'âge le long d'une branche, dit fragment marqué *biaisé* (voir Section 2.2), ayant pour générateur  $\mathcal{A}_{H_B} g(x) = g'(x) + H_B(x)(g(0) - g(x))$ , où le taux de saut du processus  $\tilde{\chi}$  n'est pas  $B$  lui-même mais un taux *biaisé*  $H_B$  caractérisé par

$$f_{H_B}(a) = m e^{-\lambda_B a} f_B(a),$$

s'écrivant par conséquent

$$H_B(a) = \frac{m e^{-\lambda_B a} f_B(a)}{1 - \int_0^a m e^{-\lambda_B s} f_B(s) ds}.$$

L'étude se réduit donc à celle du processus de Markov unidimensionnel  $\tilde{\chi}$ , non réversible, que nous montrons être ergodique. La vitesse de convergence vers l'équilibre de ce processus, disons  $\exp(-\rho_B T)$ , entre en compétition avec la croissance globale du système mesurée par  $\exp(\lambda_B T)$ . La condition  $\rho_B \geq \lambda_B/2$  est alors essentielle, quoique ardue à satisfaire, pour l'obtention du Résultat 1. Elle est cependant valable pour une classe restreinte de taux de division  $B$  minorés et majorés. Un résultat général permet d'identifier la vitesse  $\exp(\min\{\lambda_B/2, \rho_B\}T)$  pour les moyennes empiriques  $\mathcal{E}^T(\partial\mathcal{J}_T, g)$  et  $\mathcal{E}^T(\mathring{\mathcal{J}}_T, g)$ .

### Estimation du taux de division

Concernant l'étude statistique, à la lumière de (13) nous proposons l'estimateur du taux de division au point  $a > 0$  suivant :

$$\hat{B}_T(a) = \frac{\hat{f}_T(a)}{1 - \hat{F}_T(a)},$$

avec

$$\hat{f}_T(a) = \frac{1}{|\mathring{\mathcal{J}}_T|} \sum_{u \in \mathring{\mathcal{J}}_T} \hat{m}_T^{-1} e^{\hat{\lambda}_T \zeta_u} K_h(a - \zeta_u) \quad \text{et} \quad \hat{F}_T(a) = \frac{1}{|\mathring{\mathcal{J}}_T|} \sum_{u \in \mathring{\mathcal{J}}_T} \hat{m}_T^{-1} e^{\hat{\lambda}_T \zeta_u} \mathbf{1}_{\{\zeta_u \leq a\}}$$

où  $h > 0$  est une fenêtre de lissage bien choisie et  $K_h(\cdot) = h^{-1} K(h^{-1}\cdot)$  avec  $K : \mathbb{R} \rightsquigarrow \mathbb{R}$  un noyau. L'estimateur de  $B$  dépend de deux paramètres additionnels du modèle, chacun estimé à la vitesse de type paramétrique  $\exp(\lambda_B T/2)$ . Le premier est le nombre moyen d'enfants  $m$  estimé par  $\hat{m}_T$  de façon classique à l'aide des nombres d'enfants  $(\nu_u, u \in \mathring{\mathcal{J}}_T)$  observés. Le second est le paramètre de

Malthus  $\lambda_B$ , paramètre qui dépend de la fonction  $B$  inconnue elle-même. L'estimateur  $\widehat{\lambda}_T$  utilise les durées de vie tronquées  $(\zeta_u^T, u \in \partial\mathcal{J}_T)$ , en sus des durées de vie complètes utilisées par ailleurs pour construire l'estimateur  $\widehat{B}_T(a)$  en lui-même. Le Résultat 1 s'applique à l'estimateur  $\widehat{\lambda}_T$  et nous obtenons une vitesse de convergence de type paramétrique.

Rappelons-nous (16) donnant un équivalent du nombre moyen d'observations : la vitesse cible de convergence pour l'estimateur  $\widehat{B}_T(a)$  de  $B(a)$  ainsi construit est alors  $\exp(\lambda_B \frac{s}{2s+1} T)$ , où  $s$  est la régularité de  $B$ , supposée connue dans ce travail. C'est ce que nous démontrons à travers bornes supérieure et inférieure :

**Résultat 2.** *D'une part, pour une classe de taux de division  $B$  minorés et majorés (de sorte que  $\rho_B \geq \lambda_B$ ), de régularité hölderienne  $s > 1/2$ , quand  $C \rightarrow \infty$ ,*

$$\limsup_{T \rightarrow \infty} \sup_B \mathbb{P}(e^{\lambda_B \frac{s}{2s+1} T} |\widehat{B}_T(a) - B(a)| \geq C) \rightarrow 0,$$

*et, d'autre part, il existe une constante  $C' > 0$  telle que*

$$\liminf_{T \rightarrow \infty} \inf_{\widehat{B}_T} \sup_B \mathbb{P}(e^{\lambda_B \frac{s}{2s+1} T} |\widehat{B}_T(a) - B(a)| \geq C') > 0,$$

*où l'infimum porte sur tous les estimateurs  $\widehat{B}_T$  construits à partir de l'observation du processus jusqu'en  $T$ .*

La démonstration de la première partie du Résultat 2 repose sur un arbitrage entre termes de biais et de variance, le terme de variance étant contrôlé l'aide du Résultat 1, légèrement affiné pour obtenir un contrôle précis en les normes  $L^1, L^2$  et  $L^\infty$  de la fonction test  $g$ . L'établissement de la borne inférieure repose sur le processus de rapport de vraisemblance étudié par E. Löcherbach dans [Löc02a] pour un système de particules général. Le Chapitre I fournit également la vitesse de convergence de  $\widehat{B}_T(a)$  dans le régime  $\rho_B < \lambda_B$  (cependant, cette vitesse n'est vraisemblablement pas optimale).

Fait remarquable ici, non seulement la vitesse de convergence mais également son ordre de grandeur dépendent de  $B$  lui-même, à travers l'exposant  $\lambda_B$ . Pour atteindre cette vitesse de convergence, le noyau doit avoir pour fenêtre

$$h = \widehat{h}_T = \exp\left(-\widehat{\lambda}_T \frac{1}{2s+1} T\right),$$

où  $\widehat{\lambda}_T$  est un estimateur de  $\lambda_B$ . La fenêtre de lissage est donc aléatoire et cela requiert l'étude de la régularité de  $\widehat{B}_T$  vu comme processus en la fenêtre  $h$ . La mise en place d'une procédure d'estimation s'adaptant automatiquement à la régularité de  $B$  (question (A.iii)), sans nécessiter sa connaissance préalable, serait une prolongation naturelle et directe de ce travail. Soulignons que la mise en place d'une procédure adaptative nécessite d'étudier les déviations de la moyenne empirique  $\mathcal{E}^T(\mathcal{J}_T, g)$ .

Lorsque seules les observations tronquées  $(\zeta_u^T, u \in \partial\mathcal{J}_T)$  sont observées, un estimateur convergent à vitesse  $\exp(\lambda_B \frac{s}{2s+3} T)$  peut être construit. L'optimalité de cette vitesse pour l'observation à une unique date  $T$  fixée demeure à établir.

Nous voyons ce travail comme une première étape significative dans l'étude du problème de la reconstruction du taux de division d'un processus branchant observé en temps continu. La simplicité du processus en âge a le mérite de permettre l'identification claire du spectre des difficultés, dont nous donnons la décomposition page 15. La difficulté **1**) se traduit dans nos résultats par des critères de tension et non des critères plus forts, dans  $L^2$  par exemple. La difficulté **2**) concernant les dépendances complexes entre les observations est en grande partie résorbée à l'aide des many-to-one formulae, qui réduisent l'étude à un processus de Markov unidimensionnel. Le contrôle de la vitesse de convergence à l'équilibre de ce processus – convergence devant être suffisamment rapide – est la difficulté résiduelle. Le biais de sélection mentionné en **3**) apparaît ici de façon relativement simple à travers la fonction poids  $x \rightsquigarrow m \exp(-\lambda_B x)$ , explicite d'une part et pouvant être estimée aisément d'autre part, alors que rien de tel ne peut être espéré pour des modèles plus complexes.

À ce stade, la question **(A.ii)** n'est que partiellement résolue puisque le problème de l'estimation du taux de division pour le modèle structuré en *taille* demeure ouvert. Pour ce modèle, mentionnons l'étude récente de V. H. Hoang [Hoa15] sur l'estimation du noyau de fragmentation. Cette étude est menée en temps continu, sous l'hypothèse d'un taux de division constant, ce qui permet de s'affranchir des difficultés précédemment décrites. La question **(A.iii)** concernant l'adaptation pour l'estimation du taux de division reste ouverte également, pour les deux modèles, structurés en âge ou taille. Compléter la vision d'ensemble de la Table 1 est une perspective de travail naturelle. L'enrichissement du modèle structuré en âge pourra également être envisagé, avec par exemple des taux de division et taux de mort différents.

## 4 DEUXIÈME PARTIE : ÉTUDE DE LA TRANSMISSION D'UN TRAIT

L'objet central n'est plus spécifiquement le taux de branchement des processus introduits. Nous nous plaçons dans un cadre général nous permettant d'appréhender la transmission d'un trait quantitatif biologique entre cellules, d'une génération à une autre. Les Chapitres II et III abordent l'un et l'autre cet axe d'étude **(B)**.

### 4.1 CADRE GÉNÉRAL ET EXEMPLES PARADIGMATIQUES

#### *Chaînes de Markov bifurquantes*

Le cadre général que nous nous donnons est celui des chaînes de Markov bifurquantes, BMC selon l'acronyme anglais, qui peuvent être vues comme la généralisation des chaînes de Markov, adaptées à une structure d'arbre binaire. Formalisons les choses à la manière de J. Guyon dans [Guy07]. Nous rappelons que  $\mathbb{T}$  désigne l'arbre binaire infini où chaque nœud est identifié à une cellule de la population étudiée. Nous désignons par  $\mathbb{G}_m = \{0, 1\}^m$  l'ensemble des cellules appartenant à la génération  $m$  et par  $\mathbb{T}_m$  l'ensemble des cellules jusqu'à la génération  $m$ . Pour chaque cellule  $u$ , nous nous intéressons à un trait biologique quantitatif  $X_u \in \mathcal{S}$  non spécifié, où  $\mathcal{S}$  désigne un espace d'état général. Nous nous donnons un noyau markovien noté  $\mathcal{P}$  de  $(\mathcal{S}, \mathfrak{S})$  dans  $(\mathcal{S} \times \mathcal{S}, \mathfrak{S} \otimes \mathfrak{S})$ , que nous appelons  $\mathbb{T}$ -transition. Le processus  $(X_u, u \in \mathbb{T})$  est appelé chaîne de Markov bifurquante si

$$\mathbb{E} \left[ \prod_{u \in \mathbb{G}_m} g_u(X_u, X_{u0}, X_{u1}) \middle| \mathcal{F}_m \right] = \prod_{u \in \mathbb{G}_m} \mathcal{P} g_u(X_u)$$



pour toute famille  $(g_u, u \in \mathbb{G}_m)$  de fonctions mesurables bornées sur  $\mathcal{S}^3$  où la filtration  $\mathcal{F}_m$  est, par exemple, la filtration naturelle engendrée par  $(X_u, u \in \mathbb{T}_m)$ , et  $\mathcal{P}g(x) = \int_{\mathcal{S}^2} g(x, x_0, x_1) \mathcal{P}(x, dx_0 dx_1)$ . Cette propriété est fondamentale et généralise la propriété de branchement. En effet, dans ce cadre, deux cellules sœurs peuvent être corrélées conditionnellement à leur parent. Nous soulignons la relation  $\mathcal{P}(x, dx_0 dx_1) = \mathbb{P}((X_{u_0}, X_{u_1}) \in dx_0 dx_1 | X_u = x)$ .

Le fragment marqué tel que défini précédemment (Section 2.2) joue un rôle fondamental dans l'analyse. Les marques d'une branche tirée aléatoirement dans l'arbre (uniformément à chaque branchement) forment une chaîne de Markov de transition

$$\mathcal{Q} = \frac{1}{2}(\mathcal{P}_0 + \mathcal{P}_1)$$

où  $\mathcal{P}_0$  et  $\mathcal{P}_1$  sont les transitions marginales de la  $\mathbb{T}$ -transition  $\mathcal{P}$ . Nous supposons que la chaîne de transition  $\mathcal{Q}$  est ergodique, et nous notons  $\nu$  la mesure invariante associée.

Nous nous demandons alors :

**(B.i)** À partir de toutes les observations  $(X_u, u \in \mathbb{T}_n)$ , comment reconstruire la transition moyenne  $\mathcal{Q}$ , ainsi que la loi à l'équilibre correspondante  $\nu$  ?

**(B.ii)** À partir des observations  $(X_u, u \in \mathbb{T}_n)$ , comment reconstruire la  $\mathbb{T}$ -transition  $\mathcal{P}$  ?

**(B.iii)** Est-il possible de reconstruire des paramètres caractérisant la  $\mathbb{T}$ -transition dans certains modèles spécifiques ?

Dans le cas des chaînes de Markov, le problème analogue à **(B.i)** a été traité dans les travaux de S. Cléménçon et C. Lacour [Clé00, Lac07]. Bien que les objets  $\mathcal{Q}$  et  $\nu$  que nous visons à reconstruire décrivent une chaîne de Markov, celle du fragment marqué en l'occurrence, nous souhaitons les reconstruire à travers l'observation de toute la chaîne de Markov *bifurquante* qui les induit.

Les trois approximations suivantes attestent de la possibilité d'estimer les trois objets d'intérêt dès lors que  $(X_u, u \in \mathbb{T}_n)$  est observé, et les questions **(B.i)**, **(B.ii)** trouvent bien leur sens. Outre l'approximation

$$\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} g(X_u) \underset{n \rightarrow \infty}{\approx} \nu(g) \tag{17}$$

déjà rencontrée, J. Guyon établit dans [Guy07],

$$\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} g(X_{u^-}, X_u) \underset{n \rightarrow \infty}{\approx} \nu(\mathcal{Q}g), \tag{18}$$

où  $u^-$  désigne le parent de  $u$ , ainsi que le résultat général englobant les deux précédents,

$$\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} g(X_u, X_{u_0}, X_{u_1}) \underset{n \rightarrow \infty}{\approx} \nu(\mathcal{P}g), \tag{19}$$

l'ensemble  $\mathbb{T}_n$  pouvant invariablement être remplacé par  $\mathbb{G}_n$ .

### Exemples paradigmatiques

Nous nous posons de plus la question **(B.iii)** dans les deux exemples qui suivent. L'espace d'état  $\mathcal{S}$  est à présent choisi égal à, ou inclus dans,  $(0, \infty)$ .

**Exemple 1. Modèle de croissance-fragmentation.** Le processus discret des tailles à la naissance  $(\xi_u, u \in \mathbb{T})$  tel que décrit en Section 2 forme, nous l'avions annoncé, une BMC de  $\mathbb{T}$ -transition

$$\mathcal{P}_B(x, dx_0 dx_1) = Q_B(x, x_0) \delta_{x_0}(dx_1) dx_0$$

avec

$$Q_B(x, y) = \frac{B(2y)}{\tau y} \exp\left(-\int_{x/2}^y \frac{B(2z)}{\tau z} dz\right) \mathbf{1}_{\{y \geq x/2\}},$$

expression obtenue à l'aide des équations (3) et (4) (voir [DHKR15]). La transition moyenne n'est autre que  $\mathcal{Q} = Q_B$ , de mesure invariante  $\nu = \nu_B$ . Nous rappelons la formule (11), fondamentale pour établir une stratégie de reconstruction de  $B$ . L'estimation de  $\nu_B$  présente donc le grand intérêt de permettre celle du taux de division  $B$ . La réponse à la question **(B.iii)** est alors ici affirmative. Dans ce modèle spécifiquement, les objectifs **(A)** et **(B)** se trouvent être liés, et une réponse au premier problème, dans un cadre inférentiel généalogique, découle naturellement d'une réponse au second.

**Exemple 2. Modèle bifurquant auto-régressif.** Le modèle bifurquant auto-régressif linéaire a été introduit par Cowan et Staudte en 1986 dans [CS86] afin d'étudier la bactérie *E. coli*. Ces modèles ont notamment servi à étudier le vieillissement des cellules [SMPT05, GBP+05]. Le trait biologique privilégié dans ce cadre est alors le taux de croissance. Nous supposons que les quantités  $X_{u0}$  et  $X_{u1}$  associées aux deux descendants  $u0$  et  $u1$  de  $u$ , dépendent de leurs ancêtres seulement à travers  $X_u$ . Nous considérons le modèle bifurquant auto-régressif non-linéaire défini par

$$X_{u0} = f_0(X_u) + \varepsilon_{u0} \quad \text{et} \quad X_{u1} = f_1(X_u) + \varepsilon_{u1}, \quad (20)$$

pour tout  $u \in \mathbb{T}$ , où  $f_0$  et  $f_1$  sont des fonctions inconnues dites d'auto-régression. La racine  $X_\emptyset$  est distribuée selon une mesure initiale  $\mu$  quelconque et la famille  $((\varepsilon_{u0}, \varepsilon_{u1}), u \in \mathbb{T})$  est composée de variables aléatoires indépendantes et identiquement distribuées selon une densité  $G$ . Ce modèle s'insère également dans le cadre général des BMC précédemment introduites. La  $\mathbb{T}$ -transition est ici donnée par

$$\mathcal{P}(x, dx_0 dx_1) = G(x_0 - f_0(x), x_1 - f_1(x)) dx_0 dx_1, \quad (21)$$

où  $G$  est la densité bivariée du bruit.

La littérature concernant les processus bifurquants auto-régressifs est déjà vaste : citons en particulier J.-F. Delmas et L. Marsalle [DM10], B. Bercu, B. de Saporta et A. Gégout-Petit [BdSG09], S. V. Bitseki Penda et H. Djellout [BD14]. En vue des applications biologiques, des données manquantes ont été autorisées par B. de Saporta, Gégout-Petit et Marsalle [dSGPM12]. V. Blandin [Bla13] notamment a considéré des coefficients aléatoires.

Aucune étude n'ayant été menée sur les transitions  $\mathcal{P}$  et  $\mathcal{Q}$  elles-mêmes, ou sur la mesure invariante  $\nu$ , les résultats apportés par le Chapitre II sont nouveaux dans ce contexte. L'estimation de la  $\mathbb{T}$ -transition ne permet pas l'identification des paramètres fonctionnels  $(f_0, f_1)$ , pour  $G$  quelconque. Contrairement au cas du modèle précédent, la réponse à la question (B.iii) est donc négative dans ce modèle. C'est pour cette raison que nous prolongeons l'étude du processus bifurquant auto-régressif dans le Chapitre III. Le cas de fonctions d'auto-régression linéaires,  $f_\iota(x) = \alpha_\iota + \beta_\iota x$  pour  $\iota \in \{0, 1\}$ , a été étudié dans les travaux cités précédemment, avec généralisation à des processus auto-régressifs bifurquants d'ordre quelconque. Nous nous plaçons au contraire dans le Chapitre III dans un cadre général, non-paramétrique.

## 4.2 SYNTHÈSE DES RÉSULTATS NOUVEAUX

### Résultats généraux

**Préliminaire probabiliste.** Nous détaillons ici les résultats probabilistes établis dans le Chapitre II, nous permettant dans un second temps de construire une procédure répondant aux questions (B.i) et (B.ii). Nous nous intéressons dans un premier temps aux déviations des moyennes empiriques par rapport à leurs limites en probabilité respectives identifiées par J. Guyon et rappelées précédemment. Nous supposons que la chaîne de transition  $\mathcal{Q}$  est géométriquement ergodique, et ce uniformément par rapport au point initial  $x \in \mathcal{S}$ ,

$$|\mathcal{Q}^m g(x) - \nu(g)| \leq R|g|_\infty \rho^m, \quad m \geq 0, \quad (22)$$

où  $R > 0$  est une constante et  $\rho$  un taux d'ergodicité suffisamment petit,

$$\rho < 1/2, \quad (23)$$

hypothèses essentielles pour nos techniques de preuve, et nous permettant d'atteindre le

**Résultat 3.** *Sous les hypothèses (22) et (23), l'inégalité de déviations suivante est valide,*

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} g(X_u, X_{u0}, X_{u1}) - \nu(\mathcal{P}g) \geq \delta\right) \leq \exp\left(\frac{-\kappa n^{-1} |\mathbb{T}_n| \delta^2}{\Sigma_n(g) + |g|_\infty \delta}\right)$$

où  $\kappa > 0$  ne dépend pas de  $n$  et  $\Sigma_n(g)$  est un terme de variance valant  $\Sigma_n(g) = |\mathcal{P}g|_1 + \min_{1 \leq \ell \leq n-1} (|\mathcal{P}g|_1^2 2^\ell + |\mathcal{P}g|_\infty^2 2^{-\ell})$ .

Les moyennes empiriques sur  $\mathbb{G}_n$ , l'ensemble des particules de la génération  $n$ , ont également été étudiées et nous obtenons des résultats du type précédent. Ces inégalités sont de type Bernstein et viennent améliorer les inégalités de type Hoeffding obtenues par S. V. Biteski Penda, H. Djellout et A. Guillin dans [BDG14]. L'étude fine du terme de variance  $\Sigma_n(g)$  est cruciale en vue des applications statistiques.

La différence fondamentale entre chaînes de Markov et leur généralisation en BMC qui intervient ici, tout comme dans les études précédentes, est liée au contrôle de la vitesse de convergence vers l'équilibre de la transition moyenne  $\mathcal{Q}$ . La majoration par 1/2 du taux  $\rho$  est cruciale pour le Résultat 3, alors qu'aucune condition de cette nature n'apparaît dans l'étude des chaînes classiques. La vérification d'une telle condition dans les applications peut être une tâche délicate. Nous

recommandons l'emploi des résultats de M. Hairer et J. Mattingly à cet effet, [HM11].

L'hypothèse d'uniformité par rapport au point initial pour le résultat d'ergodicité est cruciale dans notre méthode de démonstration mais restreint sensiblement les applications. Dans le cas du modèle de croissance-fragmentation, nous supposons le taux de division  $B$  à support borné. Dans le cas du modèle bifurquant auto-régressif, l'hypothèse (22) est vérifiée dès que les fonctions d'auto-régression sont bornées. Des techniques radicalement différentes devront être mises en œuvre pour s'affranchir de cette hypothèse (technique de couplage, utilisation de la distance de Wasserstein, voir S. V. Bitseki Penda *et al.* [BEG15]).

**Estimation statistique.** Le Chapitre II se consacre dans un second temps à l'estimation adaptative des densités de la mesure invariante  $x \rightsquigarrow \nu(x)$ , de la transition moyenne  $(x, y) \rightsquigarrow \mathcal{Q}(x, y)$  et de la  $\mathbb{T}$ -transition  $(x, y, z) \rightsquigarrow \mathcal{P}(x, y, z)$ , répondant ainsi aux questions (B.i) et (B.ii). Le Résultat 3 précédent est développé dans un cadre le plus général possible, mais nous nous restreignons à présent au cas où  $\mathcal{S} \subseteq \mathbb{R}$  et où  $\nu(dx)$ ,  $Q(x, dy)$  et  $\mathcal{P}(x, dydz)$  sont dominées par la mesure de Lebesgue (nous notons les densités correspondantes de la même façon).

Nous construisons des estimateurs de  $\nu$ ,  $\mathcal{Q}$  et  $\mathcal{P}$  à partir des observables  $(X_u, u \in \mathbb{T}_n)$  à l'aide de techniques d'estimation par seuillage dans une base d'ondelettes – voir D. Donoho *et al.* [DJKP95], G. Kerkycharian et D. Picard [KP00] ainsi que les références afférentes. Citons également l'ouvrage de W. Härdle, G. Kerkycharian, D. Picard et A. Tsybakov [HKPT98] (en particulier pour l'implémentation numérique). Les fonctions  $\nu(x)$ ,  $f_Q(x, y) = \nu(x)Q(x, y)$  et  $f_{\mathcal{P}}(x, y, z) = \nu(x)\mathcal{P}(x, y, z)$  sont approchées via les décompositions

$$\nu(x) = \sum_{\lambda} \langle \nu, \psi_{\lambda}^1 \rangle \psi_{\lambda}^1(x), \quad f_Q(x, y) = \sum_{\lambda} \langle f_Q, \psi_{\lambda}^2 \rangle \psi_{\lambda}^2(x, y), \quad f_{\mathcal{P}}(x, y, z) = \sum_{\lambda} \langle f_{\mathcal{P}}, \psi_{\lambda}^3 \rangle \psi_{\lambda}^3(x, y, z),$$

où  $(\psi_{\lambda}^d)_{\lambda}$  pour  $d = 1, 2, 3$  est une collection de fonctions ondelettes. Forts de l'approximation précédente (17), nous estimons  $\nu(x)$  par

$$\hat{\nu}_n(x) = \sum_{|\lambda| \leq J} \mathcal{T}_{\lambda, \eta}(\hat{\nu}_{\lambda, n}) \psi_{\lambda}^1(x) \quad \text{avec} \quad \hat{\nu}_{\lambda, n} = \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} \psi_{\lambda}^1(X_u)$$

où  $J$  est le niveau maximal de résolution,  $\mathcal{T}_{\lambda, \eta}$  est l'opérateur de seuillage dur et  $\eta > 0$  le niveau de seuillage. À l'aide de l'approximation (18), nous estimons  $\mathcal{Q}(x, y)$  par  $\hat{\mathcal{Q}}_n(x, y) = \frac{\hat{f}_n(x, y)}{\hat{\nu}_n(x)\sqrt{\varpi}}$  (avec  $\varpi > 0$ ), où

$$\hat{f}_n(x, y) = \sum_{|\lambda| \leq J} \mathcal{T}_{\lambda, \eta}(\hat{f}_{\lambda, n}) \psi_{\lambda}^2(x, y) \quad \text{avec} \quad \hat{f}_{\lambda, n} = \frac{1}{|\mathbb{T}_n^*|} \sum_{u \in \mathbb{T}_n^*} \psi_{\lambda}^2(X_{u^-}, X_u).$$

Enfin, nous fondant sur l'approximation (19), nous estimons  $\mathcal{P}(x, y, z)$  par  $\hat{\mathcal{P}}_n(x, y, z) = \frac{\hat{f}_n(x, y, z)}{\hat{\nu}_n(x)\sqrt{\varpi}}$ , où

$$\hat{f}_n(x, y, z) = \sum_{|\lambda| \leq J} \mathcal{T}_{\lambda, \eta}(\hat{f}_{\lambda, n}) \psi_{\lambda}^3(x, y, z) \quad \text{avec} \quad \hat{f}_{\lambda, n} = \frac{1}{|\mathbb{T}_{n-1}|} \sum_{u \in \mathbb{T}_{n-1}} \psi_{\lambda}^3(X_u, X_{u0}, X_{u1}).$$

Les performances de ces estimateurs sont évaluées à l'aide du résultat suivant.

**Résultat 4.** *Sous les hypothèses (22) et (23), pour des choix appropriés de  $J$  et  $\eta$ , la vitesse de convergence des estimateurs  $\widehat{\nu}_n$ ,  $\widehat{\mathcal{Q}}_n$  et  $\widehat{\mathcal{P}}_n$ , pour la perte  $L^p$  avec  $p \geq 1$ , uniformément sur toute boule de Besov, est, à termes logarithmiques près,  $|\mathbb{T}_n|^{-\alpha_d(s, \pi, p)}$  où  $\alpha_d(s, \pi, p)$  dépend de la régularité  $s$  de  $\nu$ ,  $\mathcal{Q}$  ou  $\mathcal{P}$  mesurée en norme  $L^\pi$ , ainsi que de la perte  $L^p$  choisie et de la dimension  $d$  du problème d'estimation ( $d = 1, 2, 3$  pour respectivement  $\nu, \mathcal{Q}, \mathcal{P}$ ) et s'exprime de façon classique,  $\alpha_d(s, p, \pi) = \min \left\{ \frac{s}{2s+d}, \frac{s+d(1/p-1/\pi)}{2s+d(1-2/\pi)} \right\}$ .*

*Cette vitesse de convergence est optimale au sens minimax, avec un terme logarithmique additif dans le régime dit sparse.*

L'article de référence concernant l'établissement de la borne supérieure du Résultat 4 est celui de G. Kerkycharian et D. Picard [KP00]. Le contrôle de  $\Sigma_n(\psi_\lambda^d)$  est crucial dans cette preuve. Concernant la borne inférieure, outre l'utilisation de techniques désormais classiques, nous généralisons à notre cadre bifurquant les techniques employées dans un cadre dépendant par M. Hoffmann, S. Cléménçon et C. Lacour [Hof99, Clé00, Lac07].

### Sur les modèles de croissance-fragmentation

L'estimation *adaptive* du taux de division dans un modèle de croissance-fragmentation, soit un modèle structuré en taille, est l'une des applications majeures de ce travail. Nous rappelons que les auteurs de [DHKR15] ont les premiers proposé une estimation du taux de division  $B$ , fonction de la taille de la cellule. Comme nous l'avons exposé en Section 2.4, la procédure est cependant *non-adaptive*.

Nous rappelant la représentation fondamentale (11) démontrée dans [DHKR15], nous proposons d'estimer  $B$  au point  $x \in \mathcal{D}$ , pour  $\mathcal{D}$  compact d'estimation, par

$$\widehat{B}_n(x) = \frac{\tau x}{2} \frac{\widehat{\nu}_n(x/2)}{\left( \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} \mathbf{1}_{\{x/2 \leq \xi_u < x\}} \right) \vee \varpi}, \quad (24)$$

avec  $\varpi > 0$  un seuil garantissant que l'estimateur est bien défini. Les propriétés de convergence de  $\widehat{\nu}_n$  énoncées dans le Résultat 4 se transmettent à  $\widehat{B}_n$ , le dénominateur de (24) convergeant à vitesse paramétrique, et nous démontrons le

**Résultat 5.** *Pour une classe de taux de division  $B$  choisie de manière à satisfaire les hypothèses (22) et (23) uniformément, classe intersectée avec une boule de Besov contenant les taux de division  $B$  de régularité  $s$  mesurée en norme  $L^\pi$ , nous avons*

$$\sup_B \left( \mathbb{E} \left[ \|\widehat{B}_n - B\|_{L^p(\mathcal{D})}^p \right] \right)^{1/p} \lesssim \left( \frac{\log |\mathbb{T}_n|}{|\mathbb{T}_n|} \right)^{\alpha_1(s, p, \pi)},$$

avec  $\alpha_1(s, p, \pi) = \min \left\{ \frac{2s}{2s+1}, \frac{s+1/p-1/\pi}{2s+1-2/\pi} \right\}$ . D'autre part, selon le régime, dense ou sparse, nous avons

$$\inf_{\widehat{B}_n} \sup_B \left( \mathbb{E} \left[ \|\widehat{B}_n - B\|_{L^p(\mathcal{D})}^p \right] \right)^{1/p} \gtrsim |\mathbb{T}_n|^{-\alpha_1(s, p, \pi)} \quad \text{ou} \quad \left( \frac{\log |\mathbb{T}_n|}{|\mathbb{T}_n|} \right)^{\alpha_1(s, p, \pi)},$$

où l'infimum porte sur tous les estimateurs  $\widehat{B}_n$  construits à partir de  $(X_u, u \in \mathbb{T}_n)$ .

Ce résultat apporte une réponse à la question (A.i) formulée précédemment puisque l'estimateur  $\widehat{B}_n$  du taux de division, fonction de la taille, que nous proposons est adaptatif. Outre

l'adaptation que nous atteignons grâce à cette procédure, nous généralisons concomitamment la perte  $L^2$  en une perte  $L^p$  et la classe fonctionnelle de Hölder en celle de Besov, dans le contrôle d'erreur (12). Le prix pour satisfaire l'hypothèse (22) dans ce modèle est de restreindre l'espace d'état  $\mathcal{S}$  à un ensemble borné.

### Sur les modèles bifurquants auto-régressifs non-linéaires

Dans le Chapitre III, l'intérêt porte directement sur les fonctions d'auto-régression  $f_0$  et  $f_1$  du modèle bifurquant auto-régressif décrit par (20), et non plus sur les caractéristiques du fragment marqué de transition  $\mathcal{Q}$  et mesure invariante  $\nu$  ou sur la  $\mathbb{T}$ -transition  $\mathcal{P}$  elle-même (définie par (21)). Dans ce travail nous conduisons une étude non-paramétrique, c'est-à-dire que nous ne spécifions aucune forme *a priori* pour les fonctions d'auto-régression, contrairement aux études passées sus-citées, qui considèrent des fonctions d'auto-régression linéaires.

Nous proposons d'estimer  $(f_0(x), f_1(x))$  les fonctions d'auto-régression au point  $x$  à partir des observations  $(X_u, u \in \mathbb{T}_{n+1})$ , caractéristiques des particules des  $n + 1$  premières générations, par

$$\left( \widehat{f}_{\iota,n}(x) = \frac{\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_h(x - X_u) X_{uu}}{\left( \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_h(x - X_u) \right) \vee \varpi}, \iota \in \{0, 1\} \right).$$

où  $h > 0$  est une fenêtre de lissage bien choisie,  $K_h(\cdot) = h^{-1}K(h^{-1}\cdot)$  avec  $K : \mathbb{R} \rightsquigarrow \mathbb{R}$  un noyau et  $\varpi > 0$ . Nous mentionnons les inégalités de déviations pour ces estimateurs venant comme application dans le travail récent de S. V. Bitseki Penda, M. Escobar-Bach et A. Guillin [BEG15]. Nous étudions dans le Chapitre III le comportement à la fois non-asymptotique et asymptotique de ces estimateurs de type Nadaraya-Watson. Nous obtenons la majoration du risque  $\sup_{(f_0, f_1)} \mathbb{E}[(\widehat{f}_{0,n}(x) - f_0(x))^2 + (\widehat{f}_{1,n}(x) - f_1(x))^2]$  par  $|\mathbb{T}_n|^{\frac{-2s}{2s+1}}$ , à constante près, pour une fenêtre de lissage  $h_n$  bien calibrée, en  $|\mathbb{T}_n|^{-1/(2s+1)}$ , avec  $s$  la régularité hölderienne des fonctions  $f_0$  et  $f_1$  (régularités supposées égales pour simplifier), où le supremum est pris sur une classe appropriée de fonctions  $(f_0, f_1)$ , sous-linéaires. L'hypothèse de sous-linéarité des fonctions  $f_0$  et  $f_1$ , assurant une propriété d'ergodicité non uniforme en le point initiale (moins contraignante que (22)), est ici suffisante dans nos démonstrations.

L'étude asymptotique comprend la convergence presque sûre de  $\widehat{f}_{0,n}(x)$  et  $\widehat{f}_{1,n}(x)$  ainsi que leur normalité asymptotique. Les estimateurs, recentrés et normalisés par  $(|\mathbb{T}_n|h_n)^{1/2}$ , présentent un comportement asymptotiquement normal. Si  $h_n = |\mathbb{T}_n|^{-\alpha}$  avec  $\alpha > 1/(2s + 1)$ , la loi limite est centrée, mais ce choix de fenêtre ne correspond pas au choix optimal précédent. En revanche, si  $h_n = |\mathbb{T}_n|^{-1/(2s+1)}$ , la loi limite est décentrée par un biais, inconnu, d'expression que nous explicitons. Pour finir, l'étude asymptotique est renforcée par un principe de déviations modérées portant sur  $\widehat{f}_{0,n}(x)$  et  $\widehat{f}_{1,n}(x)$ . La démonstration du principe de déviations modérées repose sur une inégalité de déviations, telle celle énoncée dans le Résultat 3 (et les fonctions  $f_0$  et  $f_1$  doivent pour ce résultat, pour celui-ci seulement, être supposées bornées, afin de satisfaire l'hypothèse (22)).

La construction d'une procédure adaptative en la régularité  $s$  des fonctions d'auto-régression pourrait être envisagée via un protocole utilisant des ondelettes adaptées au design (voir V. Delouille et R. von Sachs, [DvS05]) ou une méthode de type Goldenschluger-Lepski (voir A. Goldenschluger et O. Lepski [GL11]). Nous pourrions également généraliser le modèle pour inclure de

l'hétéroscédasticité via deux fonctions de variance  $\sigma_0(\cdot)$  et  $\sigma_1(\cdot)$  intervenant dans les équations d'auto-régression. Dans le cas d'un processus auto-régressif classique (non bifurquant) l'estimation adaptative de la fonction d'auto-régression ainsi que celle de la variance sur des espaces de Besov a été réalisée par M. Hoffmann [Hof99].

Nous nous intéressons également à l'asymétrie éventuelle entre les équations auto-régressives et nous proposons un test d'égalité entre  $f_0$  et  $f_1$ . Plus précisément, pour  $x_1, \dots, x_k$  points distincts, nous construisons un test asymptotique visant à distinguer

$$\mathcal{H}_0 : \forall l \in \{1, \dots, k\}, f_0(x_l) = f_1(x_l) \quad \text{de} \quad \mathcal{H}_1 : \exists l \in \{1, \dots, k\}, f_0(x_l) \neq f_1(x_l).$$

L'analyse de la normalité asymptotique précédente, réalisé minutieusement selon la valeur de la fenêtre  $h_n$ , permet de construire aisément, en suivant H. J. Bierens [Bie97], deux nouveaux estimateurs,  $\bar{f}_{0,n}(x)$  et  $\bar{f}_{1,n}(x)$  de  $f_0(x)$  et  $f_1(x)$ , satisfaisant

$$|\mathbb{T}_n|^{\frac{s}{2s+1}} \begin{pmatrix} \bar{f}_{0,n}(x) - f_0(x) \\ \bar{f}_{1,n}(x) - f_1(x) \end{pmatrix} \xrightarrow{d} \mathcal{N}_2(\mathbf{0}_2, \Sigma_2(x)),$$

où la matrice de variance-covariance asymptotique  $\Sigma_2(x)$  est obtenue explicitement en fonction de  $\|K\|_2$  la norme  $L^2$  du noyau  $K$ ,  $\sigma_0^2 = \mathbb{E}[\varepsilon_{u0}^2]$ ,  $\sigma_1^2 = \mathbb{E}[\varepsilon_{u1}^2]$  et  $\varrho = \sigma_0^{-1}\sigma_1^{-1}\mathbb{E}[\varepsilon_{u0}\varepsilon_{u1}]$ . La construction du test est fondée sur ces estimateurs et nous obtenons le résultat suivant :

**Résultat 6.** *Construisons la statistique de test*

$$W_n(x_1, \dots, x_k) = \frac{|\mathbb{T}_n|^{\frac{2s}{2s+1}}}{(\sigma_0^2 + \sigma_1^2 - 2\sigma_0\sigma_1\varrho)\|K\|_2^2} \sum_{l=1}^k \hat{\nu}_n(x_l) (\bar{f}_{0,n}(x_l) - \bar{f}_{1,n}(x_l))^2$$

où  $\hat{\nu}_n(\cdot) = |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} K_{h_n}(\cdot - X_u)$ , avec  $h_n = |\mathbb{T}_n|^{-1/(2s+1)}$ .

Quand  $n \rightarrow \infty$ , sous  $\mathcal{H}_0$ ,  $W_n(x_1, \dots, x_k)$  converge en loi vers la loi du chi-2 à  $k$  degrés de liberté tandis que, sous  $\mathcal{H}_1$ , elle tend presque-sûrement vers  $+\infty$ .

Une étude par simulations numériques montre d'assez bonnes performances, à la fois pour la procédure d'estimation et pour le test proposé. Une étude de la transmission du taux de croissance chez la bactérie *E. coli*, dans le prolongement des travaux déjà cités – [DM10, dSGPM12] par exemple –, semble donc une perspective intéressante. D'un point de vue théorique, il serait au préalable souhaitable de construire un test d'égalité des fonctions d'auto-régression sur un intervalle, et non sur une grille de points. Cela requiert des résultats asymptotiques uniformes sur un intervalle.

## 5 TROISIÈME PARTIE : ÉTUDE COMPARATIVE DE POPULATIONS

Le Chapitre IV, que nous résumons ici, s'intégrant dans l'axe (C), est résolument tourné vers les applications. Il est consacré à l'étude d'un problème directement issu de la biologie, formulé par Lydia Robert (INRA). Les résultats que nous obtenons serviront également à calibrer des expériences futures.

## 5.1 QUESTION BIOLOGIQUE

L'enjeu est ici de prendre en compte la variabilité individuelle entre cellules. La croissance de la taille des cellules, dans certains milieux contrôlés, se fait exponentiellement selon un taux de croissance pouvant varier d'un individu à un autre, comme cela a été établi expérimentalement par Kiviet *et al.* [KNW+14].

Nous considérons alors une généralisation du modèle structuré en taille précédent. Rappelons que dans celui-ci l'évolution déterministe du trait biologique pris en considération suit une dynamique commune à l'ensemble des cellules : la taille d'une cellule augmente exponentiellement avec un taux fixé, unique au sein de la population. Ce modèle peut être étendu au cas où la taille de chaque cellule croît avec un taux qui lui est propre : le taux de croissance doit alors être ajouté comme une variable structurante supplémentaire, comme cela a été fait par M. Doumic *et al.* dans l'article [DHKR15]. Ainsi chaque cellule suivra une évolution déterministe de forme fixée (exponentielle par exemple) mais avec un taux qui lui est propre. Remarquons que cette nouvelle variable se distingue des précédentes, taille ou âge, dans la mesure où elle n'évolue pas au cours du temps, mais est donnée une fois pour toute à la naissance.

L'objectif que nous nous fixons ici est d'étudier l'influence de la variabilité individuelle du taux de croissance sur l'évolution globale de la population, de façon théorique dans un premier temps, expérimentale ensuite :

**(C.i)** La variabilité entre cellules à l'échelle microscopique pénalise-t-elle la croissance globale de la population au niveau macroscopique ? Ou bien est-elle en mesure de la favoriser ?

**(C.ii)** Compte tenu des possibilités expérimentales, quelle procédure mettre en place pour l'estimation du paramètre de Malthus, quantifiant la croissance exponentielle de la population ?

Concernant la question **(C.i)**, l'opinion dominante, fondée biologiquement, tranche en faveur d'une pénalisation, la variabilité individuelle devant uniquement permettre la survie d'une population de cellules soumises à des conditions extrêmes. Elle n'a pourtant jamais été validée ou invalidée mathématiquement. Nous souhaitons établir une comparaison entre les vitesses de croissance des deux populations de cellules suivantes :

- 1) Une **population de référence** composée de cellules possédant toutes le même taux de croissance  $\bar{v}$ .
- 2) Une **population avec variabilité** où chaque cellule possède son propre taux de croissance, le taux de croissance moyen des cellules de la population étant maintenu égal à  $\bar{v}$ .

Précisons que le taux de division est supposé de même forme pour ces deux populations.



Une étude théorique et une étude numérique sont conduites conjointement dans le Chapitre IV. Il s'agira dans un second temps d'établir des protocoles expérimentaux, en collaboration avec Lydia Robert (INRA), pour confronter ces avancées mathématiques à l'expérience. Nous souhaitons obtenir des réponses concernant, en particulier, le modèle structuré en taille et incrément de taille, modèle introduit récemment pour l'étude de la bactérie *E. coli* par S. Taheri-Araghi *et al.* [TAB+15] décrit ci-dessous (voir aussi Hall, Wake et Gandar [HWG91]), auquel nous ajoutons de la variabilité dans le taux de croissance. Le Chapitre IV traite également du modèle structuré en taille ainsi que d'un modèle récent introduit par A. Amir [Ami14].

## 5.2 RÉSULTATS THÉORIQUES SUR LE MODÈLE JOUET EN ÂGE AVEC VARIABILITÉ

Un modèle avec variabilité jouet est tout d'abord considéré. Une réponse analytique à la question de départ (C.i) est alors atteignable. Dans le modèle simple structuré en âge, nous autorisons la variabilité individuelle à travers un taux de vieillissement propre à chaque cellule (en lieu et place d'un taux de croissance individuel). L'âge ne désigne donc plus ici le temps écoulé depuis la naissance mais un âge physiologique, proportionnel seulement au temps écoulé depuis la naissance. Une cellule  $u$  vieillit selon un taux de vieillissement qui lui est propre  $\tau_u$ , hérité de son parent  $u^-$ ,

$$\mathbb{P}(\tau_u \in dv' | \tau_{u^-} = v) = \rho(v, v') dv' \quad (25)$$

et son âge physiologique au temps  $t \in [b_u, b_u + \zeta_u)$  est donné par  $\Delta_u^t = \tau_u(t - b_u)$ , où  $b_u$  est la date de naissance de  $u$ . L'âge physiologique à la division de la cellule  $u$  vaut

$$\Delta_u = \tau_u \zeta_u$$

où  $\zeta_u$  désigne toujours la durée de vie de  $u$ . Nous supposons que les âges physiologiques  $(\Delta_u, u \in \mathbb{T})$  des cellules de la population sont indépendants et identiquement distribués selon

$$\mathbb{P}(\Delta_u \in da) = B(a) \exp\left(-\int_0^a B(s) ds\right) da = f_B(a) da. \quad (26)$$

Le taux de division  $B$  ne caractérise plus à lui seul le modèle, la distribution  $\rho$  des taux de vieillissement est également déterminante. Le cas du modèle structuré en âge sans variabilité présenté en Section 3 se retrouve à travers le choix  $\rho(v, dv') = \delta_1(dv')$ , masse de Dirac au point 1. Dans ce cas, âge physiologique et temps écoulé depuis la naissance coïncident,  $\Delta_u = \zeta_u$ .

Les deux caractéristiques individuelles structurant le modèle sont ici l'âge physiologique  $a \geq 0$  et le taux de vieillissement  $v > 0$ . Considérons le pendant déterministe du modèle, donné par l'équation d'évolution

$$\begin{cases} \frac{\partial}{\partial t} n(t, a, v) + v \frac{\partial}{\partial a} n(t, a, v) + \gamma(a, v) n(t, a, v) = 0, \\ vn(t, a = 0, v) = 2 \iint_{\mathcal{S}} \gamma(a, v) n(t, a, v') \rho(v, v') dv' da, \quad n(t = 0, a, v) = n^{\text{in}}(a, v), \end{cases} \quad (27)$$

où  $n(t, a, v)$  représente la quantité de cellules d'âge physiologique  $a$  et de taux de vieillissement  $v$  au temps  $t$ . Cette équation peut se lire comme un bilan de masse, de façon analogue à l'équation (6)

sans variabilité. Elle peut être vue comme l'équation satisfaite par la mesure empirique moyenne (voir le Chapitre IV), ce qui permet de faire le lien avec l'approche stochastique précédente. Le taux de division, noté ici  $\gamma$ , dépend en toute généralité des deux variables  $a$  et  $v$ .

Le modèle décrit par (27) s'inscrit dans la lignée du modèle introduit par M. Rotenberg [Rot83], étudié par S. Mischler, B. Perthame et L. Ryzhik [MPR02]. La description donnée par (27) présente la généralisation suivante : l'hérédité dans la transmission du taux de vieillissement (via le noyau markovien  $\rho$ ) est autorisée. Nous nous intéressons dans un premier temps aux vecteurs propres (direct et adjoint) associés à l'équation d'évolution (27).

**Résultat 7.** *Pour  $\gamma$  et  $\rho$  suffisamment réguliers, il existe un unique triplet  $(\lambda_{\gamma,\rho}, N_{\gamma,\rho}, \phi_{\gamma,\rho})$  résolvant les problèmes (direct et adjoint) aux valeurs propres associés à (27), avec  $\lambda_{\gamma,\rho} > 0$ ,  $N_{\gamma,\rho}$  et  $\phi_{\gamma,\rho}$  différentiables en la première variable  $a$  et continues en la seconde variable  $v$ .*

Le paramètre de Malthus  $\lambda_{\gamma,\rho}$ , correctement défini par le Résultat 7, intervient dans l'approximation, pour  $t$  grand,  $n(t, a, v) \approx e^{\lambda_{\gamma,\rho} t} N_{\gamma,\rho}(a, v)$ , où  $N_{\gamma,\rho}$  est le profil stationnaire (voir l'approximation 3), Section 2.2). L'approximation doit être lue pour une norme pondérée par le vecteur propre adjoint  $\phi_{\gamma,\rho}$ .

Le taux de division  $\gamma(a, v)$  est à présent choisi comme  $\gamma(a, v) = vB(a)$  et cela se traduit d'un point de vue stochastique par la propriété (26). Nous notons à présent le paramètre de Malthus  $\lambda_{B,\rho}$  pour souligner la forme particulière du taux de division  $\gamma$ . Il se trouve être caractérisé de façon simple, comme nous le mentionnons par la suite. Nous négligeons l'hérédité éventuelle dans la transmission du taux de vieillissement, ce qui revient à choisir  $\rho(v, v') dv' = \rho(v') dv'$  dans (25).

Pour un taux de division  $B$  commun, l'objectif est de comparer l'évolution des deux populations suivantes – évolution mesurée par leur paramètre de Malthus :

- 1) Population de référence.** Toutes les cellules vieillissent au taux de vieillissement  $\bar{v}$  (i.e.  $\rho(dv) = \delta_{\bar{v}}(dv)$ ). Le paramètre de Malthus associé à cette population, noté  $\lambda_{B,\bar{v}}$ , est solution de

$$2 \int_0^\infty \exp\left(-\frac{\lambda_{B,\bar{v}} a}{\bar{v}}\right) f_B(a) da = 1.$$

- 2) Population avec variabilité.** Les taux de vieillissement des cellules sont indépendamment distribués selon une densité  $\rho(v) dv$  non-dégénérée telle que

$$\int_{\mathcal{V}} v \rho(v) dv = \bar{v}.$$

Le paramètre de Malthus  $\lambda_{B,\rho}$  mesurant la vitesse de croissance de cette population est solution de

$$2 \iint_{\mathcal{S}} \exp\left(-\frac{\lambda_{B,\rho} a}{v}\right) f_B(a) \rho(v) dv da = 1.$$

Pour une distribution  $\rho$  fixée, de moyenne  $\bar{v}$  fixée, la comparaison de  $\lambda_{B,\rho}$  à  $\lambda_{B,\bar{v}}$  selon les propriétés de  $B$  est l'objet du résultat suivant, répondant à la question (C.i).

**Résultat 8.** Pour les choix  $\gamma(a, v) = vB(a)$  et  $\rho(v, v') = \rho(v')$  dans (27), sous l'hypothèse que le taux de division  $B$  est différentiable,

- (i) Si la densité  $f_B$  est décroissante, alors  $\lambda_{B, \rho} < \lambda_{B, \bar{v}}$ .
- (ii) Si la densité  $f_B$  est croissante bornée, alors  $\lambda_{B, \rho} > \lambda_{B, \bar{v}}$ .

Comme le prouve le Résultat 8, la variabilité individuelle dans le taux de vieillissement peut pénaliser (cas (i)) ou avantager (cas (ii)) la croissance d'une population, selon le mécanisme de division sous-jacent, *i.e.* selon la forme de la densité  $f_B$ .

De façon complémentaire, pour une suite de densités  $\rho_\alpha$  de moyenne  $\bar{v}$  convergeant (faiblement) vers la masse de Dirac au point  $\bar{v}$  quand  $\alpha \rightarrow 0$ , les perturbations à l'ordre 2 de  $\alpha \rightsquigarrow \lambda_{B, \rho_\alpha}$  autour de  $\lambda_{B, \bar{v}}$  ont également été étudiées. Un tel résultat repose sur une technique introduite par P. Michel [Mic06] qui passe par une étude plus précise des vecteurs propres associés à  $\lambda_{B, \rho_\alpha}$ .

### 5.3 RÉSULTATS NUMÉRIQUES SUR LE MODÈLE EN TAILLE ET INCRÉMENT DE TAILLE AVEC VARIABILITÉ

Nous nous tournons à présent vers un modèle plus à même de décrire la croissance d'une population de bactéries *E. coli*.

#### *Descriptions stochastique et déterministe*

Nous travaillons ici sur un modèle qui ressemble au modèle structuré en taille déjà décrit, sensiblement enrichi. La croissance des cellules est toujours supposée exponentielle et chaque cellule possède son propre taux de croissance  $\tau_u$ . Nous autorisons par la suite la transmission héréditaire de ce taux de croissance à travers le noyau markovien  $\rho(v, v')dv' = \mathbb{P}(\tau_{u^-} \in dv' | \tau_u = v)$  où  $u^-$  désigne le parent de la cellule  $u$ . Une cellule de taille à la naissance  $\xi_u$  se divise, après une durée de vie  $\zeta_u$ , en deux cellules de tailles identiques égales à  $\frac{1}{2}\xi_u \exp(\tau_u \zeta_u)$ . La quantité nouvelle importante est l'incrément de taille, différence entre la taille courante de la cellule et sa taille à la naissance. À la division, l'incrément de taille de la cellule  $u$  prend la valeur

$$\Delta_u = \xi_u(e^{\tau_u \zeta_u} - 1).$$

Notons que l'incrément de taille est assimilable à un âge physiologique (puisque cet incrément est nul pour toute cellule naissante). Il est fondé biologiquement de supposer que les incréments de tailles à la division des cellules sont mutuellement indépendants et indépendants des autres quantités du modèle (taille à la naissance, taux de croissance),

$$\mathbb{P}(\Delta_u \in ds) = f_B(s)ds, \tag{28}$$

voir [TAB+15] (Figures S6, S8 et S12). Le modèle est entièrement déterminé par le taux de division  $B$  et par le noyau markovien  $\rho$ .

D'un point de vue déterministe, ce modèle se traduit par une équation structurée en trois

variables que sont la taille  $x > 0$ , l'incrément de taille  $a \geq 0$  et le taux de croissance  $v > 0$ ,

$$\begin{cases} \frac{\partial}{\partial t} n(t, x, a, v) + \frac{\partial}{\partial x} (vx n(t, x, a, v)) + \frac{\partial}{\partial a} (vx n(t, x, a, v)) + \gamma(x, a, v) n(t, x, a, v) = 0, \\ v' x n(t, x, a = 0, v') = 4 \int_0^\infty \gamma(2x, a) n(t, 2x, a, v) \rho(v, v') da dv, \\ n(t = 0, x, a, v) = n^{\text{in}}(x, a, v), \end{cases} \quad (29)$$

où  $n(t, x, a, v)$  représente la quantité de cellules de caractéristiques  $(x, a, v)$  au temps  $t$ . Le choix du taux de division  $\gamma(x, a, v) = vx B(a)$  permet d'obtenir la propriété fondamentale (28). Le problème (29) ne permet pas une caractérisation simple de la valeur propre  $\lambda_{B, \rho}$  pouvant aboutir à un résultat du type du Résultat 8, par les mêmes techniques de démonstration. Le cadre stochastique peut suppléer à l'analyse déterministe et nous permettre d'obtenir des résultats préliminaires à la question (C.i).

### Synthèse des résultats nouveaux

**Modélisation de l'hérédité.** Nous autorisons, contrairement au modèle jouet précédent, la transmission héréditaire du trait de variabilité. L'hérédité dans la transmission du taux de croissance peut être modélisé par un processus bifurquant auto-régressif (nous référons au Chapitre IV pour une seconde modélisation) : toute cellule  $u \in \mathbb{T}$ , de taux de croissance  $\tau_u$ , donne naissance à deux cellules dont les taux de croissance respectifs sont donnés par

$$\tau_{u0} = (1 - \kappa)\bar{v} + \kappa\tau_u + \eta_{u0} \quad \text{and} \quad \tau_{u1} = (1 - \kappa)\bar{v} + \kappa\tau_u + \eta_{u1}, \quad (30)$$

où  $(\eta_{u0}, u \in \mathbb{T})$  et  $(\eta_{u1}, u \in \mathbb{T})$  sont deux suites de bruit, de même loi centrée, indépendantes, ce qui revient à négliger l'éventuelle corrélation directe qui pourrait exister entre les taux de croissance  $\tau_{u0}$  et  $\tau_{u1}$  de deux cellules sœurs. Le processus BAR est par ailleurs choisi d'une forme simple : symétrique d'une part, linéaire d'autre part. Le coefficient  $\kappa \in (0, 1)$  mesure la force de l'hérédité, le cas sans hérédité aucune correspondant à  $\kappa = 0$ . À l'équilibre,  $\mathbb{E}[\tau_u] = \bar{v}$  et  $\mathbb{E}[(\tau_u - \bar{v})^2] = \eta^2(1 - \kappa^2)^{-1}$ , où  $\eta^2$  est le niveau de bruit. Le coefficient de variation de la distribution des taux de croissance à l'équilibre

$$CV_\kappa = \frac{\eta(1 - \kappa^2)^{-1/2}}{\bar{v}}$$

est une quantification naturelle de la force de la variabilité : plus  $CV_\kappa$  est grand, plus les taux de croissance au sein de la population sont dispersés.

**Protocole numérique.** Nous paramétrons le modèle par le taux de division  $B$  et le paramètre  $\kappa$  (la loi du bruit  $(\eta_u, u \in \mathbb{T})$  étant un paramètre de nuisance). Nous notons  $\lambda_{B, \kappa}$  le paramètre de Malthus relatif à  $B$  et  $\kappa$ . L'observation de  $(\xi_u^{T/2}, u \in \partial\mathcal{T}_{T/2})$  et de  $(\xi_u^T, u \in \partial\mathcal{T}_T)$ , où  $\xi_u^t$  est la taille à la date  $t$  de la cellule  $u$  (voir (2) avec  $\tau = \tau_u$ ), permet l'estimation de  $\lambda_{B, \kappa}$  via l'approximation 2) de la Section 2.2. En effet, en notant  $\mathfrak{M}_t$  la somme des tailles courantes de toutes les cellules

vivantes à l'instant  $t$ , cette approximation, nous permet d'utiliser l'estimateur

$$\hat{\lambda}_T = \frac{2}{T} \ln \left( \frac{\mathfrak{M}_T}{\mathfrak{M}_{T/2}} \right).$$

La simulation d'un arbre en temps continu de 0 à  $T$  permet donc l'estimation de  $\lambda_{B,\kappa}$ , et la simulation de  $M$  arbres en temps continu de 0 à  $T$  permet l'obtention d'un intervalle de confiance estimé.

**Conclusions numériques.** À l'aide du protocole précédent, pour un taux de division  $B$  fixé, expérimentalement plausible, pour différents  $\kappa$  fixés, nous évaluons la courbe  $CV_\kappa \rightsquigarrow \lambda_{B,\kappa}$ , où l'échelle des abscisses mesure la dispersion des taux de croissance au sein de la population. La Figure 3 est une synthèse des résultats numériques obtenus :

- 1) **En l'absence d'hérédité.** La courbe correspondant au cas  $\kappa = 0$  est, d'une part, significativement en-deçà de la valeur de référence ici égale à 1 et, d'autre part, significativement décroissante. Cela signifie donc que, d'une part, la variabilité dans le taux de croissance pénalise la croissance et que, d'autre part, cette pénalité est d'autant plus grande que les taux de croissance sont dispersés au sein de la population, ce qui conforte l'induite biologique.
- 2) **En présence d'hérédité.** Pour une hérédité relativement faible ( $\kappa = 0.25$ ) les deux conclusions précédentes demeurent valables. Pour une hérédité plus forte ( $\kappa = 0.50$ ), le paramètre de Malthus semble maintenu constant égal à la valeur de référence 1, quelle que soit la variabilité dans le taux de croissance. Pour une valeur fixée de  $CV_\kappa$ , *i.e.* un niveau de variabilité fixé, nous observons une croissance significative de  $\lambda_{B,\kappa}$  avec la force de l'hérédité mesurée par  $\kappa$ . Pour résumer, nous inférons

$$\lambda_{B,\kappa=0} < \lambda_{B,\kappa=0.25} < \lambda_{B,\kappa=0.50} \approx 1,$$

où 1 est ici la valeur de référence. Pour un  $\kappa$  plus grand nous sommes tentés de penser qu'il est possible que  $\lambda_{B,\kappa}$  dépasse  $\bar{v} = 1$ . Cependant, les essais pour  $\kappa = 0.75$  que nous avons réalisés se montrent rien de significatif par rapport au cas  $\kappa = 0.50$ .

La variabilité peut donc pénaliser ou non la croissance globale de la population. Une seconde modélisation de l'hérédité dans le taux de croissance, passant par la distinction de deux sous-populations, l'une avec un taux de croissance élevé, l'autre avec un taux de croissance faible, montre que la croissance globale de la population peut également être accélérée. Cette conclusion est remarquable dans la mesure où elle va à l'encontre des intuitions biologiques.

Par ailleurs, notre étude indique que les variations du paramètre de Malthus pourront être mesurées expérimentalement.

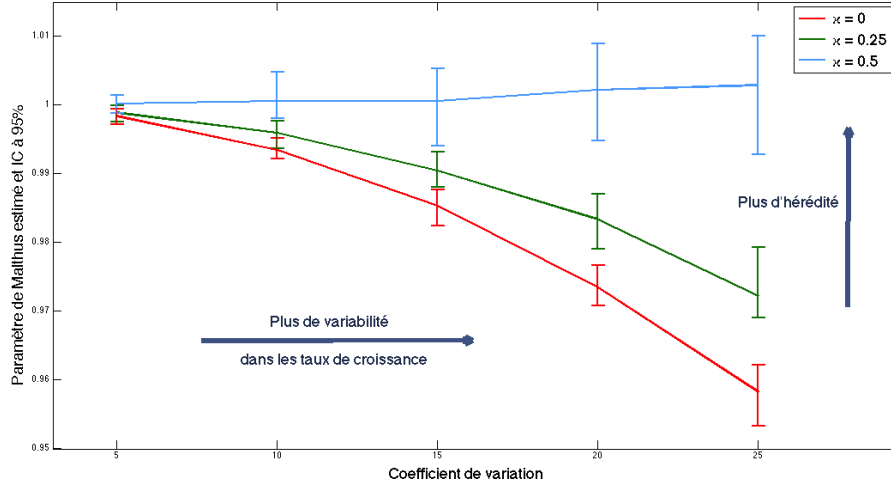


FIGURE 3 – Modèle structuré en taille et incrément de taille avec variabilité dans le taux de croissance. Taux de division  $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$ , les taux de croissance  $(\tau_u, u \in \mathbb{T})$  suivent la dynamique auto-régressive (30). Courbe estimée  $CV_\kappa \rightsquigarrow \lambda_{B,\kappa}$  (moyenne et intervalle de confiance à 95% calculés à partir de  $M = 50$  arbres en temps continu) pour différentes valeurs de  $\kappa$ . Courbe rouge (bas) :  $\kappa = 0$  (pas d'hérédité). Courbe verte (milieu) :  $\kappa = 0.25$ . Courbe bleue (haut) :  $\kappa = 0.50$ .

## 6 PERSPECTIVES

Outre les perspectives déjà mentionnées, concernant les modèles structurés en âge ou en taille (voir Table 1), deux autres axes majeurs de travail se dégagent.

### 6.1 ESTIMATION DU PARAMÈTRE DE MALTHUS À PARTIR D'UNE BRANCHE

Une quantité centrale, aussi bien pour le biologiste que pour le mathématicien, nous l'avons vu, est le paramètre de Malthus, qui caractérise la croissance exponentielle de la population de cellules en temps long. La question (C.ii) formulée précédemment (page 27) requiert encore des approfondissements : l'enjeu est de proposer une procédure fiable d'estimation du paramètre de Malthus, en adéquation avec les possibilités offertes par les protocoles expérimentaux.

Des estimateurs naturels sont fondés sur l'observation de l'évolution de toute la population en des temps grands (choisis précédemment comme  $T/2$  et  $T$  par simplicité). Cependant, d'un point de vue expérimental, l'observation de l'intégralité de la population de cellules reste difficile : elle ne peut se faire que pour 9 générations au plus (voir Stewart *et al.* [SMPT05]), soit pour des populations d'au plus 500 cellules. Constituer un échantillon de taille plus importante nécessite de répéter cette expérience : une dizaine de fois est raisonnable, ce qui conduit, par agrégation, à l'observation de  $n = 5\,000$  cellules environ. L'estimation du paramètre de Malthus qui en résulte peut être trop médiocre dès lors qu'il s'agit de comparer différentes populations à travers ce paramètre.

L'évolution récente de la technologie permet à présent d'accéder à l'observation de lignées très longues de cellules – voir l'étude expérimentale par Wang *et al.* [WRP+10]. En effet, jusqu'à 150 générations peuvent être observées le long d'une branche de l'arbre généalogique de la population.

L'observation de mille lignées est ici réalisable, ce qui permet de constituer, par agrégation, un échantillon de grande taille, de  $n = 150\,000$  cellules environ. Nous souhaitons alors conduire l'estimation du paramètre de Malthus à partir d'un tel schéma d'observation. Intuitivement, ceci est réalisable dès lors qu'il est possible d'estimer le taux de division  $B$  à partir de l'observation d'une branche de  $n$  cellules. En effet,  $\lambda_B$  est une fonctionnelle de  $B$ , définie comme la valeur propre dominante du problème aux valeurs propres associé à l'équation d'évolution décrivant macroscopiquement la croissance de la population (voir par exemple les problèmes aux valeurs propres (8) et (9) associés respectivement aux équations d'évolution (6) et (7)). Le problème relève alors de l'estimation du rayon spectral d'un opérateur continu bruité.

## 6.2 PROBLÈME INVERSE DANS LE MODÈLE TAILLE-INCRÉMENT

Considérons le modèle structuré en taille et incrément de taille décrit par (29), en négligeant la variabilité par le choix  $\rho(dv) = \delta_\tau(dv)$  pour  $\tau$  fixé, connu. Dans ce cadre un problème particulièrement intéressant est le suivant. Il s'agit de reconstruire le taux de division,  $B$ , comme fonction de l'incrément de taille, à partir de l'observation de la taille de cellules vivantes à un instant  $T$ . Le cadre inférentiel est donc celui approchant l'observation en  $T$ .

Pour le modèle structuré en taille uniquement, avec un taux de division  $B$  dépendant de la taille, ce problème a été entièrement traité (voir Section 2.4). Nous soulignons que dans le problème que nous nous proposons d'étudier, les incréments de taille des cellules, dont dépend le taux de division, ne sont pas observés (seules les tailles le sont) et c'est ce qui constitue sa nouveauté. Une étude préliminaire montre que nous aboutissons à un problème de déconvolution.

L'utilisation biologique d'une telle procédure d'estimation pourrait être la suivante. Les modèles que nous avons décrits jusqu'ici sont essentiellement descriptifs, avant d'être explicatifs. D'ailleurs, pour établir des comparaisons entre modèles, nous regardons si un modèle reproduit mieux les faits observés (distributions de tailles à la naissance, corrélations entre grandeurs mesurées, *etc.*) que l'autre. Ainsi, par exemple, L. Robert *et al.* [RHK+14] établissent que le modèle structuré en taille est plus pertinent que le modèle structuré en âge pour l'étude de la division d'*E. coli*. (Voir aussi la méthodologie utilisée dans [Ami14] pour la construction du modèle, ou encore dans [ONC14].)

Une compréhension plus profonde de la division cellulaire repose sur la mise en lumière de mécanismes moléculaires sous-jacents. Dans le but de trouver des mécanismes moléculaires expliquant les phénomènes biologiques, des banques de mutants existent. Un mutant est une cellule sur laquelle est opéré un changement génétique, par insertion, délétion, *etc.* Pour *E. coli*, un exemple de banque existante est une banque de délétion où chaque gène non-vitaux est systématiquement délété, ce qui conduit à la création de presque 4 000 mutants (voir [BAH+06]). Un gène important pour la division cellulaire pourrait être détecté à partir d'un changement de forme du taux de division  $B$  (par rapport au taux de division  $B$  de la bactérie non-mutante). Trouver des gènes importants passe nécessairement par l'étude d'une banque de mutants de taille considérable. C'est pourquoi nous visons à rendre possible l'estimation du taux de division  $B$  à partir du protocole expérimental le plus simple et rapide possible. Le protocole le plus rapide se trouve précisément être le prélèvement de cellules dans la population à un instant  $T$  et la mesure de leur taille.

Cette problématique spécifique pose également un problème de test : il s'agirait de tester si deux taux de division, le premier obtenu pour une population de bactéries *E. coli* elle-même, le second obtenu pour une population composée de mutants d'*E. coli*, sont significativement différents.

## 7 COMPOSITION DE LA THÈSE

Cette thèse se compose de quatre chapitres dont la rédaction repose sur les travaux suivants :

- [Chapitre I] *Nonparametric estimation of the division rate of an age-dependent branching process*, avec M. Hoffmann, en révision pour *Stochastic Processes and their Applications*.
  
- [Chapitre II] *Adaptive estimation for bifurcating Markov chains*, avec S. V. Bitseki Penda et M. Hoffmann, soumis.
  
- [Chapitre III] *Nonparametric estimation of the autoregressive functions in bifurcating autoregressive models*, avec S. V. Bitseki Penda, soumis.
  
- [Chapitre IV] *Influence of variability on the Malthus parameter*, avec M. Doumic et L. Robert, en préparation.



---

## Références du Chapitre Introductif

---

- [Ami14] A. Amir. *Cell size regulation in bacteria*. Physical Review Letters, 112 (2014), 208102.
- [Aza14a] R. Azaïs. *A recursive nonparametric estimator for the transition kernel of a piecewise-deterministic Markov process*. ESAIM : Probability and Statistics, 18 (2014), 726–749.
- [Aza14b] R. Azaïs. *Nonparametric estimation of the conditional distribution of the inter-jumping times for piecewise-deterministic Markov processes*. Scandinavian Journal of Statistics, 41 (2014), 950–969.
- [ABG+14] R. Azaïs, J.-B. Bardet, A. Genadot, N. Krell et P.-A. Zitt. *Piecewise deterministic Markov process - recent results*. ESAIM : Proceedings, 44 (2014), 276–290.
- [BAH+06] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner et H. Mori. *Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants : the Keio collection*. Molecular Systems Biology, 2 (2006), 11 pp.
- [BMR] F. Baccelli, D. R. McDonald et J. Reynier. *A mean-field model for multiple TCP connections through a buffer implementing red*. Performance Evaluation, 49 (2002), 77–97.
- [BST+11] H. T. Banks, K. L. Sutton, W. C. Thompson, G. Bocharov, D. Roose, T. Schenkel et A. Meyerhans. *Estimation of cell proliferation dynamics using CFSE data*, Bulletin of Mathematical Biology, 73 (2011), 116–150.
- [BDMT11] V. Bansaye, J.-F. Delmas, L. Marsalle et V. C. Tran. *Limit theorems for Markov processes indexed by continuous time Galton-Watson trees*. The Annals of Applied Probability, 21 (2011), 2263–2314.
- [BA67] G. I. Bell et E. C. Anderson. *Cell growth and division : I. a mathematical model with applications to cell volume distributions in mammalian suspension cultures*. Biophysical Journal, 7 (1967), 329–351.
- [BdSG09] B. Bercu, B. de Saporta et A. Gégout-Petit. *Asymptotic analysis for bifurcating autoregressive processes via a martingale approach*. Electronic Journal of Probability, 87 (2009), 2492–2526.
- [Ber06] J. Bertoin. *Random fragmentation and coagulation processes*. Cambridge University Press, 2006.
- [Bie97] H. J. Bierens. *Kernel estimator of regressions functions*. Advances in econometrics : Fifth world congress, Vol. 1 (1987).
- [BD14] S. V. Bitseki Penda et H. Djellout. *Deviation inequalities and moderate deviations for estimators of parameters in bifurcating autoregressive models*. Annales de l’IHP-PS, 50 (2014), 806–844.
- [BDG14] S. V. Bisteki-Penda, H. Djellout et A. Guillin. *Deviations inequalities, moderate deviations and some limits theorems for bifurcating Markov chains with application*. Annals of Applied Probability 24 (2014), 235–291.
- [BEG15] S. V. Bitseki Penda, M. Escobar-Bach and A. Guillin. *Transportation cost-information and concentration inequalities for bifurcating Markov chains*. arXiv :1501.06693 (2015).
- [Bla13] V. Blandin. *Asymptotic results for random coefficient bifurcating autoregressive processes*. Statistics 48 (2014), 1202–1232.
- [BDE14] T. Bourgeron, M. Doumic et M. Escobedo. *Estimating the division rate of the growth-fragmentation equation with a self-similar kernel*. Inverse Problems, 30 (2014), 30pp.

- [CLO+09] V. Calvez, N. Lenuzza, D. Oelz, J.-P. Deslys, P. Laurent, F. Mouthon et B. Perthame. *Size distribution dependence of prion aggregates infectivity*. *Mathematical Biosciences*, 217 (2009), 88–99.
- [Clé00] S. Clémenton. *Méthodes d'ondelettes pour la statistique non paramétrique des chaînes de Markov*. Université Paris 7 (2000).
- [Clo11] B. Cloez. *Limit theorems for some branching measure-valued processes*, hal-00598030 (2011).
- [CS86] R. Cowan et R. G. Staudte. *The bifurcating autoregressive model in cell lineage studies*. *Biometrics*, 42 (1986), 769–783.
- [DvS05] V. Delouille et R. von Sachs. *Estimation of nonlinear autoregressive models using design-adapted wavelets*. *Annals of the Institute of Statistical Mathematics*, 57 (2005), 235–253.
- [DM10] J.-F. Delmas et L. Marsalle. *Detection of cellular aging in Galton-Watson process*. *Stochastic Processes and their Applications*, 12 (2010), 2495–2519.
- [dSGPM12] B. de Saporta, A. Gégout-Petit et L. Marsalle. *Asymmetry tests for bifurcating autoregressive processes with missing data*. *Statistics & Probability Letters*, 82 (2012), 1439–1444.
- [DJKP95] D. L. Donoho, I. M. Johnstone, G. Kerkycharian et D. Picard. *Wavelet shrinkage : Asymptopia ?* *Journal of the Royal Statistical Society*, 57 (1995), 301–369.
- [DHKR15] M. Doumic, M. Hoffmann, N. Krell et L. Robert. *Statistical estimation of a growth-fragmentation model observed on a genealogical tree*. *Bernoulli*, 21 (2015), 1760–1799.
- [DHR12] M. Doumic, M. Hoffmann, P. Reynaud-Bouret et V. Rivoirard. *Nonparametric estimation of the division rate of a size-structured population*. *SIAM Journal on Numerical Analysis*, 50 (2012), 925–950.
- [Dou07] M. Doumic. *Analysis of a population model structured by the cells molecular content*. *Mathematical Modelling of Natural Phenomena*, 2 (2007), 121–152.
- [DPZ09] M. Doumic, B. Perthame et J. Zubelli. *Numerical Solution of an Inverse Problem in Size-Structured Population Dynamics*, *Inverse Problems*, 25 (2009), 25pp.
- [DT13] M. Doumic et L. M. Tine. *Estimating the Division Rate for the Growth-Fragmentation Equation*. *Journal of Mathematical Biology*, 67 (2013), 69–103.
- [Efr13] S. Efromovich. *Density estimation for biased data*. *Annals of Statistics*, 32 (2004), 1137–1161.
- [GL11] A. Goldenshluger et O. Lepski. *Bandwidth selection in kernel density estimation : oracle inequalities and adaptive minimax optimality*. *The Annals of Statistics*, 39 (2011), 1608–1632.
- [Guy07] J. Guyon. *Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging*. *The Annals of Applied Probability*, 17 (2007), 1538–1569.
- [GBP+05] J. Guyon, A. Bize, G. Paul, E. Stewart, J.-F. Delmas et F. Taddéi. *Statistical study of cellular aging*. *ESAIM : Proceedings*, 14 (2005), 100–114.
- [Hoa15] V. H. Hoang. *Estimating the division kernel of a size-structured population*. arXiv :1509.02872 (2015).
- [HM11] M. Hairer et J. Mattingly. *Yet Another Look at Harris' Ergodic Theorem for Markov Chains*. *Seminar on Stochastic Analysis, Random Fields and Applications VI*, *Progress in Probability*, 63 (2011), 109–117.
- [HWG91] A. J. Hall, G. C. Wake et P. W. Gandar. *Steady size distributions for cells in one-dimensional plant tissues*. *Journal of Mathematical Biology*, 30 (1991), 101–123.
- [HKPT98] W. Härdle, G. Kerkycharian, D. Picard et A. Tsybakov. *Wavelets, Approximation and Statistical Applications*. *Lecture Notes in Statistics*, Springer, 1998.
- [Har63] T. Harris. *The theory of branching processes*. Springer-Verlag, New-York, 1963.
- [Hof99] M. Hoffmann. *On nonparametric estimation in nonlinear AR(1)-models*. *Statistics & Probability Letters*, 44 (1999), 29–45.
- [HK11] M. Hoffmann et N. Krell. *Statistical analysis of self-similar fragmentation chains*. *Bernoulli*, 17 (2011), 395–423.

- [HHL02] R. Höpfner, M. Hoffmann et E. Löcherbach. *Nonparametric estimation of the death rate in branching diffusions*. Scandinavian Journal of Statistics, 29 (2002), 665–690.
- [KP00] G. Kerkycharian et D. Picard. *Thresholding algorithms, maxisets and well-concentrated bases*. Test, 9 (2000), 283–344.
- [KNW+14] D. J. Kiviet, P. Nghe, N. Walker, S. Boulineau, V. Sunderlikova et S. J. Tans. *Stochasticity of metabolism and growth at the single-cell level*. Nature, 514 (2014), 376–379.
- [Kre14] N. Krell. *Statistical estimation of jump rates for a specific class of Piecewise Deterministic Markov Processes*. arXiv :1406.2845v2 (2014).
- [Lac07] C. Lacour. *Estimation non paramétrique adaptative pour les chaînes de Markov et les chaînes de Markov cachées*. Université Paris-Descartes (2007).
- [Löc02a] E. Löcherbach. *Likelihood ratio processes for Markovian particle systems with killing and jumps*. Statistical inference for stochastic processes, 5 (2002a), 153–177.
- [Löc02b] E. Löcherbach. *LAN and LAMN for systems of interacting diffusions with branching and immigration*. Annales of the Institute Henri Poincaré, 38 (2002b), 59–90.
- [MHT66] A. G. Marr, R. J. Harvey et W. C. Trentini. *Growth and division of Escherichia coli*. Journal of Bacteriology, 91(1966), 2388–2389.
- [MD86] J. A. J. Metz et O. Dieckmann. *Formulating models for structured populations*. In *The dynamics of physiologically structured populations (Amsterdam, 1983)*, Lecture Notes in Biomathematics, 68 (1986), 78–135.
- [Mic06] P. Michel. *Optimal proliferation rate in a cell division model*. Mathematical Modelling of Natural Phenomena, 1 (2006), 23–44.
- [MPR02] S. Mischler, B. Perthame et L. Ryzhik. *Stability in a nonlinear population maturation model*. Mathematical Models and Methods in Applied Sciences, 12 (2002), 1751–1772.
- [MS13] S. Mischler et J. Scher. *Spectral analysis of semigroups and growth-fragmentation equations*. 2013, to appear in Annales de l’Institut Henri Poincaré (C) Analyse Non Linéaire.
- [ONC14] M. Osella, E. Nugent et M. Cosentino Lagomarsino. *Concerted control of Escherichia coli cell division*. PNAS, 111 (2014), 3431–3435.
- [PPS14] K. Pakdaman, B. Perthame et D. Salort. *Adaptation and fatigue model for neuron networks and large time asymptotics in a nonlinear fragmentation equation*. The Journal of Mathematical Neuroscience, (2014), 4–14
- [PPS10] K. Pakdaman, B. Perthame et D. Salort. *Dynamics of structured neuron population*. Nonlinearity, 23 (2010), 55–75.
- [Per07] B. Perthame. *Transport equations arising in biology*. Birkhäuser Frontiers in mathematics edition, 2007.
- [PZ07] B. Perthame et J. Zubelli. *On the inverse problem for a size-structured population model*, 23 (2007), 1037–1052
- [RHK+14] L. Robert, M. Hoffmann, N. Krell, S. Aymerich, J. Robert et M. Doumic. *Division control in Escherichia Coli is based on a size-sensing rather than a timing mechanism*. BMC Biology, 12 (2014).
- [Rot83] M. Rotenberg. *Transport theory for growing cell populations*. Journal of Theoretical Biology, 103 (1983), 181–199.
- [SRBA14] I. Soifer, L. Robert, N. Barkai et A. Amir. *Single-cell analysis of growth in budding yeast and bacteria reveals a common size regulation strategy*, arXiv :1410.4771 (2014).
- [SMPT05] E. J. Stewart, R. Madden, G. Paul et F. Taddéi. *Aging and death in an organism that reproduces by morphologically symmetric division*, PLoS Biology, 3 (2005), 295–300.
- [Tra08] V. C. Tran. *Large population limit and time behaviour of a stochastic particle model describing an age-structured population*. ESAIM : Probability and Statistics, 12 (2008), 345–386.

- [TAB+15] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola et S. Jun. *Cell-size control and homeostasis in bacteria*. *Current Biology*, 25 (2015).
- [WRP+10] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright et S. Jun. *Robust growth of Escherichia coli*. *Current Biololy*, 20 (2010), 1099–1103.

# CHAPITRE I

---

## Processus d'âge branchant : estimation non-paramétrique du taux de division

---

DANS un cadre d'observation en temps continu, ce premier chapitre est une contribution à l'objectif **(A)** *Comprendre le mécanisme de division à travers la reconstruction du taux de division en tant que fonction de la variable structurante*, formulé en Chapitre Introductif (page 2). En temps continu, divers écueils apparaissent dans la reconstruction du taux de division, y compris pour des modèles dont l'étude serait extrêmement simple en temps discret.

Notre intérêt se porte sur un taux de division dépendant de l'âge, l'âge étant ici l'unique variable structurante du modèle. Le processus que nous étudions est connu sous le nom de processus d'âge branchant ou processus de Bellman-Harris. Des particules, sans interactions entre elles, vivent chacune pendant une durée aléatoire distribuée selon  $B(\cdot) \exp(-\int_0^\cdot B(s)ds)$  avant de laisser place à un nombre aléatoire d'enfants. L'évolution du processus est observée en temps continu pendant l'intervalle  $[0, T]$ . De façon équivalente ici, les durées de vie des particules nées entre la date 0 et la date  $T$  fixée sont observées (ainsi que les nombres d'enfants). L'estimation du taux de division  $B$  est alors conduite et les résultats saillants du chapitre sur la question sont synthétisés dans la Table 1 page 13 en Chapitre Introductif.

Ce travail met en lumière les difficultés intrinsèques liées aux schémas d'observation en temps continu, parmi lesquelles un nombre d'observations aléatoire non-ancillaire, un biais de sélection ainsi qu'une dépendance complexe entre les observations.

## Abstract

---

We study the nonparametric estimation of the branching rate  $B(x)$  of a supercritical Bellman-Harris population: a particle with age  $x$  has a random lifetime governed by  $B(x)$ ; at its death time, it gives rise to  $k \geq 2$  children with lifetimes governed by the same division rate and so on. We observe in continuous time the process over  $[0, T]$ . Asymptotics are taken as  $T \rightarrow \infty$ ; the data are stochastically dependent and one has to face simultaneously censoring, bias selection and non-ancillarity of the number of observations. In this setting, under appropriate ergodicity properties, we construct a kernel-based estimator of  $B(x)$  that achieves the rate of convergence  $\exp(-\lambda_B \frac{\beta}{2\beta+1} T)$ , where  $\lambda_B$  is the Malthus parameter and  $\beta > 0$  is the smoothness of the function  $B(x)$  in a vicinity of  $x$ . We prove that this rate is optimal in a minimax sense and we relate it explicitly to classical nonparametric models such as density estimation observed on an appropriate (parameter dependent) scale. We also shed some light on the fact that estimation with kernel estimators based on data alive at time  $T$  only is not sufficient to obtain optimal rates of convergence, a phenomenon which is specific to nonparametric estimation and that has been observed in other related growth-fragmentation models.

---

## Contents

---

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>INTRODUCTION</b> . . . . .                                       | <b>43</b> |
| 1.1      | Motivation . . . . .  | 43        |
| 1.2      | Objectives and results . . . . .                                    | 44        |
| <b>2</b> | <b>RATE OF CONVERGENCE FOR BIASED EMPIRICAL MEASURES</b> . . . . .  | <b>46</b> |
| 2.1      | Continuous time rooted trees . . . . .                              | 46        |
| 2.2      | The limiting objects . . . . .                                      | 48        |
| 2.3      | Convergence results for biased empirical measures . . . . .         | 49        |
| <b>3</b> | <b>STATISTICAL ESTIMATION</b> . . . . .                             | <b>51</b> |
| 3.1      | Construction of an estimation procedure . . . . .                   | 51        |
| 3.2      | Convergence results for $\hat{B}_T(x)$ . . . . .                    | 52        |
| 3.3      | Discussion of the results . . . . .                                 | 53        |
| <b>4</b> | <b>NUMERICAL IMPLEMENTATION</b> . . . . .                           | <b>56</b> |
| <b>5</b> | <b>PROOFS</b> . . . . .   | <b>58</b> |
| 5.1      | Preliminaries . . . . .   | 58        |
| 5.2      | Proof of Theorems 3 and 4 . . . . .                                 | 60        |
| 5.3      | Proof of Proposition 5 . . . . .                                    | 64        |
| 5.4      | Proof of Theorem 7 . . . . .  | 64        |
| 5.5      | Proof of Theorem 8 . . . . .  | 73        |
| 5.6      | Proof of Proposition 9 . . . . .                                    | 75        |
| <b>6</b> | <b>APPENDIX</b> . . . . .   | <b>76</b> |
| 6.1      | Heuristics for the convergences to the limits (8) and (9) . . . . . | 76        |
| 6.2      | Proof of Proposition 10 . . . . .                                   | 78        |
| 6.3      | Proof of (26) and (27) of Proposition 11 . . . . .                  | 80        |
| 6.4      | Proof of Lemma 16 . . . . .   | 81        |
|          | <b>REFERENCES</b> . . . . .   | <b>82</b> |

---

# 1 INTRODUCTION

## 1.1 MOTIVATION

Structured models have been paid particular attention over the last few years, both from a probabilistic and an applied analysis angle, in particular with a view toward a better understanding of population evolution in mathematical biology (see for instance the textbook by Perthame [Per07] and the references therein). In this context, a more specific focus and need for statistical methods has emerged recently (*e.g.* Doumic *et al.* [DPZ09, DHRR12, DHKR15] and the references therein) and this is the topic of the present chapter. If  $x$  denotes a so-called structuring variable – for instance age, size, any measure of variability or DNA content of a cell or bacteria, and if  $n(t, x)$  denotes the number or density of cells at time  $t$  of a population starting from a single ancestor at time  $t = 0$ , a sound mathematical model can be obtained by specifying an evolution equation for  $n(t, x)$ .

Consider for instance the paradigmatic problem of age-dependent cell division, where the evolution of  $n(t, x)$  is given by the simplest transport-fragmentation equation

$$\begin{cases} \frac{\partial}{\partial t}n(t, x) + \frac{\partial}{\partial x}n(t, x) + B(x)n(t, x) = 0 \\ n(t, 0) = m \int_0^\infty B(y)n(t, y)dy, \quad t > 0, \quad n(0, x) = \delta_0, \end{cases} \quad (1)$$

where  $\delta_0$  denotes the Dirac mass at point 0. In this model, each cell dies according to a division rate  $x \rightsquigarrow B(x)$  that depends on its age  $x$  only (a living cell of age  $x$  has probability  $B(x)dx$  of dying in the interval  $[x, x + dx]$ ) and, at its time of death, it gives rise to  $m \geq 2$  children at its time of death. The parameters  $(m, B)$  specify the so-called age-dependent model.

In this seemingly simple context, we wish to draw statistical inference on the division rate function  $x \rightsquigarrow B(x)$  and on  $m$  in the most rigorous way, when we observe the evolution of the population through time and when the shape of the function  $B$  can be arbitrary, to within a prescribed smoothness class, *i.e.* in a nonparametric setting. In order to do so, we transfer the deterministic description (1) into a probabilist model that consists of a system of (non-interacting) particles specified by a probability distribution  $p$  on the integers (the offspring distribution) and a probability density  $f$  on  $[0, \infty)$ . A particle has a random lifetime drawn according to  $f(x)dx$ ; at the time of its death, it gives rise to  $k$  children with probability  $p_k$  (with  $p_0 = p_1 = 0$ ), each child having independent lifetimes distributed as  $f(x)dx$ , and so on. The resulting process is a classical supercritical Bellman-Harris, see for instance the textbooks of Harris [Har63] or Athreya and Ney [AN72]. It is described by a piecewise deterministic Markov process

$$X(t) = (X_1(t), X_2(t), \dots), \quad t \geq 0, \quad (2)$$

with values in  $\bigcup_{k \geq 1} [0, \infty)^k$ , where the  $X_i(t)$ 's denote the (ordered) ages of the living particles at time  $t$ . The formal link between  $X(t)$  and  $n(t, x)$  is obtained via  $n(t, x) = \mathbb{E}[\sum_{i=1}^\infty \delta_{X_i(t)=x}]$  which has to be understood in a weak (measure) sense, *i.e.* the empirical measure (in expectation) of the particle system and solves Equation (1), we refer to [Oel90].



The correspondence between  $(m, B)$  and  $(f, p)$  is given by

$$B(x) = \frac{f(x)}{1 - \int_0^x f(s)ds}, \quad x \in [0, \infty), \quad \text{and} \quad m = \sum_{k \geq 2} kp_k, \quad (3)$$

provided everything is well defined. Under fairly reasonable assumptions described below, it is one-to-one between  $B$  and  $f$ , but not between  $m$  and  $p$ . We are interested in the nonparametric estimation of  $x \rightsquigarrow B(x)$ , which is nothing but the hazard rate function of the lifetime density  $f$  of each particle, and also in the mean offspring  $m$ , the whole distribution  $p$  being considered as a nuisance parameter.

## 1.2 OBJECTIVES AND RESULTS

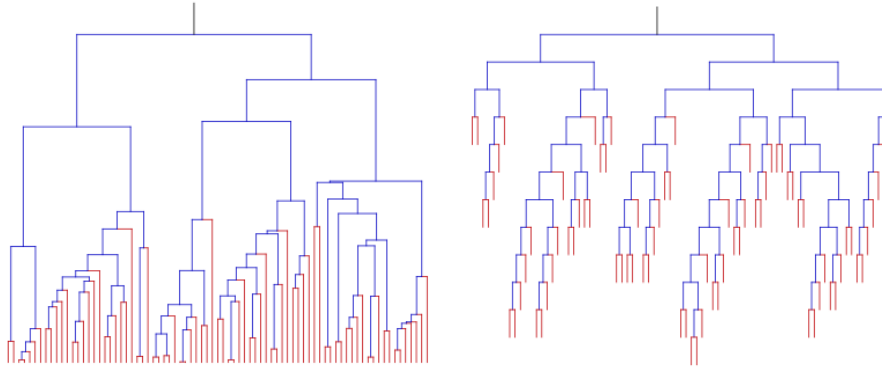


Figure 4: *The effect of bias selection. Simulation of a binary ( $p_2 = 1$  so  $m = 2$ ) age-dependent tree with  $B$  given in Section 4, up to time  $T = 8$  ( $|\mathcal{T}_T| = 145$ ). Left: the size of each segment represents the lifetime of an individual. Individuals alive at time  $T$  are represented in red. Right: genealogical representation of the same realisation of the tree.*

### Observation schemes

We assume we observe the whole trajectory  $(X(t), t \in [0, T])$ , where  $T > 0$  is a fixed (large) terminal time. Asymptotics are taken as  $T \rightarrow \infty$ . If we denote by  $\mathcal{T}_T$  the population of individuals that are born before  $T$  and observed up to time  $T$  and if  $(\zeta_u^T, u \in \mathcal{T}_T)$  denotes the values of the ages of the different individuals of  $\mathcal{T}_T$  (at their time of death or at time  $T$ ), we wish to draw inference on  $B(x)$  based on

$$\{X(t), t \in [0, T]\} = \{\zeta_u^T, u \in \mathcal{T}_T\}.$$

Although the lifetimes of the individuals are independent (and identically distributed) with common density  $f$ , this is no longer the case for the population  $(\zeta_u^T, u \in \mathcal{T}_T)$  considered as a whole: the tree structure plays a crucial role and we have to face several non-trivial difficulties:

- 1) *Bias selection*: particles with small lifetimes are more often observed than particles with large lifetimes since the observation of the process is stopped along all the branches at the fixed time  $T$ , as illustrated in Figure 4.

- 2) *Censoring*: if  $\partial\mathcal{T}_T \subset \mathcal{T}_T$  denotes the population of individuals alive at time  $T$  (in red in Figure 4), they are censored in our observation scheme (we observe their lifetime only up to time  $T$ ) but contribute to the whole estimation process at the same level as the population  $\overset{\circ}{\mathcal{T}}_T \subset \mathcal{T}_T$  of individuals born and dead before  $T$ : due to the supercriticality of the process ( $m > 1$ ) we have  $|\mathcal{T}_T| \approx |\overset{\circ}{\mathcal{T}}_T| \approx |\partial\mathcal{T}_T|$  as  $T$  grows to infinity, and this affects the statistical analysis, see Section 2.2 below.
- 3) *Non-ancillarity*: the number of observations  $|\mathcal{T}_T|$  that governs the amount of statistical information is random and its distribution depends on  $B$ : we essentially have less observations if  $B$  is small (particles split at a slow rate) than if  $B$  is large (particles split at a fast rate). This means that  $|\mathcal{T}_T|$  is not ancillary in the terminology of Fisher: it is not possible to ignore its randomness (by conditioning upon its value for instance) without losing some statistical information. We refer to the Encyclopedia of Statistics [Kie72] for more details.

### Main results

We first study in Section 2 the behaviour of empirical measures of the form

$$\mathcal{E}^T(\mathcal{V}, g) = |\mathcal{V}|^{-1} \sum_{u \in \mathcal{V}} g(\zeta_u^T), \quad \text{with } \mathcal{V} = \overset{\circ}{\mathcal{T}}_T \text{ or } \partial\mathcal{T}_T$$

for suitable test functions  $g$ . From the classical study of critical branching processes, it is known that  $|\overset{\circ}{\mathcal{T}}_T| \approx |\partial\mathcal{T}_T| \approx e^{\lambda_B T}$ , where  $\lambda_B > 0$  is the Malthus parameter associated to the model (Harris [Har63] and (6) below). Both  $\mathcal{E}^T(\overset{\circ}{\mathcal{T}}_T, g)$  and  $\mathcal{E}^T(\partial\mathcal{T}_T, g)$  converge to their respective limits with rate  $\exp(-\lambda_B T/2)$ , with some uniformity in  $B$  and  $g$  as shown in Theorem 3 and 4 below. For the proof, we heavily rely on the recent studies of Cloez [Clo11] and Bansaye *et al.* [BDMT11], two key references for this chapter, adjusting the tools developed in [BDMT11] to the non-Markovian case: the essential ingredient is the use of many-to-one formulae that reduce the problem to studying the evolution of a particle picked at random along the genealogical tree (Propositions 10 and 11). The rate of convergence to equilibrium of this tagged particle, which governs the rates of convergence for statistical estimators, is obtained by a simple coupling argument (Proposition 12).

These preliminary results enable us to address the main issue of the chapter: we construct in Section 3 a nonparametric estimator  $\widehat{B}_T(x)$  of  $B(x)$  that achieves the rate of convergence  $\exp(-\lambda_B \frac{\beta}{2\beta+1} T)$  for pointwise error and uniformly over functions  $B$  with local smoothness of order  $\beta > 0$  (Theorem 7). We show that this rate is optimal in a minimax sense in Theorem 8, thanks to statistical tools developed in Löcherbach [Löc02a]. This result is obtained under the restriction that convergence to equilibrium of a tagged particle is faster than the growth of the tree. Otherwise, we still have a rate of convergence, but we do not have (nor believe in) its optimality. We bypass the aforementioned bias selection difficulty 1) by weighting a kernel estimator by a de-biasing factor that depends on preliminary estimators of  $\lambda_B$  and  $m$ . These estimators (essentially) converge with rate  $\exp(-\lambda_B T/2)$  as shown in Proposition 5. As for the censoring part 2), we base our nonparametric kernel estimator on  $\mathcal{E}^T(\overset{\circ}{\mathcal{T}}_T, g)$  and *not on*  $\mathcal{E}^T(\partial\mathcal{T}_T, g)$ , since that latter quantity would lead to a suboptimal rate of convergence as discussed in Section 3.3. Finally, the non-ancillarity issue 3) is solved by specifying a random bandwidth for the kernel that also depends on the preliminary estimation of  $\lambda_B$ . This last point requires extra efforts in order

to show a form of stability that is detailed in Proposition 17.

The statistical study of branching processes goes back to Athreya and Keiding [AK77] for deriving maximum likelihood theory in the case of a parametric (constant) division rate, relying on the fact that the number of living cells is then a Markov process, a property we lose here for a non-constant division rate  $x \rightsquigarrow B(x)$ . The textbook of Guttorp [Gut91] gives an account of existing parametric methods in the 1990's. In the early 2000's the regularity in the sense of the LAN and LAMN property was established in the comprehensive study of Löcherbach [Löc02a, Löc02b], see also Hyrien [Hyr07] for statistical computational methods and Johnson *et al.* [JSvR79] for Bayesian analysis, and Delmas and Marsalle [DM10] in discrete time. In nonparametric estimation, only few results exist; we mention the case when dynamics between jumps is driven by a diffusion in Höpfner *et al.* [HHL02]. To the best of our knowledge, our study provides with the first fully nonparametric approach in continuous time in supercritical branching processes which are piecewise deterministic. Admittedly, the Bellman-Harris model is a toy model for the study of population dynamics, but we believe that the present contribution sheds some light in the intrinsic difficulties that need to be solved in more elaborate models like cell equation for which only simplified statistical models have been considered so far (in discrete time or under additional deterministic or stochastic noise like in *e.g.* [DPZ09, DHRR12, DHKR15]). Concerning bias selection, density estimation when observing a biased sample has been studied at length framework by Efromovich [Efr04].

### *Organisation of the chapter*

In Section 2, we define our rigorous statistical framework by means of continuous time rooted trees (Section 2.1) and study the convergence properties of the biased empirical measures  $\mathcal{E}^T(\mathcal{J}_T, g)$  and  $\mathcal{E}^T(\partial\mathcal{J}_T, g)$  in Section 2.3. We start by deriving heuristically the respective limits of the empirical measures in Section 2.2 (that can also be found in Cloez [Clo11] and Bansaye *et al.* [BDMT11]) in order to shed some light on the specific methods of proof in the subsequent study of rate of convergence. We construct in Section 3 the estimators of  $m$ ,  $\lambda_B$  and  $B(x)$  and state our statistical results together with a discussion on the extensions and limitations of our findings. Section 4 tackles the problem of numerical implementation on simulated data, advocating for a reasonable use of our estimators in practice. Section 5 is devoted to the proofs. An appendix (Section 6) contains auxiliary useful results.

## 2 RATE OF CONVERGENCE FOR BIASED EMPIRICAL MEASURES

### 2.1 CONTINUOUS TIME ROOTED TREES

It will prove more convenient to work with a representation of  $(X(t))_{t \geq 0}$  in terms of a continuous time rooted tree. We need some notation and closely follow Bansaye *et al.* [BDMT11]. Let

$$\mathcal{U} = \bigcup_{k \geq 0} (\mathbb{N}^*)^k$$

with  $\mathbb{N}^* = \{1, 2, \dots\}$  and  $(\mathbb{N}^*)^0 = \{\emptyset\}$  denote the infinite genealogical tree. We use throughout the following standard notation: for  $u = (u_1, u_2, \dots, u_m)$  and  $v = (v_1, \dots, v_n)$  in  $\mathcal{U}$ , we write  $uv = (u_1, \dots, u_m, v_1, \dots, v_n)$  for the concatenation, we identify  $\emptyset u$ ,  $u\emptyset$  and  $u$ , we write  $u \preceq v$  if

there exists  $w$  such that  $uw = v$  and  $u \prec v$  if  $u \preceq v$  and  $w \neq \emptyset$ . For  $u = (u_1, u_2, \dots, u_m)$ , we also write  $|u| = m$ .

Given a family  $(\nu_u, u \in \mathcal{U})$  of integers representing the number of children of the individuals  $u \in \mathcal{U}$ , we construct an ordered rooted tree  $\mathcal{T} \subset \mathcal{U}$  as follows:

- i)  $\emptyset \in \mathcal{T}$ ,
- ii) If  $v \in \mathcal{T}$ ,  $u \preceq v$  implies  $u \in \mathcal{T}$ ,
- iii) For every  $u \in \mathcal{T}$ , we have  $uj \in \mathcal{T}$  if and only if  $1 \leq j \leq \nu_u$ .

For a family  $(\zeta_u, u \in \mathcal{U})$  of nonnegative numbers representing the lifetimes of the individuals  $u \in \mathcal{U}$ , we set

$$b_u = \sum_{v \prec u} \zeta_v \quad \text{and} \quad d_u = b_u + \zeta_u \quad (4)$$

for the times of birth and death of the individual  $u \in \mathcal{U}$ . Let  $\mathbb{U} = \mathcal{U} \times [0, \infty)$ . A continuous time rooted tree is then a subset  $\mathbb{T}$  of  $\mathbb{U}$  such that

- (i)  $(\emptyset, 0) \in \mathbb{T}$ ,
- (ii) The projection  $\mathcal{T}$  of  $\mathbb{T}$  on  $\mathcal{U}$  is an ordered rooted tree,
- (iii) There exists a family  $(\zeta_u, u \in \mathcal{U})$  of nonnegative numbers such that  $(u, s) \in \mathbb{T}$  if and only if  $b_u \leq s < d_u$ , where  $(b_u, d_u)$  are defined by (4).

We now work on some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . In this setting, we have the following

**Definition 1** (The Bellman-Harris model). *A random continuous time rooted tree is a Bellman-Harris model with offspring distribution  $p = (p_k)_{k \geq 1}$  and division rate  $B : [0, \infty) \rightarrow [0, \infty)$  if*

- (i) *The family of the number of children  $(\nu_u, u \in \mathcal{U})$  are independent random variables with common distribution  $p$ .*
- (ii) *The family of lifetimes  $(\zeta_u, u \in \mathcal{U})$  are independent random variables such that*

$$\mathbb{P}(\zeta_u \geq x) = \exp\left(-\int_0^x B(y)dy\right), \quad x \geq 0, \quad (5)$$

with

$$\int_0^\infty B(x)dx = \infty,$$

- (iii) *The families of random variables  $(\nu_u, u \in \mathcal{U})$  and  $(\zeta_u, u \in \mathcal{U})$  are independent.*

Going back to the process  $(X(t))_{t \geq 0}$  defined in (2), we have an identity between point measures on  $(0, \infty)$  that reads

$$\sum_{i \geq 1} \mathbf{1}_{\{X_i(t) > 0\}} \delta_{X_i(t)} = \sum_{u \in \mathcal{T}} \mathbf{1}_{\{t \in [b_u, d_u)\}} \delta_{t-b_u}.$$

The following assumption will be in force in the chapter:

**Assumption 2.** *The offspring distribution  $p = (p_k)_{k \geq 0}$  satisfies*

$$p_0 = p_1 = 0, \quad 2 \leq m = \sum_{k \geq 2} k p_k < \infty, \quad \sum_{k \geq 2} k^2 p_k < \infty \quad \text{and} \quad \bar{m} = \sum_{i \neq j} \sum_{k \geq i \vee j} p_k < \infty.$$

The technical condition  $\bar{m} < \infty$  is needed for the so-called many-to-one formulae, see Proposition 11 below.

## 2.2 THE LIMITING OBJECTS

In order to extract information about  $x \rightsquigarrow B(x)$ , we consider the empirical distribution function over the lifetimes indexed by some  $\mathcal{V}_T \subset \mathcal{T}_T$  for a test function  $g$ , that is

$$\mathcal{E}^T(\mathcal{V}_T, g) = |\mathcal{V}_T|^{-1} \sum_{u \in \mathcal{V}_T} g(\zeta_u^T),$$

and expect a law of large number as  $T \rightarrow \infty$ . Without much of a surprise, it turns out that depending whether  $\zeta_u^T = \zeta_u$  or not, *i.e.* if the data are still alive at time  $T$ , therefore censored or not, we have a different limit. More precisely, define

$$\mathring{\mathcal{T}}_T = \{u \in \mathcal{T}, b_u < T \text{ and } d_u \leq T\} \quad \text{and} \quad \partial\mathcal{T}_T = \{u \in \mathcal{T}, b_u \leq T < d_u\},$$

*i.e.* the set of particles that are born and that die before  $T$ , and the set of particles alive at time  $T$ , so that  $\mathcal{T}_T = \mathring{\mathcal{T}}_T \cup \partial\mathcal{T}_T$ . We need some notation. Introduce the *Malthus parameter*  $\lambda_B > 0$  defined as the (necessarily unique) solution to

$$\int_0^\infty B(x) e^{-\lambda_B x - \int_0^x B(y) dy} dx = \frac{1}{m}. \quad (6)$$

To a division rate function  $x \rightsquigarrow B(x)$  satisfying the properties of Definition 1, we associate its *density lifetime*

$$f_B(x) = B(x) \exp\left(-\int_0^x B(y) dy\right), \quad x \geq 0$$

and its biased density lifetime

$$f_{H_B}(x) = m e^{-\lambda_B x} f_B(x), \quad x \geq 0,$$

which in turns uniquely defines a biased division rate

$$H_B(x) = \frac{m e^{-\lambda_B x} f_B(x)}{1 - m \int_0^x e^{-\lambda_B y} f_B(y) dy}. \quad (7)$$

Finally, we define the limiting measures

$$\partial\mathcal{E}_B(g) = \lambda_B \frac{m}{m-1} \int_0^\infty g(x) e^{-\lambda_B x} e^{-\int_0^x B(y) dy} dx \quad (8)$$

and

$$\mathring{\mathcal{E}}_B(g) = m \int_0^\infty g(x) e^{-\lambda_B x} f_B(x) dx = \int_0^\infty g(x) f_{H_B}(x) dx. \quad (9)$$

It is known that  $\mathcal{E}^T(\partial\mathcal{T}_T, g) \rightarrow \partial\mathcal{E}_B(g)$  and  $\mathcal{E}^T(\mathring{\mathcal{J}}_T, g) \rightarrow \mathring{\mathcal{E}}_B(g)$  in probability as  $T \rightarrow \infty$ , see Appendix 6.1 for heuristics and references. We establish in Theorems 3 and 4 in the next Section 2.3 a rate of convergence with some uniformity in  $B$ . The rate is linked to  $\lambda_B$  and the geometric ergodicity of an auxiliary one-dimensional Markov process with infinitesimal generator

$$\mathcal{A}_{H_B}g(x) = g'(x) + H_B(x)(g(0) - g(x)) \quad (10)$$

densely defined on continuous functions vanishing at infinity and that represents the value of a branch along the tree picked uniformly at random at each branching event.

### 2.3 CONVERGENCE RESULTS FOR BIASED EMPIRICAL MEASURES

#### Notation

For constants  $b, C > 0$ , introduce the sets

$$\mathcal{L}_C = \left\{ g : [0, \infty) \rightarrow \mathbb{R}, \sup_x |g(x)| \leq C \right\}$$

and

$$\mathcal{B}_{b,C} = \left\{ B : [0, \infty) \rightarrow [0, \infty), \forall x \geq 0 : b \leq B(x) \leq b \max\{C, 1\} \right\}.$$

For a family  $\Gamma_T = (\Gamma_T(\gamma))_{T \geq 0}$  of real-valued random variables, with distribution depending on some parameter  $\gamma \in \mathcal{G}$  we say that  $\Gamma_T$  is  $\mathcal{G}$ -tight for the parameter  $\gamma$  if

$$\sup_{T > 0, \gamma \in \mathcal{G}} \mathbb{P}(|\Gamma_T(\gamma)| \geq K) \rightarrow 0 \text{ as } K \rightarrow \infty.$$

#### Results

We have a trade-off between the growth rate  $\lambda_B$  of the tree  $\mathbb{E}[|\mathcal{T}_T|] \approx e^{\lambda_B T}$  and the convergence to equilibrium of the Markov process with infinitesimal generator  $\mathcal{A}_{H_B}$  defined in (10) above. More, precisely, we show in Proposition 12 below the estimate

$$\left| P_{H_B}^t g(x) - \int_0^\infty g(y) \mu_B(y) dy \right| \leq 2 \sup_y |g(y)| e^{-\rho_B t} \text{ for every } x \in (0, \infty).$$

Here,  $(P_{H_B}^t)_{t \geq 0}$  denotes the semigroup associated to  $\mathcal{A}_{H_B}$  and  $\mu_B$  its unique invariant probability, and

$$\rho_B = \inf_x H_B(x)$$

where  $H_B(x)$  is the biased division rate defined in (7) above. The rate of convergence of the biased empirical measures  $\mathcal{E}^T(\mathring{\mathcal{J}}_T, g)$  and  $\mathcal{E}^T(\partial\mathcal{T}_T, g)$  to their limits  $\partial\mathcal{E}_B(g)$  and  $\mathring{\mathcal{E}}_B(g)$  respectively defined

by (8) and (9) are governed by  $\lambda_B$  and  $\rho_B$ : define

$$v_T(B) = \begin{cases} e^{-\min\{\rho_B, \lambda_B/2\}T} & \text{if } \lambda_B \neq 2\rho_B, \\ T^{1/2}e^{-\lambda_B T/2} & \text{if } \lambda_B = 2\rho_B. \end{cases} \quad (11)$$

We have:

**Theorem 3** (Rate of convergence for particles living at time  $T$ ). *Work under Assumption 2. For every  $b, C, C' > 0$ ,*

$$v_T(B)^{-1}(\mathcal{E}^T(\partial\mathcal{J}_T, g) - \partial\mathcal{E}_B(g))$$

is  $\mathcal{B}_{b,C} \times \mathcal{L}_{C'}$ -tight for the parameter  $(B, g)$ .

**Theorem 4** (Rate of convergence for particles dying before  $T$ ). *In the same setting as Theorem 3,*

$$v_T(B)^{-1}(\mathcal{E}^T(\overset{\circ}{\mathcal{J}}_T, g) - \overset{\circ}{\mathcal{E}}_B(g))$$

is  $\mathcal{B}_{b,C} \times \mathcal{L}_{C'}$ -tight for the parameter  $(B, g)$ .

Several comments are in order:

*About the rate of convergence and the class  $\mathcal{B}_{b,C}$*

the restriction  $B \in \mathcal{B}_{b,C}$  enables us to obtain uniform convergence results. This is important for the subsequent statistical analysis. However, this can be relaxed if only  $\mathcal{L}_{C'}$ -tightness is sought, provided  $B$  complies to the conditions of Definition 1 and Assumption 2 and  $\rho_B > 0$ . In the same direction, the rate  $v_T(B)$  can be improved replacing  $\rho_B = \inf_x H_B(x)$  in (11) by

$$\rho_B^* = \sup \left\{ \rho, \forall x, t > 0 : |P_{H_B}^t g(x) - \int_0^\infty g(y)\mu_B(y)dy| \leq 2 \sup_y |g(y)|e^{-\rho t} \right\}, \quad (12)$$

and we have in particular  $\rho_B^* \geq \rho_B$ .

*About the tightness*

what we need in order to handle the random normalisation in  $\mathcal{E}^T(\overset{\circ}{\mathcal{J}}_T, g)$  is actually the convergence of  $e^{\lambda_B T} |\overset{\circ}{\mathcal{J}}_T|^{-1}$ . This convergence still holds in probability but not necessarily in  $L^2(\mathbb{P})$ , so we only have tightness in Theorems 3 (and 4 for the same reason). However, if we replace  $\mathcal{E}^T(\overset{\circ}{\mathcal{J}}_T, g)$  by

$$\frac{1}{\mathbb{E}[|\overset{\circ}{\mathcal{J}}_T|]} \sum_{u \in \overset{\circ}{\mathcal{J}}_T} g(\zeta_u^T),$$

then we have a bound in  $L^2(\mathbb{P})$  together with a control on  $g$ , see Proposition 15 below. Such a finer control is mandatory for the subsequent statistical analysis, since we need to pick a function  $g$  that depends on  $T$  and that mimics the behaviour of the Dirac mass  $\delta_x$ , see Section 3 below.

### 3 STATISTICAL ESTIMATION

#### 3.1 CONSTRUCTION OF AN ESTIMATION PROCEDURE

*Estimation of  $m$  and  $\lambda_B$*

To a particle sitting at node  $u \in \mathring{\mathcal{T}}_T$ , we associate its number of children  $\nu_u$  (see Definition 1). Note that the knowledge of  $\mathcal{T}_T$  enables us to reconstruct  $\nu_u$  for every  $u \in \mathring{\mathcal{T}}_T$ . This enables us to define an estimator for  $m$  by setting

$$\hat{m}_T = |\mathring{\mathcal{T}}_T|^{-1} \sum_{u \in \mathring{\mathcal{T}}_T} \nu_u \quad (13)$$

on the set  $|\mathring{\mathcal{T}}_T| \neq 0$  and 2 otherwise. In order to estimate  $\lambda_B$ , we first observe that for  $\text{Id}(x) = x$ , we can write

$$\begin{aligned} \mathring{\mathcal{E}}_B(\text{Id}) &= m \int_0^\infty x(B(x) + \lambda_B) e^{-\int_0^x (B(y) + \lambda_B) dy} dx - m\lambda_B \int_0^\infty x e^{-\lambda_B x} e^{-\int_0^x B(y) dy} dx \\ &= m \int_0^\infty e^{-\int_0^x (B(y) + \lambda_B) dy} dx - m\lambda_B \frac{m-1}{m\lambda_B} \partial \mathcal{E}_B(\text{Id}) = m \frac{m-1}{m\lambda_B} - (m-1) \partial \mathcal{E}_B(\text{Id}), \end{aligned}$$

the last equality being obtained integrating by parts. So we obtain the following representation

$$\lambda_B = \left( \frac{1}{m-1} \mathring{\mathcal{E}}_B(\text{Id}) + \partial \mathcal{E}_B(\text{Id}) \right)^{-1}$$

and this yields the estimator

$$\hat{\lambda}_T = \left( \frac{1}{\hat{m}_T - 1} |\mathring{\mathcal{T}}_T|^{-1} \sum_{u \in \mathring{\mathcal{T}}_T} \zeta_u + |\partial \mathcal{T}_T|^{-1} \sum_{u \in \partial \mathcal{T}_T} \zeta_u^T \right)^{-1}. \quad (14)$$

The following convergence result for  $\hat{\lambda}_T$  is then a consequence of Theorems 3 and 4.

**Proposition 5.** *In the same setting as Theorem 3 with  $v_T(B)$  given in (11) above, we have that*

$$e^{\lambda_B T/2} (\hat{m}_T - m) \quad \text{and} \quad T^{-1} v_T(B)^{-1} (\hat{\lambda}_T - \lambda_B)$$

are  $\mathcal{B}_{b,C}$ -tight for the parameter  $B$ .

*Reconstruction formula for  $B(x)$*

An estimator  $\hat{B}_T : [0, \infty) \rightarrow \mathbb{R}$  of  $B$  is a random function

$$\hat{B}_T(x) = \hat{B}_T(x, (X(t))_{t \in [0, T]}), \quad x \in [0, \infty)$$

that is measurable as a function of  $(X(t))_{t \in [0, T]}$  but also as a function of  $x$ . By (3), we have

$$B(x) = \frac{f_B(x)}{1 - \int_0^x f_B(y) dy}$$



and from the definition  $\mathring{\mathcal{E}}_B(g) = m \int_0^\infty g(x) e^{-\lambda_B x} f_B(x) dx$  we obtain the formal reconstruction formula

$$B(x) = \frac{\mathring{\mathcal{E}}_B(m^{-1} e^{\lambda_B \cdot} \delta_x(\cdot))}{1 - \mathring{\mathcal{E}}_B(m^{-1} e^{\lambda_B \cdot} \mathbf{1}_{\{\cdot \leq x\}})} \quad (15)$$

where  $\delta_x(\cdot)$  denotes the Dirac function at  $x$ . Therefore, substituting  $m$  and  $\lambda_B$  by the estimators defined in (13) and (14) and taking  $g$  as a weak approximation of  $\delta_x$ , we obtain a strategy for estimating  $B(x)$  replacing furthermore  $\mathring{\mathcal{E}}_B(\cdot)$  by its empirical version  $\mathcal{E}^T(\mathring{\mathcal{J}}_T, \cdot)$ .

### Construction of a kernel estimator and function spaces

Let  $K : \mathbb{R} \rightarrow \mathbb{R}$  be a kernel function. For  $h > 0$ , set  $K_h(x) = h^{-1} K(h^{-1}x)$ . In view of (15), we define the estimator

$$\widehat{B}_T(x) = \frac{\mathcal{E}^T(\mathring{\mathcal{J}}_T, \widehat{m}_T^{-1} e^{\widehat{\lambda}_T \cdot} K_h(x - \cdot))}{1 - \mathcal{E}^T(\mathring{\mathcal{J}}_T, \widehat{m}_T^{-1} e^{\widehat{\lambda}_T \cdot} \mathbf{1}_{\{\cdot \leq x\}})}$$

on the set  $\mathcal{E}^T(\mathring{\mathcal{J}}_T, \widehat{m}_T^{-1} e^{\widehat{\lambda}_T \cdot} \mathbf{1}_{\{\cdot \leq x\}}) \neq 1$  and 0 otherwise. Thus  $\widehat{B}_T(x)$  is specified by the choice of the kernel  $K$  and the bandwidth  $h > 0$ . Note that the observations  $(\zeta_u, u \in \partial\mathcal{J}_T)$  only occur in the estimator  $\widehat{\lambda}_T$  of  $\lambda_B$ .

We need the following property on  $K$ :

**Assumption 6.** *The kernel  $K : \mathbb{R} \rightarrow \mathbb{R}$  is differentiable with compact support and for some integer  $n_0 \geq 1$ , we have  $\int_{-\infty}^\infty x^k K(x) dx = \mathbf{1}_{\{k=0\}}$  for  $k = 1, \dots, n_0$ .*

Assumption 6 will enable us to have nice approximation results over smooth functions  $B$ , described in the following way: for a compact interval  $\mathcal{D} \subset (0, \infty)$  and  $\beta > 0$ , with  $\beta = \lfloor \beta \rfloor + \{\beta\}$ ,  $0 < \{\beta\} \leq 1$  and  $\lfloor \beta \rfloor$  an integer, let  $\mathcal{H}_{\mathcal{D}}^\beta$  denote the Hölder space of functions  $g : \mathcal{D} \rightarrow \mathbb{R}$  possessing a derivative of order  $\lfloor \beta \rfloor$  that satisfies

$$|g^{\lfloor \beta \rfloor}(y) - g^{\lfloor \beta \rfloor}(x)| \leq c(g) |x - y|^{\{\beta\}}. \quad (16)$$

The minimal constant  $c(g)$  such that (16) holds defines a semi-norm  $|g|_{\mathcal{H}_{\mathcal{D}}^\beta}$ . We equip the space  $\mathcal{H}_{\mathcal{D}}^\beta$  with the norm  $\|g\|_{\mathcal{H}_{\mathcal{D}}^\beta} = \sup_x |g(x)| + |g|_{\mathcal{H}_{\mathcal{D}}^\beta}$  and the balls

$$\mathcal{H}_{\mathcal{D}}^\beta(L) = \{g : \mathcal{D} \rightarrow \mathbb{R}, \|g\|_{\mathcal{H}_{\mathcal{D}}^\beta} \leq L\}, L > 0.$$

### 3.2 CONVERGENCE RESULTS FOR $\widehat{B}_T(x)$

We are ready to give our main result, namely a rate of convergence of  $\widehat{B}_T(x)$  for  $x$  restricted to a compact interval  $\mathcal{D}$ , uniformly over Hölder balls  $\mathcal{H}_{\mathcal{D}}^\beta(L)$  of (known) smoothness  $\beta$  intersected with  $\mathcal{B}_{b,C}$ . Define

$$w_T(B) = T^{\mathbf{1}_{\{\lambda_B = 2\rho_B\}}} \exp\left(-\min\{\lambda_B, 2\rho_B\} \frac{\beta - (\lambda_B/\rho_B - 1)_+/2}{2\beta + 1} T\right) \quad (17)$$

and note that when  $\rho_B \geq \lambda_B$ , we have  $w_T(B) = e^{-\lambda_B \frac{\beta}{2\beta+1} T} \approx \mathbb{E}[|\mathcal{J}_T|]^{-\beta/(2\beta+1)}$ .

**Theorem 7** (Upper rate of convergence). *Specify  $\widehat{B}_T$  with a kernel satisfying Assumption 6 for some  $n_0 > 1$  and*

$$h = \widehat{h}_T = \exp\left(-\widehat{\lambda}_T \frac{1}{2\beta+1} T\right) \quad (18)$$

for some  $\beta \in [1/2, n_0)$ . For every  $b, C > 0, L > 0$ , every compact interval  $\mathcal{D}$  in  $(0, \infty)$  (with non-empty interior) and every  $x \in \mathcal{D}$ ,

$$w_T(B)^{-1}(\widehat{B}_T(x) - B(x))$$

is  $\mathcal{B}_{b,C} \cap \mathcal{H}_{\mathcal{D}}^{\beta}(L)$ -tight for the parameter  $B$ .

We have a partial optimality result in a minimax sense. Define

$$\mathcal{B}_{b,C}^+ = \{B \in \mathcal{B}_{b,C}, \lambda_B \leq \rho_B\} \text{ and } \mathcal{B}_{b,C}^- = \{B \in \mathcal{B}_{b,C}, \rho_B \leq \lambda_B\}$$

so that  $\mathcal{B}_{b,c} = \mathcal{B}_{b,C}^+ \cup \mathcal{B}_{b,C}^-$ . We then have the following

**Theorem 8** (Lower rate of convergence over  $\mathcal{B}_{b,C}^+$ ). *Let  $\mathcal{D}$  be a compact interval in  $(0, \infty)$ . For every  $x \in \mathcal{D}$  and every positive  $b, C, \beta, L$ , there exists  $C' > 0$  such that*

$$\liminf_{T \rightarrow \infty} \inf_{\widehat{B}_T} \sup_B \mathbb{P}(e^{\lambda_B \frac{\beta}{2\beta+1} T} |\widehat{B}_T(x) - B(x)| \geq C') > 0,$$

where the supremum is taken among all  $B \in \mathcal{B}_{b,C}^+ \cap \mathcal{H}_{\mathcal{D}}^{\beta}(L)$  and the infimum is taken among all estimators.

We observe a conflict between the rate growth of the tree  $\lambda_B$  and its convergence rate to equilibrium  $\rho_B$ . On  $\mathcal{B}_{b,C}^+$  we retrieve the expected usual optimal rate of convergence  $\exp(-\lambda_B \frac{\beta}{2\beta+1} T) \approx \mathbb{E}[|\mathcal{J}_T|]^{-\beta/(2\beta+1)}$  whereas if  $\rho_B \leq \lambda_B$ , we obtain the deteriorated rate  $\exp(-\min\{\lambda_B, 2\rho_B\}(\beta - \frac{1}{2}(\frac{\lambda_B}{\rho_B} - 1))/(2\beta+1)T)$  and this rate is presumably not optimal, as discussed at length in Section 3.3 below.

### 3.3 DISCUSSION OF THE RESULTS

#### Rates of convergence

The ‘‘parametric case’’ for a constant division rate  $B(x) = b$  with  $b > 0$  has a statistical simpler structure, but also a nice probabilistic feature since the process  $t \rightsquigarrow |\partial\mathcal{T}_t|$ , *i.e.* the number of cells alive at time  $t$  is Markov. In that setting, explicit (asymptotic) information bounds are available (Athreya and Keiding [AK77]). In particular, the model is regular with asymptotic Fisher information of order  $e^{\lambda_B T}$ , thus the best-achievable (normalised) rate of convergence is  $e^{-\lambda_B T/2}$ . This is consistent with the minimax rate  $\exp(-\lambda_B \frac{\beta}{2\beta+1} T)$  that we obtain for the class  $\mathcal{H}_{\mathcal{D}}^{\beta}(L) \cap \mathcal{B}_{b,C}^+$ , and we retrieve the parametric rate by formally setting  $\beta = \infty$  in the previous formula.

However, this rate is strongly parameter dependent in the sense that it also depends on  $B$  via  $\lambda_B$ . This dependence is severe, since it appears at the same level as the smoothness exponent  $\beta/(2\beta+1)$  in the rate exponent  $\frac{\beta}{2\beta+1} \lambda_B$ . For instance, in the simplest case of a constant function

$B(x) = b$  for every  $x \geq 0$ , we have  $\lambda_B = (m-1)b$ , and we see that  $B$  ( $b$  here) plays at the same level as  $\beta/(2\beta+1)$ . This also has a non-trivial technical cost in establishing rates of convergence for the estimator  $\widehat{B}_T(x)$ : in order to minimise the bias-variance tradeoff, the (log)-bandwidth has to be chosen as  $-\lambda_B \frac{1}{2\beta+1} T(1+o(1))$  exactly, and this is achieved by the plug-in rule  $-\widehat{\lambda}_T \frac{1}{2\beta+1} T$  thanks to Proposition 17. We then have to carefully check that our estimator is not too sensitive to this further approximation, and this requires the analysis of the smoothness of the process  $h \rightsquigarrow \widehat{B}_{T,h}(x)$  where  $h$  is the bandwidth of  $\widehat{B}_T(x)$ , as shown in Proposition 17.

### Fast convergence to equilibrium in $\mathcal{B}_{b,C}^+$ versus slow convergence in $\mathcal{B}_{b,C}^-$

While we have an optimal rate of convergence over  $\mathcal{B}_{b,B}^+$ , the situation is unclear over  $\mathcal{B}_{b,C}^-$ . First, the convergence rate to equilibrium  $\rho_B$  should be replaced by an estimator and that would lead to extraneous difficulties. Even if we knew  $\rho_B$ , optimising the bias-variance trade-off in the proof of Theorem 7 would not lead to the expected rate  $\exp(-\min\{\lambda_B, 2\rho_B\} \frac{\beta}{2\beta+1} T)$  but to an intermediate rate that reads

$$\exp\left(-\min\{\lambda_B, 2\rho_B\} \frac{\min\{\max\{\rho_B/\lambda_B, 1/2\}, 1\}\beta}{2 \min\{\max\{\rho_B/\lambda_B, 1/2\}, 1\}\beta + 1} T\right), \quad (19)$$

and that continuously deteriorates as  $\rho_B$  separates  $\lambda_B$  from below. Let us also mention that the classes  $\mathcal{B}_{b,C}^+$  and  $\mathcal{B}_{b,C}^-$  are never trivial. To that end, define

$$\mathbb{B}_{b,m} = \{B \in \mathcal{B}_{b,m/(m-1)}, \forall x \geq 0 : B'(x) - B(x)^2 \leq 0\} \quad (20)$$

where  $m = \sum_{k \geq 2} k p_k$  is the mean number of children at each branching event.

**Proposition 9.** *For any  $b > 0$ , we have  $\mathbb{B}_{b,m} \subset \mathcal{B}_{b,m/(m-1)}^+$ . For every  $C > 2m(m+2)b/(m-1)$ ,  $\beta > 0$  and any compact interval  $\mathcal{D} \subset (0, \infty)$ , there exists  $B \in \mathcal{H}_{\mathcal{D}}^\beta$  such that  $B \in \mathcal{B}_{b,C}^-$  and  $B \notin \mathcal{B}_{b,C}^+$ .*

In the proof of Proposition 9 below we show a versatility in the choice of functions  $B$  that yield either fast or slow rate of convergence to equilibrium. Finally, one could (at least formally) replace  $\rho_B$  by  $\rho_B^*$ , the optimal geometric rate of convergence to equilibrium defined in (12) above, but that would only improve on the rate of convergence (19) replacing  $\rho_B$  by  $\rho_B^*$  which we do not know how to estimate, neither analytically nor statistically and the obtained result would still presumably not be optimal. This suggests a totally different estimation strategy – that we do not have at the moment – whenever convergence to equilibrium is slow.

### Other loss functions

If  $\mathcal{K} \subset \mathring{\mathcal{D}}$  is a closed interval ( $\mathring{\mathcal{D}}$  denotes the interior of  $\mathcal{D}$ ), then Theorem 7 also holds uniformly in  $x \in \mathcal{K}$ . So we also have that

$$w_T(B)^{-2} \int_{\mathcal{K}} (\widehat{B}_T(x) - B(x))^2 dx$$

is  $\mathcal{B}_{b,C} \cap \mathcal{H}_{\mathcal{D}}^\beta(L)$ -tight for the parameter  $B$ . For integrated squared error-loss, we could weaken the smoothness constraint  $B \in \mathcal{H}_{\mathcal{D}}^\beta(L)$  to Sobolev smoothness (see *e.g.* [Tsy09]) when the smoothness

is measured in  $L^2$ -norm. An extension of Theorem 8 can be obtained likewise.

### Smoothness adaptation

Our estimator  $\widehat{B}_T(x)$  is not  $\beta$ -adaptive, in the sense that the choice of the  $\mathcal{B}_{b,C}^+$ -optimal (log) bandwidth  $-\widehat{\lambda}_T \frac{1}{2\beta+1} T$  still depends on  $\beta$ , which is unknown in principle. In the numerical implementation Section 4 below, we address this issue from a practical point of view. However, a theoretical result is still needed. The classical analysis of adaptive (or other) kernel methods à la Lepski for instance shows that this boils down to proving concentration inequalities of the type

$$\mathbb{P}\left(|\mathcal{E}^T(\mathring{\mathcal{J}}_T, g_h) - \mathring{\mathcal{E}}_B(g_h)| \geq e^{\lambda_B T/2} c(q, T)\right) \leq e^{-q\lambda_B T}, \quad q > 0, \quad (21)$$

where, for  $0 < h^{-1} \leq e^{\lambda_B T}$ , the test function  $g_h$  has the form  $g_h(y) = h^{-1/2} g(h^{-1}(x - y))$  with  $x \in \mathcal{D}$  and  $g \in \mathcal{L}_C$ . The threshold  $c(q, T)$  should be of order  $q\lambda_B T$  and would inflate the risk by a slow term (of order  $T$ ). By a suitable choice of  $q$ , it would then be possible to obtain adaptation for  $\beta$  in compact intervals. Concentration inequalities like (21) have been explored in [BPDG14] in discrete time. To the best of our knowledge, such inequalities are not yet available in continuous time and lie beyond the scope of the chapter.

### Information from $\mathring{\mathcal{J}}_T$ versus $\partial\mathcal{J}_T$

In the regime  $B \in \mathcal{B}_{b,C}^+$ , having

$$\partial\mathcal{E}_B(g) = \lambda_B \frac{m}{m-1} \int_0^\infty g(x) e^{-\lambda_B x} \exp\left(-\int_0^x B(y) dy\right) dx$$

and ignoring the fact that the constants  $m$  and  $\lambda_B$  are unknown (or rather knowing that they can be estimated at the superoptimal rate  $e^{\lambda_B T/2}$ ), we can anticipate that by picking a suitable test function  $g$  mimicking a delta function  $g(x) \approx \delta_x$ , the information about  $B(x)$  can only be inferred through  $\exp(-\int_0^x B(y) dy)$ , which imposes to further take a derivative hence some ill-posedness.

We can briefly make all these arguments more precise (still in the regime  $B \in \mathcal{B}_{b,C}^+$ ): we assume that we have estimators of  $\widehat{m}_T$  of  $m$  and  $\widehat{\lambda}_T$  of  $\lambda_B$  (using the ones defined in (13) and (14) or by any other means) that converge with rate  $T^{-1} e^{\lambda_B T/2}$  as in Proposition 5. Consider the quantity

$$\widehat{f}_{h,T}(x) = -\mathcal{E}^T\left(\partial\mathcal{J}_T, \frac{\widehat{m}_T - 1}{\widehat{\lambda}_T \widehat{m}_T} (K_h)'(x - \cdot)\right)$$

for a kernel satisfying Assumption 6. By Theorem 3 and integrating by part, we readily see that

$$\widehat{f}_{h,T} \rightarrow -\partial\mathcal{E}_B\left(\frac{m-1}{\lambda_B m} (K_h)'(x - \cdot)\right) = \int_0^\infty K_h(x-y) f_{B+\lambda_B}(y) dy \quad (22)$$

in probability as  $T \rightarrow \infty$ , where  $f_{B+\lambda_B}$  is the density associate to the division rate  $x \rightsquigarrow B(x) + \lambda_B$ . On the one hand, it is not difficult to show that Proposition 15 (used in the proof of Theorem 7 below) is valid when substituting  $\mathring{\mathcal{J}}_T$  by  $\partial\mathcal{J}_T$ , so we expect (although not formally established) the rate of convergence in (22) to be of order  $h^{-3/2} e^{\lambda_B T/2}$  since we take the derivative of the kernel

$K_h$ . On the other hand, the limit  $\int_0^\infty K_h(x-y)f_{B+\lambda_B}(y)dy$  approximates  $f_{B+\lambda_B}(x)$  with an error of order  $h^\beta$  if  $B \in \mathcal{H}_D^\beta$ . Balancing the two error terms in  $h$ , we see that we can estimate  $f_{B+\lambda_B}(x)$  with an error of (presumably optimal) order  $\exp(-\lambda_B \frac{\beta}{2\beta+3}T)$ . Due to the fact that the denominator in representation (3) can be estimated with parametric error rate  $\exp(-\lambda_B T/2)$  (possibly up to polynomially slow terms in  $T$ ), we end up with the rate of estimation  $\exp(-\lambda_B \frac{\beta}{2\beta+3}T)$  for  $B(x)$  as well, and that can be related to an ill-posed problem of order 1 (see for instance [Tsy09]).

This phenomenon, namely the structure of an ill-posed problem of order 1 in restriction to data alive at time  $T$ , has already been observed in other settings: for the estimation of a size-division rate from living cells at a given large time in Doumic *et al.* [DPZ09, DHRR12] or for the estimation of the dislocation measure for a homogeneous fragmentation in Hoffmann and Krell [HK11]. Note also that this phenomenon does not appear in parametric estimation, since the number of data in  $\mathring{\mathcal{J}}_T$  and  $\partial\mathcal{J}_T$  are of the same order of magnitude (or put differently, the rates in Theorems 3 and 4 are the same and govern the rate of estimation of a one dimensional parameter).

## 4 NUMERICAL IMPLEMENTATION

We assume that each cell  $u \in \mathcal{U}$  has exactly two children at each division ( $p_2 = 1$ ). This can model the evolution of a population of cells reproducing by binary divisions, as described deterministically by (1). We pick a trial division rate  $B$  defined analytically by

$$B(x) = \begin{cases} \frac{1}{3}x^3 - \frac{7}{8}x^2 + \frac{5}{8}x + \frac{4}{10} & \text{if } 0 \leq x \leq \frac{3}{2} \\ \frac{119}{160} - \frac{1}{4} \exp\left(-\left(x - \frac{3}{2}\right)\right) & \text{if } x > \frac{3}{2} \end{cases}$$

and represented in Figure 5 (bold red line). We have  $b \leq B(x) \leq \frac{m}{m-1}b$  for any  $x \geq 0$  for  $b = 0.4$  and  $m = 2$  and the lifetime density  $f_B$  is non-increasing (except in a vicinity of zero). Given  $T > 0$  we simulate the lifetime of the rooted cell  $\zeta_\emptyset$  with probability density  $f_B$  and set  $d_\emptyset = \zeta_\emptyset$ . For  $u \in \mathcal{U}$  such that  $d_u > T$ , we do not simulate the lifetimes of its descendants since they are not in the observation scheme  $\mathring{\mathcal{J}}_T \cup \partial\mathcal{J}_T$ . For  $u \in \mathcal{U}$  such that  $d_u \leq T$  we simulate  $\zeta_{u0}$  and  $\zeta_{u1}$  independently with probability density  $f_B$ ; we set  $d_{u0} := d_u + \zeta_{u0}$  and  $d_{u1} := d_u + \zeta_{u1}$ . Using R software, we generate  $M = 100$  trees up to time  $T = 23$ , so that the mean number of observations  $|\mathring{\mathcal{J}}_T|$  is sufficiently large. (Note that for a binary tree, we always have the identity  $|\partial\mathcal{J}_T| = |\mathring{\mathcal{J}}_T| + 1$ .) Figure 4 represents a typical observation scheme with continuous or discrete representation. The (random) number of observations fluctuates a lot as shown in Table 2 where some elementary statistics are given.

| Min.  | 1st Qu. | Med.   | Mean    | 3rd Qu. | Max.    | Std.    |
|-------|---------|--------|---------|---------|---------|---------|
| 3 726 | 43 930  | 96 480 | 115 760 | 144 100 | 561 200 | 102 408 |

Table 2: Fluctuations of the number of observations  $|\mathring{\mathcal{J}}_T|$  for  $M = 100$  Monte-Carlo continuous time trees observed up to time  $T = 23$ .

We take a Gaussian kernel  $K(x) = (2\pi)^{-1/2} \exp(-x^2/2)$  and the bandwidth  $\hat{h}_T$  is chosen here according to the rule-of-thumb  $1.06\hat{\sigma}|\mathring{\mathcal{J}}_T|^{-1/5}$  where  $\hat{\sigma}$  is the empirical standard deviation of

$(\zeta_u, u \in \mathcal{J}_T^\circ)$ . We also implemented standard cross-validation with less success. We evaluate  $\widehat{B}_T$  on a regular grid of  $\mathcal{D} = [0.25, 0.5]$  with mesh  $\Delta x = 0.01$ . For each sample we compute the empirical error

$$e_i = \frac{\|\widehat{B}_T^{(i)} - B\|_{\Delta x}}{\|B\|_{\Delta x}}, \quad i = 1, \dots, M,$$

where  $\|\cdot\|_{\Delta x}$  denotes the discrete norm over the numerical sampling.

Table 3 displays the mean-empirical error  $\bar{e} = M^{-1} \sum_{i=1}^M e_i$  together with the empirical standard deviation  $(M^{-1} \sum_{i=1}^M (e_i - \bar{e})^2)^{1/2}$ .

| $T$                    | 13     | 15     | 17     | 19     | 21     | 23      |
|------------------------|--------|--------|--------|--------|--------|---------|
| Mean $ \mathcal{J}_T $ | 652    | 1 847  | 5 202  | 14 634 | 41 151 | 115 760 |
| $\bar{e}$              | 0.1624 | 0.1046 | 0.0735 | 0.0448 | 0.0307 | 0.0178  |
| Std. dev.              | 0.1052 | 0.0764 | 0.0599 | 0.0260 | 0.0197 | 0.0092  |

Table 3: Mean empirical relative error  $\bar{e}$  and its standard deviation, with respect to  $T$ , for the division rate  $B$  reconstructed over the interval  $\mathcal{D} = [0.25, 2.5]$  by the estimator  $\widehat{B}_T$ .

The comparison of the density of interest  $f_B$  and the biased density  $f_{H_B}$  on Figure 5 highlights the bias selection since  $f_{H_B}$  gives more weight to small lifetimes than  $f_B$ .

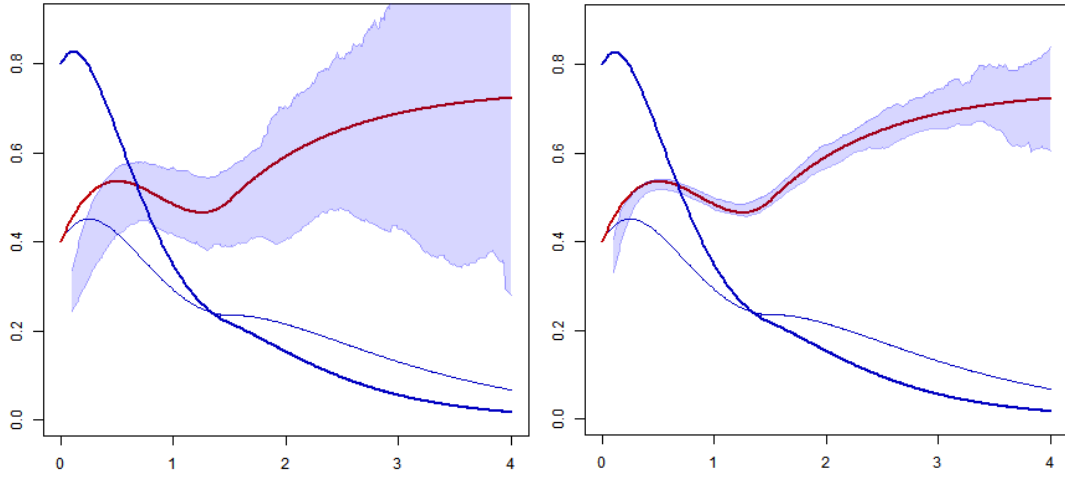


Figure 5: Reconstruction of  $B$  over  $\mathcal{D} = [0.1, 4]$  with 95%-level confidence bands constructed over  $M = 100$  Monte-Carlo continuous time trees. In bold red line:  $x \rightsquigarrow B(x)$ ; in bold blue line:  $f_{H_B}$ ; in thin blue line:  $f_B$  (on the same  $y$ -axis scale). Left:  $T = 15$ . Right:  $T = 23$ .

The error deteriorates as  $x$  grows since the biased density  $f_{H_B}$  (bold blue line - we approximate the Malthus parameter using (6) and we find  $\lambda_B \approx 0.5173$ ) decreases, see Figure 5. The larger  $T$ , the better the reconstruction at a visual level, as shown on Figure 5 where 95%-level confidence bands are built so that for each point  $x$ , the lower and upper bounds include 95% of the estimators  $(\widehat{B}_T^{(i)}(x), i = 1 \dots M)$ . Close to 0,  $B(x)$  does not lie in the confidence band: our estimator exhibits a large bias there, and this is presumably due to a boundary effect.

The error is close to  $\exp(-2\lambda_B T/5)$  as expected: indeed, for a kernel of order  $n_0$ , the bias term in density estimation is of order  $h^{\beta \wedge (n_0+1)}$ . Given that  $B$  is smooth in our example, we rather expect  $\exp(-\lambda_B \frac{(n_0+1)}{2(n_0+1)+1} T) = \exp(-2\lambda_B T/5)$  for the Gaussian kernel with  $n_0 = 1$  that we use here, and this is consistent with what we observe in Figure 6.

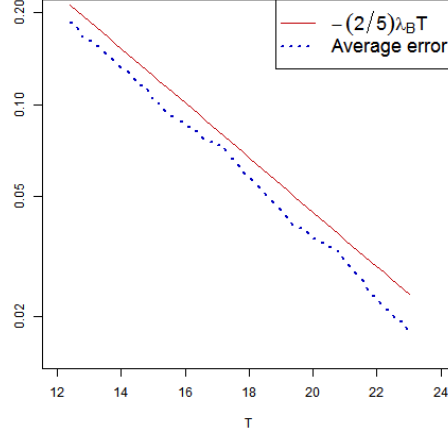


Figure 6: The log-average relative empirical error over  $M = 100$  Monte-Carlo continuous time trees vs.  $T$  (i.e. the log-rate) for  $x \rightsquigarrow B(x)$  reconstructed over  $\mathcal{D} = [0.25, 2.5]$  with  $x \rightsquigarrow \hat{B}_T(x)$  (dashed blue line) compared to the expected log-rate (solid red line).

## 5 PROOFS

For a locally integrable  $B : [0, \infty) \rightarrow [0, \infty)$  such that  $\int^\infty B(y)dy = \infty$ , recall that we set

$$f_B(x) = B(x)e^{-\int_0^x B(y)dy}, \quad x \geq 0.$$

Recall that  $H_B$  is characterised by

$$f_{H_B}(x) = me^{-\lambda_B x} f_B(x), \quad x \geq 0.$$

### 5.1 PRELIMINARIES

#### Many-to-one formulae

For  $u \in \mathcal{U}$ , we write  $\zeta_u^t$  for the age of the cell  $u$  at time  $t \in I_u = [b_u, d_u)$ , i.e.  $\zeta_u^t = (t - b_u)\mathbf{1}_{\{t \in I_u\}}$ . We extend  $\zeta_u^t$  over  $[0, b_u)$  by setting  $\zeta_u^t = \zeta_{u(t)}^t$ , where  $u(t)$  is the ancestor of  $u$  living at time  $t$ , defined by  $u(t) = v$  if  $v \preceq u$  and  $(v, t) \in \mathbb{T}$ . For  $t \geq d_u$  we set  $\zeta_u^t = \zeta_u$ . Note that  $\zeta_u^T = \zeta_u$  on the event  $u \in \mathring{\mathcal{J}}_T$ .

Let  $(\chi_t)_{t \geq 0}$  and  $(\tilde{\chi}_t)_{t \geq 0}$  denote the one-dimensional Markov processes with infinitesimal generators (densely defined on continuous functions vanishing at infinity)  $\mathcal{A}_B$  and  $\mathcal{A}_{H_B}$  respectively,

where

$$\mathcal{A}_B g(x) = g'(x) + B(x)(g(0) - g(x))$$

and such that  $\mathbb{P}(\chi_0 = 0) = \mathbb{P}(\tilde{\chi}_0 = 0) = 1$ . We also denote by  $(P_{H_B}^t)_{t \geq 0}$  the Markov semigroup associated to  $\mathcal{A}_{H_B}$ .

**Proposition 10** (Many-to-one formulae). *For any  $g \in \mathcal{L}_C$ , we have*

$$\mathbb{E} \left[ \sum_{u \in \partial \mathcal{T}_T} g(\zeta_u^T) \right] = \frac{e^{\lambda_B T}}{m} \mathbb{E} \left[ g(\tilde{\chi}_T) B(\tilde{\chi}_T)^{-1} H_B(\tilde{\chi}_T) \right], \quad (23)$$

and

$$\mathbb{E} \left[ \sum_{u \in \dot{\mathcal{T}}_T} g(\zeta_u^T) \right] = \mathbb{E} \left[ \sum_{u \in \dot{\mathcal{T}}_T} g(\zeta_u) \right] = \frac{1}{m} \int_0^T e^{\lambda_B s} \mathbb{E} \left[ g(\tilde{\chi}_s) H_B(\tilde{\chi}_s) \right] ds. \quad (24)$$

In order to compute rates of convergence, we will also need many-to-one formulae over pairs of individuals. We can pick two individuals in the same lineage or over forks, *i.e.* over pairs of individuals that are not in the same lineage. If  $u, v \in \mathcal{U}$ ,  $u \wedge v$  denote their most recent common ancestor. Define

$$\mathcal{FU} = \{(u, v) \in \mathcal{U}^2, |u \wedge v| < |u| \wedge |v|\} \text{ and } \mathcal{FT} = \mathcal{FU} \cap \mathcal{T}^2.$$

Introduce also  $\bar{m} = \sum_{i \neq j} \sum_{k \geq i \vee j} p_k$  which is finite by Assumption 2.

**Proposition 11** (Many-to-one formulae over pairs). *For any  $g \in \mathcal{L}_C$ , we have*

$$\mathbb{E} \left[ \sum_{\substack{u, v \in \partial \mathcal{T}_T, \\ u \neq v}} g(\zeta_u^T) g(\zeta_v^T) \right] = \frac{\bar{m}}{m^3} \int_0^T e^{\lambda_B s} \left( e^{\lambda_B(T-s)} P_{H_B}^{T-s} \left( g \frac{H_B}{B} \right) (0) \right)^2 P_{H_B}^s H_B(0) ds, \quad (25)$$

$$\mathbb{E} \left[ \sum_{(u, v) \in \mathcal{FT} \cap \dot{\mathcal{T}}_T^2} g(\zeta_u) g(\zeta_v) \right] = \frac{\bar{m}}{m^3} \int_0^T e^{\lambda_B s} \left( \int_0^{T-s} e^{\lambda_B t} P_{H_B}^t (g H_B)(0) dt \right)^2 P_{H_B}^s H_B(0) ds, \quad (26)$$

and

$$\mathbb{E} \left[ \sum_{\substack{u, v \in \dot{\mathcal{T}}_T, \\ u < v}} g(\zeta_u) g(\zeta_v) \right] = \int_0^T e^{\lambda_B s} \left( \int_0^{T-s} e^{\lambda_B t} P_{H_B}^t (g H_B)(0) dt \right) P_{H_B}^s (g H_B)(0) ds. \quad (27)$$

The identity (25) is a particular case of Lemma 3.9 of Cloez [Clo11]. In order to obtain identity (26), we closely follow the method of Bansaye *et al.* [BDMT11]. Although the setting in [BDMT11] is much more general than ours, it formally only applies for exponential renewal times (corresponding to constant functions  $B$ ) so we need to slightly accommodate their proof. The same ideas enable us to prove (27). This is set out in details in the appendix.



### Geometric ergodicity of the auxiliary Markov process

Define the probability measure

$$\mu_B(x)dx = c_B \exp\left(-\int_0^x H_B(y)dy\right)dx \text{ for } x \geq 0.$$

We have the fast convergence of  $P_{H_B}^T$  toward  $\mu_B$  as  $T \rightarrow \infty$ . More precisely,

**Proposition 12.** *Let  $\rho_B = \inf_x H_B(x)$ . For any  $B \in \mathcal{B}_{b,C}$ ,  $g \in \mathcal{L}_{C'}$ ,  $t \geq 0$  and  $x \in (0, \infty)$ , we have*

$$\left|P_{H_B}^t g(x) - \int_0^\infty g(y)\mu_B(y)dy\right| \leq 2 \sup_y |g(y)| \exp(-\rho_B t).$$

*Proof.* First, one readily checks that the process  $(\tilde{\chi}_t)_{t \geq 0}$  is Harris recurrent (since 0 is an atom). Then by Kac's formula we derive that the unique invariant probability measure  $\mu_B(\cdot)$  is proportional to  $\mathbb{E}\left[\int_0^{\tau_0} \mathbf{1}(\tilde{\chi}_t)dt\right]$  where  $\tau_0$  is the duration before the first jump. It is now sufficient to show

$$\|Q_B^{x,t} - \mu_B\|_{TV} \leq \exp(-\rho_B t)$$

where  $Q_B^{x,t}$  denotes the law of the Markov process with infinitesimal generator  $\mathcal{A}_{H_B}$  started from  $x$  at time  $t = 0$  and  $\|\cdot\|_{TV}$  is the total variation norm between probability measures. Let  $N(ds dt)$  be a Poisson random measure with intensity  $ds \otimes dt$  on  $[0, \infty) \times [0, \infty)$ . Define on the same probability space two random processes  $(Y_t)_{t \geq 0}$  and  $(Z_t)_{t \geq 0}$  such that

$$\begin{aligned} Y_t &= x + t - \int_0^t \int_0^\infty Y_{s-} \mathbf{1}_{\{z \leq H_B(Y_{s-})\}} N(dz ds), \quad t \geq 0, \\ Z_t &= Z_0 + t - \int_0^t \int_0^\infty Z_{s-} \mathbf{1}_{\{z \leq H_B(Z_{s-})\}} N(dz ds), \quad t \geq 0, \end{aligned}$$

where  $Z_0$  is a random variable with distribution  $\mu_B$ . We have that both  $(Y_t)_{t \geq 0}$  and  $(Z_t)_{t \geq 0}$  are Markov processes driven by the same Poisson random measure with generator  $\mathcal{A}_{H_B}$ . Moreover, if  $N$  has a jump in  $[0, t) \times [0, \inf_x H_B(x)]$ , then  $Y_t$  and  $Z_t$  both necessarily start from 0 after this jump and coincide further on. It follows that

$$\mathbb{P}(Y_t \neq Z_t) \leq \mathbb{P}\left(\int_0^t \int_0^{\inf_x H_B(x)} N(ds dt) = 0\right) = \exp(-\inf_x H_B(x)t) = \exp(-\rho_B t).$$

Observing that  $Y_t$  and  $Z_t$  have distribution  $Q_B^{x,t}$  and  $\mu_B$  respectively, we conclude thanks to the fact that  $\|Q_B^{x,t} - \mu_B\|_{TV} \leq \mathbb{P}(Y_t \neq Z_t)$ .  $\square$

## 5.2 PROOF OF THEOREMS 3 AND 4

In order to ease notation, when no confusion is possible, we abbreviate  $\mathcal{B}_{b,C}$  by  $\mathcal{B}$  and  $\mathcal{L}_C$  by  $\mathcal{L}$ .

*Proof of Theorem 3.* Writing

$$e^{\min\{\lambda_B/2, \rho_B\}T} (\mathcal{E}^T(\partial \mathcal{J}_T, g) - \partial \mathcal{E}_B(g)) = \frac{e^{\lambda_B T}}{|\partial \mathcal{J}_T|} e^{(\min\{\lambda_B/2, \rho_B\} - \lambda_B)T} \sum_{u \in \partial \mathcal{J}_T} (g(\zeta_u^T) - \partial \mathcal{E}_B(g)),$$

Theorem 3 is then a consequence of the following two facts: first we claim that

$$e^{\lambda_B T} |\partial \mathcal{T}_T|^{-1} \rightarrow W_B \text{ in probability as } T \rightarrow \infty, \quad (28)$$

uniformly in  $B \in \mathcal{B}$ , where the random variable  $W_B$  satisfies  $\mathbb{P}(W_B > 0) = 1$ , and second, for  $B \in \mathcal{B}$  and  $g \in \mathcal{L}$ , we claim that the following estimate holds:

$$\mathbb{E} \left[ \left( \sum_{u \in \partial \mathcal{T}_T} (g(\zeta_u^T) - \partial \mathcal{E}_B(g)) \right)^2 \right] \lesssim e^{(2\lambda_B - \min\{\lambda_B, 2\rho_B\})T}, \quad (29)$$

where  $\lesssim$  means up to a constant (possibly varying from line to line) that only depends on  $\mathcal{B}$  and  $\mathcal{L}$  and up to a multiplicative slow term of order  $T$  in the case  $\lambda_B = 2\rho_B$ .

*Step 1.* The convergence (28) is a consequence of the following lemma:

**Lemma 13.** *For every  $B \in \mathcal{B}$ , there exists  $\widetilde{W}_B$  with  $\mathbb{P}(\widetilde{W}_B > 0) = 1$  such that*

$$\mathbb{E} \left[ \left( \frac{|\partial \mathcal{T}_T|}{\mathbb{E}[|\partial \mathcal{T}_T|]} - \widetilde{W}_B \right)^2 \right] \rightarrow 0 \text{ as } T \rightarrow \infty, \quad (30)$$

uniformly in  $B \in \mathcal{B}$  and

$$\kappa_B^{-1} e^{\lambda_B T} \mathbb{E}[|\partial \mathcal{T}_T|] \rightarrow 1 \text{ as } T \rightarrow \infty, \quad (31)$$

uniformly in  $B \in \mathcal{B}$ , where  $\kappa_B^{-1} = \lambda_B \frac{m}{m-1} \int_0^\infty \exp(-\int_0^x H_B(y) dy) dx$ .

Lemma 13 is well known, and follows from classical renewal arguments, see Chapter 6 in the book of Harris [Har63]. Only the uniformity in  $B \in \mathcal{B}$  requires an extra argument, but with a uniform version of the key renewal theorem of [Tsi13], it readily follows from the proof of Harris, so we omit it. Note that (30) and (31) entail the convergence  $e^{\lambda_B T} |\partial \mathcal{T}_T|^{-1} \rightarrow \kappa_B \widetilde{W}_B^{-1} = W_B$  in probability as  $T \rightarrow \infty$  uniformly in  $B \in \mathcal{B}$ , and this entails (28).

*Step 2.* We now turn to the proof of (29). Without loss of generality, we may (and will) assume that  $\partial \mathcal{E}_B(g) = 0$ . We have

$$\mathbb{E} \left[ \left( \sum_{u \in \partial \mathcal{T}_T} g(\zeta_u^T) \right)^2 \right] = \mathbb{E} \left[ \sum_{u \in \partial \mathcal{T}_T} g(\zeta_u^T)^2 \right] + \mathbb{E} \left[ \sum_{\substack{u, v \in \partial \mathcal{T}_T, \\ u \neq v}} g(\zeta_u^T) g(\zeta_v^T) \right] = I + II,$$

say. By (23) in Proposition 10, we write

$$\begin{aligned} I &= \frac{e^{\lambda_B T}}{m} \mathbb{E} \left[ g(\widetilde{\chi}_T)^2 B(\widetilde{\chi}_T)^{-1} H_B(\widetilde{\chi}_T) \right] \\ &\leq \frac{e^{\lambda_B T}}{m} \int_0^\infty g(x)^2 \frac{H_B(x)}{B(x)} \mu_B(x) dx + \frac{e^{\lambda_B T}}{m} \left| P_{H_B}^T(g^2 \frac{H_B}{B})(0) - \int_0^\infty g(x)^2 \frac{H_B(x)}{B(x)} \mu_B(x) dx \right|. \end{aligned}$$

Since  $g \in \mathcal{L}$  and  $B \in \mathcal{B}$ , we successively have

$$m^{-1} \int_0^\infty g(x)^2 \frac{H_B(x)}{B(x)} \mu_B(x) dx \lesssim 1 \text{ and } g(x)^2 \frac{H_B(x)}{B(x)} \lesssim 1.$$

Note that for  $B \in \mathcal{B}$ , we have

$$H_B(x) = \frac{B(x)}{\int_x^\infty B(y)e^{-\lambda_B(y-x)} \exp(-\int_x^y B(u)du)dy} \leq \frac{b^2 \max\{C, 1\}}{\lambda_B + b \max\{C, 1\}}.$$

We also have  $\lambda_B \leq \lambda_{\tilde{B}}$  as soon as  $B(x) \leq \tilde{B}(x)$  for all  $x$  (see for instance the proof of Proposition 9) so  $\inf_{B \in \mathcal{B}} \lambda_B > 0$  and the uniformity in the above estimates follows likewise. Applying Proposition 12 we derive

$$\left| P_{H_B}^T(g^2 \frac{H_B}{B})(0) - \int_0^\infty g(x)^2 \frac{H_B(x)}{B(x)} \mu_B(x) dx \right| \lesssim 1,$$

and we conclude that  $I \lesssim e^{\lambda_B T} \leq e^{(2\lambda_B - \min\{\lambda_B, 2\rho_B\})T}$ . By (25) of Proposition 11 we have

$$II = \frac{\bar{m}e^{2\lambda_B T}}{m^3} \int_0^T e^{-\lambda_B s} \left( P_{H_B}^{T-s}(g \frac{H_B}{B})(0) \right)^2 P_{H_B}^s H_B(0) ds.$$

Since  $B \in \mathcal{B}$  and  $g \in \mathcal{L}$ , the estimates  $P_{H_B}^s H_B(0) \lesssim 1$  and  $|g(x)| \frac{H_B(x)}{B(x)} \lesssim 1$  hold true. Applying Proposition 12 to the test function  $g(x) \frac{H_B(x)}{B(x)}$  which has vanishing integral under  $\mu_B$ , we obtain

$$\left| P_{H_B}^{T-s}(g \frac{H_B}{B})(0) \right| \lesssim e^{-\rho_B(T-s)}$$

hence

$$|II| \lesssim e^{2\lambda_B T} \int_0^T e^{-\lambda_B s} e^{-2\rho_B(T-s)} ds \lesssim \begin{cases} e^{\lambda_B T} & \text{if } 2\rho_B \geq \lambda_B \\ e^{2(\lambda_B - \rho_B)T} & \text{if } 2\rho_B < \lambda_B, \end{cases}$$

up to a multiplicative slow term of order  $T$  when  $2\rho_B = \lambda_B$ . Note also that the estimate is uniform in  $B \in \mathcal{B}$  since  $\inf_{B \in \mathcal{B}} \lambda_B > 0$  and  $\inf_{B \in \mathcal{B}} \rho_B > 0$ . We conclude  $|II| \lesssim e^{(2\lambda_B - \min\{\lambda_B, 2\rho_B\})T}$ .  $\square$

*Proof of Theorem 4.* The proof goes along the same line but is slightly more intricate. First, we implicitly work on the event  $\{|\mathring{\mathcal{J}}_T| \geq 1\}$  which has probability that goes to 1 as  $T \rightarrow \infty$ , uniformly in  $B \in \mathcal{B}$ . We again write

$$e^{\min\{\lambda_B/2, \rho_B\}T} (\mathcal{E}(\mathring{\mathcal{J}}_T, g) - \mathring{\mathcal{E}}_B(g)) = \frac{e^{\lambda_B T}}{|\mathring{\mathcal{J}}_T|} e^{(\min\{\lambda_B/2, \rho_B\} - \lambda_B)T} \sum_{u \in \mathring{\mathcal{J}}_T} (g(\zeta_u^T) - \mathring{\mathcal{E}}_B(g)),$$

and we claim that

$$e^{\lambda_B T} |\mathring{\mathcal{J}}_T|^{-1} \rightarrow W'_B > 0 \text{ in probability as } T \rightarrow \infty, \quad (32)$$

uniformly in  $B \in \mathcal{B}$ , where  $W'_B$  satisfies  $\mathbb{P}(W'_B > 0) = 1$  and that the following estimate holds:

$$\mathbb{E} \left[ \left( \sum_{u \in \mathring{\mathcal{J}}_T} (g(\zeta_u^T) - \mathring{\mathcal{E}}_B(g)) \right)^2 \right] \lesssim e^{(2\lambda_B - \min\{\lambda_B, 2\rho_B\})T}, \quad (33)$$

uniformly in  $B \in \mathcal{B}$  and  $g \in \mathcal{L}$ . In the same way as in the proof of Theorem 3, (32) is a consequence of the following classical result, which can be obtained in the same way as for Lemma 13 and proof

which we omit.

**Lemma 14.** *For every  $B \in \mathcal{B}$ , there exists  $\widetilde{W}'_B > 0$  with  $\mathbb{P}(\widetilde{W}'_B > 0) = 1$  such that*

$$\mathbb{E}\left[\left(\frac{|\mathring{\mathcal{J}}_T|}{\mathbb{E}[|\mathring{\mathcal{J}}_T|]} - \widetilde{W}'_B\right)^2\right] \rightarrow 0 \text{ as } T \rightarrow \infty,$$

uniformly in  $B \in \mathcal{B}$  and

$$(\kappa'_B)^{-1} e^{\lambda_B T} \mathbb{E}[|\mathring{\mathcal{J}}_T|] \rightarrow 1 \text{ as } T \rightarrow \infty,$$

uniformly in  $B \in \mathcal{B}$ , where  $(\kappa'_B)^{-1} = \lambda_B m \int_0^\infty \exp(-\int_0^x H_B(y) dy) dx$ .

It remains to prove (33). We again assume without loss of generality that  $\mathring{\mathcal{E}}_B(g) = 0$  and we plan to use the following decomposition:

$$\mathbb{E}\left[\left(\sum_{u \in \mathring{\mathcal{J}}_T} g(\zeta_u)\right)^2\right] = I + II + III, \quad (34)$$

with

$$I = \mathbb{E}\left[\sum_{u \in \mathring{\mathcal{J}}_T} g(\zeta_u)^2\right],$$

$$II = \mathbb{E}\left[\sum_{(u,v) \in \mathcal{F}\mathcal{T} \cap \mathring{\mathcal{J}}_T^2} g(\zeta_u)g(\zeta_v)\right]$$

and

$$III = 2\mathbb{E}\left[\sum_{\substack{u,v \in \mathring{\mathcal{J}}_T, \\ u < v}} g(\zeta_u)g(\zeta_v)\right].$$

*Step 1.* By (24) of Proposition 10, we have

$$I = \frac{1}{m} \int_0^T e^{\lambda_B s} \mathbb{E}[g(\widetilde{\chi}_s)^2 H_B(\widetilde{\chi}_s)] ds,$$

In the same way as for the term  $I$  in the proof of Theorem 3, we readily check that  $g \in \mathcal{L}$  and  $B \in \mathcal{B}$  guarantee that  $\mathbb{E}[g(\widetilde{\chi}_s)^2 H_B(\widetilde{\chi}_s)] \lesssim 1$  therefore

$$I \lesssim e^{\lambda_B T} \leq e^{(2\lambda_B - \min\{\lambda_B, 2\rho_B\})T}.$$

*Step 2.* By (26) of Proposition 11, we have

$$II = \frac{\bar{m}}{m^3} \int_0^T e^{\lambda_B s} \left( \int_0^{T-s} e^{\lambda_B t} P_{H_B}^t(gH_B)(0) dt \right)^2 P_{H_B}^s(H_B)(0) ds.$$

We work as for the term  $II$  in the proof of Theorem 3: we successively have  $P_{H_B}^s(H_B)(0) \lesssim 1$  and  $|P_{H_B}^t(gH_B)(0)| \lesssim \exp(-\rho_B t)$  by Proposition 12 and the fact that  $gH_B$  has vanishing integral under  $\mu_B$ . Therefore

$$|II| \lesssim \int_0^T e^{\lambda_B s} \left( \int_0^{T-s} e^{(\lambda_B - \rho_B)t} dt \right)^2 ds \lesssim \begin{cases} e^{\lambda_B T} & \text{if } 2\rho_B \geq \lambda_B \\ e^{2(\lambda_B - \rho_B)T} & \text{if } 2\rho_B < \lambda_B \end{cases}$$

up to a multiplicative slow term of order  $T$  when  $2\rho_B = \lambda_B$ . We conclude

$$|II| \lesssim e^{(2\lambda_B - \min\{\lambda_B, 2\rho_B\})T}$$

likewise.

*Step 3.* By (27) of Proposition 11, we have

$$|III| \leq \int_0^T e^{\lambda_B s} \left| \int_0^{T-s} e^{\lambda_B t} P_{H_B}^t(gH_B)(0) dt \right| P_{H_B}^s(|gH_B)(0)| ds.$$

In the same way as for the term  $II$ , we have  $|P_{H_B}^s(|gH_B)(0)| \lesssim 1$  and  $|P_{H_B}^t(gH_B)(0)| \lesssim \exp(-\rho_B t)$ . Therefore

$$|III| \lesssim \int_0^T e^{\lambda_B s} \left( \int_0^{T-s} e^{(\lambda_B - \rho_B)t} dt \right) ds \lesssim e^{\lambda_B T} \leq e^{(2\lambda_B - \min\{\lambda_B, 2\rho_B\})T}.$$

□

### 5.3 PROOF OF PROPOSITION 5

Conditional on  $\mathring{\mathcal{J}}_T$ , the random variables  $(\nu_u, u \in \mathring{\mathcal{J}}_T)$  are independent, with common distribution  $p_k$ . It follows that

$$\mathbb{E}[(\widehat{m}_T - m)^2 | \mathring{\mathcal{J}}_T] \leq |\mathring{\mathcal{J}}_T|^{-1} \sum_k k^2 p_k.$$

Since  $e^{\lambda_B T} |\mathring{\mathcal{J}}_T|^{-1}$  is  $\mathcal{B}$ -tight thanks to Lemma 14, we obtain the result for  $e^{\lambda_B T/2}(\widehat{m}_T - m)$ .

The  $\mathcal{B}$ -tightness of  $T^{-1}v_T(B)^{-1}(\widehat{\lambda}_T - \lambda_B)$  is a consequence of Theorem 3 and 4, together with the convergence of the preliminary estimators  $\widehat{m}_T$ . For  $M > 0$ ,

$$\begin{aligned} \mathcal{E}^T(\text{Id}, \partial\mathcal{J}_T) - \partial\mathcal{E}_B(\text{Id}) &= (\mathcal{E}^T(\min\{\text{Id}, M\}, \partial\mathcal{J}_T) - \partial\mathcal{E}_B(\min\{\text{Id}, M\})) \\ &\quad + \left( \mathcal{E}^T((\cdot - M)\mathbf{1}_{\{\cdot > M\}}, \partial\mathcal{J}_T) - \partial\mathcal{E}_B((\cdot - M)\mathbf{1}_{\{\cdot > M\}}) \right) = I + II \end{aligned}$$

say. We choose  $M = M_T = 2T$  and we apply Theorem 3 for the test functions  $g_T(x) = \min\{x, M_T\}/M_T$  which are uniformly bounded in  $T$  to get the  $\mathcal{B}$ -tightness of  $T^{-1}v_T(B)^{-1}I$ . Since  $\zeta_u \leq T$  when  $u \in \partial T_T$ , we also have  $|II| = \partial\mathcal{E}_B((\cdot - 2T)\mathbf{1}_{\{\cdot > 2T\}})$  and we deduce that  $v_T(B)^{-1}II$  is  $\mathcal{B}$ -tight. We study in the same way  $\mathcal{E}^T(\text{Id}, \mathring{\mathcal{J}}_T)$  to conclude.

### 5.4 PROOF OF THEOREM 7

The proof of Theorem 7 goes along the classical line of a bias-variance analysis in nonparametrics (see for instance the classical textbook [Tsy09]). However, we have two kind of extra difficulties: first we have to get rid of the random bandwidth  $\widehat{h}_T = \exp(-\widehat{\lambda}_T \frac{1}{2\beta+1} T)$  defined in (18) (actually the most delicate part of the proof) and second, we have to get rid of the preliminary estimators  $\widehat{m}_T$  and  $\widehat{\lambda}_T$ .

The point  $x \in (0, \infty)$  where we estimate  $B(x)$  is fixed throughout, and further omitted in the notation. We first need a slight extension of Theorem 4 – actually of the estimate (33) – in order to accommodate test functions  $g = g_T$  such that  $g_T \rightarrow \delta_x$  weakly as  $T \rightarrow \infty$ . For a function  $g : [0, \infty) \rightarrow \mathbb{R}$  let

$$|g|_1 = \int_0^\infty |g(y)| dy, \quad |g|_2^2 = \int_0^\infty g(y)^2 dy \quad \text{and} \quad |g|_\infty = \sup_y |g(y)|$$

denote the usual  $L^p$ -norms over  $[0, \infty)$  for  $p = 1, 2, \infty$ . Define also

$$\Phi_T(B, g) = \begin{cases} |g|_2^2 + \inf_{0 \leq v \leq T} (|g|_1^2 e^{\lambda_B v} + |g|_\infty^2 e^{(2(\lambda_B - \rho_B) + -\lambda_B)v}) + |g|_1 |g|_\infty & \text{if } \lambda_B \leq 2\rho_B \\ |g|_2^2 + |g|_\infty^2 e^{(\lambda_B - 2\rho_B)T} + |g|_1 |g|_\infty & \text{if } \lambda_B > 2\rho_B. \end{cases} \quad (35)$$

**Proposition 15.** *In the same setting as Theorem 4, we have, for any  $g \in \mathcal{L}$ ,*

$$\mathbb{E} \left[ \left( \sum_{u \in \mathring{\mathcal{J}}_T} (g(\zeta_u^T) - \mathring{\mathcal{E}}_B(g)) \right)^2 \right] \lesssim e^{(\lambda_B - \rho_B)T} |g|_\infty^2 + e^{\lambda_B T} \Phi_T(B, g), \quad (36)$$

where the symbol  $\lesssim$  means here uniformly in  $B \in \mathcal{B}$  and independently of  $g$ .

Let us briefly comment on Proposition 15. If  $g$  is bounded and compactly supported with  $\int g = 1$ , consider the function  $g_{h_T}(y) = h_T^{-1} g(h_T^{-1}(x - y))$  that mimics the Dirac function  $\delta_x$  for  $h_T \rightarrow 0$ . It is noteworthy that in the left-hand side of (36),  $g_{h_T}(\zeta_u^T)^2$  is of order  $h_T^{-2}$  while the right-hand side is of order  $e^{\lambda_B T} h_T^{-1}$  if we pick  $\omega = h_T^{-1}$  (allowed as soon as  $h_T^{-1} \leq e^{\lambda_B T}$ ). We can thus expect to gain a crucial factor  $h_T$  thanks to averaging over  $\mathring{\mathcal{J}}_T$ .

*Proof.* We carefully revisit the estimate (33) in the proof of Theorem 4 keeping up with the same notation and assuming with no loss of generality that  $\mathring{\mathcal{E}}_B(g) = 0$ . Recall decomposition (34).

*Step 1.* For the term  $I$ , we insert  $\int_0^\infty g(y)^2 H_B(y) \mu_B(y) dy = m c_B \int_0^\infty g(y)^2 e^{-\lambda_B y} f_B(y) dy$  to obtain  $I = IV + V$ , where

$$IV \lesssim e^{\lambda_B T} \int_0^\infty g(y)^2 e^{-\lambda_B y} f_B(y) dy$$

and

$$|V| \leq \frac{1}{m} \int_0^T e^{\lambda_B s} \left| P_{H_B}^s (g^2 H_B)(0) - \int_0^\infty g(y)^2 H_B(y) \mu_B(y) dy \right| ds.$$

Clearly,  $|IV| \lesssim e^{\lambda_B T} |g|_2^2$ . By Proposition 12, we further infer

$$|V| \lesssim |g|_\infty^2 \int_0^T e^{\lambda_B s} e^{-\rho_B s} ds \lesssim |g|_\infty^2 e^{(\lambda_B - \rho_B)T}.$$

*Step 2.* For the term  $II$ , using  $P_{H_B}^s (H_B)(0) \lesssim 1$  we now obtain

$$II \lesssim e^{\lambda_B T} \int_0^T e^{-\lambda_B s} \left( \int_0^s e^{\lambda_B t} P_{H_B}^t (g H_B)(0) dt \right)^2 ds.$$

A new difficulty appears here, since the crude bound

$$|P_{H_B}^t(gH_B)(0)| \lesssim |g|_\infty \exp(-\rho_B t) \quad (37)$$

given by Proposition 12 does not yield to the correct order for small value of  $t$  because of the term  $|g|_\infty$ . We need the following refinement (for small values of  $t$ ), based on a renewal argument and proved in Appendix:

**Lemma 16.** *For every  $t \geq 0$  and  $g \in \mathcal{L}$ , we have*

$$|P_{H_B}^t(gH_B)(0)| \lesssim |g(t)|e^{-\lambda_B t} + |g|_1$$

uniformly in  $B \in \mathcal{B}$ .

Let  $v \in [0, T]$  be arbitrary. For  $0 \leq s \leq v$ , by Lemma 16 we obtain

$$\mathcal{J}_s = \left( \int_0^s e^{\lambda_B t} |P_{H_B}^t(gH_B)(0)| dt \right)^2 \lesssim \left( \int_0^s |g(t)| dt + |g|_1 \int_0^s e^{\lambda_B t} dt \right)^2 \lesssim |g|_1^2 e^{2\lambda_B s}.$$

For  $s \geq v$ , we have by (37)

$$\mathcal{J}_s \lesssim \mathcal{J}_v + |g|_\infty^2 \left( \int_v^s e^{(\lambda_B - \rho_B)t} dt \right)^2 \lesssim \mathcal{J}_v + |g|_\infty^2 (e^{2(\lambda_B - \rho_B)s} \mathbf{1}_{\{\lambda_B > \rho_B\}} + (s - v)^2 \mathbf{1}_{\{\lambda_B \leq \rho_B\}} \mathbf{1}_{\{s \geq v\}}).$$

On the one hand,  $\int_0^v e^{-\lambda_B s} \mathcal{J}_s ds \lesssim |g|_1^2 e^{\lambda_B v}$  and on the other hand  $\int_v^T e^{-\lambda_B s} \mathcal{J}_s ds$  is less than

$$\begin{aligned} & \mathcal{J}_v \int_v^T e^{-\lambda_B s} ds + |g|_\infty^2 \left( \int_v^T e^{-\lambda_B s} e^{2(\lambda_B - \rho_B)s} ds + \int_v^T e^{-\lambda_B s} (s - v)^2 ds \mathbf{1}_{\{\lambda_B \leq \rho_B\}} \right) \\ & \lesssim \begin{cases} |g|_1^2 e^{\lambda_B v} + |g|_\infty^2 e^{-\lambda_B v} & \text{if } \lambda_B \leq \rho_B \\ |g|_1^2 e^{\lambda_B v} + |g|_\infty^2 e^{(\lambda_B - 2\rho_B)v} & \text{if } \rho_B \leq \lambda_B \leq 2\rho_B \\ |g|_1^2 e^{\lambda_B v} + |g|_\infty^2 e^{(\lambda_B - 2\rho_B)T} & \text{if } \lambda_B \geq 2\rho_B, \end{cases} \end{aligned}$$

whence for every  $v \in [0, T]$ , we derive

$$|II| \lesssim e^{\lambda_B T} \left( |g|_1^2 e^{\lambda_B v} + |g|_\infty^2 (e^{(-\lambda_B + 2(\lambda_B - \rho_B)_+)v} \mathbf{1}_{\{\lambda_B \leq 2\rho_B\}} + e^{(\lambda_B - 2\rho_B)T} \mathbf{1}_{\{\lambda_B > 2\rho_B\}}) \right).$$

*Step 3.* Finally going back to Step 3 in the proof of Theorem 4 we readily obtain

$$\begin{aligned} |III| & \lesssim \int_0^T e^{\lambda_B s} P_{H_B}^s (|g|_{H_B})(0) \int_0^{T-s} e^{\lambda_B t} |P_{H_B}^t(gH_B)(0)| dt ds \\ & \lesssim \int_0^T e^{\lambda_B s} (|g(s)|e^{-\lambda_B s} + |g|_1) |g|_\infty \int_0^{T-s} e^{\lambda_B t} e^{-\rho_B t} dt ds \end{aligned}$$

by applying Lemma 16 for the term involving  $P_{H_B}^s$  and the estimate (37) for the term involving  $P_{H_B}^t$ , therefore  $|III| \lesssim e^{\lambda_B T} |g|_1 |g|_\infty$ .  $\square$

Proposition 15 enables us to obtain the next result which is the key ingredient to get rid of the random bandwidth  $\widehat{h}_T$ , thanks to the fact that it is concentrated around its estimated value  $h_T(\beta) = e^{-\frac{1}{2\beta+1}\lambda_B T}$ . To that end, define, for  $C > 0$

$$\mathcal{C}_C = \left\{ g : \mathbb{R} \rightarrow \mathbb{R}, \text{supp}(g) \subset [0, C] \text{ and } \sup_y |g(y)| \leq C \right\}.$$

Denote by  $\mathcal{C}_C^1$  (later abbreviated by  $\mathcal{C}^1$ ) the subset of  $\mathcal{C}_C$  of functions that are moreover differentiable, with derivative uniformly bounded by  $C$ . For  $h > 0$  we set  $g_h(y) = h^{-1}g(h^{-1}(x-y))$ . Finally, for  $a, b \geq 0$  we set  $[a \pm b] = [(a-b)_+, a+b]$ . Recall from Section 3.2 that  $v_T(B) = e^{-\min\{\rho_B, \lambda_B/2\}T}$  if  $\lambda_B \neq 2\rho_B$  and  $T^{1/2}e^{-\lambda_B T/2}$  otherwise.

**Proposition 17.** *Assume that  $\beta \geq 1/2$ . Define  $\varpi_B = \min\{\max\{1, \lambda_B/\rho_B\}, 2\}$ . For every  $\kappa > 0$ ,*

$$v_T(B)^{-1} \sup_{h \in [h_T(\beta)(1 \pm \kappa T^2 v_T(B))]} \left| \mathcal{E}^T(\overset{\circ}{\mathcal{J}}_T, h^{\varpi_B/2} f g_h) - \overset{\circ}{\mathcal{E}}_B(h^{\varpi_B/2} f g_h) \right|$$

is  $\mathcal{B} \times \mathcal{L} \times \mathcal{C}^1$ -tight for the parameter  $(B, f, g)$ .

*Proof. Step 1.* Define  $\overline{f g_h} = f g_h - \overset{\circ}{\mathcal{E}}_B(f g_h)$ . Writing

$$\begin{aligned} & v_T(B)^{-1} \left( \mathcal{E}^T(\overset{\circ}{\mathcal{J}}_T, h^{\varpi_B/2} f g_h) - \overset{\circ}{\mathcal{E}}_B(h^{\varpi_B/2} f g_h) \right) \\ &= \frac{e^{\lambda_B T}}{|\overset{\circ}{\mathcal{J}}_T|} e^{(\min\{\rho_B, \lambda_B/2\} - \lambda_B)T} (T^{-1/2})^{\mathbf{1}_{\{\lambda_B = 2\rho_B\}}} \sum_{u \in \overset{\circ}{\mathcal{J}}_T} h^{\varpi_B/2} \overline{f g_h}(\zeta_u), \end{aligned}$$

we see as in the proof of Theorem 4 that thanks to Lemma 14, it is enough to prove the  $\mathcal{B}$ -tightness of

$$\sup_{h \in [h_T(\beta)(1 \pm \kappa T^2 v_T(B))]} |V_h^T| = \sup_{s \in [0, 1]} |V_{h_s}^T|,$$

where

$$V_h^T = e^{(\min\{\rho_B, \lambda_B/2\} - \lambda_B)T} (T^{-1/2})^{\mathbf{1}_{\{\lambda_B = 2\rho_B\}}} \sum_{u \in \overset{\circ}{\mathcal{J}}_T} h^{\varpi_B/2} \overline{f g_h}(\zeta_u),$$

and

$$h_s = h_T(\beta)(1 - \kappa T^2 v_T(B)) + 2s\kappa h_T(\beta) T^2 v_T(B), \quad s \in [0, 1].$$

*Step 2.* We claim that

$$\begin{cases} \sup_{T > 0} \mathbb{E}[(V_{h_0}^T)^2] < \infty \\ \mathbb{E}[(V_{h_t}^T - V_{h_s}^T)^2] \leq C'(t-s)^2 \quad \text{for } s, t \in [0, 1], \end{cases} \quad (38)$$

for some constant  $C' > 0$  that does not depend on  $T$  nor  $B \in \mathcal{B}$ . Then, by Kolmogorov continuity criterion, this implies in particular that

$$\sup_{T > 0} \sup_{B \in \mathcal{B}} \mathbb{E} \left[ \sup_{s \in [0, 1]} |V_{h_s}^T| \right] < \infty$$

hence the result (see for instance [Sch00]) to track the constant and obtain a uniform version of the



continuity criterion). We have

$$V_{h_t}^T - V_{h_s}^T = e^{(\min\{\rho_B, \frac{\lambda_B}{2}\} - \lambda_B)T} (T^{-\frac{1}{2}} \mathbf{1}_{\{\lambda_B = 2\rho_B\}}) \sum_{u \in \mathring{\mathcal{J}}_T} \left( \Delta_{s,t}(h^{\varpi_B/2} f g_h)(\zeta_u) - \mathring{\mathcal{E}}_B(\Delta_{s,t}(h^{\varpi_B/2} f g_h)) \right)$$

where  $\Delta_{s,t}(h^{\varpi_B/2} f g_h)(y) = h_t^{\varpi_B/2} f(y) g_{h_t}(y) - h_s^{\varpi_B/2} f(y) g_{h_s}(y)$ . By Proposition 15, we derive that  $\mathbb{E}[(V_{h_t}^T - V_{h_s}^T)^2]$  is less than

$$\begin{cases} e^{-\lambda_B T} |\Delta_{s,t}(h^{\varpi_B/2} f g_h)|_\infty^2 + \Phi_T(B, \Delta_{s,t}(h^{\varpi_B/2} f g_h)) & \text{if } \lambda_B \leq \rho_B \\ e^{-\rho_B T} |\Delta_{s,t}(h^{\varpi_B/2} f g_h)|_\infty^2 + \Phi_T(B, \Delta_{s,t}(h^{\varpi_B/2} f g_h)) & \text{if } \rho_B \leq \lambda_B \leq 2\rho_B \\ e^{-(\lambda_B - \rho_B)T} |\Delta_{s,t}(h^{\varpi_B/2} f g_h)|_\infty^2 + e^{-(\lambda_B - 2\rho_B)T} \Phi_T(B, \Delta_{s,t}(h^{\varpi_B/2} f g_h)) & \text{if } \lambda_B \geq 2\rho_B \end{cases}$$

(we ignore the slow term in the limiting case  $\lambda_B = 2\rho_B$ ) and the remainder of the proof amounts to check that each term in the estimate above has order  $(t-s)^2$  uniformly in  $T$  and  $B \in \mathcal{B}$ .

*Step 3.* For every  $y$ , we have

$$\Delta_{s,t}(h^{\varpi_B/2} f g_h)(y) = (h_t - h_s) \partial_h (h^{\varpi_B/2} f(y) g_h(y))|_{h=h^*(y)}$$

for some  $h^*(y) \in [\min\{h_t, h_s\}, \max\{h_t, h_s\}]$ . Observe now that since  $g \in \mathcal{C}^1$  and  $f \in \mathcal{L}$ , we have

$$\partial_h (h^{\varpi_B/2} f g_h(y)) = \left(\frac{\varpi_B}{2} - 1\right) h^{\frac{\varpi_B}{2}-2} f(y) g(h^{-1}(x-y)) - h^{\frac{\varpi_B}{2}-3} (x-y) f(y) g'(h^{-1}(x-y))$$

therefore, for small enough  $h$  (which is always the case for  $T$  large enough, uniformly in  $B \in \mathcal{B}$ ) and since  $|x-y| \lesssim h$  thanks to the fact that  $g$  is compactly supported, we obtain

$$|\partial_h (h^{\varpi_B/2} f g_h(y))| \lesssim h^{\varpi_B/2-2} \mathbf{1}_{[0,C]}(h^{-1}(x-y)).$$

Assume with no loss of generality that  $s \leq t$  so that  $h_s \leq h(y)^* \leq h_t$ . It follows that

$$\begin{aligned} |\Delta_{s,t}(h^{\varpi_B/2} f g_h)(y)| &\lesssim (h_t - h_s) h^*(y)^{\varpi_B/2-2} \mathbf{1}_{[0,C]}(h^*(y)^{-1}(x-y)) \\ &\leq (h_t - h_s) h_s^{\varpi_B/2-2} \mathbf{1}_{[0,C]}(h_t^{-1}(x-y)). \end{aligned}$$

Using that  $h_t - h_s = 2(t-s)\kappa T^2 h_T(\beta) v_T(B)$ , we successively obtain

$$\begin{aligned} |\Delta_{s,t}(h^{\varpi_B/2} f g_h)|_\infty^2 &\lesssim (h_t - h_s)^2 h_s^{\varpi_B-4} \lesssim (t-s)^2 T^4 v_T(B)^2 h_T(\beta)^{\varpi_B-2}, \\ |\Delta_{s,t}(h^{\varpi_B/2} f g_h)|_2^2 &\lesssim (h_t - h_s)^2 h_s^{\varpi_B-4} h_t \lesssim (t-s)^2 T^4 v_T(B)^2 h_T(\beta)^{\varpi_B-1}, \\ |\Delta_{s,t}(h^{\varpi_B/2} f g_h)|_1^2 &\lesssim (h_t - h_s)^2 h_s^{\varpi_B-4} h_t^2 \lesssim (t-s)^2 T^4 v_T(B)^2 h_T(\beta)^{\varpi_B}, \\ |\Delta_{s,t}(h^{\varpi_B/2} f g_h)|_1 |\Delta_{s,t}(h^{\varpi_B/2} f g_h)|_\infty &\lesssim (t-s)^2 T^4 v_T(B)^2 h_T(\beta)^{\varpi_B-1}. \end{aligned}$$

*Step 4.* Recall that  $h_T(\beta) = e^{-\lambda_B T/(2\beta+1)}$ .

When  $\lambda_B \leq \rho_B$ , we have  $v_T(B) = e^{-\lambda_B T/2}$  and  $\varpi_B = 1$ . By definition of  $\Phi_T$  in (35) together

with the estimates of Steps 2 and 3, we obtain

$$\begin{aligned} \mathbb{E}[(V_{h_t}^T - V_{h_s}^T)^2] &\lesssim e^{-\lambda_B T} |\Delta_{s,t}(h^{1/2} f g_h)|_\infty^2 + \Phi_T(B, \Delta_{s,t}(h^{1/2} f g_h)) \\ &\lesssim (t-s)^2 T^4 (e^{\lambda_B(\frac{1}{2\beta+1}-2)T} + e^{-\lambda_B T} + e^{-\lambda_B(\frac{1}{2\beta+1}+1)T} e^{\lambda_B v} + e^{\lambda_B(\frac{1}{2\beta+1}-1)T} e^{-\lambda_B v}) \end{aligned}$$

which is of order  $(t-s)^2$  uniformly in  $T > 0$  by picking  $v = 0$  for instance.

When  $\rho_B \leq \lambda_B \leq 2\rho_B$ , we still have  $v_T(B) = e^{-\lambda_B T/2}$  but now  $\varpi_B = \lambda_B/\rho_B$ . It follows that  $\mathbb{E}[(V_{h_t}^T - V_{h_s}^T)^2]$  is of order

$$\begin{aligned} &e^{-\rho_B T} |\Delta_{s,t}(h^{\lambda_B/2\rho_B} f g_h)|_\infty^2 + \Phi_T(B, \Delta_{s,t}(h^{\lambda_B/2\rho_B} f g_h)) \\ &\lesssim (t-s)^2 T^4 (e^{\lambda_B(\frac{2-\lambda_B/\rho_B}{2\beta+1}-1)T} (e^{-\rho_B T} + e^{(\lambda_B-2\rho_B)v}) + e^{\lambda_B(\frac{1-\lambda_B/\rho_B}{2\beta+1}-1)T} + e^{-\lambda_B(\frac{\lambda_B/\rho_B}{2\beta+1}+1)T} e^{\lambda_B v}) \end{aligned}$$

and this last term is again of order  $(t-s)^2$  uniformly in  $T > 0$  by noting that  $1 \leq \lambda_B/\rho_B \leq 2$  and picking  $v = 0$  for instance.

Finally, when  $2\rho_B \leq \lambda_B$ , we have  $v_T(B) = e^{-\rho_B T}$  and  $\varpi_B = 2$ . This entails

$$\begin{aligned} \mathbb{E}[(V_{h_t}^T - V_{h_s}^T)^2] &\lesssim e^{-(\lambda_B-\rho_B)T} |\Delta_{s,t}(h f g_h)|_\infty^2 + e^{-(\lambda_B-2\rho_B)T} \Phi_T(B, \Delta_{s,t}(h f g_h)) \\ &\lesssim (t-s)^2 T^4 (e^{-(\lambda_B+\rho_B)T} + e^{-\lambda_B(\frac{1}{2\beta+1}+1)T} + e^{-2\rho_B T}) \end{aligned}$$

and these terms are all again of order  $(t-s)^2$  uniformly in  $T$ .

*Step 5.* It remains to show  $\sup_{T>0} \mathbb{E}[(V_{h_0}^T)^2] < \infty$  in order to complete the proof of (38). By Step 2 and the definition of  $\varpi_B$ , we readily have

$$\mathbb{E}[(V_{h_0}^T)^2] \lesssim \begin{cases} e^{-\lambda_B T} |h_0^{1/2} f g_{h_0}|_\infty^2 + \Phi_T(B, h_0^{1/2} f g_{h_0}) & \text{if } \lambda_B \leq \rho_B \\ e^{-\rho_B T} |h_0^{\lambda_B/2\rho_B} f g_{h_0}|_\infty^2 + \Phi_T(B, h_0^{\lambda_B/2\rho_B} f g_{h_0}) & \text{if } \rho_B \leq \lambda_B \leq 2\rho_B \\ e^{-(\lambda_B-\rho_B)T} |h_0 f g_{h_0}|_\infty^2 + e^{-(\lambda_B-2\rho_B)T} \Phi_T(B, h_0 f g_{h_0}) & \text{if } \lambda_B \geq 2\rho_B. \end{cases}$$

When  $\lambda_B \leq \rho_B$ , since  $h_0$  is of order  $h_T(\beta)$ , we have

$$\mathbb{E}[(V_{h_0}^T)^2] \lesssim e^{-\lambda_B T} h_T(\beta)^{-1} + 1 + h_T(\beta) e^{\lambda_B v} + h_T(\beta)^{-1} e^{-\lambda_B v}$$

for every  $v \in [0, T]$ , and the choice  $v = \frac{1}{2\beta+1}T$  entails  $\mathbb{E}[(V_{h_0}^T)^2] \lesssim 1$ . When  $\rho_B \leq \lambda_B \leq 2\rho_B$ , we have

$$\mathbb{E}[(V_{h_0}^T)^2] \lesssim e^{-\rho_B T} h_T(\beta)^{\frac{\lambda_B}{\rho_B}-2} + h_T(\beta)^{\frac{\lambda_B}{\rho_B}-1} + h_T(\beta)^{\frac{\lambda_B}{\rho_B}} e^{\lambda_B v} + h_T(\beta)^{\frac{\lambda_B}{\rho_B}-2} e^{(\lambda_B-2\rho_B)v}.$$

The first term is bounded as soon as  $\beta \geq 1/2$  and the choice  $v = \frac{\lambda_B}{\rho_B(2\beta+1)}T$  for the last two terms entails  $\mathbb{E}[(V_{h_0}^T)^2] \lesssim 1$ . Finally, when  $2\rho_B \leq \lambda_B$  we have

$$\mathbb{E}[(V_{h_0}^T)^2] \lesssim e^{-(\lambda_B-\rho_B)T} + 1$$

and this term is bounded likewise. Eventually (38) is established and Proposition 17 is proved.  $\square$

We now get rid of the preliminary estimators  $\widehat{m}_T$  and  $\widehat{\lambda}_T$ . Remember that the target rate of convergence for  $\widehat{B}_T(x)$  is  $w_T(B) = T^{\mathbf{1}_{\{\lambda_B=2\rho_B\}}} \exp\left(-\min\{\lambda_B, 2\rho_B\} \frac{\beta - (\lambda_B/\rho_B - 1) + /2}{2\beta + 1} T\right)$ .

**Lemma 18.** *Assume that  $\beta > 1$ . Let either  $G_T(y) = g_{\widehat{h}_T}(y)$  with  $g \in \mathcal{C}^1$  or  $G_T(y) = \mathbf{1}_{\{y \leq x\}}$  for  $y \in [0, \infty)$ . Then*

$$w_T(B)^{-1} \left( \mathcal{E}^T(\mathring{\mathcal{J}}_T, \widehat{m}_T^{-1} e^{\widehat{\lambda}_T \cdot} G_T) - \mathcal{E}^T(\mathring{\mathcal{J}}_T, m^{-1} e^{\lambda_B \cdot} G_T) \right)$$

is  $\mathcal{B}$ -tight for the parameter  $B$ .

*Proof.* For  $u \in \mathring{\mathcal{J}}_T$  and its lifetime  $\zeta_u$ , define

$$\gamma_T(u) = w_T(B)^{-1} \left( \widehat{m}_T^{-1} e^{\widehat{\lambda}_T \zeta_u} - m^{-1} e^{\lambda_B \zeta_u} \right) G_T(\zeta_u).$$

Lemma 18 amounts to show that  $|\mathring{\mathcal{J}}_T|^{-1} \sum_{u \in \mathring{\mathcal{J}}_T} \gamma_T(u)$  is  $\mathcal{B}$ -tight. Set  $h_T(\beta) = \exp(-\lambda_B \frac{1}{2\beta+1} T)$  and note that

$$w_T(B)^{-1} = (T^{-1/2})^{\mathbf{1}_{\{\lambda_B=2\rho_B\}}} e^{\min\{\rho_B, \lambda_B/2\}T} h_T(\beta)^{\varpi_B/2} = v_T(B)^{-1} h_T(\beta)^{\varpi_B/2},$$

where  $\varpi_B = \min\{\max\{1, \lambda_B/\rho_B\}, 2\}$ . We first treat the case  $G_T(y) = g_{\widehat{h}_T}(y)$ .

*Step 1.* By Proposition 5, we have

$$\widehat{\lambda}_T = \lambda_B + T v_T(B) r_T \quad \text{and} \quad \widehat{m}_T^{-1} = m^{-1} + e^{-\lambda_B T/2} r'_T,$$

where both  $r_T$  and  $r'_T$  are  $\mathcal{B}$ -tight. We then have the decomposition

$$\begin{aligned} \gamma_T(u) &= w_T(B)^{-1} \widehat{m}_T^{-1} (e^{\widehat{\lambda}_T \zeta_u} - e^{\lambda_B \zeta_u}) g_{\widehat{h}_T}(\zeta_u) + w_T(B)^{-1} (\widehat{m}_T^{-1} - m^{-1}) e^{\lambda_B \zeta_u} g_{\widehat{h}_T}(\zeta_u) \\ &= T h_T(\beta)^{\varpi_B/2} \widehat{m}_T^{-1} r_T \zeta_u e^{\vartheta_T \zeta_u} g_{\widehat{h}_T}(\zeta_u) + w_T(B)^{-1} e^{-\lambda_B T/2} e^{\lambda_B \zeta_u} r'_T g_{\widehat{h}_T}(\zeta_u) \\ &= I + II, \end{aligned}$$

say, with  $\vartheta_T \in [\min\{\lambda_B, \widehat{\lambda}_T\}, \max\{\lambda_B, \widehat{\lambda}_T\}]$ . Since  $g \in \mathcal{C}^1 \subset \mathcal{C}$  and  $\widehat{m}_T^{-1}$ ,  $\vartheta_T$  and  $\widehat{h}_T$  are  $\mathcal{B}$ -tight, we can write

$$|I| \leq T h_T(\beta)^{\varpi_B/2} \widehat{m}_T^{-1} r_T (C \widehat{h}_T + x) e^{\vartheta_T (C \widehat{h}_T + x)} |g_{\widehat{h}_T}(\zeta_u)| = T h_T(\beta)^{\varpi_B/2} |g_{\widehat{h}_T}(\zeta_u)| \widetilde{r}_T$$

and

$$|II| \leq h_T(\beta)^{\varpi_B/2} e^{\lambda_B (C \widehat{h}_T + x)} r'_T |g_{\widehat{h}_T}(\zeta_u)| = h_T(\beta)^{\varpi_B/2} |g_{\widehat{h}_T}(\zeta_u)| \widetilde{r}'_T$$

where  $\widetilde{r}_T$  and  $\widetilde{r}'_T$  are  $B \in \mathcal{B}$ -tight.

*Step 2.* We are left to proving the tightness of  $T h_T(\beta)^{\varpi_B/2} |g_{\widehat{h}_T}(\zeta_u)|$  when averaging over  $\mathring{\mathcal{J}}_T$  that is to say the tightness of  $T h_T(\beta)^{\varpi_B/2} \mathcal{E}^T(\mathring{\mathcal{J}}_T, |g_{\widehat{h}_T}|)$ . We plan to use Proposition 17. For  $\kappa > 0$ , on the event

$$\mathcal{A}_{T,\kappa} = \{\widehat{h}_T \in \mathcal{J}_{T,\kappa}\}, \quad \mathcal{J}_{T,\kappa} = [h_T(\beta)(1 \pm \kappa T^2 v_T(B))],$$

we have

$$Th_T(\beta)^{\varpi_B/2} \mathcal{E}^T(\mathring{\mathcal{J}}_T, |g_{\hat{h}_T}|) \leq III + IV,$$

with

$$III = Th_T(\beta)^{\varpi_B/2} \sup_{h \in \mathcal{J}_{T,\kappa}} \mathring{\mathcal{E}}_B(|g_h|)$$

and

$$\begin{aligned} IV &= Th_T(\beta)^{\varpi_B/2} (h_T(\beta)(1 - \kappa T^2 v_T(B)))^{-\varpi_B/2} \sup_{h \in \mathcal{J}_{T,\kappa}} |\mathcal{E}^T(\mathring{\mathcal{J}}_T, h^{\varpi_B/2} |g_h|) - \mathring{\mathcal{E}}_B(h^{\varpi_B/2} |g_h|)| \\ &\leq T \sup_{h \in \mathcal{J}_{T,\kappa}} |\mathcal{E}^T(\mathring{\mathcal{J}}_T, h^{\varpi_B/2} |g_h|) - \mathring{\mathcal{E}}_B(h^{\varpi_B/2} |g_h|)|. \end{aligned}$$

Concerning the main term  $III$ , we write

$$\mathring{\mathcal{E}}_B(|g_h|) = m \int_0^\infty h^{-1} |g(h^{-1}(x-y))| e^{-\lambda_B y} f_B(y) dy \leq m \sup_y (e^{-\lambda_B y} f_B(y)) \int_0^\infty |g(y)| dy \lesssim 1$$

since  $B \in \mathcal{B}$ , so we have a bound that does not depend on  $h$  and we readily conclude  $III \lesssim 1$  on  $\mathcal{A}_{T,\kappa}$ . For the remainder term  $IV$ , we apply Proposition 17 and obtain the  $\mathcal{B}$ -tightness of  $IV$  (that actually goes to 0 at a fast rate) on  $\mathcal{A}_{T,\kappa}$ .

*Step 3.* It remains to control the probability of  $\mathcal{A}_{T,\kappa}$ . By Proposition 5, we have  $\hat{\lambda}_T = \lambda_B + Tv_T(B)r_T$ , where  $r_T$  is  $\mathcal{B}$ -tight. It follows that

$$\begin{aligned} \mathbb{P}(\mathcal{A}_{T,\kappa}^c) &= \mathbb{P}(|\hat{h}_T - h_T(\beta)| \geq \kappa h_T(\beta) T^2 v_T(B)) \\ &= \mathbb{P}(|1 - e^{-(\hat{\lambda}_T - \lambda_B) \frac{1}{2\beta+1} T}| \geq \kappa T^2 v_T(B)) \\ &= \mathbb{P}(|\frac{1}{2\beta+1} r_T e^{-\vartheta_T \frac{1}{2\beta+1} T}| \geq \kappa) \end{aligned}$$

where both  $|\vartheta_T| \leq |\hat{\lambda}_T - \lambda_B|$  and  $r_T$  are tight, and this term can be made arbitrarily small by taking  $\kappa$  large enough.

The case  $G_T(y) = \mathbf{1}_{\{y \leq x\}}$  is obtained in the same way and is actually much simpler, since there is no factor  $\hat{h}_T^{-1}$  in the Step 2 which is therefore straightforward and there is also no need for a Step 3. We omit the details.  $\square$

*Proof of Theorem 7.* We are ready to prove the main result of the chapter. The key ingredient is Proposition 17.

*Step 1.* In view of Lemma 18 with test function  $g = K$ , it is now sufficient to prove Theorem 7 replacing  $\hat{B}_T(x)$  by  $\tilde{B}_T(x)$ , where

$$\tilde{B}_T(x) = \frac{\mathcal{E}^T(\mathring{\mathcal{J}}_T, m^{-1} e^{\lambda_B \cdot} K_{\hat{h}_T}(x - \cdot))}{1 - \mathcal{E}^T(\mathring{\mathcal{J}}_T, m^{-1} e^{\lambda_B \cdot} \mathbf{1}_{\{\cdot \leq x\}})}.$$

Since  $(x, y) \rightsquigarrow x/(1-y)$  is Lipschitz continuous on compact sets that are bounded away from

$\{y = 1\}$ , this simply amounts to show the  $\mathcal{B}$ -tightness of

$$w_T(B)^{-1} \left( \mathcal{E}^T(\mathring{\mathcal{J}}_T, m^{-1}e^{\lambda_B \cdot} \mathbf{1}_{\{\cdot \leq x\}}) - \mathring{\mathcal{E}}_B(m^{-1}e^{\lambda_B \cdot} \mathbf{1}_{\{\cdot \leq x\}}) \right) \quad (39)$$

and

$$w_T(B)^{-1} \left( \mathcal{E}^T(\mathring{\mathcal{J}}_T, m^{-1}e^{\lambda_B \cdot} K_{\hat{h}_T}(x - \cdot)) - f_B(x) \right), \quad (40)$$

where  $w_T(B)^{-1} = (T^{-1/2})^{\mathbf{1}_{\{\lambda_B = 2\rho_B\}}} e^{\min\{\rho_B, \lambda_B/2\}T} h_T(\beta)^{\varpi_B/2} = v_T(B)^{-1} h_T(\beta)^{\varpi_B/2}$ . We readily obtain the  $\mathcal{B}$ -tightness of (39) by applying Theorem 4 with test function  $g(y) = m^{-1}e^{\lambda_B y} \mathbf{1}_{\{y \leq x\}}$  since  $v_T(B) \ll w_T(B)$  (we even have convergence to 0).

*Step 2.* We turn to the main term (40). For  $h > 0$ , introduce the notation

$$K_h f_B(x) = \mathring{\mathcal{E}}_B(m^{-1}e^{\lambda_B \cdot} K_h) = \int_0^\infty K_h(x-y) f_B(y) dy.$$

For  $\kappa > 0$  on the event  $\mathcal{A}_{T,\kappa} = \{\hat{h}_T \in \mathcal{J}_{T,\kappa}\}$  with  $\mathcal{J}_{T,\kappa} = [h_T(\beta)(1 \pm \kappa T^2 v_T(B))]$  introducing the approximation term  $K_h f_B(x)$ , we obtain a bias-variance bound that reads

$$|\mathcal{E}^T(\mathring{\mathcal{J}}_T, m^{-1}e^{\lambda_B \cdot} K_{\hat{h}_T}) - f_B(x)| \leq I + II,$$

with

$$I = \sup_{h \in \mathcal{J}_{T,\kappa}} |K_h f_B(x) - f_B(x)|$$

and

$$II = \sup_{h \in \mathcal{J}_{T,\kappa}} |\mathcal{E}^T(\mathring{\mathcal{J}}_T, m^{-1}e^{\lambda_B \cdot} K_h) - \mathring{\mathcal{E}}_B(m^{-1}e^{\lambda_B \cdot} K_h)|.$$

The term  $I$  is treated by the following classical argument in nonparametric estimation: since  $B \in \mathcal{H}_{\mathcal{D}}^\beta(L)$  we also have  $f_B \in \mathcal{H}_{\mathcal{D}}^\beta(L')$  for another constant  $L'$  that only depends on  $\mathcal{D}, L$  and  $\beta$ . Write  $\beta = \lfloor \beta \rfloor + \{\beta\}$  with  $\lfloor \beta \rfloor$  a non-negative integer,  $\{\beta\} > 0$ . By a Taylor expansion up to order  $\lfloor \beta \rfloor$  (recall that the number  $n_0$  of vanishing moments of  $K$  in Assumption 6 satisfies  $n_0 > \beta$ ), we obtain

$$I \lesssim \sup_{h \in \mathcal{J}_{T,\kappa}} h^\beta = (h_T(\beta)(1 + \kappa T^2 v_T(B)))^\beta \lesssim w_T(B)$$

see for instance, Proposition 1.2 in Tsybakov [Tsy09]. This term has the right order whenever  $\lambda_B \leq \rho_B$  and is negligible otherwise.

*Step 3.* We further bound the term  $II$  on  $\mathcal{A}_{T,\kappa}$  as follows:

$$|II| \leq (h_T(\beta)(1 - \kappa T^2 v_T(B)))^{-\varpi_B/2} \sup_{h \in \mathcal{J}_{T,\kappa}} |\mathcal{E}^T(\mathring{\mathcal{J}}_T, h^{\varpi_B/2} m^{-1}e^{\lambda_B \cdot} K_h) - \mathring{\mathcal{E}}_B(h^{\varpi_B/2} m^{-1}e^{\lambda_B \cdot} K_h)|.$$

By assumption, we have  $\beta \geq 1/2$ , so by Proposition 17 applied to  $f(y) = m^{-1}e^{\lambda_B y} \mathbf{1}_{\{y \leq x+C\}} \in \mathcal{L}_{C+x}$  and  $g = K \in \mathcal{C}_{C+x}$  we conclude that  $v_T(B)^{-1} h_T(\beta)^{\varpi_B/2} |II|$  is  $\mathcal{B}$ -tight. The fact that  $v_T(B)^{-1} h_T(\beta)^{\varpi_B/2} = w_T(B)^{-1}$  enables us to conclude.

*Step 4.* It remains to control the probability of  $\mathcal{A}_{T,\kappa}$ . This is done exactly in the same way as for Step 3 in the proof of Lemma 18.  $\square$

## 5.5 PROOF OF THEOREM 8

We will prove actually a slightly stronger result, by restricting the supremum in  $B$  over a neighbourhood of an arbitrary function  $B_0$ , provided  $B_0$  is an element of the set  $\mathbb{B}_{b,m}$  defined in (20) and slightly smoother in  $\mathcal{H}_{\mathcal{D}}^{\beta}$  norm (and not identically equal to the maximal element of  $\mathbb{B}_{b,m}$ ). (Remember also that  $\mathbb{B}_{b,m} \subset \mathcal{B}_{b,m/(m-1)}^+$  by Proposition 9.)

Remember that the evolution of the Bellman-Harris model can be described by a piecewise deterministic Markov process

$$X(t) = (X_1(t), X_2(t), \dots), \quad t \geq 0$$

with values in  $\mathcal{S} = \bigcup_{k \geq 1} [0, \infty)^k$  and where the  $X_i(t)$  denote the (ordered) ages of the living particles at time  $t$ . Following Löcherbach [Löc02a], we set  $\mathbb{D}([0, \infty), \mathcal{S})$  for the Skorokhod space of càdlàg functions  $\varphi : [0, \infty) \rightarrow \mathcal{S}$  and introduce the subset  $\Omega \subset \mathbb{D}([0, \infty), \mathcal{S})$  of functions  $\varphi$  such that:

- (i) There is an increasing sequence of jump times  $T_0 = 0 < T_1 < T_2 < \dots$  such that the restriction  $\varphi|_{[T_k, T_{k+1})}$  is continuous with values in  $[0, \infty)^{l_{k,\varphi}}$  for some  $l_{k,\varphi} \geq 0$  and every  $k \geq 0$ .
- (ii) We have  $\ell(\varphi(T_k)) \neq \ell(\varphi(T_{k+1}))$  for every  $k \geq 0$ , where we set  $\ell(x) = \sum_{k \geq 0} k \mathbf{1}_{\{x \in [0, \infty)^k\}}$  for  $x \in \mathcal{S}$ .

We endow  $\Omega$  with its Borel sigma-field  $\mathcal{F}$ , its canonical process  $X_t(\varphi) = (\varphi_1(t), \varphi_2(t), \dots)$  and its canonical filtration  $(\mathcal{F}_t)_{t \geq 0}$  (modified in order to be right-continuous). By Proposition 3.3 of Löcherbach [Löc02a], there is a unique probability measure  $\mathbb{P}_B$  on  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0})$  such that  $X$  is strongly Markov under  $\mathbb{P}_B$  with  $\mathbb{P}_B(X(0) = 0) = 1$  (*i.e.* we start with one common ancestor with age 0 at time 0) and such that the random continuous time rooted tree associated to  $X$  via

$$\sum_{i \geq 1} \mathbf{1}_{\{X_i(t) > 0\}} \delta_{X_i(t)} = \sum_{u \in \mathcal{T}} \mathbf{1}_{\{t \in [b_u, d_u)\}} \delta_{t-b_u}$$

is a Harris-Bellman process according to Definition 1. The strategy for proving the lower bound is a classical two point information inequality: we nevertheless need to be careful since the target lower bound rate  $e^{-\lambda_B \frac{\beta}{2\beta+1} T}$  is parameter dependent in a non-trivial way.

*Step 1.* Let  $\delta > 0$ . Fix  $B_0 \in \mathbb{B}_{b,m} \cap \mathcal{H}_{\mathcal{D}}^{\beta}(L - \delta)$  and  $x \in \mathcal{D}$ . Then, for large enough  $T$ , setting  $h_T(B) = e^{-\lambda_B \frac{1}{2\beta+1} T}$ , we construct a perturbation  $B_T$  of  $B_0$  defined by

$$B_T(y) = B_0(y) + ah_T(B_0)^{\beta+1} K_{h_T(B_0)}(y - x), \quad y \in [0, \infty),$$

for some nonnegative smooth kernel  $K$  with compact support such that  $K(0) = 1$  and for some  $a = a_{\delta,K} > 0$  chosen in such a way that  $B_T \in \mathbb{B}_{b,m} \cap \mathcal{H}_{\mathcal{D}}^{\beta}(L)$  for every  $T \geq 0$ . Such a choice is always possible (if  $B_0 \neq \max\{C, 1\}$  identically in a neighbourhood of  $x$ , which we may and

will assume from now on) thanks to the assumption  $\|B_0\|_{\mathcal{H}_D^\beta} \leq L - \delta$ ; it suffices then to impose  $\|ah_T^{\beta+1}K_{h_T}(\cdot - x)\|_{\mathcal{H}_D^\beta} \leq \delta$  which is easily obtained by picking  $a_{\delta,K}$  sufficiently small.

Also, by construction, we have  $B_0(y) \leq B_T(y)$  for every  $y \geq 0$  hence  $\lambda_{B_0} \leq \lambda_{B_T}$ , compare the proof of Proposition 12 (ii) and at  $y = x$ , the lower estimate  $|B_0(x) - B_T(x)| = a_{\delta,K}h_T^\beta(B_0)$  holds, and this quantity is of order  $e^{-\lambda_{B_0}\frac{\beta}{2\beta+1}T}$ .

*Step 2.* Abusing notation slightly, we further write  $\mathbb{P}_B$  for  $\mathbb{P}_{B|\mathcal{F}_T}$ , *i.e.* the measure in restriction to the  $\sigma$ -field generated by the observation  $(X(t))_{0 \leq t \leq T}$ . Since  $B_0, B_T \in \mathbb{B}_{b,m} \cap \mathcal{H}_D^\beta(L)$ , for an arbitrary estimator  $\widehat{B}_T(x)$  and any constant  $C' > 0$  the maximal risk is bounded below by

$$\begin{aligned} & \max_{B \in \{B_0, B_T\}} \mathbb{P}_B(e^{\lambda_B \frac{\beta}{2\beta+1}T} |\widehat{B}_T(x) - B(x)| \geq C') \\ & \geq \frac{1}{2} \left( \mathbb{P}_{B_0}(e^{\lambda_{B_0} \frac{\beta}{2\beta+1}T} |\widehat{B}_T(x) - B_0(x)| \geq C') + \mathbb{P}_{B_T}(e^{\lambda_{B_T} \frac{\beta}{2\beta+1}T} |\widehat{B}_T(x) - B_T(x)| \geq C') \right) \\ & \geq \frac{1}{2} \mathbb{E}_{B_0} \left[ \mathbf{1}_{\{e^{\lambda_{B_0} \frac{\beta}{2\beta+1}T} |\widehat{B}_T(x) - B_0(x)| \geq C'\}} + \mathbf{1}_{\{e^{\lambda_{B_T} \frac{\beta}{2\beta+1}T} |\widehat{B}_T(x) - B_T(x)| \geq C'\}} \right] - \frac{1}{2} \|\mathbb{P}_{B_0} - \mathbb{P}_{B_T}\|_{TV}. \end{aligned}$$

By triangle inequality, we have

$$\begin{aligned} & e^{\lambda_{B_0} \frac{\beta}{2\beta+1}T} |\widehat{B}_T(x) - B_0(x)| + e^{\lambda_{B_T} \frac{\beta}{2\beta+1}T} |\widehat{B}_T(x) - B_T(x)| \\ & \geq e^{\min\{\lambda_{B_0}, \lambda_{B_T}\} \frac{\beta}{2\beta+1}T} |B_0(x) - B_T(x)| \geq a_{K,\delta} \end{aligned}$$

by Step 1, so if we pick  $C' < a_{K,\delta}/2$ , one of the two indicators within the expectation above must be equal to one with full  $\mathbb{P}_{B_0}$ -probability. In that case

$$\max_{B \in \{B_0, B_T\}} \mathbb{P}_B(e^{\lambda_B \frac{\beta}{2\beta+1}T} |\widehat{B}_T(x) - B(x)| \geq C') \geq \frac{1}{2}(1 - \|\mathbb{P}_{B_0} - \mathbb{P}_{B_T}\|_{TV})$$

and Theorem 8 is thus proved if  $\limsup_{T \rightarrow \infty} \|\mathbb{P}_{B_0} - \mathbb{P}_{B_T}\|_{TV} < 1$ .

*Step 3.* By Pinsker's inequality, we have  $\|\mathbb{P}_{B_0} - \mathbb{P}_{B_T}\|_{TV} \leq \frac{\sqrt{2}}{2} \left( \mathbb{E}_{B_0} \left[ \log \frac{d\mathbb{P}_{B_0}}{d\mathbb{P}_{B_T}} \right] \right)^{1/2}$ . By Theorem 3.5 in [Löc02a], the measures  $\mathbb{P}_{B_0}$  and  $\mathbb{P}_{B_T}$  are equivalent on  $\mathcal{F}_T$  and we have

$$\log \left( \frac{d\mathbb{P}_{B_T}}{d\mathbb{P}_{B_0}} \right) = \sum_{u \in \mathring{\mathcal{J}}_T} \log \left( \frac{B_T}{B_0}(\zeta_u) \right) - \int_0^T \sum_{u \in \partial \mathcal{T}_s} (B_T - B_0)(\zeta_u^s) ds,$$

where  $\zeta_u^t$  denotes the age of the cell  $u$  at time  $t \in I_u = [b_u, d_u)$ . Using  $-\log(1+x) \leq x^2 - x$  if  $x \geq -1/2$  and setting  $\varepsilon_T(y) = a_{K,\delta}h_T(B_0)^{\beta+1}K_{h_T(B_0)}(y-x)$ , we further infer

$$\begin{aligned} \|\mathbb{P}_{B_0} - \mathbb{P}_{B_T}\|_{TV}^2 & \leq \frac{1}{2} \left( \mathbb{E}_{B_0} \left[ \sum_{u \in \mathring{\mathcal{J}}_T} \frac{\varepsilon_T^2(\zeta_u)}{B_0^2} \right] - \mathbb{E}_{B_0} \left[ \sum_{u \in \mathring{\mathcal{J}}_T} \frac{\varepsilon_T}{B_0}(\zeta_u) \right] + \int_0^T \mathbb{E}_{B_0} \left[ \sum_{u \in \partial \mathcal{T}_s} \varepsilon_T(\zeta_u^s) \right] ds \right) \\ & = \frac{1}{2m} \int_0^T e^{\lambda_{B_0}s} \mathbb{E}_{B_0} \left[ \frac{\varepsilon_T^2}{B_0^2}(\tilde{\chi}_s) H_{B_0}(\tilde{\chi}_s) \right] ds \end{aligned}$$

by (23) and (24) in Proposition 10 and the fact that the last two terms cancel. We now use the same kind of estimates as in the proof of Proposition 15, Step 1 with test function  $g = \varepsilon_T/B_0$  to

finally get

$$\|\mathbb{P}_{B_0} - \mathbb{P}_{B_T}\|_{TV}^2 \lesssim e^{\lambda_{B_0} T} |B_0^{-1} \varepsilon_T|_2^2 + |B_0^{-1} \varepsilon_T|_\infty^2 \lesssim a_{K,\delta}^2$$

and this term can be made arbitrarily small by picking  $a_{K,\delta}$  small enough.

## 5.6 PROOF OF PROPOSITION 9

Pick  $B \in \mathbb{B}_{b,m}$ . We need to prove that  $\lambda_B \leq \rho_B = \inf_x H_B(x)$ . By representation (3), we have

$$\begin{aligned} H_B(x) &= \frac{me^{-\lambda_B x} f_B(x)}{1 - m \int_0^x e^{-\lambda_B y} f_B(y) dy} \\ &= \frac{me^{-\lambda_B x} B(x) e^{-\int_0^x B(y) dy}}{1 - m \int_0^x e^{-\lambda_B y} B(y) e^{-\int_0^y B(u) du} dy}. \end{aligned}$$

Set

$$G_B(x) = me^{-\lambda_B x} B(x) e^{-\int_0^x B(y) dy} - \lambda_B (1 - m \int_0^x e^{-\lambda_B y} B(y) e^{-\int_0^y B(u) du} dy).$$

The statement  $\lambda_B \leq \rho_B$  is equivalent to proving that  $\inf_{x \geq 0} G_B(x) \geq 0$ . We first claim that

$$B(x) \leq \tilde{B}(x) \text{ for every } x \in (0, \infty) \text{ implies } \lambda_B \leq \lambda_{\tilde{B}}.$$

Indeed, in that case, one can construct on the same probability space two random variables  $\tau_B$  with density  $f_B$  and  $\tau_{\tilde{B}}$  with density  $f_{\tilde{B}}$  such that  $\tau_B \geq \tau_{\tilde{B}}$ . It follows that  $\phi_B(\lambda) = \mathbb{E}[e^{-\lambda \tau_B}] \leq \phi_{\tilde{B}}(\lambda) = \mathbb{E}[e^{-\lambda \tau_{\tilde{B}}}]$  for every  $\lambda \geq 0$ . Also,  $\phi_B$  and  $\phi_{\tilde{B}}$  are both non-increasing, vanishing at infinity, and  $\phi_B(0) = \phi_{\tilde{B}}(0) = 1 > \frac{1}{m}$ . Consequently, the values  $\lambda_B$  and  $\lambda_{\tilde{B}}$  such that  $\phi_B(\lambda_B) = \phi_{\tilde{B}}(\lambda_{\tilde{B}}) = \frac{1}{m}$  necessarily satisfy  $\lambda_B \leq \lambda_{\tilde{B}}$  hence the claim. Now, for constant functions  $B(x) = \alpha$ , we clearly have  $\lambda_B = (m-1)\alpha$  and this enables us to infer

$$\lambda_B \leq (m-1) \sup_x B(x).$$

Remember now that  $B \in \mathbb{B}_{b,m}$  implies  $b \leq B(x) \leq \frac{m}{m-1}b$  for every  $x \geq 0$ . Therefore

$$\lambda_B \leq (m-1) \frac{m}{m-1} b = mb \leq mB(0) \quad (41)$$

and  $G_B(0) = mB(0) - \lambda_B \geq 0$  follows. Moreover, one readily checks that

$$G'_B(x) = me^{-\lambda_B x} e^{-\int_0^x B(y) dy} (B'(x) - B(x)^2) \leq 0$$

since  $B'(x) - B(x)^2 \leq 0$  as soon as  $B \in \mathbb{B}_{b,m}$ . So  $G_B$  is non-increasing,  $G_B(0) \geq 0$  and its infimum is thus attained for  $x \rightarrow \infty$ . Since  $G_B(\infty) = 0$ , we conclude  $\inf_{x \geq 0} G_B(x) \geq 0$ .

We finally briefly indicate how to show that  $\mathcal{B}_{b,C}^-$  is non-trivial when  $C > mb/(m-1)$ . To that end, pick  $0 < x_0 \leq x_1$ ,  $mb/(m-1) < c \leq C$  and let  $B(x) = b$  for  $x \leq x_0$ ,  $B(x) = c$  for  $x \geq x_1$  and any smooth continuation between  $x_0$  and  $x_1$  bounded above by  $C$  and below by  $b$ . Then, having  $b, c$  such that  $2m(m+2)b/(m-1) < c$  and suitable choices for  $x_0$  and  $x_1$  implies  $\rho_B < \lambda_B/2$ . Having  $2mb/(m-1) > c$  and suitable choices for  $x_0, x_1$  implies  $\rho_B < \lambda_B \leq 2\rho_B$ . The



computations, based on the same kind of estimates, are rather tedious but not difficult. We omit the details.

## 6 APPENDIX

### 6.1 HEURISTICS FOR THE CONVERGENCES TO THE LIMITS (8) AND (9)

*Information from  $\mathcal{E}^T(\partial\mathcal{T}_T, g)$*

Heuristically, we postulate for large  $T$  the approximation

$$\mathcal{E}^T(\partial\mathcal{T}_T, g) \sim \frac{1}{\mathbb{E}[|\partial\mathcal{T}_T|]} \mathbb{E} \left[ \sum_{u \in \partial\mathcal{T}_T} g(\zeta_u^T) \right].$$

Then, a classical result based on renewal theory (see Theorem 17.1 pp 142-143 of [Har63]) gives the estimate

$$\mathbb{E}[|\partial\mathcal{T}_T|] \sim \kappa_B e^{\lambda_B T}, \quad (42)$$

where  $\lambda_B > 0$  is the Malthus parameter defined in (6) and  $\kappa_B > 0$  is an explicitly computable constant (that also depends on  $m$ , see [Har63] and also Lemma 13 below). As for the numerator, call  $\chi_t$  the age of a particle at time  $t$  along a branch of the tree picked at random uniformly at each branching event. The process  $(\chi_t)_{t \geq 0}$  is Markov process with values in  $[0, \infty)$  with infinitesimal generator

$$\mathcal{A}_B g(x) = g'(x) + B(x)(g(0) - g(x)) \quad (43)$$

densely defined on continuous functions vanishing at infinity. Assume for simplicity that each cell  $u \in \mathcal{U}$  has exactly  $m$  children at each division. It is then relatively straightforward to obtain the identity

$$\mathbb{E} \left[ \sum_{u \in \partial\mathcal{T}_T} g(\zeta_u^T) \right] = \mathbb{E}[m^{N_T} g(\chi_T)], \quad (44)$$

where  $N_t = \sum_{s \leq t} \mathbf{1}_{\{\chi_s - \chi_{s-} < 0\}}$  is the counting process associated to  $(\chi_t)_{t \geq 0}$ , see Proposition 10 in a general setting. Putting together (42) and (44), we thus expect

$$\mathcal{E}^T(\partial\mathcal{T}_T, g) \sim \kappa_B^{-1} e^{-\lambda_B T} \mathbb{E}[m^{N_T} g(\chi_T)],$$

and we anticipate that the term  $e^{-\lambda_B T}$  should somehow be compensated by the term  $m^{N_T}$  within the expectation. To that end, following Cloez [Clo11] (and also in Bansaye *et al.* [BDMT11] when  $B$  is constant) one introduces an auxiliary “biased” Markov process  $(\tilde{\chi}_t)_{t \geq 0}$ , with generator  $\mathcal{A}_{H_B}$  for a biasing function  $H_B(x)$  characterised by

$$f_{H_B}(x) = m e^{-\lambda_B x} f_B(x), \quad x \geq 0, \quad (45)$$

where  $f_B(x) = B(x) \exp(-\int_0^x B(y) dy)$  denotes the density associated to the division rate  $B$ , as follows from (3) or (5). This implies

$$H_B(x) = \frac{m e^{-\lambda_B x} f_B(x)}{1 - m \int_0^x e^{-\lambda_B y} f_B(y) ds}.$$

Furthemore, this choice (and this choice only, see Proposition 10) enables us to obtain

$$e^{-\lambda_B T} \mathbb{E}[m^{N_T} g(\chi_T)] = m^{-1} \mathbb{E}[g(\tilde{\chi}_T) B(\tilde{\chi}_T)^{-1} H_B(\tilde{\chi}_T)] \quad (46)$$

with  $\tilde{\chi}_0 = 0$  under  $\mathbb{P}$ . Moreover  $(\tilde{\chi}_t)_{t \geq 0}$  is geometrically ergodic, with invariant probability  $c_B \exp(-\int_0^x H_B(y) dy) dx$  (see Proposition 12). We further anticipate

$$\begin{aligned} \mathbb{E}[g(\tilde{\chi}_T) B(\tilde{\chi}_T)^{-1} H_B(\tilde{\chi}_T)] &\sim c_B \int_0^\infty g(x) B(x)^{-1} H_B(x) e^{-\int_0^x H_B(y) dy} dx \\ &= m c_B \int_0^\infty g(x) e^{-\lambda_B x} B(x)^{-1} f_B(x) dx \end{aligned}$$

assuming everthing is well-defined, since  $H_B(x) \exp(-\int_0^x H_B(y) dy) = f_{H_B}(x) = m e^{-\lambda_B} f_B(x)$  by (45). Finally, we have  $\kappa_B^{-1} c_B = \lambda_B \frac{m}{m-1}$  by Lemma 13 which enables us to conclude

$$\mathcal{E}^T(\partial \mathcal{J}_T, g) \sim \partial \mathcal{E}_B(g),$$

where

$$\partial \mathcal{E}_B(g) = \lambda_B \frac{m}{m-1} \int_0^\infty g(x) e^{-\lambda_B x} e^{-\int_0^x B(y) dy} dx.$$

Unfortunately, the statistical information extracted from  $\mathcal{E}^T(\partial \mathcal{J}_T, g)$  does not enable us to obtain classical optimal rates of convergence, since the form of  $\partial \mathcal{E}_B(g)$  involves an antiderivative of  $B$  leading to so-called ill-posedness. This is discussed at length in Section 3.3. We thus investigate in a second step the statistical information we can get from  $\mathring{\mathcal{J}}_T$ .

*Information from  $\mathcal{E}^T(\mathring{\mathcal{J}}_T, g)$*

The situation is a bit different if we allow for data in  $\mathring{\mathcal{J}}_T$ . Note first that  $\zeta_u^T = \zeta_u$  on the event  $u \in \mathring{\mathcal{J}}_T$ . We also have in that case a many-to-one formula that now reads

$$\mathbb{E}\left[\sum_{u \in \mathring{\mathcal{J}}_T} g(\zeta_u^T)\right] = \mathbb{E}\left[\sum_{u \in \mathring{\mathcal{J}}_T} g(\zeta_u)\right] = m^{-1} \int_0^T e^{\lambda_B s} \mathbb{E}[g(\tilde{\chi}_s) H_B(\tilde{\chi}_s)] ds, \quad (47)$$

where  $(\tilde{\chi}_t)_{t \geq 0}$  is the one-dimensional auxiliary Markov process with generator  $\mathcal{A}_{H_B}$ , see (43), where  $H_B$  is characterised by (45) above. Assuming again ergodicity, we approximate the right-hand side of (47) and obtain

$$\begin{aligned} \mathbb{E}\left[\sum_{u \in \mathring{\mathcal{J}}_T} g(\zeta_u)\right] &\sim c_B m^{-1} \frac{e^{\lambda_B T}}{\lambda_B} \int_0^\infty g(x) H_B(x) e^{-\int_0^x H_B(u) du} dx \\ &= c_B \frac{e^{\lambda_B T}}{\lambda_B} \int_0^\infty g(x) e^{-\lambda_B x} f_B(x) dx \end{aligned}$$

since  $H_B(x) \exp(-\int_0^x H_B(y) dy) = f_{H_B}(x) = m e^{-\lambda_B} f_B(x)$  by (45). We again have an approximation of the type (42) with another constant  $\kappa'_B$ , see Lemma 14 and we eventually expect

$$\mathcal{E}^T(\mathring{\mathcal{J}}_T, g) \sim \mathring{\mathcal{E}}_B(g),$$

where

$$\dot{\mathcal{E}}_B(g) = \frac{c_B}{\lambda_B \kappa'_B} \int_0^\infty g(x) e^{-\lambda_B x} f_B(x) dx = m \int_0^\infty g(x) e^{-\lambda_B x} f_B(x) dx$$

as  $T \rightarrow \infty$ , where the last equality stems from the identity  $c_B = \lambda_B \kappa'_B m$  that can be readily derived by picking  $g = 1$  and using (45) together with the fact that  $f_{H_B}$  is a density function.

## 6.2 PROOF OF PROPOSITION 10

We start with a continuous time rooted tree which is a Bellman Harris process in the sense of Definition 1, so we have random variables  $(\zeta_u, \nu_u, u \in \mathcal{U})$  satisfying properties (i), (ii) and (iii) of the definition. For  $u \in \mathcal{U}$ , and  $t \geq 0$ , let  $\Lambda_t^u = \sum_{v \prec u(t)} \log(\nu_v)$ ,  $t \geq 0$  denote the process that encodes the birth times and the numbers of children of the ancestors of  $u$ . Let  $\vartheta = (\vartheta_k)_{k \geq 0}$  with  $\vartheta_k \in \mathcal{U}$  be such that  $|\vartheta_k| = k$  for  $k \geq 1$  (with  $\vartheta_0 = \emptyset$ ) and  $\vartheta_k \preceq \vartheta_l$  for  $k \leq l$ . We associate to  $\vartheta$  a counting process  $(N_t)_{t \geq 0}$  via the relationship

$$b_{\vartheta_{N_t}} \leq t < d_{\vartheta_{N_t}}, \quad t \geq 0.$$

This enables us to further obtain a ‘‘tagged process of age’’ such that  $\chi_t = \zeta_{\vartheta_{N_t}}^t$  for  $t \in I_{\vartheta_{N_t}}$  and also a process  $(\Lambda_t)_{t \geq 0}$  that encodes the genealogy of the tagged branch

$$\Lambda_t = \sum_{k=1}^{N_t} \log(\nu_{\vartheta_k}), \quad t \geq 0.$$

*Step 1.* Let us pick  $\vartheta$  at random along the genealogical tree  $\mathcal{T}$ . This means that if  $\mathcal{H}_n$  denotes the sigma-field generated by  $(\zeta_u, \nu_u, u \in \mathcal{T}, |u| \leq n)$ , then on the event  $\{t \in I_u\}$  (i.e. the particle  $u$  is living at time  $t$ ), we have (or rather, we set)

$$\mathbb{P}(\vartheta_{N_t} = u | \mathcal{H}_{|u|}) = \prod_{v \prec u} \frac{1}{\nu_v} = e^{-\Lambda_t^u}.$$

It is not difficult to see that  $(\chi_t)_{t \geq 0}$  is a Markov process with generator  $\mathcal{A}_B$ . By definition of  $(\chi_t)_{t \geq 0}$  and  $(\Lambda_t)_{t \geq 0}$ , it follows that  $\mathbb{E}[e^{\Lambda_T} g(\chi_T)]$  can be rewritten as

$$\sum_{u \in \mathcal{U}} \mathbb{E}[e^{\Lambda_T} g(\chi_T) \mathbf{1}_{\{T \in I_u, u = \vartheta_{N_T}\}}] = \sum_{u \in \mathcal{U}} \mathbb{E}[e^{\Lambda_t^u} g(\zeta_u^T) \mathbf{1}_{\{T \in I_u, u = \vartheta_{N_T}\}}] = \sum_{u \in \mathcal{U}} \mathbb{E}[g(\zeta_u^T) \mathbf{1}_{\{T \in I_u\}}],$$

where the last equality is obtained by conditioning with respect to  $\mathcal{H}_{|u|}$ .

*Step 2.* For  $j \geq 1$ , let  $\tau_j = \inf\{t \geq 0, N_t \geq j\} - \inf\{t \geq 0, N_t \geq j-1\}$  denote the durations between the jumps of  $(\chi_t)_{t \geq 0}$ , so that

$$e^{\Lambda_T} g(\chi_T) = \sum_{k=0}^{\infty} e^{\sum_{j=1}^k \log(\nu_{\vartheta_j})} g(T - \sum_{j=1}^k \tau_j) \mathbf{1}_{\{\sum_{j=1}^k \tau_j \leq T < \sum_{j=1}^{k+1} \tau_j\}}.$$

By properties (i)-(iii) of Definition 1, the  $\tau_i$  are independent with common distribution  $f_B(x)dx$ , and independent of the  $\nu_{\vartheta_k}$  that are independent with common distribution  $(p_k)_{k \geq 1}$ . We thus infer

that  $\mathbb{E}[e^{\Lambda_T} g(\chi_T)]$  is equal to

$$\sum_{k=0}^{\infty} \sum_{h_j \geq 1, j \leq k} e^{\sum_{j=1}^k \log(h_j)} \prod_{j=1}^k p_{h_j} \int_{[0, \infty)^{k+1}} g\left(T - \sum_{j=1}^k t_j\right) \mathbf{1}_{\{\sum_{j=1}^k t_j \leq T < \sum_{j=1}^{k+1} t_j\}} \prod_{j=1}^{k+1} f_B(t_j) dt_1 \dots dt_{k+1}.$$

We set  $\bar{F}_B(x) = 1 - \int_0^x f_B(y) dy$  and  $q_k = m^{-1} k p_k$ , so that  $(q_k)_{k \geq 1}$  defines a probability distribution. Using  $f_{H_B}(x) = m e^{-\lambda_B x} f_B(x)$ , we can rewrite the preceding formula so that

$$\begin{aligned} e^{-\lambda_B T} \mathbb{E}[e^{\Lambda_T} g(\chi_T)] &= \sum_{k=0}^{\infty} \sum_{h_j \geq 1, j \leq k} \prod_{j=1}^k q_{h_j} \int_{[0, \infty)^k} g\left(T - \sum_{j=1}^k t_j\right) \mathbf{1}_{\{T - \sum_{j=1}^k t_j \geq 0\}} e^{-\lambda_B (T - \sum_{j=1}^k t_j)} \\ &\quad \times \bar{F}_B\left(T - \sum_{j=1}^k t_j\right) \prod_{j=1}^k f_{H_B}(t_j) dt_1 \dots dt_k. \end{aligned}$$

*Step 3.* Putting  $W_B(x) = m e^{-\lambda_B x} \bar{F}_B(x) / \bar{F}_{H_B}(x)$ , we finally obtain the representation

$$e^{-\lambda_B T} \mathbb{E}[e^{\Lambda_T} g(\chi_T)] = \frac{1}{m} \mathbb{E}\left[g(\tilde{\chi}_T) W_B(\tilde{\chi}_T)\right],$$

where  $(\tilde{\chi}_t)_{t \geq 0}$  is a Markov process with generator  $\mathcal{A}_{H_B}$  that can be constructed in the same way as  $(\chi_t)_{t \geq 0}$ , substituting  $f_B$  by  $f_{H_B}$ . Straightforward computations give  $W_B(x) = \frac{H_B(x)}{B(x)}$ . Putting together all the three steps, we have proved

$$\sum_{u \in \mathcal{U}} \mathbb{E}[g(\zeta_u^T) \mathbf{1}_{\{T \in I_u\}}] = \mathbb{E}[e^{\Lambda_T} g(\chi_T)] = \frac{e^{\lambda_B T}}{m} \mathbb{E}\left[g(\tilde{\chi}_T) \frac{H_B(\tilde{\chi}_T)}{B(\tilde{\chi}_T)}\right].$$

Noticing that  $\sum_{u \in \mathcal{U}} \mathbb{E}[g(\zeta_u^T) \mathbf{1}_{\{T \in I_u\}}]$  is nothing but  $\mathbb{E}[\sum_{u \in \partial \mathcal{T}_T} g(\zeta_u^T)]$  establishes (23).

*Step 4.* By definition of the set  $\mathring{\mathcal{T}}_T$ ,

$$\mathbb{E}\left[\sum_{u \in \mathring{\mathcal{T}}_T} g(\zeta_u)\right] = \sum_{u \in \mathcal{U}} \mathbb{E}\left[g(\zeta_u) \mathbf{1}_{\{b_u + \zeta_u \leq T\}} \mathbf{1}_{\{u \in \mathcal{T}\}}\right].$$

We denote by  $\mathcal{F}_t$  the sigma-field generated by  $(\zeta_u^s, u \in \partial \mathcal{T}_s, s \leq t)$  and we note that  $d_u \mathbf{1}_{\{u \in \mathcal{T}\}}$  is a stopping time for the filtration  $(\mathcal{F}_t)_{t \geq 0}$ . Conditioning w.r.t  $\mathcal{F}_{b_u}$ , using that the  $\zeta_u$  are independent of  $\mathcal{F}_{b_u}$ , we successively obtain

$$\begin{aligned} \mathbb{E}\left[\sum_{u \in \mathring{\mathcal{T}}_T} g(\zeta_u)\right] &= \sum_{u \in \mathcal{U}} \mathbb{E}\left[\mathbf{1}_{\{u \in \mathcal{T}\}} \int_0^{\infty} g(x) \mathbf{1}_{\{b_u + x \leq T\}} B(x) e^{-\int_0^x B(y) dy} dx\right] \\ &= \sum_{u \in \mathcal{U}} \mathbb{E}\left[\mathbf{1}_{\{u \in \mathcal{T}\}} \int_0^{\infty} \left(\int_0^y g(x) B(x) \mathbf{1}_{\{b_u + x \leq T\}} dx\right) B(y) e^{-\int_0^y B(z) dz} dy\right] \\ &= \sum_{u \in \mathcal{U}} \mathbb{E}\left[\mathbf{1}_{\{u \in \mathcal{T}\}} \int_0^{\zeta_u} g(x) B(x) \mathbf{1}_{\{b_u + x \leq T\}} dx\right]. \end{aligned}$$

Next, using that  $\zeta_u^s = s - b_u$  for  $s \in I_u$ , we get

$$\mathbb{E}\left[\sum_{u \in \dot{\mathcal{T}}_T} g(\zeta_u)\right] = \sum_{u \in \mathcal{U}} \mathbb{E}\left[\mathbf{1}_{\{u \in \mathcal{T}\}} \int_{b_u}^{d_u} g(\zeta_u^s) B(\zeta_u^s) \mathbf{1}_{\{s \leq T\}} ds\right].$$

Observing that  $\{s \in I_u\} = \{u \in \partial \mathcal{T}_s\}$ , we finally infer

$$\mathbb{E}\left[\sum_{u \in \dot{\mathcal{T}}_T} g(\zeta_u)\right] = \int_0^\infty \mathbb{E}\left[\sum_{u \in \partial \mathcal{T}_s} g(\zeta_u^s) B(\zeta_u^s) \mathbf{1}_{\{s \leq T\}} ds\right].$$

Using (23) completes the proof of (24).

### 6.3 PROOF OF (26) AND (27) OF PROPOSITION 11

Whenever  $(u, v) \in \mathcal{FU}$  there exist  $w, \tilde{u}$  and  $\tilde{v} \in \mathcal{U}$  together with integers  $i \neq j$ , such that  $u = wi\tilde{u}$  and  $v = wj\tilde{v}$ . Conditioning w.r.t  $\mathcal{F}_{d_w}$ , using the branching property between descendants of  $w$  and the strong Markov property at time  $d_w$ , we have

$$\begin{aligned} \mathbb{E}\left[\sum_{(u,v) \in \mathcal{FT} \cap \dot{\mathcal{T}}_T^2} g(\zeta_u)g(\zeta_v)\right] &= \sum_{(u,v) \in \mathcal{FU}} \mathbb{E}\left[g(\zeta_u) \mathbf{1}_{\{d_u < T\}} \mathbf{1}_{\{u \in \mathcal{T}\}} g(\zeta_v) \mathbf{1}_{\{d_v < T\}} \mathbf{1}_{\{v \in \mathcal{T}\}}\right] \\ &= \sum_{w \in \mathcal{U}} \sum_{i \neq j} \mathbb{E}\left[\mathbb{E}\left[\sum_{\tilde{u} \in \mathcal{U}} g(\zeta_{wi\tilde{u}}) \mathbf{1}_{\{d_{wi\tilde{u}} < T\}} \mathbf{1}_{\{wi\tilde{u} \in \mathcal{T}\}} \mid \mathcal{F}_{d_w}\right] \right. \\ &\quad \left. \times \mathbb{E}\left[\sum_{\tilde{v} \in \mathcal{U}} g(\zeta_{wj\tilde{v}}) \mathbf{1}_{\{d_{wj\tilde{v}} < T\}} \mathbf{1}_{\{wj\tilde{v} \in \mathcal{T}\}} \mid \mathcal{F}_{d_w}\right]\right] \\ &= \sum_{w \in \mathcal{U}} \sum_{i \neq j} \mathbb{E}\left[\mathbf{1}_{\{wi \in \mathcal{T}, wj \in \mathcal{T}\}} \left(\mathbb{E}\left[\sum_{u \in \mathcal{T}} g(\zeta_u) \mathbf{1}_{\{d_u < T-t\}}\right]_{t=d_w}\right)^2 \mathbf{1}_{\{d_w < T\}}\right]. \end{aligned}$$

Notice that  $\{wi \in \mathcal{T}, wj \in \mathcal{T}\} = \{w \in \mathcal{T}\} \cap \{i \leq \nu_w, j \leq \nu_w\}$ , and  $\nu_w$  is independent of  $d_w$  and has distribution  $(p_k)_{k \geq 1}$ . We conclude by using (24) of Proposition 10 (slightly generalized for test functions that depend on  $d_u$  and  $\zeta_u$ ).

Let us now turn to (27). For  $u, v \in \mathcal{T}$  with  $u \prec v$ , we have  $uiw = v$  for some  $w \in \mathcal{T}$  and some integer  $i$ . It follows that

$$\begin{aligned} \mathbb{E}\left[\sum_{\substack{u, v \in \dot{\mathcal{T}}_T, \\ u \prec v}} g(\zeta_u)g(\zeta_v)\right] &= \sum_{u \in \mathcal{U}} \sum_i \mathbb{E}\left[g(\zeta_u) \mathbf{1}_{\{d_u < T\}} \mathbf{1}_{\{u \in \mathcal{T}\}} \mathbb{E}\left[\sum_{w \in \mathcal{U}} g(\zeta_{uiw}) \mathbf{1}_{\{d_{uiw} < T\}} \mathbf{1}_{\{uiw \in \mathcal{T}\}} \mid \mathcal{F}_{d_u}\right]\right] \\ &= \sum_{u \in \mathcal{U}} \sum_i \mathbb{E}\left[g(\zeta_u) \mathbf{1}_{\{0 \leq d_u < T\}} \mathbf{1}_{\{ui \in \mathcal{T}\}} \mathbb{E}\left[\sum_{w \in \mathcal{T}} g(\zeta_w) \mathbf{1}_{\{d_w < T-s\}}\right]_{s=d_u} \mathbf{1}_{\{d_u < T\}}\right] \end{aligned}$$

conditioning with respect to  $\mathcal{F}_{d_u}$  on  $\{d_u < T\}$  and applying the branching property. Next, we have

$$\mathbb{E}\left[\sum_{w \in \mathcal{T}} g(\zeta_w) \mathbf{1}_{\{d_w < T-s\}}\right] = \mathbb{E}\left[\sum_{w \in \dot{\mathcal{T}}_{T-s}} g(\zeta_w)\right] = \frac{1}{m} \int_0^{T-s} e^{\lambda B^t} P_{H_B}^t(gH_B)(0) dt$$

by (24) of Proposition 10. Since  $\{ui \in \mathcal{T}\} = \{i \leq \nu_u\}$ , and  $\nu_u$  is independent of  $\zeta_u$  and  $d_u$  and has

distribution with expectation  $m$ , we obtain

$$\mathbb{E}\left[\sum_{\substack{u,v \in \tilde{\mathcal{T}}_T, \\ u \prec v}} g(\zeta_u)g(\zeta_v)\right] = \mathbb{E}\left[\sum_{u \in \tilde{\mathcal{T}}_T} g(\zeta_u) \int_0^{T-d_u} e^{\lambda_B t} P_{H_B}^t(gH_B)(0) dt\right]$$

and we conclude by using once more (24) of Proposition 10 (slightly generalized for test functions that depend on  $d_u$  and  $\zeta_u$ ).

#### 6.4 PROOF OF LEMMA 16

Let  $\tau$  denote the first jump time of the process  $(\tilde{\chi}_t)_{t \geq 0}$ . Conditioning on  $\{\tau > t\}$  and applying the strong Markov property yields

$$P_{H_B}^t(gH_B)(0) = g(t)H_B(t)\mathbb{P}(\tau > t) + \int_0^t P_{H_B}^{t-u}(gH_B)(0)f_{H_B}(u)du.$$

The function  $t \rightsquigarrow u(t) = P_{H_B}^t(gH_B)(0)$  satisfies a renewal equation of the form  $u = u_0 + u \star f_{H_B}$ , with locally bounded initial condition  $u_0 = gH_B\mathbb{P}(\tau > \cdot)$  and renewal distribution  $f_{H_B}(y)dy$ . Its unique solution is given by

$$P_{H_B}^t(gH_B)(0) = g(t)H_B(t)\mathbb{P}(\tau > t) + \int_0^t g(t-s)H_B(t-s)\mathbb{P}(\tau > t-s)d\mathbb{E}[\tilde{N}_s],$$

where  $\tilde{N}_t = \sum_{s \leq t} \mathbf{1}_{\{\tilde{\chi}_s - \tilde{\chi}_{s-} < 0\}}$  is the counting process associated to  $(\tilde{\chi}_t)_{t \geq 0}$ . By construction, we have  $\mathbb{E}[\tilde{N}_t] = \mathbb{E}\left[\int_0^t H_B(\tilde{\chi}_s)ds\right]$  and  $\mathbb{P}(\tau > t) = \int_t^\infty f_{H_B}(y)dy = m \int_t^\infty e^{-\lambda_B y} f_B(y)dy \leq me^{-\lambda_B t}$ , therefore

$$|P_{H_B}^t(gH_B)(0)| \leq |g(t)|e^{-\lambda_B t}m|H_B|_\infty + |H_B|_\infty^2 \int_0^t |g(u)|du$$

and we obtain the desired estimate thanks to the fact that  $H_B$  is uniformly bounded over  $\mathcal{B}$ .

---

## Références du Chapitre I

---

- [AK77] K. B. Athreya and N. Keiding. *Estimation theory for continuous-time branching processes*. Sankhya: The Indian Journal of Statistics, Series A 39 (1977), 101–123.
- [AN72] K. B. Athreya and P. Ney. *Branching processes*. Springer-Verlag, New-York, 1972.
- [BDMT11] V. Bansaye, J.-F. Delmas, L. Marsalle and V. C. Tran. *Limit theorems for Markov processes indexed by continuous time Galton-Watson trees*. The Annals of Applied Probability, 21 (2011), 2263–2314.
- [BPDG14] S. V. Bisteki-Penda, H. Djellout and A. Guillin. *Deviations inequalities, moderate deviations and some limits theorems for bifurcating Markov chains with application*. Annals of Applied Probability 24 (2014), 235–291.
- [Clo11] B. Cloez. *Limit theorems for some branching measure-valued processes*, hal-00598030 (2011).
- [DM10] J.-F. Delmas and L. Marsalle. *Detection of cellular aging in a Galton-Watson process*. Stochastic Processes and their Applications, 120 (2010), 2495–2519.
- [DHKR15] M. Doumic, M. Hoffmann, N. Krell and L. Robert. *Statistical estimation of a growth-fragmentation model observed on a genealogical tree*. Bernoulli, 21 (2015), 1760–1799.
- [DHR12] M. Doumic, M. Hoffmann, P. Reynaud-Bouret and V. Rivoirard. *Nonparametric estimation of the division rate of a size-structured population*. SIAM Journal on Numerical Analysis, 50 (2012), 925–950.
- [DPZ09] M. Doumic, B. Perthame and J. P. Zubelli. *Numerical solution of an inverse problem in size-structured population dynamics*. Inverse Problems, 25 (2009), 25pp.
- [Efr04] S. Efromovich. *Density estimation for biased data*. Annals of Statistics, 32 (2004), 1137–1161.
- [Gut91] P. Guttorp. *Statistical Inference for Branching Processes*. Wiley, 1991.
- [Har63] T. Harris. *The theory of branching processes*. Springer-Verlag, New-York, 1963.
- [HK11] M. Hoffmann and N. Krell. *Statistical analysis of self-similar fragmentation chains*. Bernoulli, 17 (2011), 395–423.
- [HHL02] R. Höpfner, M. Hoffmann and E. Löcherbach. *Nonparametric estimation of the death rate in branching diffusions*. Scandinavian Journal of Statistics, 29 (2002), 665–690.
- [Hyr07] O. Hyrien. *Pseudo-likelihood estimation for discretely observed multitype Bellman-Harris branching processes*. Journal of Statistical Planning and Inference, 137 (2007), 1375–1388.
- [JSvR79] R. Johnson, V. Susarla and J. van Ryzin. *Bayesian nonparametric estimation for age-dependent branching processes*. Stochastic Processes and their Applications, 9 (1979), 307–318.
- [Kie72] J. Kiefer. *Conditional inference*. In: Encyclopedia of Statistical Science, Volume 2 (1972), 103–109, John Wiley, New York.
- [Löc02a] E. Löcherbach. *Likelihood ratio processes for Markovian particle systems with killing and jumps*. Statistical inference for stochastic processes, 5 (2002a), 153–177.
- [Löc02b] E. Löcherbach. *LAN and LAMN for systems of interacting diffusions with branching and immigration*. Annales de l'Institut Henri Poincaré, 38 (2002b), 59–90.

- [Oel90] K. Oelschläger. *Limit Theorems for Age-Structured Populations*. The Annals of Probability, 18 (1990), 290–318.
- [Per07] B. Perthame. *Transport equations arising in biology*. Birkhäuser Frontiers in mathematics edition, 2007.
- [Sch00] R. L. Schilling. *Sobolev embeddings for stochastic processes*. Expositiones Mathematicae, 18 (2000), 239–242.
- [Tsi13] B. Tsirelson *From uniform renewal theorem to uniform large and moderate deviations for renewal-reward processes*. Electronic Communication in Probability, 18 (2013), 1–13.
- [Tsy09] A. Tsybakov. *Introduction to nonparametric estimation*. Springer series in statistics, Springer-Verlag, New-York, 2009.



## CHAPITRE II

---

### Chaîne de Markov bifurquante : estimation adaptative des transitions et mesure invariante

---

La présente étude est conduite en temps discret, ce qui diffère radicalement du chapitre précédent (en particulier, nous nous affranchissons par là du biais de sélection). Nous nous plaçons dans un cadre général, celui des chaînes de Markov bifurquantes. Celles-ci sont le fruit d'une adaptation des chaînes de Markov à une structure d'arbre binaire (noté  $\mathbb{T}$ ). Ce cadre est adapté à l'étude du problème **(B)** *Étudier la transmission du trait individuel d'une cellule à ses descendants, étudier le trait d'une cellule typique*, posé page 2.

Les caractéristiques de la chaîne de Markov bifurquante que nous visons à reconstruire sont au nombre de trois : sa transition, dite  $\mathbb{T}$ -transition pour souligner la structure d'arbre, ainsi que la transition dite moyenne, décrivant une lignée typique de cellules, et sa mesure invariante. Les observables sont formées des traits des cellules appartenant aux  $n$  premières générations. Les procédures que nous mettons en place présentent l'avantage d'être adaptatives et nous évaluons leurs performances à partir de critères les plus généraux possibles.

Le modèle de croissance-fragmentation est un exemple clé s'insérant dans le cadre général du chapitre. Nous considérons le processus discret formé des tailles à la naissance d'une population de cellules croissant exponentiellement vite et se divisant aléatoirement en deux cellules de tailles égales, selon un taux de division  $B$  qui dépend de la taille courante de la cellule. Pour ce modèle spécifiquement, la reconstruction adaptative, par l'observation des  $n$  premières générations, du paramètre fonctionnel clé  $B$  découle de celle des caractéristiques de la chaîne de Markov bifurquante sous-jacente (de l'estimation de la mesure invariante associée à la transition moyenne en particulier). Ce chapitre apporte en conséquence une contribution à la question **(A)** page 2, toute autre que celle du Chapitre I.

## **Abstract**

---

In a first part, we prove Bernstein-type deviation inequalities for bifurcating Markov chains (BMC) under a geometric ergodicity assumption, completing former results of Guyon and Bitseki Penda, Djellout and Guillin. These preliminary results are the key ingredient to implement nonparametric wavelet thresholding estimation procedures: in a second part, we construct nonparametric estimators of the transition density of a BMC, of its mean transition density and of the corresponding invariant density, and show smoothness adaptation over various multivariate Besov classes under  $L^p$ -loss error, for  $1 \leq p < \infty$ . We prove that our estimators are (nearly) optimal in a minimax sense. As an application, we obtain new results for the estimation of the splitting size-dependent rate of growth-fragmentation models and we extend the statistical study of bifurcating autoregressive processes.

---

## Contents

---

|          |   |            |
|----------|---|------------|
| <b>1</b> | <b>INTRODUCTION</b> . . . . .   | <b>87</b>  |
| 1.1      | Bifurcating Markov chains . . . . .   | 87         |
| 1.2      | Objectives . . . . .  | 88         |
| 1.3      | Main results and organisation of the chapter . . . . .                                    | 89         |
| <b>2</b> | <b>DEVIATIONS INEQUALITIES FOR EMPIRICAL MEANS</b> . . . . .                              | <b>91</b>  |
| <b>3</b> | <b>STATISTICAL ESTIMATION</b> . . . . .   | <b>93</b>  |
| 3.1      | Atomic decompositions and wavelets . . . . .  | 94         |
| 3.2      | Estimation of the invariant density $\nu$ . . . . .                                       | 95         |
| 3.3      | Estimation of the density of the mean transition $\mathcal{Q}$ . . . . .                  | 96         |
| 3.4      | Estimation of the density of the $\mathbb{T}$ -transition $\mathcal{P}$ . . . . .         | 97         |
| <b>4</b> | <b>APPLICATIONS</b> . . . . .   | <b>98</b>  |
| 4.1      | Estimation of the size-dependent splitting rate in a growth-fragmentation model . . . . . | 98         |
| 4.2      | Bifurcating autoregressive process . . . . .  | 100        |
| 4.3      | Numerical illustration . . . . .  | 102        |
| <b>5</b> | <b>PROOFS</b> . . . . .   | <b>105</b> |
| 5.1      | Proof of Theorem 4 (i) . . . . .  | 105        |
| 5.2      | Proof of Theorem 4 (ii) . . . . .   | 107        |
| 5.3      | Proof of Theorem 5 (i) . . . . .  | 109        |
| 5.4      | Proof of Theorem 5 (ii) . . . . .   | 111        |
| 5.5      | Proof of Theorem 8 . . . . .  | 111        |
| 5.6      | Preparation for the proof of Theorem 9 . . . . .  | 112        |
| 5.7      | Proof of Theorem 9, upper bound . . . . .   | 113        |
| 5.8      | Proof of Theorem 9, lower bound . . . . .   | 114        |
| 5.9      | Proof of Theorem 10 . . . . .   | 116        |
| 5.10     | Proof of Theorem 12 . . . . .   | 117        |
| <b>6</b> | <b>APPENDIX</b> . . . . .   | <b>118</b> |
| 6.1      | Proof of Lemma 11 . . . . .   | 118        |
| 6.2      | Proof of Lemma 14 . . . . .   | 119        |
| 6.3      | Proof of Lemma 15 . . . . .   | 120        |
|          | <b>REFERENCES</b> . . . . .   | <b>121</b> |

---

## 1 INTRODUCTION

### 1.1 BIFURCATING MARKOV CHAINS

Bifurcating Markov Chains (BMC) are Markov chains indexed by a tree (Athreya and Kang [AK98], Benjamini and Peres [BP94], Takacs [Tak01]) that are particularly well adapted to model and understand dependent data mechanisms involved in cell division. To that end, bifurcating autoregressive models (a specific class of BMC, also considered in the chapter) were first introduced by Cowan and Staudte [CS86]. More recently Guyon [Guy07] systematically studied BMC in a general framework. In continuous time, BMC encode certain piecewise deterministic Markov processes on trees that serve as the stochastic realisation of growth-fragmentation models (see *e.g.* Doumic *et al.* [DHKR15], Robert *et al.* [RHK+14] for modelling cell division in *Escherichia coli* and the references therein).

For  $m \geq 0$ , let  $\mathbb{G}_m = \{0, 1\}^m$  (with  $\mathbb{G}_0 = \{\emptyset\}$ ) and introduce the infinite genealogical tree

$$\mathbb{T} = \bigcup_{m=0}^{\infty} \mathbb{G}_m.$$

For  $u \in \mathbb{G}_m$ , set  $|u| = m$  and define the concatenation  $u0 = (u, 0) \in \mathbb{G}_{m+1}$  and  $u1 = (u, 1) \in \mathbb{G}_{m+1}$ . A bifurcating Markov chain is specified by **1**) a measurable state space  $(\mathcal{S}, \mathfrak{S})$  with a Markov kernel (later called  $\mathbb{T}$ -transition)  $\mathcal{P}$  from  $(\mathcal{S}, \mathfrak{S})$  to  $(\mathcal{S} \times \mathcal{S}, \mathfrak{S} \otimes \mathfrak{S})$  and **2**) a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_m)_{m \geq 0}, \mathbb{P})$ . Following Guyon, we have the

**Definition 1.** A bifurcating Markov chain is a family  $(X_u)_{u \in \mathbb{T}}$  of random variables with value in  $(\mathcal{S}, \mathfrak{S})$  such that  $X_u$  is  $\mathcal{F}_{|u|}$ -measurable for every  $u \in \mathbb{T}$  and

$$\mathbb{E} \left[ \prod_{u \in \mathbb{G}_m} g_u(X_u, X_{u0}, X_{u1}) \middle| \mathcal{F}_m \right] = \prod_{u \in \mathbb{G}_m} \mathcal{P}g_u(X_u)$$

for every  $m \geq 0$  and any family of (bounded) measurable functions  $(g_u)_{u \in \mathbb{G}_m}$ , where  $\mathcal{P}g(x) = \int_{\mathcal{S} \times \mathcal{S}} g(x, y, z) \mathcal{P}(x, dy dz)$  denotes the action of  $\mathcal{P}$  on  $g$ .

The distribution of  $(X_u)_{u \in \mathbb{T}}$  is thus entirely determined by  $\mathcal{P}$  and an initial distribution for  $X_\emptyset$ . Informally, we may view  $(X_u)_{u \in \mathbb{T}}$  as a population of individuals, cells or particles indexed by  $\mathbb{T}$  and governed by the following dynamics: to each  $u \in \mathbb{T}$  we associate a trait  $X_u$  (its size, lifetime, growth rate, DNA content and so on) with value in  $\mathcal{S}$ . At its time of death, the particle  $u$  gives rise to two children  $u0$  and  $u1$ . Conditional on  $X_u = x$ , the trait  $(X_{u0}, X_{u1}) \in \mathcal{S} \times \mathcal{S}$  of the offspring of  $u$  is distributed according to  $\mathcal{P}(x, dy dz)$ .

For  $n \geq 0$ , let  $\mathbb{T}_n = \bigcup_{m=0}^n \mathbb{G}_m$  denote the genealogical tree up to the  $n$ -th generation. Assume we observe  $\mathbb{X}^n = (X_u)_{u \in \mathbb{T}_n}$ , *i.e.* we have  $2^{n+1} - 1$  random variables with value in  $\mathcal{S}$ . There are several objects of interest that we may try to infer from the data  $\mathbb{X}^n$ . Similarly to fragmentation processes (see *e.g.* Bertoin [Ber06]) a key role for both asymptotic and non-asymptotic analysis of bifurcating Markov chains is played by the so-called *tagged-branch chain*, as shown by Guyon [Guy07] and Bitseki Penda *et al.* [BDG14]. The tagged-branch chain  $(Y_m)_{m \geq 0}$  corresponds to a

lineage picked at random in the population  $(X_u)_{u \in \mathbb{T}}$ : it is a Markov chain with value in  $\mathcal{S}$  defined by  $Y_0 = X_\emptyset$  and for  $m \geq 1$ ,

$$Y_m = X_{\emptyset \epsilon_1 \dots \epsilon_m},$$

where  $(\epsilon_m)_{m \geq 1}$  is a sequence of independent Bernoulli variables with parameter  $1/2$ , independent of  $(X_u)_{u \in \mathbb{T}}$ . It has transition

$$\Omega = (\mathcal{P}_0 + \mathcal{P}_1) / 2,$$

obtained from the marginal transitions

$$\mathcal{P}_0(x, dy) = \int_{z \in \mathcal{S}} \mathcal{P}(x, dy dz) \quad \text{and} \quad \mathcal{P}_1(x, dz) = \int_{y \in \mathcal{S}} \mathcal{P}(x, dy dz)$$

of  $\mathcal{P}$ . Guyon proves in [Guy07] that if  $(Y_m)_{m \geq 0}$  is ergodic with invariant measure  $\nu$ , then the convergence

$$\frac{1}{|\mathbb{G}_n|} \sum_{u \in \mathbb{G}_n} g(X_u) \rightarrow \int_{\mathcal{S}} g(x) \nu(dx) \quad (1)$$

holds almost-surely as  $n \rightarrow \infty$  for appropriate test functions  $g$ . Moreover, we also have convergence results of the type

$$\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} g(X_u, X_{u0}, X_{u1}) \rightarrow \int_{\mathcal{S}} \mathcal{P}g(x) \nu(dx) \quad (2)$$

almost-surely as  $n \rightarrow \infty$ . These results are appended with central limit theorems (Theorem 19 of [Guy07]) and Hoeffding-type deviations inequalities in a non-asymptotic setting (Theorem 2.11 and 2.12 of Bitseki Penda *et al.* [BDG14]).

## 1.2 OBJECTIVES

The observation of  $\mathbb{X}^n$  enables us to identify  $\nu(dx)$  as  $n \rightarrow \infty$  thanks to (1). Consequently, convergence (2) reveals  $\mathcal{P}$  and therefore  $\Omega$  is identified as well, at least asymptotically. The purpose of the present work is at least threefold:

- 1) Construct – under appropriate regularity conditions – estimators of  $\nu, \Omega$  and  $\mathcal{P}$  and study their rates of convergence as  $n \rightarrow \infty$  under various loss functions. When  $\mathcal{S} \subseteq \mathbb{R}$  and when  $\mathcal{P}$  is absolutely continuous with respect to the Lebesgue measure, we estimate the corresponding density functions under various smoothness class assumptions and build *smoothness adaptive* estimators, *i.e.* estimator that achieve an optimal rate of convergence without prior knowledge of the smoothness class.
- 2) Apply these constructions to investigate further specific classes of BMC. These include binary growth-fragmentation processes, where we subsequently estimate adaptively the splitting rate of a size-dependent model, thus extending previous results of Doumic *et al.* [DHKR15] and bifurcating autoregressive processes, where we complete previous studies of Bitseki Penda *et al.* [BEG15].

- 3)** For the estimation of  $\nu, \mathcal{Q}$  and  $\mathcal{P}$  and the subsequent estimation results of **2)**, prove that our results are sharp in a minimax sense.

Our smoothness adaptive estimators are based on wavelet thresholding for density estimation (Donoho *et al.* [DJKP95] in the generalised framework of Kerkyacharian and Picard [KP00]). Implementing these techniques requires concentration properties of empirical wavelet coefficients. To that end, we prove new deviation inequalities for bifurcating Markov chains that we develop independently in a more general setting, when  $\mathcal{S}$  is not necessarily restricted to  $\mathbb{R}$ .

We stress that when  $\mathcal{P}_0 = \mathcal{P}_1$ , we have  $\mathcal{Q} = \mathcal{P}_0 = \mathcal{P}_1$  as well and we retrieve the usual framework of nonparametric estimation of Markov chains when the observation is based on  $(Y_i)_{1 \leq i \leq n}$  solely. We are therefore in the line of combining and generalising the study of Cléménçon [Clé00] and Lacour [Lac07, Lac08] that both consider adaptive estimation for Markov chains when  $\mathcal{S} \subseteq \mathbb{R}$ .

### 1.3 MAIN RESULTS AND ORGANISATION OF THE CHAPTER

In Section 2, we generalise the Hoeffding-type deviations inequalities of Bitseki Penda *et al.* [BDG14] for BMC to Bernstein-type inequalities: when  $\mathcal{P}$  is uniformly geometrically ergodic (Assumption 3 below), we prove in Theorem 5 deviations of the form

$$\mathbb{P}\left(\frac{1}{|\mathbb{G}_n|} \sum_{u \in \mathbb{G}_n} g(X_u, X_{u0}, X_{u1}) - \int \mathcal{P}g d\nu \geq \delta\right) \leq \exp\left(-\frac{\kappa|\mathbb{G}_n|\delta^2}{\Sigma_n(g) + |g|_\infty \delta}\right)$$

and

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} g(X_u, X_{u0}, X_{u1}) - \int \mathcal{P}g d\nu \geq \delta\right) \leq \exp\left(-\frac{\tilde{\kappa}n^{-1}|\mathbb{T}_n|\delta^2}{\Sigma_n(g) + |g|_\infty \delta}\right),$$

where  $\kappa, \tilde{\kappa} > 0$  only depend on  $\mathcal{P}$  and  $\Sigma_n(g)$  is a variance term which depends on a combination of the  $L^p$ -norms of  $g$  for  $p = 1, 2, \infty$  with respect to a common dominating measure for the family  $\{\mathcal{Q}(x, dy), x \in \mathcal{S}\}$ . The precise results are stated in Theorems 4 and 5.

Section 3 is devoted to the statistical estimation of  $\nu, \mathcal{Q}$  and  $\mathcal{P}$  when  $\mathcal{S} \subseteq \mathbb{R}$  and the family  $\{\mathcal{P}(x, dy dz), x \in \mathcal{S}\}$  is dominated by the Lebesgue measure on  $\mathbb{R}^2$ . In that setting, abusing notation slightly, we have  $\nu(dx) = \nu(x)dx$ ,  $\mathcal{Q}(x, dy) = \mathcal{Q}(x, y)dy$  and  $\mathcal{P}(x, dy dz) = \mathcal{P}(x, y, z)dydz$  for some functions  $x \rightsquigarrow \nu(x)$ ,  $(x, y) \rightsquigarrow \mathcal{Q}(x, y)$  and  $(x, y, z) \rightsquigarrow \mathcal{P}(x, y, z)$  that we reconstruct nonparametrically. Our estimators are constructed in several steps:

- i)** We approximate the functions  $\nu(x)$ ,  $f_{\mathcal{Q}}(x, y) = \nu(x)\mathcal{Q}(x, y)$  and  $f_{\mathcal{P}}(x, y, z) = \nu(x)\mathcal{P}(x, y, z)$  by atomic representations

$$\begin{aligned} \nu(x) &\approx \sum_{\lambda \in \mathcal{V}^1(\nu)} \langle \nu, \psi_\lambda^1 \rangle \psi_\lambda^1(x), \\ f_{\mathcal{Q}}(x, y) &\approx \sum_{\lambda \in \mathcal{V}^2(f_{\mathcal{Q}})} \langle f_{\mathcal{Q}}, \psi_\lambda^2 \rangle \psi_\lambda^2(x, y), \\ f_{\mathcal{P}}(x, y, z) &\approx \sum_{\lambda \in \mathcal{V}^3(f_{\mathcal{P}})} \langle f_{\mathcal{P}}, \psi_\lambda^3 \rangle \psi_\lambda^3(x, y, z), \end{aligned}$$

where  $\langle \cdot, \cdot \rangle$  denotes the usual  $L^2$ -inner product (over  $\mathbb{R}^d$ , for  $d = 1, 2, 3$  respectively) and  $(\psi_\lambda^d, \lambda \in \mathcal{V}^d(\cdot))$  is a collection of functions (wavelets) in  $L^2(\mathbb{R}^d)$  that are localised in time and frequency, indexed by a set  $\mathcal{V}^d(\cdot)$  that depends on the signal itself (the precise meaning of the symbol  $\approx$  and the properties of the  $\psi_\lambda$ 's are stated precisely in Section 3.1).

ii) We estimate

$$\begin{aligned} \langle \nu, \psi_\lambda^1 \rangle & \text{ by } |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} \psi_\lambda^1(X_u), \\ \langle f_\Omega, \psi_\lambda^2 \rangle & \text{ by } |\mathbb{T}_n^*|^{-1} \sum_{u \in \mathbb{T}_n^*} \psi_\lambda^2(X_{u^-}, X_u), \\ \langle f_{\mathcal{P}}, \psi_\lambda^3 \rangle & \text{ by } |\mathbb{T}_{n-1}|^{-1} \sum_{u \in \mathbb{T}_{n-1}} \psi_\lambda^3(X_u, X_{u0}, X_{u1}), \end{aligned}$$

where  $X_{u^-}$  denotes the trait of the parent of  $u$  and  $\mathbb{T}_n^* = \mathbb{T}_n \setminus \mathbb{G}_0$ , and specify a selection rule for  $\mathcal{V}^d(\cdot)$  (with the dependence in the unknown function somehow replaced by an estimator). The rule is dictated by hard thresholding over the estimation of the coefficients that are kept only if they exceed some noise level, tuned with  $|\mathbb{T}_n|$  and prior knowledge on the unknown function, as follows by standard density estimation by wavelet thresholding (Donoho *et al.* [DJKP96], Kerkycharian and Picard [KP00]).

iii) Denoting by  $\widehat{\nu}_n(x)$ ,  $\widehat{f}_n(x, y)$  and  $\widehat{f}_n(x, y, z)$  the estimators of  $\nu(x)$ ,  $f_\Omega(x, y)$  and  $f_{\mathcal{P}}(x, y, z)$  respectively constructed in Step ii), we finally take as estimators for  $\mathcal{Q}(x, y)$  and  $\mathcal{P}(x, y, z)$  the quotient estimators

$$\widehat{\mathcal{Q}}_n(x, y) = \frac{\widehat{f}_n(x, y)}{\widehat{\nu}_n(x)} \quad \text{and} \quad \widehat{\mathcal{P}}_n(x, y, z) = \frac{\widehat{f}_n(x, y, z)}{\widehat{\nu}_n(x)}$$

provided  $\widehat{\nu}_n(x)$  exceeds a minimal threshold.

Beyond the inherent technical difficulties of the approximation Steps i) and iii), the crucial novel part is the estimation Step ii), where Theorems 4 and 5 are used to estimate precisely the probability that the thresholding rule applied to the empirical wavelet coefficient is close in effect to thresholding the true coefficients.

When  $\nu$ ,  $\mathcal{Q}$  or  $\mathcal{P}$  (identified with their densities with respect to appropriate dominating measures) belong to an isotropic Besov ball of smoothness  $s$  measured in  $L^\pi$  over a domain  $\mathcal{D}^d$  in  $\mathbb{R}^d$ , with  $s > d/\pi$  and  $d = 1, 2, 3$  respectively, we prove in Theorems 8, 9 and 10 that if  $\mathcal{Q}$  is uniformly geometrically ergodic, then our estimators achieve the rate  $|\mathbb{T}_n|^{-\alpha_d(s, p, \pi)}$  in  $L^p(\mathcal{D})$ -loss, up to additional  $\log |\mathbb{T}_n|$  terms, where

$$\alpha_d(s, p, \pi) = \min \left\{ \frac{s}{2s + d}, \frac{s + d(1/p - 1/\pi)}{2s + d(1 - 2/\pi)} \right\}$$

is the usual exponent for the minimax rate of estimation of a  $d$ -variate function with order of smoothness  $s$  measured in  $L^\pi$  in  $L^p$ -loss error.

This rate is nearly optimal in a minimax sense for  $d = 1$ , as follows from particular case  $\mathcal{Q}(x, dy) = \nu(dy)$  that boils down to density estimation with  $|\mathbb{T}_n|$  data: the optimality is then a direct consequence of Theorem 2 in Donoho *et al.* [DJKP96]. As for the case  $d = 2$  and  $d = 3$ , the structure of BMC comes into play and we need to prove a specific optimality result, stated in Theorems 9 and 10. We rely on classical lower bound techniques for density estimation and Markov chains (Hoffmann [Hof99], Cléménçon [Clé00], Lacour [Lac07, Lac08]).

We apply our generic results in Section 4 to two illustrative examples. We consider in Section 4.1 the growth-fragmentation model as studied in Doumic *et al.* [DHKR15], where we estimate the size-dependent splitting rate of the model as a function of the invariant measure of an associated BMC in Theorem 12. This enables us to extend the recent results of Doumic *et al.* in several directions: adaptive estimation, extension of the smoothness classes and the loss functions considered, and also a proof of a minimax lower bound. In Section 4.2, we show how bifurcating autoregressive models (BAR) as developed for instance in Bercu *et al.* [BdSGP09] and in Chapter III which follows are embedded into our generic framework of estimation. A numerical illustration highlights the feasibility of our procedure in practice and is presented in Section 4.3. The proofs are postponed to Section 5.

## 2 DEVIATIONS INEQUALITIES FOR EMPIRICAL MEANS

In the sequel, we fix a (measurable) subset  $\mathcal{D} \subseteq \mathcal{S}$  that will be later needed for statistical purposes. We need some regularity on the  $\mathbb{T}$ -transition  $\mathcal{P}$  via its mean transition  $\mathcal{Q} = \frac{1}{2}(\mathcal{P}_0 + \mathcal{P}_1)$ .

**Assumption 2.** *The family  $\{\mathcal{Q}(x, dy), x \in \mathcal{S}\}$  is dominated by a common sigma-finite measure  $\mathfrak{n}(dy)$ . We have (abusing notation slightly)*

$$\mathcal{Q}(x, dy) = \mathcal{Q}(x, y)\mathfrak{n}(dy) \text{ for every } x \in \mathcal{S},$$

for some  $\mathcal{Q} : \mathcal{S}^2 \rightarrow [0, \infty)$  such that

$$|\mathcal{Q}|_{\mathcal{D}} = \sup_{x \in \mathcal{S}, y \in \mathcal{D}} \mathcal{Q}(x, y) < \infty.$$

An invariant probability measure for  $\mathcal{Q}$  is a probability  $\nu$  on  $(\mathcal{S}, \mathfrak{S})$  such that  $\nu\mathcal{Q} = \nu$  where  $\nu\mathcal{Q}(dy) = \int_{x \in \mathcal{S}} \nu(dx)\mathcal{Q}(x, dy)$ . We set

$$\mathcal{Q}^r(x, dy) = \int_{z \in \mathcal{S}} \mathcal{Q}(x, dz)\mathcal{Q}^{r-1}(z, dy) \text{ with } \mathcal{Q}^0(x, dy) = \delta_x(dy)$$

for the  $r$ -th iteration of  $\mathcal{Q}$ . For a function  $g : \mathcal{S}^d \rightarrow \mathbb{R}$  with  $d = 1, 2, 3$  and  $1 \leq p \leq \infty$ , we denote by  $|g|_p$  its  $L^p$ -norm with respect to the measure  $\mathfrak{n}^{\otimes d}$ , allowing for the value  $|g|_p = \infty$  if  $g \notin L^p(\mathfrak{n}^{\otimes d})$ . The same notation applies to a function  $g : \mathcal{D}^d \rightarrow \mathbb{R}$  tacitly considered as a function from  $\mathcal{S}^d \rightarrow \mathbb{R}$  by setting  $g(x) = 0$  for  $x \in \mathcal{S} \setminus \mathcal{D}$ .



**Assumption 3.** *The mean transition  $\mathcal{Q}$  admits a unique invariant probability measure  $\nu$  and there exist  $R > 0$  and  $0 < \rho < 1/2$  such that*

$$|\mathcal{Q}^m g(x) - \int_{\mathcal{S}} g d\nu| \leq R|g|_{\infty} \rho^m, \quad x \in \mathcal{S}, \quad m \geq 0,$$

for every  $g$  integrable with respect to  $\nu$ .

Assumption 3 is a uniform geometric ergodicity condition that can be verified in most applications using the theory of Meyn and Tweedie [MT93]. The ergodicity rate should be small enough ( $\rho < 1/2$ ) and this point is crucial for the proofs. However this is sometimes delicate to check in applications and we refer to Hairer and Mattingly [HM11] for an explicit control of the ergodicity rate.

Our first result is a deviation inequality for empirical means over  $\mathbb{G}_n$  or  $\mathbb{T}_n$ . We need some notation. Let

$$\begin{aligned} \kappa_1 &= \kappa_1(\mathcal{Q}, \mathcal{D}) = 32 \max \{ |\mathcal{Q}|_{\mathcal{D}}, 4|\mathcal{Q}|_{\mathcal{D}}^2, 4R^2(1+\rho)^2 \}, \\ \kappa_2 &= \kappa_2(\mathcal{Q}) = \frac{16}{3} \max \{ 1 + R\rho, R(1+\rho) \}, \\ \kappa_3 &= \kappa_3(\mathcal{Q}, \mathcal{D}) = 96 \max \{ |\mathcal{Q}|_{\mathcal{D}}, 16|\mathcal{Q}|_{\mathcal{D}}^2, 4R^2(1+\rho)^2(1-2\rho)^{-2} \}, \\ \kappa_4 &= \kappa_4(\mathcal{Q}) = \frac{16}{3} \max \{ 1 + R\rho, R(1+\rho)(1-2\rho)^{-1} \}, \end{aligned}$$

where  $|\mathcal{Q}|_{\mathcal{D}} = \sup_{x \in \mathcal{S}, y \in \mathcal{D}} \mathcal{Q}(x, y)$  is defined in Assumption 2. For  $g : \mathcal{S}^d \rightarrow \mathbb{R}$ , define  $\Sigma_{1,1}(g) = |g|_2^2$  and for  $n \geq 2$ ,

$$\Sigma_{1,n}(g) = |g|_2^2 + \min_{1 \leq \ell \leq n-1} (|g|_1^2 2^\ell + |g|_{\infty}^2 2^{-\ell}). \quad (3)$$

Define also  $\Sigma_{2,1}(g) = |\mathcal{P}g^2|_1$  and for  $n \geq 2$ ,

$$\Sigma_{2,n}(g) = |\mathcal{P}g^2|_1 + \min_{1 \leq \ell \leq n-1} (|\mathcal{P}g|_1^2 2^\ell + |\mathcal{P}g|_{\infty}^2 2^{-\ell}). \quad (4)$$

**Theorem 4.** *Work under Assumptions 2 and 3. Then, for every  $n \geq 1$  and every  $g : \mathcal{D} \subseteq \mathcal{S} \rightarrow \mathbb{R}$  integrable with respect to  $\nu$ , the following inequalities hold true:*

(i) *For any  $\delta > 0$  such that  $\delta \geq 4R|g|_{\infty}|\mathbb{G}_n|^{-1}$ , we have*

$$\mathbb{P}\left(\frac{1}{|\mathbb{G}_n|} \sum_{u \in \mathbb{G}_n} g(X_u) - \int_{\mathcal{S}} g d\nu \geq \delta\right) \leq \exp\left(\frac{-|\mathbb{G}_n|\delta^2}{\kappa_1 \Sigma_{1,n}(g) + \kappa_2 |g|_{\infty} \delta}\right).$$

(ii) *For any  $\delta > 0$  such that  $\delta \geq 4R(1-2\rho)^{-1}|g|_{\infty}|\mathbb{T}_n|^{-1}$ , we have*

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} g(X_u) - \int_{\mathcal{S}} g d\nu \geq \delta\right) \leq \exp\left(\frac{-|\mathbb{T}_n|\delta^2}{\kappa_3 \Sigma_{1,n}(g) + \kappa_4 |g|_{\infty} \delta}\right).$$

**Theorem 5.** *Work under Assumptions 2 and 3. Then, for every  $n \geq 2$  and for every  $g : \mathcal{D}^3 \subseteq \mathcal{S}^3 \rightarrow \mathbb{R}$  such that  $\mathcal{P}g$  is well defined and integrable with respect to  $\nu$ , the following inequalities hold true:*

(i) For any  $\delta > 0$  such that  $\delta \geq 4R|\mathcal{P}g|_\infty|\mathbb{G}_n|^{-1}$ , we have

$$\mathbb{P}\left(\frac{1}{|\mathbb{G}_n|} \sum_{u \in \mathbb{G}_n} g(X_u, X_{u0}, X_{u1}) - \int_{\mathcal{S}} \mathcal{P}g \, d\nu \geq \delta\right) \leq \exp\left(\frac{-|\mathbb{G}_n|\delta^2}{\kappa_1 \Sigma_{2,n}(g) + \kappa_2 |g|_\infty \delta}\right).$$

(ii) For any  $\delta > 0$  such that  $\delta \geq 4(nR|\mathcal{P}g|_\infty + |g|_\infty)|\mathbb{T}_{n-1}|^{-1}$ , we have

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_{n-1}|} \sum_{u \in \mathbb{T}_{n-1}} g(X_u, X_{u0}, X_{u1}) - \int_{\mathcal{S}} \mathcal{P}g \, d\nu \geq \delta\right) \leq \exp\left(\frac{-n^{-1}|\mathbb{T}_{n-1}|\delta^2}{\kappa_1 \Sigma_{2,n-1}(g) + \kappa_2 |g|_\infty \delta}\right).$$

A few remarks are in order:

1) Theorem 4 (i) is a direct consequence of Theorem 5 (i) but Theorem 4 (ii) is not a corollary of Theorem 5 (ii): we note that a slow term of order  $n^{-1} \approx (\log |\mathbb{T}_n|)^{-1}$  comes in Theorem 5 (ii).

2) Bitseki-Penda *et al.* in [BDG14] study similar Hoeffding-type deviations inequalities for functionals of bifurcating Markov chains under ergodicity assumption and for uniformly bounded functionals. In the present work and for statistical purposes, we need Bernstein-type deviations inequalities which require a specific treatment than cannot be obtained from a direct adaptation of [BDG14]. In particular, we apply our results to multivariate wavelets test functions  $\psi_\lambda^d$  that are well localised but unbounded, and a fine control of the conditional variance  $\Sigma_{i,n}(\psi_\lambda^d)$ ,  $i = 1, 2$  is of crucial importance.

3) For  $g$  such that  $|g|_1^{-1}|g|_\infty \leq 2^{n-1}$ , the variance term  $\Sigma_{1,n}(g)$  is controlled by  $|g|_1|g|_\infty$  up to a constant. Similarly, for  $g$  such that  $|g|_1^{-1}|g|_{\infty,1} \leq 2^{n-1}$ , assuming in addition that  $|\mathcal{P}|_{\mathcal{D}} = \sup_{\mathcal{D}^3} |\mathcal{P}(x, y, z)| < \infty$ , the variance term  $\Sigma_{2,n}(g)$  is controlled by  $\max\{|\mathcal{P}|_{\mathcal{D}}, |\mathcal{P}|_{\mathcal{D}}^2\}|g|_1|g|_{\infty,1}$  up to a constant, setting  $|g|_{\infty,1} = \sup_{x \in \mathcal{S}} \int_{\mathcal{S}^2} |g(x, y, z)| \, dydz$  for any  $g : \mathcal{S}^3 \rightarrow \mathbb{R}$ .

4) Assumption 3 about the uniform geometric ergodicity is quite strong, although satisfied in the two examples developed in Section 4 (at the cost however of assuming that the splitting rate of the growth-fragmentation model has bounded support in Section 4.1). Presumably, a way to relax this restriction would be to require a weaker geometric ergodicity condition of the form

$$\left| \mathcal{Q}^m g(x) - \int_{\mathcal{S}} g \, d\nu \right| \leq R|g|_\infty V(x) \rho^m, \quad x \in \mathcal{S}, \quad m \geq 0,$$

for some Lyapunov function  $V : \mathcal{S} \rightarrow [1, \infty)$ . Analogous results could then be obtained via transportation information inequalities for bifurcating Markov chains with a similar approach as in Gao *et al.* [GGW14], but this lies beyond the scope of this work.

### 3 STATISTICAL ESTIMATION

In this section, we take  $(\mathcal{S}, \mathfrak{G}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . As in the previous section, we fix a compact interval  $\mathcal{D} \subseteq \mathcal{S}$ . The following assumption will be needed here

**Assumption 6.** *The family  $\{\mathcal{P}(x, dy dz), x \in \mathcal{S}\}$  is dominated with respect to the Lebesgue measure on  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ . We have (abusing notation slightly)*

$$\mathcal{P}(x, dy dz) = \mathcal{P}(x, y, z) dy dz \text{ for every } x \in \mathcal{S}$$

for some  $\mathcal{P} : \mathcal{S}^3 \rightarrow [0, \infty)$  such that

$$|\mathcal{P}|_{\mathcal{D}} = \sup_{(x,y,z) \in \mathcal{D}^3} |\mathcal{P}(x, y, z)| < \infty.$$

Under Assumptions 2, 3 and 6 with  $\mathfrak{n}(dy) = dy$ , we have (abusing notation slightly)

$$\mathcal{P}(x, dy dz) = \mathcal{P}(x, y, z) dy dz, \quad \mathcal{Q}(x, dy) = \mathcal{Q}(x, y) dy \text{ and } \nu(dx) = \nu(x) dx.$$

For some  $n \geq 1$ , we observe  $\mathbb{X}_n = (X_u)_{u \in \mathbb{T}_n}$  and we aim at constructing nonparametric estimators of  $x \rightsquigarrow \nu(x)$ ,  $(x, y) \rightsquigarrow \mathcal{Q}(x, y)$  and  $(x, y, z) \rightsquigarrow \mathcal{P}(x, y, z)$  for  $x, y, z \in \mathcal{D}$ . To that end, we use regular wavelet bases adapted to the domain  $\mathcal{D}^d$  for  $d = 1, 2, 3$ .

### 3.1 ATOMIC DECOMPOSITIONS AND WAVELETS

Wavelet bases  $(\psi_\lambda^d)_\lambda$  adapted to a domain  $\mathcal{D}^d$  in  $\mathbb{R}^d$ , for  $d = 1, 2, 3$  are documented in numerous textbooks, see *e.g.* Cohen [Coh00]. The multi-index  $\lambda$  concatenates the spatial index and the resolution level  $j = |\lambda|$ . We set  $\Lambda_j = \{\lambda, |\lambda| = j\}$  and  $\Lambda = \cup_{j \geq -1} \Lambda_j$ . Thus, for  $g \in L^\pi(\mathcal{D}^d)$  for some  $\pi \in (0, \infty]$ , we have

$$g = \sum_{j \geq -1} \sum_{\lambda \in \Lambda_j} g_\lambda \psi_\lambda^d = \sum_{\lambda \in \Lambda} g_\lambda \psi_\lambda^d, \quad \text{with } g_\lambda = \langle g, \psi_\lambda^d \rangle,$$

where we have set  $j = -1$  in order to incorporate the low frequency part of the decomposition and  $\langle g, \psi_\lambda^d \rangle = \int g \psi_\lambda^d$  denotes the inner product in  $L^2(\mathbb{R}^d)$ . From now on, the basis  $(\psi_\lambda^d)_\lambda$  is fixed. For  $s > 0$  and  $\pi \in (0, \infty]$ ,  $g$  belongs to  $B_{\pi, \infty}^s(\mathcal{D})$  if the following norm is finite:

$$\|g\|_{B_{\pi, \infty}^s(\mathcal{D})} = \sup_{j \geq -1} 2^{j(s+d(1/2-1/\pi))} \left( \sum_{\lambda \in \Lambda_j} |\langle g, \psi_\lambda^d \rangle|^\pi \right)^{1/\pi} \quad (5)$$

with the usual modification if  $\pi = \infty$ . Precise connection between this definition of Besov norm and more standard ones can be found in [Coh00]. Given a basis  $(\psi_\lambda^d)_\lambda$ , there exists  $\sigma > 0$  such that for  $\pi \geq 1$  and  $s \leq \sigma$  the Besov space defined by (5) exactly matches the usual definition in terms of moduli of smoothness for  $g$ . The index  $\sigma$  can be taken arbitrarily large. The additional properties of the wavelet basis  $(\psi_\lambda^d)_\lambda$  that we need are summarized in the next assumption.

**Assumption 7.** *For  $p \geq 1$ ,*

$$\|\psi_\lambda^d\|_{L^p}^p \sim 2^{|\lambda|d(p/2-1)}, \quad (6)$$

for some  $\sigma > 0$  and for all  $s \leq \sigma$ ,  $j_0 \geq 0$ ,

$$\|g - \sum_{j \leq j_0} \sum_{\lambda \in \Lambda_j} g_\lambda \psi_\lambda^d\|_{L^p} \lesssim 2^{-j_0 s} \|g\|_{B_{p, \infty}^s(\mathcal{D})}, \quad (7)$$

for any subset  $\Lambda_0 \subset \Lambda$ ,

$$\int_{\mathcal{D}} \left( \sum_{\lambda \in \Lambda_0} |\psi_\lambda^d(x)|^2 \right)^{p/2} dx \sim \sum_{\lambda \in \Lambda_0} \|\psi_\lambda^d\|_{L^p}^p. \quad (8)$$

If  $p > 1$ , for any sequence  $(u_\lambda)_{\lambda \in \Lambda}$ ,

$$\left\| \left( \sum_{\lambda \in \Lambda} |u_\lambda \psi_\lambda^d|^2 \right)^{1/2} \right\|_{L^p} \sim \left\| \sum_{\lambda \in \Lambda} u_\lambda \psi_\lambda^d \right\|_{L^p}. \quad (9)$$

The symbol  $\sim$  means inequality in both ways, up to a constant depending on  $p$  and  $\mathcal{D}$  only. The property (7) reflects that our definition (5) of Besov spaces matches the definition in term of linear approximation. Property (9) reflects an unconditional basis property, see Kerkyacharian and Picard [KP00], De Vore *et al.* [DKT98] and (8) is referred to as a superconcentration inequality, or Temlyakov property [KP00]. The formulation of (8)-(9) in the context of statistical estimation is posterior to the original papers of Donoho and Johnstone [DJ94, DJ95] and Donoho *et al.* [DJKP96, DJKP95] and is due to Kerkyacharian and Picard [KP00]. The existence of compactly supported wavelet bases satisfying Assumption 7 is discussed in Meyer [Mey90], see also Cohen [Coh00].

### 3.2 ESTIMATION OF THE INVARIANT DENSITY $\nu$

Recall that we estimate  $x \rightsquigarrow \nu(x)$  for  $x \in \mathcal{D}$ , taken as a compact interval in  $\mathcal{S} \subseteq \mathbb{R}$ . We approximate the representation

$$\nu(x) = \sum_{\lambda \in \Lambda} \nu_\lambda \psi_\lambda^1(x), \quad \nu_\lambda = \langle \nu, \psi_\lambda^1 \rangle$$

by

$$\hat{\nu}_n(x) = \sum_{|\lambda| \leq J} \hat{\nu}_{\lambda,n} \psi_\lambda^1(x),$$

with

$$\hat{\nu}_{\lambda,n} = \mathcal{T}_{\lambda,\eta} \left( \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} \psi_\lambda^1(X_u) \right),$$

and  $\mathcal{T}_{\lambda,\eta}(x) = x \mathbf{1}_{|x| \geq \eta}$  denotes the standard threshold operator (with  $\mathcal{T}_{\lambda,\eta}(x) = x$  for the low frequency part when  $\lambda \in \Lambda_{-1}$ ). Thus  $\hat{\nu}_n$  is specified by the maximal resolution level  $J$  and the threshold  $\eta$ .

**Theorem 8.** *Work under Assumptions 2 and 3 with  $\mathfrak{n}(dx) = dx$ . Specify  $\hat{\nu}_n$  with*

$$J = \log_2 \frac{|\mathbb{T}_n|}{\log |\mathbb{T}_n|} \quad \text{and} \quad \eta = c \sqrt{\log |\mathbb{T}_n| / |\mathbb{T}_n|}$$

for some  $c > 0$ . For every  $\pi \in (0, \infty]$ ,  $s \in (1/\pi, \sigma]$  and  $p \geq 1$ , for large enough  $n$  and  $c$ , the following estimate holds

$$\left( \mathbb{E} \left[ \|\hat{\nu}_n - \nu\|_{L^p(\mathcal{D})}^p \right] \right)^{1/p} \lesssim \left( \frac{\log |\mathbb{T}_n|}{|\mathbb{T}_n|} \right)^{\alpha_1(s,p,\pi)},$$

with  $\alpha_1(s,p,\pi) = \min \left\{ \frac{s}{2s+1}, \frac{s+1/p-1/\pi}{2s+1-2/\pi} \right\}$ , up to a constant that depends on  $s, p, \pi, \|\nu\|_{\mathcal{B}_{\pi,\infty}^s(\mathcal{D})}, \rho$ ,

$R$  and  $|\mathcal{Q}|_{\mathcal{D}}$  and that is continuous in its arguments.

Two remarks are in order:

1) The upper-rate of convergence is the classical minimax rate in density estimation. We infer that our estimator is nearly optimal in a minimax sense as follows from Theorem 2 in Donoho *et al.* [DJKP96] applied to the class  $\mathcal{Q}(x, y)dy = \nu(y)dy$ , *i.e.* in the particular case when we have i.i.d.  $X_u$ 's. We highlight the fact that  $n$  represents here the number of observed generations in the tree, which means that we observe  $|\mathbb{T}_n| = 2^{n+1} - 1$  traits.

2) The estimator  $\widehat{\nu}_n$  is *smooth-adaptive* in the following sense: for every  $s_0 > 0$ ,  $0 < \rho_0 < 1/2$ ,  $R_0 > 0$  and  $\mathcal{Q}_0 > 0$ , define the sets  $\mathcal{A}(s_0) = \{(s, \pi), s \geq s_0, s_0 \geq 1/\pi\}$  and

$$\mathcal{Q}(\rho_0, R_0, \mathcal{Q}_0) = \{\mathcal{Q} \text{ such that } \rho \leq \rho_0, R \leq R_0, |\mathcal{Q}|_{\mathcal{D}} \leq \mathcal{Q}_0\},$$

where  $\mathcal{Q}$  is taken among mean transitions for which Assumption 3 holds. Then, for every  $C > 0$ , there exists  $c^* = c^*(\mathcal{D}, p, s_0, \rho_0, R_0, \mathcal{Q}_0, C)$  such that  $\widehat{\nu}_n$  specified with  $c^*$  satisfies

$$\sup_n \sup_{(s, \pi) \in \mathcal{A}(s_0)} \sup_{\nu, \mathcal{Q}} \left( \frac{|\mathbb{T}_n|}{\log |\mathbb{T}_n|} \right)^{p\alpha_1(s, p, \pi)} \mathbb{E}[\|\widehat{\nu}_n - \nu\|_{L^p(\mathcal{D})}^p] < \infty$$

where the supremum is taken among  $(\nu, \mathcal{Q})$  such that  $\nu\mathcal{Q} = \nu$  with  $\mathcal{Q} \in \mathcal{Q}(\rho_0, R_0, \mathcal{Q}_0)$  and  $\|\nu\|_{\mathcal{B}_{\pi, \infty}^s(\mathcal{D})} \leq C$ . In particular,  $\widehat{\nu}_n$  achieves the (near) optimal rate of convergence over Besov balls simultaneously for all  $(s, \pi) \in \mathcal{A}(s_0)$ . Analogous smoothness adaptive results hold for Theorems 9, 10 and 12 below.

### 3.3 ESTIMATION OF THE DENSITY OF THE MEAN TRANSITION $\mathcal{Q}$

In this section we estimate  $(x, y) \rightsquigarrow \mathcal{Q}(x, y)$  for  $(x, y) \in \mathcal{D}^2$  and  $\mathcal{D}$  is a compact interval in  $\mathcal{S} \subseteq \mathbb{R}$ . In a first step, we estimate the density

$$f_{\mathcal{Q}}(x, y) = \nu(x)\mathcal{Q}(x, y)$$

of the distribution of  $(X_{u^-}, X_u)$  (when  $\mathcal{L}(X_{\emptyset}) = \nu$ ) by

$$\widehat{f}_n(x, y) = \sum_{|\lambda| \leq J} \widehat{f}_{\lambda, n} \psi_{\lambda}^2(x, y),$$

with

$$\widehat{f}_{\lambda, n} = \mathcal{T}_{\lambda, \eta} \left( \frac{1}{|\mathbb{T}_n^*|} \sum_{u \in \mathbb{T}_n^*} \psi_{\lambda}^2(X_{u^-}, X_u) \right),$$

and  $\mathcal{T}_{\lambda, \eta}(\cdot)$  is the hard-threshold operator defined in Section 3.2 and  $\mathbb{T}_n^* = \mathbb{T}_n \setminus \mathbb{G}_0$ . We can now estimate the density  $\mathcal{Q}(x, y)$  of the mean transition probability by

$$\widehat{\mathcal{Q}}_n(x, y) = \frac{\widehat{f}_n(x, y)}{\max\{\widehat{\nu}_n(x), \varpi\}} \quad (10)$$

for some threshold  $\varpi > 0$ . Thus the estimator  $\widehat{\mathcal{Q}}_n$  is specified by  $J$ ,  $\eta$  and  $\varpi$ . Define also

$$m(\nu) = \inf_{x \in \mathcal{D}} \nu(x). \quad (11)$$

**Theorem 9.** *Work under Assumptions 2 and 3 with  $\mathfrak{n}(dx) = dx$ . Specify  $\widehat{\mathcal{Q}}_n$  with*

$$J = \frac{1}{2} \log_2 \frac{|\mathbb{T}_n|}{\log |\mathbb{T}_n|} \quad \text{and} \quad \eta = c \sqrt{(\log |\mathbb{T}_n|)^2 / |\mathbb{T}_n|}$$

for some  $c > 0$  and  $\varpi > 0$ . For every  $\pi \in [1, \infty]$ ,  $s \in (2/\pi, \sigma]$  and  $p \geq 1$ , for large enough  $n$  and  $c$  and small enough  $\varpi$ , the following estimate holds

$$\left( \mathbb{E} [\|\widehat{\mathcal{Q}}_n - \mathcal{Q}\|_{L^p(\mathcal{D}^2)}^p] \right)^{1/p} \lesssim \left( \frac{(\log |\mathbb{T}_n|)^2}{|\mathbb{T}_n|} \right)^{\alpha_2(s,p,\pi)}, \quad (12)$$

with  $\alpha_2(s,p,\pi) = \min \left\{ \frac{s}{2s+2}, \frac{s/2+1/p-1/\pi}{s+1-2/\pi} \right\}$ , provided  $m(\nu) \geq \varpi > 0$  and up to a constant that depends on  $s, p, \pi, \|\mathcal{Q}\|_{\mathcal{B}_{\pi,\infty}^s(\mathcal{D}^2)}$ ,  $m(\nu)$  and that is continuous in its arguments.

This rate is moreover (nearly) optimal: define  $\varepsilon_2 = s\pi - (p - \pi)$ . We have

$$\inf_{\widehat{\mathcal{Q}}_n} \sup_{\mathcal{Q}} \left( \mathbb{E} [\|\widehat{\mathcal{Q}}_n - \mathcal{Q}\|_{L^p(\mathcal{D}^2)}^p] \right)^{1/p} \gtrsim \begin{cases} |\mathbb{T}_n|^{-\alpha_2(s,p,\pi)} & \text{if } \varepsilon_2 > 0 \\ \left( \frac{\log |\mathbb{T}_n|}{|\mathbb{T}_n|} \right)^{\alpha_2(s,p,\pi)} & \text{if } \varepsilon_2 \leq 0, \end{cases}$$

where the infimum is taken among all estimators of  $\mathcal{Q}$  based on  $(X_u)_{u \in \mathbb{T}_n}$  and the supremum is taken among all  $\mathcal{Q}$  such that  $\|\mathcal{Q}\|_{\mathcal{B}_{\pi,\infty}^s(\mathcal{D}^2)} \leq C$  and  $m(\nu) \geq C'$  for some  $C, C' > 0$ .

Note that the calibration of the threshold  $\varpi$  needed to define  $\widehat{\mathcal{Q}}_n$  requires an *a priori* bound on  $m(\nu)$ . The  $(\log |\mathbb{T}_n|)^2$  comes from the slow term in the deviations inequality of Theorem 5 (ii) and from the wavelet thresholding procedure.

### 3.4 ESTIMATION OF THE DENSITY OF THE $\mathbb{T}$ -TRANSITION $\mathcal{P}$

In this section we estimate  $(x, y, z) \rightsquigarrow \mathcal{P}(x, y, z)$  for  $(x, y, z) \in \mathcal{D}^3$  and  $\mathcal{D}$  is a compact interval in  $\mathbb{S} \subseteq \mathbb{R}$ . In a first step, we estimate the density

$$f_{\mathcal{P}}(x, y, z) = \nu(x) \mathcal{P}(x, y, z)$$

of the distribution of  $(X_u, X_{u0}, X_{u1})$  (when  $\mathcal{L}(X_\emptyset) = \nu$ ) by

$$\widehat{f}_n(x, y, z) = \sum_{|\lambda| \leq J} \widehat{f}_{\lambda,n} \psi_\lambda^3(x, y, z),$$

with

$$\widehat{f}_{\lambda,n} = \mathcal{J}_{\lambda,\eta} \left( \frac{1}{|\mathbb{T}_{n-1}|} \sum_{u \in \mathbb{T}_{n-1}} \psi_\lambda^3(X_u, X_{u0}, X_{u1}) \right),$$

and  $\mathcal{T}_{\lambda,\eta}(\cdot)$  is the hard-threshold operator defined in Section 3.2. In the same way as in the previous section, we can next estimate the density  $\mathcal{P}$  of the  $\mathbb{T}$ -transition by

$$\widehat{\mathcal{P}}_n(x, y, z) = \frac{\widehat{f}_n(x, y, z)}{\max\{\widehat{\nu}_n(x), \varpi\}} \quad (13)$$

for some threshold  $\varpi > 0$ . Thus the estimator  $\widehat{\mathcal{P}}_n$  is specified by  $J$ ,  $\eta$  and  $\varpi$ .

**Theorem 10.** *Work under Assumptions 2, 3 and 6. Specify  $\widehat{\mathcal{P}}_n$  with*

$$J = \frac{1}{3} \log_2 \frac{|\mathbb{T}_n|}{\log |\mathbb{T}_n|} \quad \text{and} \quad \eta = c \sqrt{(\log |\mathbb{T}_n|)^2 / |\mathbb{T}_n|}$$

for some  $c > 0$  and  $\varpi > 0$ . For every  $\pi \in [1, \infty]$ ,  $s \in (3/\pi, \sigma]$  and  $p \geq 1$ , for large enough  $n$  and  $c$  and small enough  $\varpi$ , the following estimate holds

$$\left( \mathbb{E} [\|\widehat{\mathcal{P}}_n - \mathcal{P}\|_{L^p(\mathcal{D}^3)}^p] \right)^{1/p} \lesssim \left( \frac{(\log |\mathbb{T}_n|)^2}{|\mathbb{T}_n|} \right)^{\alpha_3(s,p,\pi)}, \quad (14)$$

with  $\alpha_3(s,p,\pi) = \min \left\{ \frac{s}{2s+3}, \frac{s/3+1/p-1/\pi}{2s/3+1-2/\pi} \right\}$ , provided  $m(\nu) \geq \varpi > 0$  and up to a constant that depends on  $s, p, \pi, \|\mathcal{P}\|_{\mathcal{B}_{\pi,\infty}^s(\mathcal{D}^3)}$  and  $m(\nu)$  and that is continuous in its arguments.

This rate is moreover (nearly) optimal: define  $\varepsilon_3 = \frac{s\pi}{3} - \frac{p-\pi}{2}$ . We have

$$\inf_{\widehat{\mathcal{P}}_n} \sup_{\mathcal{P}} \left( \mathbb{E} [\|\widehat{\mathcal{P}}_n - \mathcal{P}\|_{L^p(\mathcal{D}^3)}^p] \right)^{1/p} \gtrsim \begin{cases} |\mathbb{T}_n|^{-\alpha_3(s,p,\pi)} & \text{if } \varepsilon_3 > 0 \\ \left( \frac{\log |\mathbb{T}_n|}{|\mathbb{T}_n|} \right)^{\alpha_3(s,p,\pi)} & \text{if } \varepsilon_3 \leq 0, \end{cases}$$

where the infimum is taken among all estimators of  $\mathcal{P}$  based on  $(X_u)_{u \in \mathbb{T}_n}$  and the supremum is taken among all  $\mathcal{P}$  such that  $\|\mathcal{P}\|_{\mathcal{B}_{\pi,\infty}^s(\mathcal{D}^3)} \leq C$  and  $m(\nu) \geq C'$  for some  $C, C' > 0$ .

## 4 APPLICATIONS

### 4.1 ESTIMATION OF THE SIZE-DEPENDENT SPLITTING RATE IN A GROWTH-FRAGMENTATION MODEL

Recently, Doumic *et al.* [DHKR15] have studied the problem of estimating nonparametrically the size-dependent splitting rate in growth-fragmentation models (see *e.g.* the textbook of Perthame [Per07]). Stochastically, these are piecewise deterministic Markov processes on trees that model the evolution of a population of cells or bacteria: to each node (or cell)  $u \in \mathbb{T}$ , we associate as trait  $X_u \in \mathcal{S} \subset (0, \infty)$  the size at birth of the cell  $u$ . The evolution mechanism is described as follows: each cell grows exponentially with a common rate  $\tau > 0$ . A cell of size  $x$  splits into two newborn cells of size  $x/2$  each (thus  $X_{u_0} = X_{u_1}$  here), with a size-dependent splitting rate  $B(x)$  for some  $B : \mathcal{S} \rightarrow [0, \infty)$ . Two newborn cells start a new life independently of each other. If  $\zeta_u$  denotes the lifetime of the cell  $u$ , we thus have

$$\mathbb{P}(\zeta_u \in [t, t + dt) | \zeta_u \geq t, X_u = x) = B(x \exp(\tau t)) dt \quad (15)$$

and

$$X_u = \frac{1}{2}X_{u-} \exp(\tau\zeta_{u-}) \quad (16)$$

so that (15) and (16) entirely determine the evolution of the population. We are interested in estimating  $x \rightsquigarrow B(x)$  for  $x \in \mathcal{D}$  where  $\mathcal{D} \subset \mathcal{S}$  is a given compact interval. The process  $(X_u)_{u \in \mathbb{T}}$  is a bifurcating Markov chain with state space  $\mathcal{S}$  and  $\mathbb{T}$ -transition any version of

$$\mathcal{P}_B(x, dy dz) = \mathbb{P}(X_{u0} \in dy, X_{u1} \in dz | X_{u-} = x).$$

Moreover, using (15) and (16), (see for instance the derivation of Equation (11) in [DHKR15]), it is not difficult to check that

$$\mathcal{P}_B(x, dy dz) = Q_B(x, dy) \otimes \delta_y(dz)$$

where  $\delta_y$  denotes the Dirac mass at  $y$  and

$$Q_B(x, dy) = \frac{B(2y)}{\tau y} \exp\left(-\int_{x/2}^y \frac{B(2s)}{\tau s} ds\right) \mathbf{1}_{\{y \geq x/2\}} dy. \quad (17)$$

If we assume moreover that  $x \rightsquigarrow B(x)$  is continuous, then we have Assumption 2 with  $\Omega = Q_B$  and  $\mathfrak{n}(dx) = dx$ .

Now, let  $\mathcal{S}$  be a bounded and open interval in  $(0, \infty)$  such that  $\inf \mathcal{S} = 0$ . Pick  $r \in \mathcal{S}$  and  $L > 0$  and introduce the function class

$$\mathcal{C}(r, L) = \left\{ B : \mathcal{S} \rightarrow [0, \infty), \int^{\sup \mathcal{S}} \frac{B(x)}{x} dx = \infty, \int_0^r \frac{B(x)}{x} dx \leq L \right\}.$$

The requirement  $\int^{\sup \mathcal{S}} \frac{B(x)}{x} dx = \infty$  is natural and leads to  $\int_{\mathcal{S}} Q_B(x, dy) = 1$  via (17). We comply with Assumption 3 for  $\Omega = Q_B$  as stated in the following

**Lemma 11.** *Let  $r \in \mathcal{S}$  such that  $r > \sup \mathcal{S}/2$  and  $0 < L < \tau \log 2$ . Then for every  $B \in \mathcal{C}(r, L)$ , the mean transition  $Q_B$  admits a unique invariant probability measure  $\nu_B$  absolutely continuous with respect to the Lebesgue measure. Moreover there exist  $R > 0$  and  $0 < \rho < 1/2$  such that*

$$\sup_{B \in \mathcal{C}(r, L)} |Q_B^m g(x) - \nu_B(g)| \leq R |g|_{\infty} \rho^m, \quad x \in \mathcal{S}, \quad m \geq 0,$$

for every  $g : \mathcal{S} \rightarrow \mathbb{R}$ .

Finally, we know by Proposition 2 in Doumic *et al.* [DHKR15] – see in particular Equation (24) – that

$$B(x) = \frac{\tau x}{2} \frac{\nu_B(x/2)}{\int_{x/2}^x \nu_B(z) dz},$$

where  $\nu_B$  denotes the unique invariant probability of the transition  $\Omega = Q_B$ . This yields a strategy for estimating  $x \rightsquigarrow B(x)$  via an estimator of  $x \rightsquigarrow \nu_B(x)$ . For a given compact interval  $\mathcal{D} \subset \mathcal{S}$ , define

$$\hat{B}_n(x) = \frac{\tau x}{2} \frac{\hat{\nu}_n(x/2)}{\left(\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} \mathbf{1}_{\{x/2 \leq X_u < x\}}\right) \vee \varpi}, \quad (18)$$



where  $\widehat{\nu}_n$  is the wavelet thresholding estimator given in Section 3.2 specified by a maximal resolution level  $J$  and a threshold  $\eta$  and  $\varpi > 0$  (that can be chosen given  $r$  and  $L$ , reproducing Lemma 3 and Lemma 4 of [DHKR15]). As a consequence of Theorem 8 we obtain the following

**Theorem 12.** *Specify  $\widehat{B}_n$  with*

$$J = \frac{1}{2} \log_2 \frac{|\mathbb{T}_n|}{\log |\mathbb{T}_n|} \quad \text{and} \quad \eta = c \sqrt{\log |\mathbb{T}_n| / |\mathbb{T}_n|}$$

for some  $c > 0$ . For every  $B \in \mathcal{C}(r, L)$ , for every  $\pi \in (0, \infty]$ ,  $s \in (1/\pi, \sigma]$  and  $p \geq 1$ , for large enough  $n$  and  $c$  and small enough  $\varpi$ , the following estimate holds

$$\left( \mathbb{E} [\| \widehat{B}_n - B \|_{L^p(\mathcal{D})}^p] \right)^{1/p} \lesssim \left( \frac{\log |\mathbb{T}_n|}{|\mathbb{T}_n|} \right)^{\alpha_1(s, p, \pi)},$$

provided that  $\inf \mathcal{D} \leq r/2$ , with  $\alpha_1(s, p, \pi) = \min \left\{ \frac{2s}{2s+1}, \frac{s+1/p-1/\pi}{2s+1-2/\pi} \right\}$ , up to a constant that depends on  $s, p, \pi, \|B\|_{\mathcal{B}_{\pi, \infty}^s(\mathcal{D})}$ ,  $r$  and  $L$  and that is continuous in its arguments.

This rate is moreover (nearly) optimal: define  $\varepsilon_1 = s\pi - \frac{1}{2}(p - \pi)$ . We have

$$\inf_{\widehat{B}_n} \sup_B \left( \mathbb{E} [\| \widehat{B}_n - B \|_{L^p(\mathcal{D})}^p] \right)^{1/p} \gtrsim \begin{cases} |\mathbb{T}_n|^{-\alpha_1(s, p, \pi)} & \text{if } \varepsilon_1 > 0 \\ \left( \frac{\log |\mathbb{T}_n|}{|\mathbb{T}_n|} \right)^{\alpha_1(s, p, \pi)} & \text{if } \varepsilon_1 \leq 0, \end{cases}$$

where the infimum is taken among all estimators of  $B$  based on  $(X_u)_{u \in \mathbb{T}_n}$  and the supremum is taken among all  $B \in \mathcal{C}(r, L)$  such that  $\|B\|_{\mathcal{B}_{\pi, \infty}^s(\mathcal{D})} \leq C$ .

Two remarks are in order:

1) We improve on the results of Doumic *et al.* [DHKR15] in two directions: we have smoothness-adaptation (in the sense described in Remark 2) after Theorem 8 in Section 3 for several loss functions over various Besov smoothness classes, while [DHKR15] constructs a non-adaptive estimator for Hölder smoothness in squared-error loss; moreover, we prove that the obtained rate is (nearly) optimal in a minimax sense.

2) We unfortunately need to work under the quite stringent restriction that  $\mathcal{S}$  is bounded in order to obtain the uniform ergodicity Assumption 3, see Remark 3) after Theorem 5 in Section 2.

## 4.2 BIFURCATING AUTOREGRESSIVE PROCESS

Bifurcating autoregressive processes (BAR), first introduced by Cowan and Staudte [CS86], are yet another stochastic model for understanding cell division. The trait  $X_u$  may represent the growth rate of a bacteria  $u \in \mathbb{T}$  in a population of *Escherichia coli* but other choices are obviously possible. Contrary to the growth-fragmentation model of Section 4.1 the trait  $(X_{u0}, X_{u1})$  of the

two newborn cells differ and are linked through the autoregressive dynamics

$$\begin{cases} X_{u0} = f_0(X_u) + \sigma_0(X_u)\varepsilon_{u0}, \\ X_{u1} = f_1(X_u) + \sigma_1(X_u)\varepsilon_{u1}, \end{cases} \quad (19)$$

initiated with  $X_\emptyset$  and where

$$f_0, f_1 : \mathbb{R} \rightarrow \mathbb{R} \text{ and } \sigma_0, \sigma_1 : \mathbb{R} \rightarrow (0, \infty)$$

are functions and  $(\varepsilon_{u0}, \varepsilon_{u1})_{u \in \mathbb{T}}$  are i.i.d. noise variables with common density function  $G : \mathbb{R}^2 \rightarrow [0, \infty)$  that specify the model.

The process  $(X_u)_{u \in \mathbb{T}}$  is a bifurcating Markov chain with state space  $\mathcal{S} = \mathbb{R}$  and  $\mathbb{T}$ -transition

$$\mathcal{P}(x, dy dz) = G\left(\sigma_0(x)^{-1}(y - f_0(x)), \sigma_1(x)^{-1}(z - f_1(x))\right) dy dz. \quad (20)$$

This model can be seen as an adaptation of nonlinear autoregressive model when the data have a binary tree structure. The original BAR process in [CS86] is defined for linear link functions  $f_0$  and  $f_1$  with  $f_0 = f_1$ . Several extensions have been studied from a parametric point of view, see *e.g.* Basawa and Huggins [BH99, BH00] and Basawa and Zhou [BZ04, BZ05]. More recently, Delmas and Marsalle [DM10] extend the study to a Galton-Watson tree and introduce asymmetry, de Saporta *et al.* [BdSGP09, dSGPM11, dSGPM12] take into account missing data while Blandin [Bla14], Bercu and Blandin [BB15], and de Saporta *et al.* [dSGPM14] study an extension with random coefficients. Bitseki Penda and Djellout [BD14] prove deviations inequalities and moderate deviations for estimators of parameters in linear BAR processes. From a nonparametric point of view, we mention the applications of [BEG15] (Section 4) where deviations inequalities are derived for the Nadaraya-Watson type estimators of  $f_0$  and  $f_1$  with constant and known functions  $\sigma_0$  and  $\sigma_1$ ). A detailed nonparametric study of these estimators is carried out in Chapter III.

We focus here on the nonparametric estimation of the characteristics of the tagged-branch chain  $\nu$  and  $\mathcal{Q}$  and on the  $\mathbb{T}$ -transition  $\mathcal{P}$ , based on the observation of  $(X_u)_{u \in \mathbb{T}_n}$  for some  $n \geq 1$ . Such an approach can be helpful for the subsequent study of goodness-of-fit tests for instance, when one needs to assess whether the data  $(X_u)_{u \in \mathbb{T}}$  are generated by a model of the form (19) or not.

We set  $G_0(x) = \int_{\mathcal{S}} G(x, y) dy$  and  $G_1(y) = \int_{\mathcal{S}} G(x, y) dx$  for the marginals of  $G$ , and define, for any  $M > 0$ ,

$$\delta(M) = \min \left\{ \inf_{|x| \leq M} G_0(x), \inf_{|x| \leq M} G_1(x) \right\}.$$

**Assumption 13.** For some  $\ell > 0$  and  $\underline{\sigma} > 0$ , we have

$$\max \left\{ \sup_x |f_0(x)|, \sup_x |f_1(x)| \right\} \leq \ell < \infty$$

and

$$\min \left\{ \inf_x \sigma_0(x), \inf_x \sigma_1(x) \right\} \geq \underline{\sigma} > 0.$$

Moreover,  $G_0$  and  $G_1$  are bounded and there exist  $\mu > 0$  and  $M > \ell/\sigma$  such that  $\delta((\mu + \ell)/\sigma) > 0$  and  $2(M\sigma - \ell)\delta(M) > 1/2$ .

Using that  $G_0$  and  $G_1$  are bounded, and (20), we readily check that Assumption 6 is satisfied. We also have Assumption 2 with  $\mathbf{n}(dx) = dx$  and

$$\mathcal{Q}(x, y) = \frac{1}{2} \left( G_0(y - f_0(x)) + G_1(y - f_1(x)) \right),$$

Assumption 13 implies Assumption 3 as well, as follows from a straightforward adaptation of Lemma 25 in Chapter III. Denoting by  $\nu$  the invariant probability of  $\mathcal{Q}$  we also have  $m(\nu) > 0$  with  $m(\nu)$  defined by (11), for every  $\mathcal{D} \subset [-\mu, \mu]$ , see the proof of Lemma 24 in Chapter III. As a consequence, the results stated in Theorems 8, 9 and 10 of Section 3 carry over to the setting of BAR processes satisfying Assumption 13. We thus readily obtain smoothness-adaptive estimators for  $\nu$ ,  $\mathcal{Q}$  and  $\mathcal{P}$  in this context and these results are new.

### 4.3 NUMERICAL ILLUSTRATION

We focus on the growth-fragmentation model and reconstruct its size-dependent splitting rate. We consider a perturbation of the baseline splitting rate  $\tilde{B}(x) = x/(5-x)$  over the range  $x \in \mathcal{S} = (0, 5)$  of the form

$$B(x) = \tilde{B}(x) + \mathbf{c} T(2^j(x - \frac{7}{2}))$$

with  $(\mathbf{c}, j) = (3, 1)$  or  $(\mathbf{c}, j) = (9, 4)$ , and where  $T(x) = (1+x)\mathbf{1}_{\{-1 \leq x < 0\}} + (1-x)\mathbf{1}_{\{0 \leq x \leq 1\}}$  is a tent shaped function. Thus the trial splitting rate with parameter  $(\mathbf{c}, j) = (9, 4)$  is more localized around  $7/2$  and higher than the one associated with parameter  $(\mathbf{c}, j) = (3, 1)$ . One can easily check that both  $\tilde{B}$  and  $B$  belong to the class  $\mathcal{C}(r, L)$  for an appropriate choice of  $(r, L)$ . For a given  $B$ , we simulate  $M = 100$  Monte Carlo trees up to the generation  $n = 15$ . To do so, we draw the size at birth of the initial cell  $X_\emptyset$  uniformly in the interval  $[1.25, 2.25]$ , we fix the growth rate  $\tau = 2$  and given a size at birth  $X_u = x$ , we pick  $X_{u0}$  according to the density  $y \rightsquigarrow Q_B(x, y)$  defined by (17) using a rejection sampling algorithm (with proposition density  $y \rightsquigarrow Q_{\tilde{B}}(x, y)$ ) and set  $X_{u1} = X_{u0}$ . Figure 7 illustrates quantitatively how fast the decorrelation on the tree occurs.

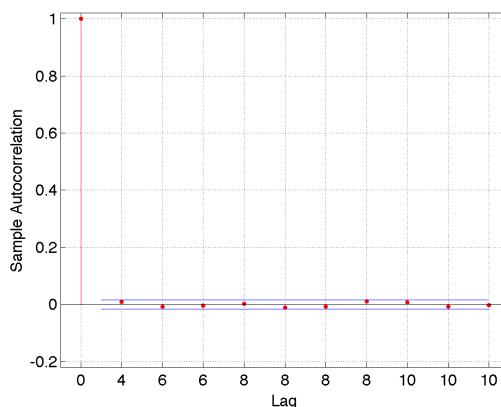


Figure 7: Sample autocorrelation of ordered  $(X_{u0}, u \in \mathbb{G}_{n-1})$  for  $n = 15$ . Note: due to the binary tree structure the lags are  $\{4, 6, 6, \dots\}$ . As expected, we observe a fast decorrelation.

Computational aspects of statistical estimation using wavelets can be found in Härdle *et al.*, Chapter 12 of [HKPT98]. We implement the estimator  $\widehat{B}_n$  defined by (18) using the Matlab wavelet toolbox. We take a wavelet filter corresponding to compactly supported Daubechies wavelets of order 8. As specified in Theorem 12, the maximal resolution level  $J$  is chosen as  $\frac{1}{2} \log_2(|\mathbb{T}_n| / \log |\mathbb{T}_n|)$  and we threshold the coefficients  $\widehat{\nu}_{\lambda,n}$  which are too small by hard thresholding. We choose the threshold proportional to  $\sqrt{\log |\mathbb{T}_n| / |\mathbb{T}_n|}$  (and we calibrate the constant to 10 or 15 for respectively the two trial splitting rates, mainly by visual inspection). We evaluate  $\widehat{B}_n$  on a regular grid of  $\mathcal{D} = [1.5, 4.8]$  with mesh  $\Delta x = (|\mathbb{T}_n|)^{-1/2}$ . For each sample we compute the empirical error

$$e_i = \frac{\|\widehat{B}_n^{(i)} - B\|_{\Delta x}}{\|B\|_{\Delta x}}, \quad i = 1, \dots, M,$$

where  $\|\cdot\|_{\Delta x}$  denotes the discrete  $L^2$ -norm over the numerical sampling and sum up the results through the mean-empirical error  $\bar{e} = M^{-1} \sum_{i=1}^M e_i$ , together with the empirical standard deviation  $(M^{-1} \sum_{i=1}^M (e_i - \bar{e})^2)^{1/2}$ .

Table 4 displays the numerical results we obtained, also giving the compression rate (columns %) defined as the number of wavelet coefficients put to zero divided by the total number of coefficients. We choose a oracle-like error as benchmark: the oracle-like estimator is computed by picking the best resolution level  $J^*$  with no coefficient thresholded. We also display the results when constructing  $\widehat{B}_n$  with  $\mathbb{G}_n$  (instead of  $\mathbb{T}_n$ ), in which case an analog of Theorem 12 holds. For the large spike, the thresholding estimator behaves quite well compared to the oracle-like for a large spike and achieves the same performance for a high spike.

|                    |                | $n = 12$           |       |                    |      | $n = 15$           |       |                    |      |
|--------------------|----------------|--------------------|-------|--------------------|------|--------------------|-------|--------------------|------|
|                    |                | Oracle-like        |       | Threshold est.     |      | Oracle-like        |       | Threshold est.     |      |
|                    |                | $\bar{e}$<br>(sd.) | $J^*$ | $\bar{e}$<br>(sd.) | %    | $\bar{e}$<br>(sd.) | $J^*$ | $\bar{e}$<br>(sd.) | %    |
| <b>Large spike</b> | $\mathbb{T}_n$ | 0.0677<br>(0.0159) | 5     | 0.1020<br>(0.0196) | 96.6 | 0.0324<br>(0.0055) | 6     | 0.0502<br>(0.0055) | 97.1 |
|                    | $\mathbb{G}_n$ | 0.0933<br>(0.0202) | 5     | 0.1454<br>(0.0267) | 97.9 | 0.0453<br>(0.0081) | 6     | 0.0728<br>(0.0097) | 96.7 |
| <b>High spike</b>  | $\mathbb{T}_n$ | 0.1343<br>(0.0180) | 7     | 0.1281<br>(0.0163) | 97.4 | 0.0586<br>(0.0059) | 8     | 0.0596<br>(0.0060) | 97.7 |
|                    | $\mathbb{G}_n$ | 0.1556<br>(0.0222) | 7     | 0.1676<br>(0.0228) | 97.7 | 0.0787<br>(0.0079) | 8     | 0.0847<br>(0.0087) | 97.9 |

Table 4: Mean empirical relative error  $\bar{e}$  and its standard deviation (over  $M = 100$  Monte-Carlo trees), with respect to  $n$ , for the trial splitting rate  $B$  specified by  $(c, j) = (3, 1)$  (large spike) or  $(c, j) = (4, 9)$  (high spike) reconstructed over the interval  $\mathcal{D} = [1.5, 4.8]$  by the estimator  $\widehat{B}_n$ . Note: for  $n = 15$ ,  $\frac{1}{2}|\mathbb{T}_n| = 32\,767$  and  $\frac{1}{2}|\mathbb{G}_n| = 16\,384$ ; for  $n = 12$ ,  $\frac{1}{2}|\mathbb{T}_n| = 4\,095$  and  $\frac{1}{2}|\mathbb{G}_n| = 2\,048$ .

Figure 8 and Figure 9 show the reconstruction of the size-dependent splitting rate  $B$  and the invariant measure  $\nu_B$  in the two cases (large or high spike) for one typical sample of size  $\frac{1}{2}|\mathbb{T}_n| = 32\,767$ .

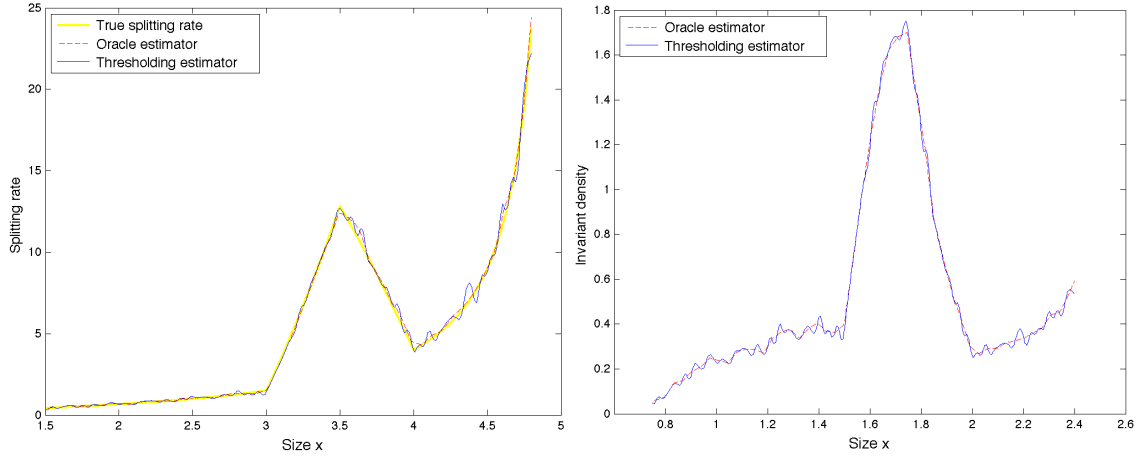


Figure 8: *Large spike: reconstruction of the trial splitting rate  $B$  specified by  $(\mathbf{c}, j) = (3, 1)$  over  $\mathcal{D} = [1.5, 4.8]$  and reconstruction of  $\nu_B$  over  $\mathcal{D}/2$  based on one sample  $(X_u, u \in \mathbb{T}_n)$  for  $n = 15$  (i.e.  $\frac{1}{2}|\mathbb{T}_n| = 32\,767$ ).*

In both cases, the spike is well reconstructed and so are the discontinuities in the derivative of  $B$ . As expected, the spike being localized around  $\frac{7}{2}$  for  $B$ , we detect it around  $\frac{7}{4}$  for the invariant measure of the sizes at birth  $\nu_B$ . The large spike concentrates approximately 50% of the mass of  $\nu_B$  whereas the large only concentrates 20% of the mass of  $\nu_B$ . (Note that the reconstruction of  $B$  on  $\mathcal{D}$  requires the reconstruction of the invariant measure  $\nu_B$  on  $\mathcal{D}/2$  only, recall (18).)

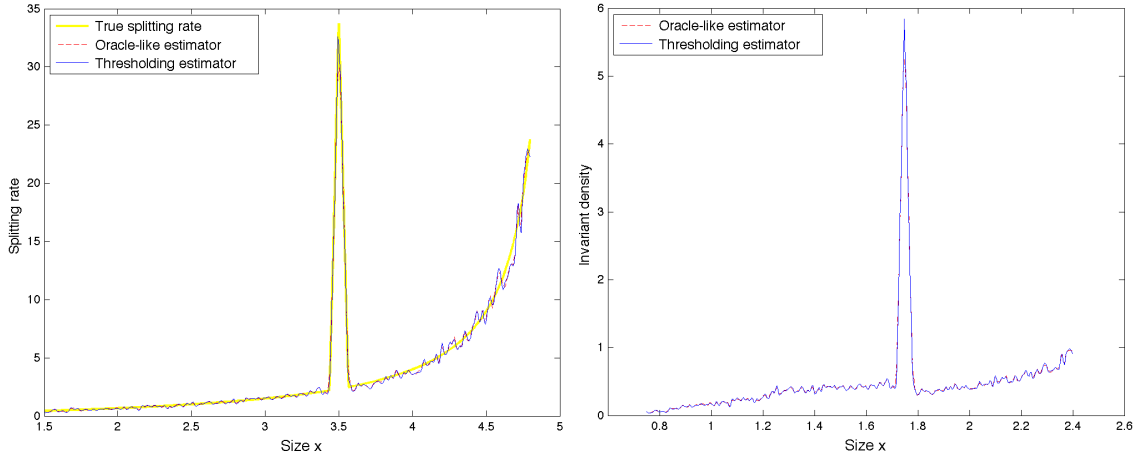


Figure 9: *High spike: reconstruction of the trial splitting rate  $B$  specified by  $(\mathbf{c}, j) = (9, 4)$  over  $\mathcal{D} = [1.5, 4.8]$  and reconstruction of  $\nu_B$  over  $\mathcal{D}/2$  based on one sample  $(X_u, u \in \mathbb{T}_n)$  for  $n = 15$  (i.e.  $\frac{1}{2}|\mathbb{T}_n| = 32\,767$ ).*

## 5 PROOFS

### 5.1 PROOF OF THEOREM 4 (I)

Let  $g : \mathcal{S} \rightarrow \mathbb{R}$  such that  $|g|_1 < \infty$ . Set  $\nu(g) = \int_{\mathcal{S}} g(x)\nu(dx)$  and  $\tilde{g} = g - \nu(g)$ . Let  $n \geq 2$ . By the usual Chernoff bound argument, for every  $\lambda > 0$ , we have

$$\mathbb{P}\left(\frac{1}{|\mathbb{G}_n|} \sum_{u \in \mathbb{G}_n} \tilde{g}(X_u) \geq \delta\right) \leq \exp(-\lambda|\mathbb{G}_n|\delta) \mathbb{E}\left[\exp\left(\lambda \sum_{u \in \mathbb{G}_n} \tilde{g}(X_u)\right)\right]. \quad (21)$$

*Step 1.* We have

$$\begin{aligned} \mathbb{E}\left[\exp\left(\lambda \sum_{u \in \mathbb{G}_n} \tilde{g}(X_u)\right) \middle| \mathcal{F}_{n-1}\right] &= \mathbb{E}\left[\prod_{u \in \mathbb{G}_{n-1}} \exp\left(\lambda(\tilde{g}(X_{u0}) + \tilde{g}(X_{u1}))\right) \middle| \mathcal{F}_{n-1}\right] \\ &= \prod_{u \in \mathbb{G}_{n-1}} \mathbb{E}\left[\exp\left(\lambda(\tilde{g}(X_{u0}) + \tilde{g}(X_{u1}))\right) \middle| \mathcal{F}_{n-1}\right] \end{aligned}$$

thanks to the conditional independence of the  $(X_{u0}, X_{u1})_{u \in \mathbb{G}_{n-1}}$  given  $\mathcal{F}_{n-1}$ , as follows from Definition 1. We rewrite this last term as

$$\prod_{u \in \mathbb{G}_{n-1}} \mathbb{E}\left[\exp\left(\lambda(\tilde{g}(X_{u0}) + \tilde{g}(X_{u1}) - 2\Omega\tilde{g}(X_u))\right) \middle| \mathcal{F}_{n-1}\right] \exp(\lambda 2\Omega\tilde{g}(X_u)),$$

inserting the  $\mathcal{F}_{n-1}$ -measurable random variable  $2\Omega\tilde{g}(X_u)$  for  $u \in \mathbb{G}_{n-1}$ . Moreover, the bifurcating structure of  $(X_u)_{u \in \mathbb{T}}$  implies

$$\mathbb{E}[\tilde{g}(X_{u0}) + \tilde{g}(X_{u1}) - 2\Omega\tilde{g}(X_u) \middle| \mathcal{F}_{n-1}] = 0, \quad u \in \mathbb{G}_{n-1}, \quad (22)$$

since  $\Omega = \frac{1}{2}(\mathcal{P}_0 + \mathcal{P}_1)$ . We will also need the following bound, proof of which is delayed until Appendix

**Lemma 14.** *Work under Assumptions 2 and 3. For all  $r = 0, \dots, n-1$  and  $u \in \mathbb{G}_{n-r-1}$ , we have*

$$|2^r(\mathcal{Q}^r\tilde{g}(X_{u0}) + \mathcal{Q}^r\tilde{g}(X_{u1}) - 2\mathcal{Q}^{r+1}\tilde{g}(X_u))| \leq c_1|g|_\infty$$

and

$$\mathbb{E}\left[\left(2^r(\mathcal{Q}^r\tilde{g}(X_{u0}) + \mathcal{Q}^r\tilde{g}(X_{u1}) - 2\mathcal{Q}^{r+1}\tilde{g}(X_u))\right)^2 \middle| \mathcal{F}_{n-r-1}\right] \leq c_2\sigma_r^2(g),$$

with  $c_1 = 4 \max\{1 + R\rho, R(1 + \rho)\}$ ,  $c_2 = 4 \max\{|\mathcal{Q}|_{\mathcal{D}}, 4|\mathcal{Q}|_{\mathcal{D}}^2, 4R^2(1 + \rho)^2\}$  and

$$\sigma_r^2(g) = \begin{cases} |g|_2^2 & r = 0, \\ \min\{|g|_1^2 2^{2r}, |g|_\infty^2 (2\rho)^{2r}\} & r = 1, \dots, n-1. \end{cases} \quad (23)$$

(Recall that  $|\mathcal{Q}|_{\mathcal{D}} = \sup_{x \in \mathcal{S}, y \in \mathcal{D}} \mathcal{Q}(x, y)$  and  $R, \rho$  are defined via Assumption 3.)

In view of (22) and Lemma 14 for  $r = 0$ , we plan to use the bound

$$\mathbb{E}[\exp(\lambda Z)] \leq \exp\left(\frac{\lambda^2 \sigma^2}{2(1 - \lambda M/3)}\right) \quad (24)$$

valid for any  $\lambda \in (0, 3/M)$ , any random variable  $Z$  such that  $|Z| \leq M$ ,  $\mathbb{E}[Z] = 0$  and  $\mathbb{E}[Z^2] \leq \sigma^2$ . Thus, for any  $\lambda \in (0, 3/c_1|g|_\infty)$  and any  $u \in \mathbb{G}_{n-1}$ , with  $Z = \tilde{g}(X_{u0}) + \tilde{g}(X_{u1}) - 2Q\tilde{g}(X_u)$ , we obtain

$$\mathbb{E} \left[ \exp \left( \lambda (\tilde{g}(X_{u0}) + \tilde{g}(X_{u1}) - 2Q\tilde{g}(X_u)) \right) \middle| \mathcal{F}_{n-1} \right] \leq \exp \left( \frac{\lambda^2 c_2 \sigma_0^2(g)}{2(1 - \lambda c_1 |g|_\infty / 3)} \right).$$

It follows that

$$\mathbb{E} \left[ \exp \left( \lambda \sum_{u \in \mathbb{G}_n} \tilde{g}(X_u) \right) \middle| \mathcal{F}_{n-1} \right] \leq \exp \left( \frac{\lambda^2 c_2 \sigma_0^2(g) |\mathbb{G}_{n-1}|}{2(1 - \lambda c_1 |g|_\infty / 3)} \right) \prod_{u \in \mathbb{G}_{n-1}} \exp(\lambda 2Q\tilde{g}(X_u)). \quad (25)$$

*Step 2.* We iterate the procedure in Step 1. Conditioning with respect to  $\mathcal{F}_{n-2}$ , we need to control

$$\mathbb{E} \left[ \prod_{u \in \mathbb{G}_{n-1}} \exp(\lambda 2Q\tilde{g}(X_u)) \middle| \mathcal{F}_{n-2} \right],$$

and more generally, for  $1 \leq r \leq n-1$ :

$$\begin{aligned} & \mathbb{E} \left[ \prod_{u \in \mathbb{G}_{n-r}} \exp(\lambda 2^r Q^r \tilde{g}(X_u)) \middle| \mathcal{F}_{n-r-1} \right] \\ &= \prod_{u \in \mathbb{G}_{n-r-1}} \mathbb{E} \left[ \exp \left( \lambda 2^r (Q^r \tilde{g}(X_{u0}) + Q^r \tilde{g}(X_{u1}) - 2Q^{r+1} \tilde{g}(X_u)) \right) \middle| \mathcal{F}_{n-r-1} \right] \exp(\lambda 2^{r+1} Q^{r+1} \tilde{g}(X_u)), \end{aligned}$$

the last equality being obtained thanks to the conditional independence of the  $(X_{u0}, X_{u1})_{u \in \mathbb{G}_{n-r-1}}$  given  $\mathcal{F}_{n-r-1}$ . We plan to use (24) again: for  $u \in \mathbb{G}_{n-r-1}$ , we have

$$\mathbb{E} [2^r (Q^r \tilde{g}(X_{u0}) + Q^r \tilde{g}(X_{u1}) - 2Q^{r+1} \tilde{g}(X_u)) \middle| \mathcal{F}_{n-r-1}] = 0$$

and the conditional variance given  $\mathcal{F}_{n-r-1}$  can be controlled using Lemma 14. Using recursively (24), for  $r = 1, \dots, n-1$ ,

$$\mathbb{E} \left[ \prod_{u \in \mathbb{G}_{n-1}} \exp(\lambda 2Q\tilde{g}(X_u)) \middle| \mathcal{F}_0 \right] \leq \prod_{r=1}^{n-1} \exp \left( \frac{\lambda^2 c_2 \sigma_r^2(g) |\mathbb{G}_{n-r-1}|}{2(1 - \lambda c_1 |g|_\infty / 3)} \right) \exp(\lambda 2^n Q^n \tilde{g}(X_\emptyset))$$

for  $\lambda \in (0, 3/c_1|g|_\infty)$ . By Assumption 3,

$$\exp(\lambda 2^n Q^n \tilde{g}(X_\emptyset)) \leq \exp(\lambda 2^n R(2|g|_\infty)\rho^n) \leq \exp(\lambda 2R|g|_\infty)$$

since  $\rho < 1/2$ . In conclusion

$$\mathbb{E} \left[ \prod_{u \in \mathbb{G}_{n-1}} \exp(\lambda 2Q\tilde{g}(X_u)) \right] \leq \exp \left( \frac{\lambda^2 c_2 \sum_{r=1}^{n-1} \sigma_r^2(g) |\mathbb{G}_{n-r-1}|}{2(1 - \lambda c_1 |g|_\infty / 3)} \right) \exp(\lambda 2R|g|_\infty).$$

*Step 3.* Let  $1 \leq \ell \leq n-1$ . By definition of  $\sigma_r^2(g)$  – recall (23) – and using the fact that  $(2\rho)^{2r} \leq 1$ ,

since moreover  $|\mathbb{G}_{n-r-1}| = 2^{n-r-1}$ , we successively obtain

$$\begin{aligned} \sum_{r=1}^{n-1} \sigma_r^2(g) 2^{n-r-1} &\leq 2^{n-1} (|g|_1^2 \sum_{r=1}^{\ell} 2^r + |g|_{\infty}^2 \sum_{r=\ell+1}^{n-1} 2^{-r} (2\rho)^{2r}) \\ &\leq 2^n (|g|_1^2 2^{\ell} + |g|_{\infty}^2 2^{-\ell}) \\ &\leq |\mathbb{G}_n| \phi_n(g) \end{aligned}$$

for an appropriate choice of  $\ell$ , with  $\phi_n(g) = \min_{1 \leq \ell \leq n-1} (|g|_1^2 2^{\ell} + |g|_{\infty}^2 2^{-\ell})$ . It follows that

$$\mathbb{E} \left[ \prod_{u \in \mathbb{G}_{n-1}} \exp(\lambda 2Q \tilde{g}(X_u)) \right] \leq \exp \left( \frac{\lambda^2 c_2 |\mathbb{G}_n| \phi_n(g)}{2(1 - \lambda c_1 |g|_{\infty}/3)} + \lambda 2R |g|_{\infty} \right). \quad (26)$$

*Step 4.* Putting together the estimates (25) and (26) and coming back to (21), we obtain

$$\mathbb{P} \left( \frac{1}{|\mathbb{G}_n|} \sum_{u \in \mathbb{G}_n} \tilde{g}(X_u) \geq \delta \right) \leq \exp \left( -\lambda |\mathbb{G}_n| \delta + \frac{\lambda^2 c_2 |\mathbb{G}_n| \Sigma_{1,n}(g)}{2(1 - \lambda c_1 |g|_{\infty}/3)} + \lambda 2R |g|_{\infty} \right)$$

with  $\Sigma_{1,n}(g) = |g|_2^2 + \phi_n(g)$  for  $n \geq 2$  and  $\Sigma_{1,1}(g) = \sigma_0^2(g) = |g|_2^2$ . Since  $\delta$  is such that  $2R |g|_{\infty} \leq |\mathbb{G}_n| \delta/2$ , we obtain

$$\mathbb{P} \left( \frac{1}{|\mathbb{G}_n|} \sum_{u \in \mathbb{G}_n} \tilde{g}(X_u) \geq \delta \right) \leq \exp \left( -\lambda |\mathbb{G}_n| \frac{\delta}{2} + \frac{\lambda^2 c_2 |\mathbb{G}_n| \Sigma_{1,n}(g)}{2(1 - \lambda c_1 |g|_{\infty}/3)} \right).$$

The admissible choice  $\lambda = \delta / (\frac{2}{3} \delta c_1 |g|_{\infty} + 2c_2 \Sigma_{1,n}(g))$  yields the result.

## 5.2 PROOF OF THEOREM 4 (II)

*Step 1.* Similarly to (21), we plan to use

$$\mathbb{P} \left( \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} \tilde{g}(X_u) \geq \delta \right) \leq \exp(-\lambda |\mathbb{T}_n| \delta) \mathbb{E} \left[ \exp \left( \lambda \sum_{u \in \mathbb{T}_n} \tilde{g}(X_u) \right) \right] \quad (27)$$

for a specific choice of  $\lambda > 0$ . We first need to control

$$\mathbb{E} \left[ \exp \left( \lambda \sum_{u \in \mathbb{T}_n} \tilde{g}(X_u) \right) \middle| \mathcal{F}_{n-1} \right] = \prod_{u \in \mathbb{T}_{n-1}} \exp(\lambda \tilde{g}(X_u)) \mathbb{E} \left[ \exp \left( \lambda \sum_{u \in \mathbb{G}_n} \tilde{g}(X_u) \right) \middle| \mathcal{F}_{n-1} \right].$$

Using (25) to control  $\mathbb{E} \left[ \exp \left( \lambda \sum_{u \in \mathbb{G}_n} \tilde{g}(X_u) \right) \middle| \mathcal{F}_{n-1} \right]$ , we obtain

$$\begin{aligned} &\mathbb{E} \left[ \exp \left( \lambda \sum_{u \in \mathbb{T}_n} \tilde{g}(X_u) \right) \middle| \mathcal{F}_{n-1} \right] \\ &\leq \exp \left( \frac{\lambda^2 c_2 \sigma_0^2(g) |\mathbb{G}_{n-1}|}{2(1 - \lambda c_1 |g|_{\infty}/3)} \right) \prod_{u \in \mathbb{G}_{n-1}} \exp(\lambda 2Q \tilde{g}(X_u)) \prod_{u \in \mathbb{T}_{n-1}} \exp(\lambda \tilde{g}(X_u)). \end{aligned}$$



*Step 2.* We iterate the procedure. At the second step, conditioning with respect to  $\mathcal{F}_{n-2}$ , we need to control

$$\mathbb{E} \left[ \prod_{u \in \mathbb{T}_{n-2}} \exp(\lambda \tilde{g}(X_u)) \prod_{u \in \mathbb{G}_{n-1}} \exp(\lambda \tilde{g}(X_u) + 2\lambda \mathcal{Q} \tilde{g}(X_u)) \middle| \mathcal{F}_{n-2} \right]$$

and more generally, at the  $(r+1)$ -th step (for  $1 \leq r \leq n-1$ ), we need to control

$$\begin{aligned} & \mathbb{E} \left[ \prod_{u \in \mathbb{T}_{n-r-1}} \exp(\lambda \tilde{g}(X_u)) \prod_{u \in \mathbb{G}_{n-r}} \exp \left( \lambda \sum_{m=0}^r 2^m \mathcal{Q}^m \tilde{g}(X_u) \right) \middle| \mathcal{F}_{n-r-1} \right] \\ &= \prod_{u \in \mathbb{T}_{n-r-2}} \exp(\lambda \tilde{g}(X_u)) \prod_{u \in \mathbb{G}_{n-r-1}} \exp \left( \lambda \sum_{m=0}^{r+1} 2^m \mathcal{Q}^m \tilde{g}(X_u) \right) \\ & \quad \times \mathbb{E} \left[ \exp(\lambda \Upsilon_r(X_u, X_{u0}, X_{u1})) \middle| \mathcal{F}_{n-r-1} \right], \end{aligned}$$

where we set

$$\Upsilon_r(X_u, X_{u0}, X_{u1}) = \sum_{m=0}^r 2^m (\mathcal{Q}^m \tilde{g}(X_{u0}) + \mathcal{Q}^m \tilde{g}(X_{u1}) - 2\mathcal{Q}^{m+1} \tilde{g}(X_u)).$$

This representation successively follows from the  $\mathcal{F}_{n-r-1}$ -measurability of the random variable  $\prod_{u \in \mathbb{T}_{n-r-1}} \exp(\lambda \tilde{g}(X_u))$ , the identity

$$\prod_{u \in \mathbb{G}_{n-r}} \exp(F(X_u)) = \prod_{u \in \mathbb{G}_{n-r-1}} \exp(F(X_{u0}) + F(X_{u1})),$$

the independence of  $(X_{u0}, X_{u1})_{u \in \mathbb{G}_{n-r-1}}$  conditional on  $\mathcal{F}_{n-r-1}$  and finally the introduction of the term  $2 \sum_{m=0}^r 2^m \mathcal{Q}^{m+1} \tilde{g}(X_u)$ .

We have, for  $u \in \mathbb{G}_{n-r-1}$

$$\mathbb{E}[\Upsilon_r(X_u, X_{u0}, X_{u1}) \middle| \mathcal{F}_{n-r-1}] = 0,$$

and we prove in Appendix the following bound

**Lemma 15.** *For any  $r = 1, \dots, n-1$ ,  $u \in \mathbb{G}_{n-r-1}$ , we have*

$$|\Upsilon_r(X_u, X_{u0}, X_{u1})| \leq c_3 |g|_\infty$$

and

$$\mathbb{E}[\Upsilon_r(X_u, X_{u0}, X_{u1})^2 \middle| \mathcal{F}_{n-r-1}] \leq c_4 \sigma_r^2(g) < \infty$$

where  $c_3 = 4R(1+\rho)(1-2\rho)^{-1}$ ,  $c_4 = 12 \max\{|\mathcal{Q}|_{\mathcal{D}}, 16|\mathcal{Q}|_{\mathcal{D}}^2, 4R^2(1+\rho)^2(1-2\rho)^{-2}\}$  and

$$\sigma_r^2(g) = |g|_2^2 + \min_{\ell \geq 1} (|g|_1^2 2^{2(\ell \wedge r)} + |g|_\infty^2 (2\rho)^{2\ell} \mathbf{1}_{\{r > \ell\}}). \quad (28)$$

(Recall that  $|\mathcal{Q}|_{\mathcal{D}} = \sup_{x \in \mathcal{S}, y \in \mathcal{D}} \mathcal{Q}(x, y)$  and  $R, \rho$  are defined via Assumption 3.)

In the same way as for Step 2 in the proof of Theorem 4(i), we apply recursively (24) for

$r = 1, \dots, n-1$  to obtain

$$\mathbb{E}\left[\exp\left(\lambda \sum_{u \in \mathbb{T}_n} \tilde{g}(X_u)\right) \middle| \mathcal{F}_0\right] \leq \prod_{r=0}^{n-1} \exp\left(\frac{c_4 \lambda^2 \sigma_r^2(g) |\mathbb{G}_{n-r-1}|}{2(1 - c'_3 \lambda |g|_\infty / 3)}\right) \exp\left(\lambda \sum_{m=0}^n 2^m \mathcal{Q}^m \tilde{g}(X_\emptyset)\right),$$

if  $\lambda \in (0, 3/c'_3 |g|_\infty)$  with  $c'_3 = \max\{c_1, c_3\} = 4 \max\{1 + R\rho, R(1 + \rho)(1 - 2\rho)^{-1}\}$  and  $\sigma_0^2(g) = |g|_2^2$  in order to include Step 1 (we use  $c_4 \geq c_2$  as well). Now, by Assumption 3, this last term can be bounded by

$$\exp\left(\lambda \sum_{m=0}^n 2^m (R |\tilde{g}|_\infty \rho^m)\right) \leq \exp\left(\lambda 2R(1 - 2\rho)^{-1} |g|_\infty\right)$$

since  $\rho < 1/2$ . Since  $|\mathbb{G}_{n-r-1}| = 2^{n-r-1}$ , by definition of  $\sigma_r^2(g)$  – recall (28) – for any  $1 \leq \ell \leq n-1$  and using moreover that  $(2\rho)^\ell \leq 1$ , we obtain

$$\begin{aligned} \sum_{r=0}^{n-1} \sigma_r^2(g) |\mathbb{G}_{n-r-1}| &\leq 2^{n-1} \left( |g|_2^2 \sum_{r=0}^{n-1} 2^{-r} + |g|_1^2 \left( \sum_{r=1}^{\ell} 2^{2r} 2^{-r} + \sum_{r=\ell+1}^{n-1} 2^{2\ell} 2^{-r} \right) + |g|_\infty^2 \sum_{r=\ell+1}^{n-1} 2^{-r} \right) \\ &\leq |\mathbb{T}_n| \Sigma_{1,n}(g), \end{aligned}$$

where  $\Sigma_{1,n}(g)$  is defined in (3). Thus

$$\mathbb{E}\left[\exp\left(\lambda \sum_{u \in \mathbb{T}_n} \tilde{g}(X_u)\right)\right] \leq \exp\left(\frac{c_4 \lambda^2 |\mathbb{T}_n| \Sigma_{1,n}(g)}{2(1 - c'_3 \lambda |g|_\infty / 3)} + \lambda 2R(1 - 2\rho)^{-1} |g|_\infty\right).$$

Step 3. Coming back to (27), for  $\delta > 0$  such that  $2R(1 - 2\rho)^{-1} |g|_\infty \leq |\mathbb{T}_n| \delta / 2$ , we obtain

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} \tilde{g}(X_u) \geq \delta\right) \leq \exp\left(-\lambda |\mathbb{T}_n| \frac{\delta}{2} + \frac{c_4 \lambda^2 |\mathbb{T}_n| \Sigma_{1,n}(g)}{2(1 - c'_3 \lambda |g|_\infty / 3)}\right).$$

We conclude in the same way as in Step 4 of the proof of Theorem 4 (i).

### 5.3 PROOF OF THEOREM 5 (I)

The strategy of proof is similar as for Theorem 4. Let  $g : \mathbb{S}^3 \rightarrow \mathbb{R}$  such that  $|g|_1 < \infty$  and set  $\tilde{g} = g - \nu(\mathcal{P}g)$ . Let  $n \geq 2$  (if  $n = 1$ , set  $\Sigma_{2,1}(g) = |\mathcal{Q}(\mathcal{P}g)|_\infty$ ). Introduce the notation  $\Delta_u = (X_u, X_{u0}, X_{u1})$  for simplicity. For every  $\lambda > 0$ , the usual Chernoff bound reads

$$\mathbb{P}\left(\frac{1}{|\mathbb{G}_n|} \sum_{u \in \mathbb{G}_n} \tilde{g}(\Delta_u) \geq \delta\right) \leq \exp(-\lambda |\mathbb{G}_n| \delta) \mathbb{E}\left[\exp\left(\lambda \sum_{u \in \mathbb{G}_n} \tilde{g}(\Delta_u)\right)\right]. \quad (29)$$

Step 1. We first need to control

$$\begin{aligned} \mathbb{E}\left[\exp\left(\lambda \sum_{u \in \mathbb{G}_n} \tilde{g}(\Delta_u)\right) \middle| \mathcal{F}_{n-1}\right] &= \mathbb{E}\left[\prod_{u \in \mathbb{G}_{n-1}} \exp\left(\lambda(\tilde{g}(\Delta_{u0}) + \tilde{g}(\Delta_{u1}))\right) \middle| \mathcal{F}_{n-1}\right] \\ &= \prod_{u \in \mathbb{G}_{n-1}} \mathbb{E}\left[\exp\left(\lambda(\tilde{g}(\Delta_{u0}) + \tilde{g}(\Delta_{u1}))\right) \middle| \mathcal{F}_{n-1}\right] \end{aligned}$$

using the independence of  $(\Delta_{u0}, \Delta_{u1})_{u \in \mathbb{G}_{n-1}}$  conditional on  $\mathcal{F}_{n-1}$ . Inserting the term  $2\mathcal{Q}(\mathcal{P}\tilde{g})(X_u)$ , this last quantity is also equal to

$$\prod_{u \in \mathbb{G}_{n-1}} \mathbb{E} \left[ \exp \left( \lambda (\tilde{g}(\Delta_{u0}) + \tilde{g}(\Delta_{u1}) - 2\mathcal{Q}(\mathcal{P}\tilde{g})(X_u)) \right) \middle| \mathcal{F}_{n-1} \right] \exp (\lambda 2\mathcal{Q}(\mathcal{P}\tilde{g})(X_u)).$$

For  $u \in \mathbb{G}_{n-1}$  we successively have

$$\mathbb{E} [\tilde{g}(\Delta_{u0}) + \tilde{g}(\Delta_{u1}) - 2\mathcal{Q}(\mathcal{P}\tilde{g})(X_u) | \mathcal{F}_{n-1}] = 0,$$

$$|\tilde{g}(\Delta_{u0}) + \tilde{g}(\Delta_{u1}) - 2\mathcal{Q}(\mathcal{P}\tilde{g})(X_u)| \leq 4(1 + R\rho)|g|_\infty$$

and

$$\mathbb{E} [(\tilde{g}(\Delta_{u0}) + \tilde{g}(\Delta_{u1}) - 2\mathcal{Q}(\mathcal{P}\tilde{g})(X_u))^2 | \mathcal{F}_{n-1}] \leq 4|\mathcal{Q}|_{\mathcal{D}}|\mathcal{P}g^2|_1,$$

with  $|\mathcal{Q}|_{\mathcal{D}} = \sup_{x \in \mathcal{S}, y \in \mathcal{D}} \mathcal{Q}(x, y)$  and  $R, \rho$  defined via Assumption 3. The first equality is obtained by conditioning first on  $\mathcal{F}_n$  then on  $\mathcal{F}_{n-1}$ . The last two estimates are obtained in the same line as the proof of Lemma 14 for  $r = 0$ , using in particular  $\mathcal{Q}(\mathcal{P}g^2)(x) = \int_{\mathcal{S}} \mathcal{P}g^2(y)\mathcal{Q}(x, y)\mathbf{n}(dy) \leq |\mathcal{Q}|_{\mathcal{D}}|\mathcal{P}g^2|_1$  since  $\mathcal{P}g^2$  vanishes outside  $\mathcal{D}$ .

Finally, thanks to (24) with  $Z = \tilde{g}(\Delta_{u0}) + \tilde{g}(\Delta_{u1}) - 2\mathcal{Q}(\mathcal{P}\tilde{g})(X_u)$ , we infer

$$\mathbb{E} \left[ \exp \left( \lambda \sum_{u \in \mathbb{G}_n} \tilde{g}(\Delta_u) \right) \middle| \mathcal{F}_{n-1} \right] \leq \exp \left( \frac{\lambda^2 4|\mathcal{Q}|_{\mathcal{D}}|\mathcal{P}g^2|_1}{2(1 - \lambda 4(1 + R\rho)|g|_\infty/3)} \right) \prod_{u \in \mathbb{G}_{n-1}} \exp (\lambda 2\mathcal{Q}(\mathcal{P}\tilde{g})(X_u)) \quad (30)$$

for  $\lambda \in (0, 3/(4(1 + R\rho)|g|_\infty))$ .

*Step 2.* We wish to control  $\mathbb{E} [\prod_{u \in \mathbb{G}_{n-1}} \exp (\lambda 2\mathcal{Q}(\mathcal{P}\tilde{g})(X_u))]$ . We are back to Step 2 and Step 3 of the proof of Theorem 4 (i), replacing  $\tilde{g}$  by  $\mathcal{P}\tilde{g}$ , which satisfies  $\nu(\mathcal{P}\tilde{g}) = 0$ . Equation (26) entails

$$\mathbb{E} \left[ \prod_{u \in \mathbb{G}_{n-1}} \exp (\lambda 2\mathcal{Q}(\mathcal{P}\tilde{g})(X_u)) \right] \leq \exp \left( \frac{\lambda^2 c_2 |\mathbb{G}_n| \phi_n(\mathcal{P}g)}{2(1 - \lambda c_1 |\mathcal{P}g|_\infty/3)} + \lambda 2R|\mathcal{P}g|_\infty \right) \quad (31)$$

with  $\phi_n(\mathcal{P}g) = \min_{1 \leq \ell \leq n-1} (|\mathcal{P}g|_1^2 2^\ell + |\mathcal{P}g|_\infty^2 2^{-\ell})$  and  $c_1 = 4 \max \{1 + R\rho, R(1 + \rho)\}$ ,  $c_2 = 4 \max \{|\mathcal{Q}|_{\mathcal{D}}, 4|\mathcal{Q}|_{\mathcal{D}}^2, 4R^2(1 + \rho)^2\}$ .

*Step 3.* Putting together (30) and (31), we obtain

$$\mathbb{E} \left[ \exp \left( \lambda \sum_{u \in \mathbb{G}_n} \tilde{g}(\Delta_u) \right) \right] \leq \exp \left( \frac{\lambda^2 c_2 |\mathbb{G}_n| \Sigma_{2,n}(g)}{2(1 - \lambda c_1 |g|_\infty/3)} + \lambda 2R|\mathcal{P}g|_\infty \right) \quad (32)$$

with  $\Sigma_{2,n}(g) = |\mathcal{P}g^2|_1 + \phi_n(\mathcal{P}g)$  and using moreover  $|g|_\infty \geq |\mathcal{P}g|_\infty$  and  $c_1 \geq 4(1 + R\rho)$ . Back to (29), since  $2R|\mathcal{P}g|_\infty \leq |\mathbb{G}_n|\delta/2$  we finally infer

$$\mathbb{P} \left( \frac{1}{|\mathbb{G}_n|} \sum_{u \in \mathbb{G}_n} g(\Delta_u) - \nu(\mathcal{P}g) \geq \delta \right) \leq \exp \left( -\lambda |\mathbb{G}_n| \frac{\delta}{2} + \frac{\lambda^2 c_2 |\mathbb{G}_n| \Sigma_{2,n}(g)}{2(1 - \lambda c_1 |g|_\infty/3)} \right).$$

We conclude in the same way as in Step 4 of the proof of Theorem 4 (i).

#### 5.4 PROOF OF THEOREM 5 (II)

In the same way as before, for every  $\lambda > 0$ ,

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_{n-1}|} \sum_{u \in \mathbb{T}_{n-1}} \tilde{g}(\Delta_u) \geq \delta\right) \leq \exp(-\lambda |\mathbb{T}_{n-1}| \delta) \mathbb{E}\left[\exp\left(\lambda \sum_{u \in \mathbb{T}_{n-1}} \tilde{g}(\Delta_u)\right)\right]. \quad (33)$$

Introduce  $\Sigma'_{2,0}(g) = |\mathcal{P}g^2|_1$  and

$$\Sigma'_{2,n}(g) = |\mathcal{P}g^2|_1 + \inf_{\ell \geq 1} (|\mathcal{P}g|_1^2 2^{\ell \wedge (n-1)} + |\mathcal{P}g|_\infty^2 2^{-\ell} \mathbf{1}_{\{\ell < n-1\}}), \text{ for } n \geq 1.$$

It is not difficult to check that (32) is still valid when replacing  $\Sigma_{2,n}$  by  $\Sigma'_{2,n}$ . We plan to successively expand the sum over the whole tree  $\mathbb{T}_{n-1}$  into sums over each generation  $\mathbb{G}_m$  for  $m = 0, \dots, n-1$ , apply Hölder inequality, apply inequality (32) repeatedly (with  $\Sigma'_{2,m}$ ) together with the bound

$$\sum_{m=0}^{n-1} |\mathbb{G}_m| \Sigma'_{2,m}(g) \leq |\mathbb{T}_{n-1}| \Sigma_{2,n-1}(g).$$

We thus obtain

$$\begin{aligned} \mathbb{E}\left[\exp\left(\lambda \sum_{u \in \mathbb{T}_{n-1}} \tilde{g}(\Delta_u)\right)\right] &= \mathbb{E}\left[\prod_{m=0}^{n-1} \exp\left(\lambda \sum_{u \in \mathbb{G}_m} \tilde{g}(\Delta_u)\right)\right] \\ &\leq \left(\mathbb{E}\left[\exp(n\lambda \tilde{g}(\Delta_\emptyset))\right]\right) \prod_{m=1}^{n-1} \mathbb{E}\left[\exp\left(n\lambda \sum_{u \in \mathbb{G}_m} \tilde{g}(\Delta_u)\right)\right]^{1/n} \\ &\leq \left(\exp(n\lambda 2|g|_\infty) \prod_{m=1}^{n-1} \exp\left(\frac{(n\lambda)^2 c_2 |\mathbb{G}_m| \Sigma'_{2,m}(g)}{2(1 - (n\lambda)c_1|g|_\infty/3)} + (n\lambda)2R|\mathcal{P}g|_\infty\right)\right)^{1/n} \\ &\leq \exp\left(\frac{\lambda^2 c_2 n |\mathbb{T}_{n-1}| \Sigma_{2,n-1}(g)}{2(1 - c_1(n\lambda)|g|_\infty/3)} + 2\lambda(nR|\mathcal{P}g|_\infty + |g|_\infty)\right). \end{aligned}$$

Coming back to (33) and using  $2(nR|\mathcal{P}g|_\infty + |g|_\infty) \leq |\mathbb{T}_{n-1}| \delta/2$ , we obtain

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_{n-1}|} \sum_{u \in \mathbb{T}_{n-1}} \tilde{g}(\Delta_u) \geq \delta\right) \leq \exp\left(-\lambda |\mathbb{T}_{n-1}| \frac{\delta}{2} + \frac{\lambda^2 c_2 n |\mathbb{T}_{n-1}| \Sigma_{2,n-1}(g)}{2(1 - (n\lambda)c_1|g|_\infty/3)}\right).$$

We conclude in the same way as in Step 4 of the proof of Theorem 4 (i).

#### 5.5 PROOF OF THEOREM 8

Put  $c(n) = (\log |\mathbb{T}_n| / |\mathbb{T}_n|)^{1/2}$  and note that the maximal resolution  $J = J_n$  is such that  $2^{J_n} \sim c(n)^{-2}$ . Theorem 8 is a consequence of the general theory of wavelet threshold estimators, see Kerkycharian and Picard [KP00]. We first claim that the following moment bounds and moderate

deviation inequalities hold: for every  $p \geq 1$ ,

$$\mathbb{E}[|\widehat{\nu}_{\lambda,n} - \nu_\lambda|^p] \lesssim c(n)^p \text{ for every } |\lambda| \leq J_n \quad (34)$$

and

$$\mathbb{P}(|\widehat{\nu}_{\lambda,n} - \nu_\lambda| \geq p\kappa c(n)) \leq c(n)^{2p} \text{ for every } |\lambda| \leq J_n \quad (35)$$

provided  $\kappa > 0$  is large enough, see Condition (37) below. In turn, we have Conditions (5.1) and (5.2) of Theorem 5.1 of [KP00] with  $\Lambda_n = J_n$  (with the notation of [KP00]). By Corollary 5.1 and Theorem 6.1 of [KP00] we obtain Theorem 8.

It remains to prove (34) and (35). We plan to apply Theorem 4(ii) with  $g = \psi_\lambda$  and  $\delta = \delta_n = p\kappa c(n)$ . First, we have  $|\psi_\lambda^1|_p \leq C_p 2^{|\lambda|(1/2-1/p)}$  for  $p = 1, 2, \infty$  by (6), so one readily checks that for

$$\kappa \geq \frac{4}{p} R(1-2\rho)^{-1} C_\infty (\log |\mathbb{T}_n|)^{-1},$$

the condition  $\delta_n \geq 4R(1-2\rho)^{-1} |\psi_\lambda^1|_\infty |\mathbb{T}_n|^{-1}$  is satisfied, and this is always true for large enough  $n$ . Furthermore, since  $2^{|\lambda|} \leq 2^{J_n} \leq c(n)^{-2}$  it is not difficult to check that

$$\Sigma_{1,n}(\psi_\lambda^1) = |\psi_\lambda^1|_2^2 + \min_{1 \leq \ell \leq n-1} (|\psi_\lambda^1|_1^2 2^\ell + |\psi_\lambda^1|_\infty^2 2^{-\ell}) \leq C \quad (36)$$

for some  $C > 0$  and thus  $\kappa_3 \Sigma_{1,n}(\psi_\lambda) \leq \kappa_3 C = C'$  say. Also  $\kappa_4 |\psi_\lambda^1|_\infty \delta_n \leq \kappa_4 C_\infty 2^{|\lambda|/2} c(n) p \kappa \leq C'' p \kappa$ , where  $C'' > 0$  does not depend on  $n$  since  $2^{|\lambda|/2} \leq c(n)^{-1}$ . Theorem 4(ii) yields

$$\mathbb{P}(|\widehat{\nu}_{\lambda,n} - \nu_\lambda| \geq p\kappa c(n)) \leq 2 \exp\left(-\frac{|\mathbb{T}_n| p^2 \kappa^2 c(n)^2}{C' + C'' p \kappa}\right) \leq c(n)^{2p}$$

for  $\kappa$  such that

$$\kappa \geq \frac{1}{2} C'' + \sqrt{(C'')^2 + \frac{4}{p} C'} \quad (37)$$

and large enough  $n$ . Thus (35) is proved. Straightforward computations show that (34) follows using  $\mathbb{E}[|\widehat{\nu}_{\lambda,n} - \nu_\lambda|^p] = \int_0^\infty p u^{p-1} \mathbb{P}(|\widehat{\nu}_{\lambda,n} - \nu_\lambda| \geq u) du$  and (35) again. The proof of Theorem 8 is complete.

## 5.6 PREPARATION FOR THE PROOF OF THEOREM 9

For  $h : \mathbb{S}^2 \rightarrow \mathbb{R}$ , define  $|h|_{\infty,1} = \sup_{x \in \mathbb{S}} \int_{\mathbb{S}} |h(x,y)| dy$ . For  $n \geq 2$ , set also

$$\Sigma_{3,n}(h) = |h|_2^2 + \min_{1 \leq \ell \leq n-1} (|h|_1^2 2^\ell + |h|_{\infty,1}^2 2^{-\ell}). \quad (38)$$

Recall that under Assumption 3 with  $\mathbf{n}(dx) = dx$ , we set  $f_\Omega(x,y) = \nu(x)\Omega(x,y)$ . Before proving Theorem 9, we first need the following preliminary estimate

**Lemma 16.** *Work under Assumption 2 with  $\mathbf{n}(dx) = dx$  and Assumption 3. Let  $h : \mathcal{D}^2 \subseteq \mathbb{S}^2 \rightarrow \mathbb{R}$*

be such that  $|hf_\Omega|_1 < \infty$ . For every  $n \geq 1$  and for any  $\delta \geq 4|h|_\infty(Rn+1)|\mathbb{T}_n^*|^{-1}$ , we have

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_n^*|} \sum_{u \in \mathbb{T}_n^*} h(X_{u^-}, X_u) - \langle h, f_\Omega \rangle \geq \delta\right) \leq \exp\left(\frac{-n^{-1}|\mathbb{T}_n^*|\delta^2}{\kappa_5 \Sigma_{3,n}(h) + \kappa_2 |h|_\infty \delta}\right)$$

where  $\mathbb{T}_n^* = \mathbb{T}_n \setminus \{\emptyset\}$  and  $\kappa_5 = \max\{|\Omega|_{\mathcal{D}}, |\Omega|_{\mathcal{D}}^2\} \kappa_1(\Omega, \mathcal{D})$ .

*Proof.* We plan to apply Theorem 5 (ii) to  $g(x, x_0, x_1) = \frac{1}{2}(h(x, x_0) + h(x, x_1))$ . Since  $\Omega = \frac{1}{2}(\mathcal{P}_0 + \mathcal{P}_1)$  we readily have  $\mathcal{P}g(x) = \int_{\mathcal{S}} h(x, y)\Omega(x, y)dy$ . Moreover, in that case,

$$\frac{1}{|\mathbb{T}_{n-1}|} \sum_{u \in \mathbb{T}_{n-1}} g(X_u, X_{u0}, X_{u1}) = \frac{1}{|\mathbb{T}_n^*|} \sum_{u \in \mathbb{T}_n^*} h(X_{u^-}, X_u)$$

and  $\int_{\mathcal{S}} \mathcal{P}g(x)\nu(x)dx = \int_{\mathcal{S} \times \mathcal{S}} h(x, y)\Omega(x, y)\nu(x)dxdy = \langle h, f_\Omega \rangle$ . We then simply need to estimate  $\Sigma_{2,n}(g)$  defined by (4). It is not difficult to check that the following estimates hold

$$|\mathcal{P}g|_1^2 \leq |\Omega|_{\mathcal{D}}^2 |h|_1^2, \quad |\mathcal{P}g|_\infty^2 \leq |\Omega|_{\mathcal{D}}^2 |h|_{\infty,1}^2 \quad \text{and} \quad |\mathcal{P}g^2|_1 \leq |\Omega|_{\mathcal{D}} |h|_2^2$$

since  $(\mathcal{P}g^2)(x) \leq \int_{\mathcal{S}} h(x, y)^2 \Omega(x, y)dy$  and  $h$  vanishes outside  $\mathcal{D}^2$ . This entails

$$\Sigma_{2,n}(g) \leq \max\{|\Omega|_{\mathcal{D}}, |\Omega|_{\mathcal{D}}^2\} \Sigma_{3,n}(h)$$

and the result follows.  $\square$

## 5.7 PROOF OF THEOREM 9, UPPER BOUND

*Step 1.* We proceed as for Theorem 8. Putting  $c(n) = (n \log |\mathbb{T}_n^*|/|\mathbb{T}_n^*|)^{1/2}$  and noting that the maximal resolution  $J = J_n$  is such that  $2^{dJ_n} \sim c(n)^{-2}$  with  $d = 2$ , we only have to prove that for every  $p \geq 1$ ,

$$\mathbb{E}[|\widehat{f}_{\lambda,n} - f_\lambda|^p] \lesssim c(n)^p \quad \text{for every } |\lambda| \leq J_n \quad (39)$$

and

$$\mathbb{P}(|\widehat{f}_{\lambda,n} - f_\lambda| \geq p\kappa c(n)) \leq c(n)^{2p} \quad \text{for every } |\lambda| \leq J_n. \quad (40)$$

We plan to apply Lemma 16 with  $h(x, y) = \psi_\lambda^d(x, y) = \psi_\lambda^2(x, y)$  and  $\delta = \delta_n = p\kappa c(n)$ . With the notation used in the proof of Theorem 8 one readily checks that for

$$\varkappa \geq \frac{4}{p}(1-2\rho)^{-1} C_\infty(Rn+1)(\log |\mathbb{T}_n^*|)^{-1}$$

the condition  $\delta_n \geq 4|\psi_\lambda^2|_\infty(Rn+1)|\mathbb{T}_n^*|^{-1}$  is satisfied, and this is always true for large enough  $n$  and

$$\varkappa \geq \frac{4}{p}(1-2\rho)^{-1} C_\infty(2R+1). \quad (41)$$

Furthermore, since  $|\psi_\lambda^d|_p \leq C_p 2^{d|\lambda|(1/2-1/p)}$  for  $p = 1, 2, \infty$  and  $2^{d|\lambda|} \leq 2^{dJ_n} \leq c(n)^{-2}$  we can easily check

$$\Sigma_{3,n}(\psi_\lambda^d) = |\psi_\lambda^d|_2^2 + \min_{1 \leq \ell \leq n-1} (|\psi_\lambda^d|_1^2 2^\ell + |\psi_\lambda^d|_{\infty,1} 2^{-\ell}) \leq C$$

for some  $C > 0$ , and thus  $\kappa_5 \Sigma_{3,n}(g) \leq \kappa_5 C = C'$  say. Also,  $\kappa_2 |\psi_\lambda^d|_\infty \delta_n \leq \kappa_2 C_\infty 2^{d|\lambda|/2} c(n) p \varkappa \leq C'' p \varkappa$ , where  $C''$  does not depend on  $n$ . Applying Lemma 16, we derive

$$\mathbb{P}(|\widehat{f}_{\lambda,n} - f_\lambda| \geq p \varkappa c(n)) \leq 2 \exp\left(-\frac{n^{-1} |\mathbb{T}_{n-1}| p^2 \varkappa^2 c(n)^2}{C' + C'' p \varkappa}\right) \leq c(n)^{2p}$$

as soon as  $\varkappa$  satisfies (41) and (37) (with appropriate changes for  $C'$  and  $C''$ ). Thus (40) is proved and (39) follows likewise. By [KP00] (Corollary 5.1 and Theorem 6.1), we obtain

$$\mathbb{E}\left(\left[\|\widehat{f}_n - f_\Omega\|_{L^p(\mathcal{D}^2)}^p\right]\right)^{1/p} \lesssim \left(\frac{n \log |\mathbb{T}_n|}{|\mathbb{T}_n|}\right)^{\alpha_2(s,p,\pi)} \quad (42)$$

as soon as  $\|f_\Omega\|_{B_{\pi,\infty}^s(\mathcal{D}^2)}$  is finite, as follows from  $f_\Omega(x,y) = \mathcal{Q}(x,y)\nu(x)$  and the fact that  $\|\nu\|_{B_{\pi,\infty}^s(\mathcal{D})}$  is finite too. The last statement can be readily seen from the representation  $\nu(x) = \int_{\mathcal{S}} \nu(y) \mathcal{Q}(y,x) dy$  and the definition of Besov spaces in terms of moduli of continuity, see *e.g.* Meyer [Mey90] or Härdle *et al.* [HKPT98], using moreover that  $\pi \geq 1$ .

*Step 2.* Since  $\mathcal{Q}(x,y) = f_\Omega(x,y)/\nu(x)$  and  $\widehat{\mathcal{Q}}_n(x,y) = \widehat{f}_n(x,y)/\max\{\widehat{\nu}_n(x), \varpi\}$ , we readily have

$$|\widehat{\mathcal{Q}}_n(x,y) - \mathcal{Q}(x,y)|^p \lesssim \frac{1}{\varpi^p} (|\widehat{f}_n(x,y) - f_\Omega(x,y)|^p + \frac{|f_\Omega|_\infty^p}{m(\nu)^p} |\max\{\widehat{\nu}_n(x), \varpi\} - \nu(x)|^p),$$

where the supremum for  $f_\Omega$  can be restricted over  $\mathcal{D}^2$ . Since  $m(\nu) \geq \varpi$ , we have  $|\max\{\widehat{\nu}_n(x), \varpi\} - \nu(x)| \leq |\widehat{\nu}_n(x) - \nu(x)|$  for  $x \in \mathcal{D}$ , therefore

$$\|\widehat{\mathcal{Q}}_n - \mathcal{Q}\|_{L^p(\mathcal{D}^2)}^p \lesssim \frac{1}{\varpi^p} (\|\widehat{f}_n - f_\Omega\|_{L^p(\mathcal{D}^2)}^p + \frac{|f_\Omega|_\infty^p}{m(\nu)^p} \|\nu - \nu_n\|_{L^p(\mathcal{D})}^p)$$

holds as well. We conclude by applying successively the estimate (42) and Theorem 8.

## 5.8 PROOF OF THEOREM 9, LOWER BOUND

We only give a brief sketch: the proof follows classical lower bounds techniques, bounding appropriate statistical distances along hypercubes, see [DJKP96, HKPT98] and more specifically [Clé00, Hof99, Lac08] for specific techniques involving Markov chains. We separate the so-called *dense* and *sparse* case.

*The dense case*  $\varepsilon_2 > 0$ . Let  $\psi_\lambda : \mathcal{D}^2 \rightarrow \mathbb{R}$  a family of (compactly supported) wavelets adapted to the domain  $\mathcal{D}$  and satisfying Assumption 7. For  $j$  such that  $|\mathbb{T}_n|^{-1/2} \lesssim 2^{-j(s+1)}$ , consider the family

$$\mathcal{Q}_{\varepsilon,j}(x,y) = |\mathcal{D}^2|^{-1} \mathbf{1}_{\mathcal{D}^2}(x,y) + \gamma |\mathbb{T}_n|^{-1/2} \sum_{\lambda \in \Lambda_j} \varepsilon_\lambda \psi_\lambda^2(x,y)$$

where  $\varepsilon \in \{-1, 1\}^{\Lambda_j}$  and  $\gamma > 0$  is a tuning parameter (independent of  $n$ ). Since  $|\psi_\lambda^2|_\infty \leq C_\infty 2^{|\lambda|} = C_\infty 2^j$  and since the number of overlapping terms in the sum is bounded (by some fixed integer  $N$ ), we have

$$\gamma |\mathbb{T}_n|^{-1/2} \left| \sum_{\lambda \in \Lambda_j} \varepsilon_\lambda \psi_\lambda^2(x,y) \right| \leq \gamma |\mathbb{T}_n|^{-1/2} N C_\infty 2^j \lesssim \gamma.$$

This term can be made smaller than  $|\mathcal{D}^2|^{-1}$  by picking  $\gamma$  sufficiently small. Hence  $\mathcal{Q}_{\varepsilon,j}(x,y) \geq 0$

and since  $\int \psi_\lambda = 0$ , the family  $\mathcal{Q}_{\epsilon,j}(x, y)$  are all admissible *mean* transitions with common invariant measure  $\nu(dx) = \mathbf{1}_{\mathcal{D}}(x)dx$  and belong to a common ball in  $\mathcal{B}_{\pi,\infty}^s(\mathcal{D}^2)$ . For  $\lambda \in \Lambda_j$ , define  $T_\lambda : \{-1, 1\}^{\Lambda_j} \rightarrow \{-1, 1\}^{|\Lambda_j|}$  by  $T_\lambda(\epsilon_\lambda) = -\epsilon_\lambda$  and  $T_\lambda(\epsilon_\mu) = \epsilon_\mu$  if  $\mu \neq \lambda$ . The lower bound in the dense case is then a consequence of the following inequality

$$\limsup_n \max_{\epsilon \in \{-1, 1\}^{\Lambda_j}, \lambda \in \Lambda_j} \|\mathbb{P}_{\epsilon,j}^n - \mathbb{P}_{T_\lambda(\epsilon),j}^n\|_{TV} < 1, \quad (43)$$

where  $\mathbb{P}_{\epsilon,j}^n$  is the law of  $(X_u)_{u \in \mathbb{T}_n}$  specified by the  $\mathbb{T}$ -transition  $\mathcal{P}_{\epsilon,j} = \mathcal{Q}_{\epsilon,j} \otimes \mathcal{Q}_{\epsilon,j}$  and the initial condition  $\mathcal{L}(X_\emptyset) = \nu$ .

We briefly show how to obtain (43). By Pinsker's inequality, it is sufficient to prove that  $\mathbb{E}_{\epsilon,j}^n \left[ \log \frac{d\mathbb{P}_{\epsilon,j}^n}{d\mathbb{P}_{T_\lambda(\epsilon),j}^n} \right]$  can be made arbitrarily small uniformly in  $n$  (but fixed). We have

$$\begin{aligned} \mathbb{E}_{\epsilon,j}^n \left[ -\log \frac{d\mathbb{P}_{T_\lambda(\epsilon),j}^n}{d\mathbb{P}_{\epsilon,j}^n} \right] &= - \sum_{u \in \mathbb{T}_n} \mathbb{E}_{\epsilon,j}^n \left[ \log \frac{\mathcal{P}_{T_\lambda(\epsilon),j}(X_u, X_{u0}, X_{u1})}{\mathcal{P}_{\epsilon,j}(X_u, X_{u0}, X_{u1})} \right] \\ &= - \sum_{u \in \mathbb{T}_{n+1}^*} \mathbb{E}_{\epsilon,j}^n \left[ \log \frac{\mathcal{Q}_{T_\lambda(\epsilon),j}(X_{u^-}, X_u)}{\mathcal{Q}_{\epsilon,j}(X_{u^-}, X_u)} \right] \\ &= -|\mathbb{T}_{n+1}^*| \int_{\mathcal{D}^2} \log \left( \frac{\mathcal{Q}_{T_\lambda(\epsilon),j}(x, y)}{\mathcal{Q}_{\epsilon,j}(x, y)} \right) \mathcal{Q}_{\epsilon,j}(x, y) \nu(dx) dy \\ &\leq |\mathbb{T}_{n+1}^*| \int_{\mathcal{D}^2} \left( \frac{\mathcal{Q}_{T_\lambda(\epsilon),j}(x, y)}{\mathcal{Q}_{\epsilon,j}(x, y)} - 1 \right)^2 \mathcal{Q}_{\epsilon,j}(x, y) \nu(dx) dy \end{aligned}$$

using  $-\log(1+z) \leq z^2 - z$  valid for  $z \geq -1/2$  and the fact that  $\nu(dx)$  is an invariant measure for both  $\mathcal{Q}_{T_\lambda(\epsilon),j}$  and  $\mathcal{Q}_{\epsilon,j}$ . Noting that

$$\mathcal{Q}_{T_\lambda(\epsilon),j}(x, y) = \mathcal{Q}_{\epsilon,j}(x, y) - 2\gamma|\mathbb{T}_n|^{-1/2} \epsilon_\lambda \psi_\lambda^2(x, y),$$

we derive

$$\left| \frac{\mathcal{Q}_{T_\lambda(\epsilon),j}(x, y)}{\mathcal{Q}_{\epsilon,j}(x, y)} - 1 \right| \leq \frac{2\gamma|\mathbb{T}_n|^{-1/2} C_\infty 2^j}{1 - \gamma|\mathbb{T}_n|^{-1/2} N C_\infty 2^j} \lesssim \gamma|\mathbb{T}_n|^{-1/2}$$

hence the squared term within the integral is of order  $\gamma^2|\mathbb{T}_n|^{-1}$  so that, by picking  $\gamma$  sufficiently small, our claim about  $\mathbb{E}_{\epsilon,j}^n \left[ \log \frac{d\mathbb{P}_{\epsilon,j}^n}{d\mathbb{P}_{T_\lambda(\epsilon),j}^n} \right]$  is proved and (43) follows.

*The sparse case*  $\epsilon_2 \leq 0$ . We now consider the family

$$\mathcal{Q}_{\lambda,j}(x, y) = |\mathcal{D}^2|^{-1} \mathbf{1}_{\mathcal{D}^2}(x, y) + \gamma \left( \frac{\log |\mathbb{T}_n|}{|\mathbb{T}_n|} \right)^{1/2} \epsilon_\lambda \psi_\lambda^2(x, y)$$

with  $\epsilon_\lambda \in \{-1, +1\}$  and  $\lambda \in \Lambda_j$ , with  $j$  such that  $\left( \frac{\log |\mathbb{T}_n|}{|\mathbb{T}_n|} \right)^{1/2} \lesssim 2^{-j(s+1-2/\pi)}$ . The lower bound then follows from the representation

$$\log \frac{d\mathbb{P}_{\lambda,j}^n}{d\mathbb{P}_\nu^n} = \mathcal{U}_\lambda^n - \omega_\lambda \log 2^j$$

where  $\mathbb{P}_{\lambda,j}^n$  and  $\mathbb{P}_\nu^n$  denote the law of  $(X_u)_{u \in \mathbb{T}_n}$  specified by the  $\mathbb{T}$ -transitions  $\mathcal{Q}_{\lambda,j} \otimes \mathcal{Q}_{\lambda,j}$  and  $\nu \otimes \nu$



respectively (and the initial condition  $\mathcal{L}(X_0) = \nu$ ); the  $\omega$ 's are such that  $\sup_n \max_{\lambda \in \Lambda_j} \omega_\lambda < 1$ , and  $\mathcal{U}_\lambda^n$  are random variables such that  $\mathbb{P}_{\lambda,j}^n(\mathcal{U}_\lambda^n \geq -C_1) \geq C_2 > 0$  for some  $C_1, C_2 > 0$ . We omit the details, see e.g. [Clé00, Hof99, Lac08].

## 5.9 PROOF OF THEOREM 10

*Proof of Theorem 10, upper bound.* We closely follow the proof of Theorem 9, choosing  $c(n) = (n \log |\mathbb{T}_{n-1}| / |\mathbb{T}_{n-1}|)^{1/2}$  and  $J = J_n$  such that  $2^{dJ_n} \sim c(n)^{-2}$  with  $d = 3$  now. With  $\delta = \delta_n = p\kappa c(n)$ , for  $\kappa \geq \frac{4}{p}(1-2\rho)^{-1}C_\infty(2R+1)$ , we have  $\delta_n \geq 4|\psi_\lambda^3|_\infty(Rn+1)|\mathbb{T}_n^*|^{-1}$ .

Furthermore, since  $|\psi_\lambda^d|_p \leq C_p 2^{d|\lambda|(1/2-1/p)}$  for  $p = 1, 2, \infty$  and  $2^{d|\lambda|} \leq 2^{dJ_n} \leq c(n)^{-2}$  it is not difficult to check that

$$\Sigma_{2,n}(\psi_\lambda) \leq \max\{|\mathcal{P}|_{\mathcal{D}}, |\mathcal{P}|_{\mathcal{D}}^2\} \left( |\psi_\lambda^d|_2^2 + \min_{1 \leq \ell \leq n-1} (|\psi_\lambda^d|_1^2 2^\ell + |\psi_\lambda^d|_{\infty,1} 2^{-\ell}) \right) \leq C$$

thanks to Assumption 6, where  $|\psi_\lambda^d|_{\infty,1} = \sup_{x \in \mathcal{D}} \int_{\mathcal{D}^2} |\psi_\lambda^d(x, y, z)| dy dz$ , and thus  $\kappa_1 \Sigma_{2,n}(g) \leq \kappa_1 C = C'$ . We also have  $\kappa_2 |\psi_\lambda^d|_\infty \delta_n \leq \kappa_2 C_\infty 2^{|\lambda|d/2} c(n) p \kappa \leq C'' p \kappa$ , where  $C''$  does not depend on  $n$ . Noting that  $f_\lambda = \langle f_{\mathcal{P}}, \psi_\lambda^d \rangle = \int \mathcal{P} \psi_\lambda^d dv$ , we apply Theorem 5 (ii) to  $g = \psi_\lambda$  and derive

$$\mathbb{P}(|\widehat{f}_{\lambda,n} - f_\lambda| \geq p\kappa c(n)) \leq 2 \exp\left(-\frac{n^{-1} |\mathbb{T}_{n-1}| p^2 \kappa^2 c(n)^2}{C' + C'' p \kappa}\right) \leq c(n)^{2p}$$

for every  $|\lambda| \leq J_n$  as soon as  $\kappa$  is large enough and the estimate

$$\mathbb{E}\left(\left[\|\widehat{f}_n - f_{\mathcal{P}}\|_{L^p(\mathcal{D}^3)}^p\right]\right)^{1/p} \lesssim \left(\frac{n \log |\mathbb{T}_n|}{|\mathbb{T}_n|}\right)^{\alpha_3(s,p,\pi)}$$

follows thanks to the theory of [KP00]. The end of the proof follows Step 2 of the proof of Theorem 9 line by line, substituting  $f_\Omega$  by  $f_{\mathcal{P}}$ .  $\square$

*Proof of Theorem 10, lower bound.* This is a slight modification of the proof of Theorem 9, lower bound. For the dense case  $\epsilon_3 > 0$ , we consider an hypercube of the form

$$\mathcal{P}_{\epsilon,j}(x, y, z) = |\mathcal{D}^3|^{-1} \mathbf{1}_{\mathcal{D}^3}(x, y, z) + \gamma |\mathbb{T}_n|^{-1/2} \sum_{\lambda \in \Lambda_j} \epsilon_\lambda \psi_\lambda^3(x, y, z)$$

where  $\epsilon \in \{-1, 1\}^{\Lambda_j}$  with  $j$  such that  $|\mathbb{T}_n|^{-1/2} \lesssim 2^{-j(s+3/2)}$  and  $\gamma > 0$  a tuning parameter, while for the sparse case  $\epsilon_3 \leq 0$ , we consider the family

$$\mathcal{P}_{\lambda,j}(x, y, z) = |\mathcal{D}^3|^{-1} \mathbf{1}_{\mathcal{D}^3}(x, y, z) + \gamma \left(\frac{\log |\mathbb{T}_n|}{|\mathbb{T}_n|}\right)^{1/2} \epsilon_\lambda \psi_\lambda^3(x, y, z)$$

with  $\epsilon_\lambda \in \{-1, +1\}$ ,  $\lambda \in \Lambda_j$ , and  $j$  such that  $\left(\frac{\log |\mathbb{T}_n|}{|\mathbb{T}_n|}\right)^{1/2} \lesssim 2^{-j(s+3(1/2-1/\pi))}$ . The proof then goes along a classical line.  $\square$

## 5.10 PROOF OF THEOREM 12

*Proof of Theorem 12, upper bound.* Set  $\widehat{v}_n(x) = \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} \mathbf{1}_{\{x/2 \leq X_u \leq x\}}$  and  $v_\nu(x) = \int_{x/2}^x \nu_B(y) dy$ . Mimicking Lemma 3 and 4 in Doumic *et al.* [DHKR15], one can check that  $\sup_{x \in \mathcal{D}} \nu_B(x) < \infty$  and  $\inf_{x \in \mathcal{D}} v_\nu(x) > 0$  uniformly in  $B \in \mathcal{C}(r, L)$  provided that  $\inf \mathcal{D} \leq r/2$ . For  $x \in \mathcal{D}$ , we have

$$\begin{aligned} |\widehat{B}_n(x) - B(x)|^p &\lesssim \frac{1}{\varpi^p} |\widehat{v}_n(x) - \nu_B(x)|^p + \frac{\sup_{x \in \mathcal{D}} \nu_B(x)^p}{\inf_{x \in \mathcal{D}} v_\nu(x)^p} |\max\{\widehat{v}_n(x), \varpi\} - v_\nu(x)|^p \\ &\lesssim |\widehat{v}_n(x) - \nu_B(x)|^p + |\widehat{v}_n(x) - v_\nu(x)|^p. \end{aligned}$$

By Theorem 4 (ii) with  $g = \mathbf{1}_{\{x/2 \leq \cdot \leq x\}}$ , one readily checks

$$\mathbb{E}[|\widehat{v}_n(x) - v_\nu(x)|^p] = \int_0^\infty p u^{p-1} \mathbb{P}(|\widehat{v}_n(x) - v_\nu(x)| \geq u) du \lesssim |\mathbb{T}_n|^{-p/2}$$

and this term is negligible. Finally, it suffices to note that  $\|\nu_B\|_{\mathcal{B}_{\pi, \infty}^s(\mathcal{D})}$  is finite as soon as  $\|B\|_{\mathcal{B}_{\pi, \infty}^s(\mathcal{D})}$  is finite. This follows from

$$\nu_B(x) = \int_{\mathfrak{s}} \nu_B(y) Q_B(y, x) dy = \frac{B(2x)}{\tau x} \int_0^{2x} \nu_B(y) \exp\left(-\int_{y/2}^x \frac{B(2z)}{\tau z} dz\right) dy.$$

We conclude by applying Theorem 8.  $\square$

*Proof of Theorem 12, lower bound.* This is again a slight modification of the proof of Theorem 9, lower bound. For the dense case  $\epsilon_1 > 0$ , we consider an hypercube of the form

$$B_{\epsilon, j}(x) = B_0(x) + \gamma |\mathbb{T}_n|^{-1/2} \sum_{\lambda_j} \epsilon_k \psi_\lambda^1(x)$$

where  $\epsilon \in \{-1, 1\}^{\Lambda_j}$  with  $j$  such that  $|\mathbb{T}_n|^{-1/2} \lesssim 2^{-j(s+1/2)}$  and  $\gamma > 0$  a tuning parameter. By picking  $B_0$  and  $\gamma$  in an appropriate way, we have that  $B_0$  and  $B_{\epsilon, j}$  belong to a common ball in  $\mathcal{B}_{\pi, \infty}^s(\mathcal{D})$  and also belong to  $\mathcal{C}(r, L)$ . The associated  $\mathbb{T}$ -transition  $\mathcal{P}_{B_{\epsilon, j}}$  defined in (17) admits as mean transition

$$\mathcal{Q}_{B_{\epsilon, j}}(x, dy) = \frac{B_{\epsilon, j}(2y)}{\tau y} \exp\left(-\int_{x/2}^y \frac{B_{\epsilon, j}(2z)}{\tau z} dz\right) \mathbf{1}_{\{y \geq x/2\}} dy$$

which has a unique invariant measure  $\nu_{B_{\epsilon, j}}$ . Establishing (43) is similar to the proof of Theorem 9, lower bound, using the explicit representation for  $\mathcal{Q}_{B_{\epsilon, j}}$  with a slight modification due to the fact that the invariant measures  $\nu_{B_{\epsilon, j}}$  and  $\nu_{B_{T_\lambda(\epsilon), j}}$  do not necessarily coincide. We omit the details.

For the sparse case  $\epsilon_1 \leq 0$ , we consider the family

$$B_{\lambda, j}(x) = B_0(x) + \gamma \left(\frac{\log |\mathbb{T}_n|}{|\mathbb{T}_n|}\right)^{1/2} \epsilon_\lambda \psi_\lambda^1(x)$$

with  $\epsilon_\lambda \in \{-1, +1\}$ ,  $\lambda \in \Lambda_j$ , with  $j$  such that  $\left(\frac{\log |\mathbb{T}_n|}{|\mathbb{T}_n|}\right)^{1/2} \lesssim 2^{-j(s+1/2-1/\pi)}$ . The proof is then similar.  $\square$

## 6 APPENDIX

### 6.1 PROOF OF LEMMA 11

Pick  $B : \mathcal{S} \rightarrow [0, \infty)$  in the class  $\mathcal{C}(r, L)$  for well-chosen  $r$  and  $L$ , as specified in Lemma 11. Recall that  $\mathcal{S}$  is a bounded open interval such that  $\inf \mathcal{S} = 0$ , and set  $\mathcal{S} = (0, x_M)$  for a positive  $x_M$ . We shall use Theorem 1.3 in Hairer and Mattingly [HM11].

We first check Assumptions 1 and 2 of [HM11]. Let us define  $V(x) = x^p$  for  $x \in \mathcal{S}$  and  $p > 1$ . For any  $x \in \mathcal{S}$ , we have  $Q_B V(x) \leq V(x_M/2)$  since the size at birth of cell  $u$  is at most  $x_M/2$ , the size at division of its parents  $u^-$  being at most  $x_M$  for our choice of  $B$ . Thus Assumption 1 of [HM11] is satisfied with  $\gamma = 0$  and  $K = (x_M/2)^p$  (with the notation of [HM11], noting that picking  $\gamma = 0$  is actually possible). Let us denote

$$\varphi_B(y) = \frac{B(2y)}{\tau y} \exp\left(-\int_0^y \frac{B(2s)}{\tau s} ds\right), \quad y \in \mathcal{S}.$$

For any  $A \in \mathfrak{S}$  and  $x \in (0, r]$ ,

$$\begin{aligned} Q_B(x, A) &\geq \int_{A \cap (\frac{x}{2}, \frac{x_M}{2})} \frac{B(2y)}{\tau y} \exp\left(-\int_{x/2}^y \frac{B(2s)}{\tau s} ds\right) \mathbf{1}_{\{y \geq x/2\}} dy \\ &\geq \int_{A \cap (\frac{x}{2}, \frac{x_M}{2})} \varphi_B(y) dy, \end{aligned}$$

since the indicator is equal to one for  $y > r/2$  and  $x \leq r$ . We obtain

$$\inf_{x \in \mathcal{C}} Q_B(x, \cdot) \geq \alpha_B \eta_B(\cdot)$$

setting  $\mathcal{C} = (0, r]$ ,  $\alpha_B = \int_{(\frac{x}{2}, \frac{x_M}{2})} \varphi_B$  and  $\eta_B(A) = \alpha_B^{-1} \int_{A \cap (\frac{x}{2}, \frac{x_M}{2})} \varphi_B$  for any  $A \in \mathfrak{S}$ . Noting that  $\mathcal{C}$  can be written  $\{x \in \mathcal{S}; V(x) \leq R\}$  with some  $R > 2K/(1 - \gamma)$  as soon as  $r > 2^{1/p-1} x_M$ , which is true by picking  $p$  large enough, we deduce that Assumption 2 of [HM11] is satisfied.

The existence and uniqueness of an invariant probability measure  $\nu_B$  follows from Theorem 1.2 of [HM11]. Moreover  $\nu_B$  is absolutely continuous with respect to the Lebesgue measure since  $Q_B(x, dy)$  itself is (see (17)). Note that  $\alpha_B$  can be rewritten  $\alpha_B = \exp(-\int_0^{r/2} \frac{B(2s)}{\tau s} ds)$ , so that

$$\inf_{B \in \mathcal{C}(r, L)} \alpha_B \geq \exp\left(-\frac{L}{\tau}\right) > \frac{1}{2}.$$

For some  $\alpha_0 \in (0, 1/2)$  such that  $\inf_{B \in \mathcal{C}(r, L)} \alpha_B > 1/2 + \alpha_0$ , we set  $\beta = \alpha_0/K$ . We apply Theorem 1.3 of [HM11] recursively, picking  $Q_B^m \delta_x$  and  $\nu_B$  for  $\mu_1$  and  $\mu_2$ , with  $m \geq 0$  and  $x \in \mathcal{S}$ . We conclude that for any function  $g : \mathcal{S} \rightarrow \mathbb{R}$  such that  $|g(x)| \leq 1 + \beta x^p$ , we have

$$|Q_B^m g(x) - \nu_B(g)| \leq C_B \rho_B^m (1 + \beta x^p), \quad x \in \mathcal{S}, \quad m \geq 0, \quad (44)$$

with  $C_B = 1 + \int_{\mathcal{S}} (1 + \beta x^p) \nu_B(x) dx$  and  $\rho_B > 0$ . Note that  $\sup_{B \in \mathcal{C}(r, L)} C_B \leq 2 + \beta x_M^p$ . In addition,

Theorem 1.3 of [HM11] gives a precise control of  $\rho_B$  with respect to  $\gamma, K, R, \alpha_B, \alpha_0, \beta$  defined previously and one can check that  $\sup_{B \in \mathcal{C}(r,L)} \rho_B < 1/2$ . Finally, we apply (44) to the test function  $g/|g|_\infty$  and use  $x < x_M$  to get the announced result, uniformly over  $B \in \mathcal{C}(r, L)$ .

## 6.2 PROOF OF LEMMA 14

*The case  $r = 0$ .* By Assumption 3,

$$|\tilde{g}(X_{u0}) + \tilde{g}(X_{u1}) - 2\mathcal{Q}\tilde{g}(X_u)| \leq 2(|\tilde{g}|_\infty + R|\tilde{g}|_\infty\rho) \leq 4(1 + R\rho)|g|_\infty.$$

This proves the first estimate in the case  $r = 0$ . For  $u \in \mathbb{G}_{n-1}$ ,

$$\begin{aligned} & \mathbb{E}[(\tilde{g}(X_{u0}) + \tilde{g}(X_{u1}) - 2\mathcal{Q}\tilde{g}(X_u))^2 | \mathcal{F}_{n-1}] \\ &= \mathbb{E}[(g(X_{u0}) + g(X_{u1}) - 2\mathcal{Q}g(X_u))^2 | \mathcal{F}_{n-1}] \\ &\leq \mathbb{E}[(g(X_{u0}) + g(X_{u1}))^2 | \mathcal{F}_{n-1}] \leq 2(\mathcal{P}_0g^2(X_u) + \mathcal{P}_1g^2(X_u)) = 4\mathcal{Q}g^2(X_u) \end{aligned}$$

and for  $x \in \mathcal{S}$ , by Assumption 2,

$$\mathcal{Q}g^2(x) = \int_{\mathcal{S}} g(y)^2 \mathcal{Q}(x, y) \mathbf{n}(dy) \leq |\mathcal{Q}|_{\mathcal{D}} |g|_2^2$$

since  $g$  vanishes outside  $\mathcal{D}$ . Thus

$$\mathbb{E}[(\tilde{g}(X_{u0}) + \tilde{g}(X_{u1}) - 2\mathcal{Q}\tilde{g}(X_u))^2 | \mathcal{F}_{n-1}] \leq 4|\mathcal{Q}|_{\mathcal{D}} |g|_2^2 \quad (45)$$

hence the result for  $r = 0$ .

*The case  $r \geq 1$ .* On the one hand, by Assumption 3,

$$\begin{aligned} |2^r (\mathcal{Q}^r \tilde{g}(X_{u0}) + \mathcal{Q}^r \tilde{g}(X_{u1}) - 2\mathcal{Q}^{r+1} \tilde{g}(X_u))| &\leq 2^r (2R|\tilde{g}|_\infty (\rho^r + \rho^{r+1})) \\ &\leq 4R(1 + \rho)|g|_\infty (2\rho)^r. \end{aligned} \quad (46)$$

On the other hand, since

$$|\mathcal{Q}g(x)| \leq \int_{\mathcal{S}} |g(y)| \mathcal{Q}(x, y) \mathbf{n}(dy) \leq |\mathcal{Q}|_{\mathcal{D}} |g|_1,$$

we also have

$$\begin{aligned} & 2^r |\mathcal{Q}^r \tilde{g}(X_{u0}) + \mathcal{Q}^r \tilde{g}(X_{u1}) - 2\mathcal{Q}^{r+1} \tilde{g}(X_u)| \\ &= 2^r |\mathcal{Q}^r g(X_{u0}) + \mathcal{Q}^r g(X_{u1}) - 2\mathcal{Q}^{r+1} g(X_u)| \leq 2^r 4|\mathcal{Q}|_{\mathcal{D}} |g|_1. \end{aligned} \quad (47)$$

Putting together these two estimates yields the result for the case  $r \geq 1$ .

### 6.3 PROOF OF LEMMA 15

By Assumption 3,

$$|\Upsilon_r(X_u, X_{u0}, X_{u1})| \leq 2 \sum_{m=0}^r 2^m R |\tilde{g}|_\infty \rho^m (1 + \rho) \leq 4R |g|_\infty (1 + \rho) (1 - 2\rho)^{-1}$$

since  $\rho < 1/2$ . This proves the first bound. For the second bound we balance the estimates (46) and (47) obtained in the proof of Lemma 14. Let  $\ell \geq 1$ . For  $u \in \mathbb{G}_{n-r-1}$ , we have

$$|\Upsilon_r(X_u, X_{u0}, X_{u1})| \leq I + II + III,$$

with

$$\begin{aligned} I &= |\tilde{g}(X_{u0}) + \tilde{g}(X_{u1}) - \mathcal{Q}\tilde{g}(X_u)|, \\ II &= \sum_{m=1}^{\ell \wedge r} 2^m |\mathcal{Q}^m \tilde{g}(X_{u0}) + \mathcal{Q}^m \tilde{g}(X_{u1}) - 2\mathcal{Q}^{m+1} \tilde{g}(X_u)|, \\ III &= \sum_{m=\ell \wedge r + 1}^r 2^m |\mathcal{Q}^m \tilde{g}(X_{u0}) + \mathcal{Q}^m \tilde{g}(X_{u1}) - 2\mathcal{Q}^{m+1} \tilde{g}(X_u)|, \end{aligned}$$

with  $III = 0$  if  $\ell > r$ . For  $u \in \mathbb{G}_{n-r-1}$ , by (45), we successively have

$$\mathbb{E}[I^2 | \mathcal{F}_{n-r-1}] \leq 4|\mathcal{Q}|_{\mathcal{D}} |g|_2^2,$$

$$II \leq 4|\mathcal{Q}|_{\mathcal{D}} |g|_1 \sum_{m=1}^{\ell \wedge r} 2^m \leq 8|\mathcal{Q}|_{\mathcal{D}} |g|_1 2^{\ell \wedge r}$$

by (47), while for  $\ell \leq r$ ,

$$III \leq 4R(1 + \rho) |g|_\infty \sum_{m=\ell+1}^r (2\rho)^m \leq 4R(1 + \rho) (1 - 2\rho)^{-1} |g|_\infty (2\rho)^{\ell+1}$$

by (46). The result follows.

---

## Références du Chapitre II

---

- [AK98] K. B. Athreya and H.-J. Kang. *Some limit theorems for positive recurrent branching Markov chains: I*. Advances in Applied Probability, 30 (1998), 693–710.
- [BH99] I. V. Basawa and R. M. Huggins. *Extensions of the bifurcating autoregressive model for cell lineage studies*. Journal of Applied Probability, 36 (1999), 1225–1233.
- [BH00] I. V. Basawa and R. M. Huggins. *Inference for the extended bifurcating autoregressive model for cell lineage studies*. Australian & New Zealand Journal of Statistics, 42 (2000), 423–432.
- [BZ04] I. V. Basawa and J. Zhou. *Non-Gaussian bifurcating models and quasi-likelihood estimation*. Journal of Applied Probability, 41 (2004), 55–64.
- [BZ05] I. V. Basawa and J. Zhou. *Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors*. Journal of Time Series Analysis, 26 (2005), 825–842.
- [BP94] I. Benjamini and Y. Peres. *Markov chains indexed by trees*. The Annals of Probability (1994), 219–243.
- [BB15] B. Bercu and V. Blandin. *A Rademacher-Menchov approach for random coefficient bifurcating autoregressive processes*. Stochastic Processes and their Applications, 125 (2015), 1218–1243.
- [BdSGP09] B. Bercu, B. de Saporta and A. Gégout-Petit. *Asymptotic analysis for bifurcating autoregressive processes via a martingale approach*. Electronic Journal of Probability, 87 (2009), 2492–2526.
- [Ber06] J. Bertoin. *Random fragmentation and coagulation processes*. Cambridge University Press, 2006.
- [BD14] S. V. Bitseki Penda and H. Djellout. *Deviation inequalities and moderate deviations for estimators of parameters in bifurcating autoregressive models*. Annales de l’IHP-PS, 50 (2014), 806–844.
- [BD14] S. V. Bitseki Penda and H. Djellout. *Deviation inequalities and moderate deviations for estimators of parameters in bifurcating autoregressive models*. Annales de l’IHP-PS, 50 (2014), 806–844.
- [BDG14] S. V. Bitseki Penda, H. Djellout and A. Guillin. *Deviation inequalities, moderate deviations and some limit theorems for bifurcating Markov chains with application*. The Annals of Applied Probability, 24 (2014), 235–291.
- [BEG15] S. V. Bitseki Penda, M. Escobar-Bach and A. Guillin. *Transportation cost-information and concentration inequalities for bifurcating Markov chains*. *arXiv:1501.06693*.
- [Bla14] V. Blandin. *Asymptotic results for random coefficient bifurcating autoregressive processes*. Statistics 48 (2014), 1202–1232.
- [Clé00] S. Cléménçon. *Adaptive Estimation of the Transition Density of a Regular Markov Chain by Wavelet Methods*. Mathematical Methods of Statistics, 9 (2000), 323–357.
- [CS86] R. Cowan and R. G. Staudte. *The bifurcating autoregressive model in cell lineage studies*. Biometrics, 42 (1986), 769–783.
- [Coh00] A. Cohen. *Wavelets in Numerical Analysis*. In: Ciarlet, P.G., Lions, J.L. (eds.) Handbook of Numerical Analysis, vol. VII. Elsevier, Amsterdam, 2000.
- [DM10] J.-F. Delmas and L. Marsalle. *Detection of cellular aging in Galton-Watson process*. Stochastic Processes and their Applications, 12 (2010), 2495–2519.

- [dSGPM11] B. de Saporta, A. Gégout-Petit and L. Marsalle. *Parameters estimation for asymmetric bifurcating autoregressive processes with missing data*. *Electronic Journal of Statistics*, 5 (2011), 1313–1353.
- [dSGPM12] B. de Saporta, A. Gégout-Petit and L. Marsalle. *Asymmetry tests for Bifurcating Autoregressive Processes with missing data*. *Statistics & Probability Letters*, 82 (2012), 1439–1444.
- [dSGPM14] B. de Saporta, A. Gégout-Petit and L. Marsalle. *Random coefficients bifurcating autoregressive processes*. *ESAIM PS*, 18 (2014), 365–399.
- [DKT98] R. DeVore, S.V. Konyagin and V.N. Temlyakov. *Hyperbolic Wavelet Approximation. Constructive Approximation*. Springer, New York, 1998.
- [DJ94] D.L. Donoho, I.M. Johnstone. *Ideal spatial adaptation via wavelet shrinkage*. *Biometrika*, 81 (1994), 425–455.
- [DJ95] D.L. Donoho, I.M. Johnstone. *Adapting to unknown smoothness via wavelet shrinkage*. *Journal of the American Statistical Association*, 90 (1995), 1200–1224.
- [DJKP95] D.L. Donoho, I.M. Johnstone, G. Kerkycharian and D. Picard. *Wavelet shrinkage: Asymptopia?* *Journal of the Royal Statistical Society*, 57 (1995), 301–369.
- [DJKP96] D.L. Donoho, I.M. Johnstone, G. Kerkycharian and D. Picard. *Density estimation by wavelet thresholding*. *The Annals of Statistics*, 54 (1996), 508–539.
- [DHKR15] M. Doumic, M. Hoffmann, N. Krell and L. Robert. *Statistical estimation of a growth-fragmentation model observed on a genealogical tree*. *Bernoulli*, 21 (2015), 1760–1799.
- [GGW14] F. Gao, A. Guillin and L. Wu. *Bernstein-type concentration inequalities for symmetric Markov processes*. *Theory of Probability and its Applications*, 58 (2014), 358–382.
- [Guy07] J. Guyon. *Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging*. *The Annals of Applied Probability*, 17 (2007), 1538–1569.
- [HM11] M. Hairer and J. Mattingly. *Yet Another Look at Harris’ Ergodic Theorem for Markov Chains*. *Seminar on Stochastic Analysis, Random Fields and Applications VI, Progress in Probability*, 63 (2011), 109–117.
- [HKPT98] W. Härdle, G. Kerkycharian, D. Picard and A. Tsybakov. *Wavelets, Approximation and Statistical Applications*. *Lecture Notes in Statistics*, Springer, 1998.
- [Hof99] M. Hoffmann. *Adaptive estimation in diffusion processes*. *Stochastic Processes and their Applications*, 79 (1999), 135–163.
- [KP00] G. Kerkycharian and D. Picard. *Thresholding algorithms, maxisets and well-concentrated bases*. *Test*, 9 (2000), 283–344.
- [Lac07] C. Lacour. *Adaptive estimation of the transition density of a Markov chain* *Annales de l’Institut Henri Poincaré, Probabilités et Statistique*, 43 (2007), 571–597.
- [Lac08] C. Lacour. *Nonparametric estimation of the stationary density and the transition density of a Markov chain* *Stochastic Process and their Applications* 118 (2008), 232–260.
- [Mey90] Y. Meyer. *Ondelettes et Opérateurs*, vol. 1. Hermann, Paris, 1990.
- [MT93] S. Meyn and R. Tweedie. *Markov chains and stochastic stability*. Springer Berlin Heidelberg, 1993.
- [Per07] B. Perthame. *Transport Equations in Biology*. *Frontiers in Mathematics*. Basel: Birkhäuser, 2007.
- [RHK+14] L. Robert, M. Hoffmann, N. Krell, S. Aymerich, J. Robert and M. Doumic. *Division control in Escherichia coli is based on a size-sensing rather than a timing mechanism*. *BMC Biology*, 02/2014 12(1): 17.
- [Tak01] C. Takacs. *Strong law of large numbers for branching Markov chains*. *Markov Processes and Related Fields*, 8 (2001), 107–116.

## CHAPITRE III

---

### Processus bifurquant auto-régressif non linéaire : estimation non-paramétrique des fonctions auto-régressives

---

**L'**OBJECTIF (B) *Étudier la transmission du trait individuel d'une cellule à ses descendants* (voir page 2) est abordé à travers une modélisation spécifique : la transmission du trait biologique d'intérêt de la cellule mère à ses deux cellules filles est ici spécifiée comme étant auto-régressive. Autrement dit, une généralisation sur un arbre d'un processus auto-régressif d'ordre 1 est étudiée. Le processus discret ainsi formé est nommé processus *bifurquant* auto-régressif.

Ce processus entre dans la classe des chaînes de Markov bifurquantes que nous venons d'étudier dans le chapitre précédent. Les résultats du Chapitre II s'appliquent ici et nous sommes d'ores et déjà en mesure de reconstruire, de façon non-paramétrique, la  $\mathbb{T}$ -transition, ainsi que la transition moyenne et la mesure invariante associées à une branche typique. Contrairement au cas des processus de croissance-fragmentation, l'accès à ces caractéristiques du processus ne permet pas d'en déduire les paramètres naturels du modèle que sont les fonctions auto-régressives.

Le but est dans ce chapitre de reconstruire les deux fonctions auto-régressives caractérisant le processus (la loi du bruit étant vue comme un paramètre de nuisance). L'observation est généalogique : les caractéristiques des cellules sont collectées depuis la cellule initiale jusqu'aux cellules de la génération  $n$  fixée. Nous ne spécifions pas *a priori* la forme de ces fonctions auto-régressives, et venons donc à ce titre prolonger la littérature existante sur les processus bifurquants auto-régressifs, traitant de liens linéaires.



## Abstract

---

Bifurcating autoregressive processes, which can be seen as an adaptation of autoregressive processes for a binary tree structure, have been extensively studied during the last decade in a parametric context. In this work we do not specify any *a priori* form for the two autoregressive functions and we use nonparametric techniques. We investigate both nonasymptotic and asymptotic behavior of the Nadaraya-Watson type estimators of the autoregressive functions. We build our estimators observing the process on a finite subtree denoted by  $\mathbb{T}_n$ , up to the depth  $n$ . Estimators achieve the classical rate  $|\mathbb{T}_n|^{-\beta/(2\beta+1)}$  in quadratic loss over Hölder classes of smoothness. We prove almost sure convergence, asymptotic normality giving the bias expression when choosing the optimal bandwidth and a moderate deviations principle. Our proofs rely on specific techniques used to study bifurcating Markov chains. Finally, we address the question of asymmetry and develop an asymptotic test for the equality of the two autoregressive functions.

---

## Contents

---

|          |   |            |
|----------|---|------------|
| <b>1</b> | <b>INTRODUCTION</b>                                       | <b>126</b> |
| 1.1      | A generalisation of the bifurcating autoregressive model  | 126        |
| 1.2      | Estimation of the autoregressive functions and objectives | 127        |
| <b>2</b> | <b>NON-ASYMPTOTIC BEHAVIOUR</b>                           | <b>128</b> |
| 2.1      | Tagged-branch chain                                       | 128        |
| 2.2      | Model constraints   | 130        |
| 2.3      | Main results  | 131        |
| <b>3</b> | <b>ASYMPTOTIC BEHAVIOUR</b>                               | <b>133</b> |
| 3.1      | Main results  | 133        |
| 3.2      | Construction of an asymmetry test                         | 135        |
| <b>4</b> | <b>DISCUSSION</b>   | <b>137</b> |
| <b>5</b> | <b>NUMERICAL IMPLEMENTATION</b>                           | <b>138</b> |
| <b>6</b> | <b>PROOFS</b>   | <b>142</b> |
| 6.1      | Preliminary   | 143        |
| 6.2      | Estimation of the density of the invariant measure        | 144        |
| 6.3      | Proof of Theorem 5  | 147        |
| 6.4      | Proof of Theorem 6  | 149        |
| 6.5      | Proof of Proposition 7                                    | 151        |
| 6.6      | Proofs of Theorems 8 and 9                                | 151        |
| 6.7      | Proof of Theorem 13                                       | 155        |
| 6.8      | Proof of Theorem 15                                       | 159        |
| <b>7</b> | <b>APPENDIX</b>   | <b>161</b> |
| 7.1      | Proof of Lemma 16   | 161        |
| 7.2      | Proof of Lemma 19   | 162        |
| 7.3      | Proof of Lemma 21   | 162        |
| 7.4      | Proof of Lemma 22   | 163        |
|          | <b>REFERENCES</b>   | <b>164</b> |

---

## 1 INTRODUCTION

### 1.1 A GENERALISATION OF THE BIFURCATING AUTOREGRESSIVE MODEL

#### *BAR process*

Roughly speaking, bifurcating autoregressive processes, BAR processes for short, are an adaptation of autoregressive processes when the index set have a binary tree structure. BAR processes were introduced by Cowan and Staudte [CS86] in 1986 in order to study cell division in *Escherichia coli* bacteria. Let  $\mathbb{T}$  be the infinite binary tree indexed in the following way

$$\mathbb{T} = \bigcup_{m=0}^{\infty} \{0, 1\}^m.$$

We can see the nodes of  $\mathbb{T}$  as representing a given population of cells, where the initial individual is denoted by  $\emptyset$ . We suppose that each individual  $u \in \mathbb{T}$  gives birth to two individuals  $u0 = (u, 0)$  and  $u1 = (u, 1)$ , called respectively type 0 pole and type 1 pole. In [CS86], the original BAR process is defined as follows. If  $X_u$  denotes the observation of some feature of cell  $u$ , then the first order BAR process is given, for all  $u \in \mathbb{T}$ , by

$$\begin{cases} X_{u0} = a + bX_u + \varepsilon_{u0}, \\ X_{u1} = a + bX_u + \varepsilon_{u1}, \end{cases}$$

drawing the initial value  $X_\emptyset$  according to a Gaussian law, where  $a$  and  $b$  are unknown real parameters, with  $|b| < 1$ , which measure heredity in the transmission of the biological feature. The noise sequence  $((\varepsilon_{u0}, \varepsilon_{u1}), u \in \mathbb{T})$  forms a sequence of independent and identically distributed bivariate centered Gaussian random variables and represents environmental effects.

Since then, several extensions of this model have been studied and various estimators for unknown parameters have been proposed. First, one can mention [Guy07]. There Guyon introduces asymmetry to take into account the fact that autoregressive parameters for poles of type 0 or 1 can differ. Introducing the bifurcating Markov chain theory, Guyon studies the asymptotic behaviour of the least squares estimators of the unknown parameters. He also introduces some asymptotic tests which allow to decide if the model is symmetric or not. Several extensions of this linear model have been proposed and studied from a parametric point of view, see for instance Basawa and Huggins [BH99, BH00] and Basawa and Zhou [BZ04, BZ05] where the BAR process is studied for non-Gaussian noise and long memory. Next, around 2010, Bercu, Blandin, Delmas, de Saporta, Gégout-Petit and Marsalle extended in different directions the study of the BAR process. Bercu *et al.* [BdSGP09] use martingale approach in order to study least squares estimators of unknown parameters for processes with memory greater than 1 and without normality assumption on the noise sequence. Even more recently, [dSGPM11, dSGPM12] takes into account missing data and [BB13, Bla13, dSGPM14] study the model with random coefficients. A number of other extensions were also surveyed, one can cite Delmas and Marsalle [DM10] for a generalisation to Galton-Watson trees and Bitseki Penda and Djellout for deviations inequalities and moderate deviations. However, to our best knowledge, there is no extensive nonparametric study for nonlinear bifurcating autoregressive processes, where there is no *a priori* specification on the two autoregressive functions. We

can mention the applications of [BEG15] (section 4) where deviations inequalities are derived for Nadaraya-Watson type estimators of the autoregressive functions.

### Nonlinear BAR process

Nonlinear bifurcating autoregressive (NBAR, for short) processes generalize bifurcating autoregressive (BAR, for short) processes. Let us now give a precise definition for NBAR processes. We now work on some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . For an individual  $u \in \mathbb{T}$ , we are interested in some feature  $X_u$  (it may be its size, its growth rate, its age, ...) with values in a generic state space  $\mathcal{S}$  (e.g. the real line) endowed with its Borel  $\sigma$ -field  $\mathfrak{S}$ . We suppose that the features  $X_{u0}$  and  $X_{u1}$  associated to its offspring, denoted by  $u0$  and  $u1$ , depend on the past only through  $X_u$ , via an autoregressive system of equations. In this setting we have the following

**Definition 1** (NBAR process). *Let  $\mu$  be a probability on  $(\mathcal{S}, \mathfrak{S})$ ,  $G$  be a probability on  $(\mathcal{S} \times \mathcal{S}, \mathfrak{S})$  with a null first order moment and  $f_0, f_1 : \mathcal{S} \rightarrow \mathcal{S}$  be two functions. We say that  $X = (X_u, u \in \mathbb{T})$  is a nonlinear bifurcating autoregressive process (NBAR process for short) with initial distribution  $\mu$  and autoregressive functions  $f_0$  and  $f_1$  if*

$$(i) \quad X_\emptyset \sim \mu,$$

(ii) For any  $u \in \mathbb{T}$ ,

$$X_{u0} = f_0(X_u) + \varepsilon_{u0} \quad \text{and} \quad X_{u1} = f_1(X_u) + \varepsilon_{u1}$$

where  $((\varepsilon_{u0}, \varepsilon_{u1}), u \in \mathbb{T})$  is a sequence of independent and identically distributed centered bivariate random variables with common density  $G$ , independent of  $X_\emptyset$ .

We denote by  $\mathbb{P}_\mu$  the law of the process  $(X_u, u \in \mathbb{T})$  built in such a way and by  $\mathbb{E}_\mu[\cdot]$  the expectation with respect to the probability  $\mathbb{P}_\mu$ .

For the sake of simplicity, along the present study, we pick  $\mathcal{S} = \mathbb{R}$  and we assume that the noise density  $G$  is continuous with respect to the Lebesgue measure, with continuous density on  $\mathcal{S}$  (still denoted by  $G$ ).

## 1.2 ESTIMATION OF THE AUTOREGRESSIVE FUNCTIONS AND OBJECTIVES

### Autoregressive functions estimators

Our aim is to estimate the unknown autoregressive functions  $f_0$  and  $f_1$  in Definition 1 from the observation of a subpopulation. For that purpose, we propose to make use of a Nadaraya-Watson type estimator. Two seminal articles in nonparametric regression function estimation were written in 1964 independently by Nadaraya [Nad64] and Watson [Wat64]. They introduced a kernel estimator, now called Nadaraya-Watson estimator, for the regression function  $\mathbb{E}[Y|X = x]$  when observing  $n$  pairs  $(X_1, Y_1), \dots, (X_n, Y_n)$  of independent random variables distributed as  $(X, Y)$ . The Nadaraya-Watson estimator was also used in the framework of autoregressive time series. In that case the aim is to reconstruct  $\mathbb{E}[X_n|X_{n-1} = x]$ , assuming that  $(X_n, n \geq 0)$  is stationary, see [Rob83, HV92]. We generalize here the use of the Nadaraya-Watson estimator to the case of an autoregressive process indexed by a binary tree.

Let  $\mathcal{D} \subset \mathcal{S}$  be a compact interval. We propose to estimate  $(f_0(x), f_1(x))$  the autoregressive functions at point  $x \in \mathcal{D}$  from the observations

$$(X_u, u \in \mathbb{T}_{n+1})$$

by

$$\left( \widehat{f}_{\iota, n}(x) = \frac{|\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) X_{uu}}{\left( |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) \right) \vee \varpi_n}, \iota \in \{0, 1\} \right), \quad (1)$$

where  $\varpi_n > 0$  and we set  $K_{h_n}(\cdot) = h_n^{-1} K(h_n^{-1} \cdot)$  for  $h_n > 0$  and a kernel function  $K : \mathbb{R} \rightarrow \mathbb{R}$  such that  $\int_{\mathbb{R}} K = 1$ .

### Objectives

Our first objective in this work is to study the estimators (1) both from non-asymptotic and asymptotic points of view. The non-asymptotic study includes the control of the quadratic loss in a minimax sense (Theorems 5 and 6) and the asymptotic study includes almost sure convergence (Proposition 7), asymptotic normality (Theorems 8 and 9) and a moderate deviations principle (Theorem 13). To this end, we shall make use of non-asymptotic behaviour for bifurcating Markov chains (see [Guy07, BDG14]) and asymptotic behaviour of martingales. We are also interested in comparing the two autoregressive functions  $f_0$  and  $f_1$  and to test whether the phenomenon studied is symmetric or not. The test we build to do so relies on the asymptotic study.

The present work is organised as follows. The results are obtained under the assumption of geometric ergodicity of the so-called tagged-branch Markov chain we define in Section 2, together with the non-asymptotic results. In Section 3, we state asymptotic results for our estimators which enable us to address the question of asymmetry and build a test to compare  $f_0$  and  $f_1$ . After discussion in Section 4, we also give there some numerical results to illustrate the estimation and test strategies (Section 5). The last part of the article, Section 6 with an appendix, is devoted to the proofs of our results.

## 2 NON-ASYMPTOTIC BEHAVIOUR

### 2.1 TAGGED-BRANCH CHAIN

Let us first introduce some usual notation for binary trees. We denote by  $\mathbb{G}_m$  the set of individuals belonging to the  $m$ -th generation ( $m \geq 0$ ),  $\mathbb{T}_r$  the set of individuals up to the  $r$ -th generation ( $r \geq 0$ ) and  $\mathbb{T}$  the infinite binary tree which represent the whole population:

$$\mathbb{G}_m = \{0, 1\}^m, \quad \mathbb{T}_r = \bigcup_{m=0}^r \mathbb{G}_m, \quad \mathbb{T} = \bigcup_{m=0}^{\infty} \mathbb{G}_m,$$

with the convention  $\mathbb{G}_0 = \{\emptyset\}$  where  $\emptyset$  is called the initial individual. The cardinality of the set  $\mathbb{G}_m$ , *e.g.*, is denoted by  $|\mathbb{G}_m|$ . We also use throughout the following standard notation: for  $u \in \mathbb{G}_m$ , its length (the generation to which it belongs) is  $|u| = m$  and we write  $u0 = (u, 0) \in \mathbb{G}_{m+1}$  and

$u1 = (u, 1) \in \mathbb{G}_{m+1}$  for its children.

The main tools in the study of the previous estimators come from the following property. A NBAR process  $(X_u, u \in \mathbb{T})$  as defined above is a bifurcating Markov chain on  $\mathbb{T}$  (see Chapter II, Definition 1) with

$$\mathcal{P}(x, dydz) = G(y - f_0(x), z - f_1(x))dydz \quad (2)$$

as  $\mathbb{T}$ -transition probability, defined on  $\mathcal{S} \times \mathfrak{S}^2$  (recall that  $G$  is the noise probability distribution). For a  $\mathfrak{S}$ -measurable function  $f : \mathcal{S}^3 \rightarrow \mathcal{S}$ , we denote by  $\mathcal{P}f$  the  $\mathfrak{S}$ -measurable function

$$x \in \mathcal{S} \rightsquigarrow \mathcal{P}f(x) = \int_{\mathfrak{S}^2} f(x, y, z) \mathcal{P}(x, dydz),$$

when it is defined. Bifurcating Markov chains are studied by Guyon in [Guy07], it allows the case where the two offspring are not conditionally independent as we can see in the Definition 1 of Chapter II. We denote by  $\mathcal{P}_0$  and  $\mathcal{P}_1$  respectively the first and the second marginal of  $\mathcal{P}$ , that is

$$\mathcal{P}_0(x, A) = \mathcal{P}(x, A \times \mathcal{S}), \quad \mathcal{P}_1(x, A) = \mathcal{P}(x, \mathcal{S} \times A)$$

for all  $x \in \mathcal{S}$  and  $A \in \mathfrak{S}$ . We denote by  $\mathcal{Q}$  the mean of  $\mathcal{P}_0$  and  $\mathcal{P}_1$ ,

$$\mathcal{Q} = (\mathcal{P}_0 + \mathcal{P}_1) / 2. \quad (3)$$

From (2), the explicit expression of  $\mathcal{Q}$  is

$$\mathcal{Q}(x, dy) = \frac{1}{2} \left( G_0(y - f_0(x)) + G_1(y - f_1(x)) \right) dy, \quad (4)$$

where

$$G_0(\cdot) = \int_{\mathcal{S}} G(\cdot, y) dy \quad \text{and} \quad G_1(\cdot) = \int_{\mathcal{S}} G(x, \cdot) dx \quad (5)$$

are the marginals of  $G$ .

Let  $Y = (Y_m)_{m \geq 0}$  be the Markov chain on  $\mathcal{S}$  with initial value  $Y_0 = X_\emptyset$  and transition probability  $\mathcal{Q}$ . The chain  $(Y_m)_{m \geq 0}$  corresponds to a lineage taken randomly (uniformly at each branching event) in the population. In the specific case of NBAR processes, the tagged-branch chain  $Y$  can be constructed as follows. Let  $(\iota_m)_{m \geq 1}$  be a sequence of independent and identically distributed random variables that have the Bernoulli distribution of parameter 1/2. Let  $(\varepsilon'_m)_{m \geq 1}$  be a sequence of independent random variables such that  $\varepsilon'_m$  given  $\iota_m = \iota$  has density  $G_\iota$  (for  $\iota \in \{0, 1\}$ ). In addition  $\iota_m$  and  $\varepsilon'_m$  are independent of  $(Y_k)_{0 \leq k \leq m-1}$ . We set

$$\begin{cases} Y_0 = X_\emptyset, \\ Y_m = f_{\iota_m}(Y_{m-1}) + \varepsilon'_m, \quad m \geq 1. \end{cases} \quad (6)$$

Then  $\mathbb{P}(Y_1 \in dy | Y_0 = x) = \mathcal{Q}(x, dy)$ . More generally, for all  $m \geq 1$ , we have  $\mathbb{P}(Y_m \in dy | Y_0 = x) = \mathcal{Q}^m(x, dy)$  where  $\mathcal{Q}^m$  denotes the  $m$ -th iterated of  $\mathcal{Q}$  recursively defined by the formulae

$$\mathcal{Q}^0(x, A) = \delta_x(A) \quad \text{and} \quad \mathcal{Q}^{m+1}(x, A) = \int_{\mathcal{S}} \mathcal{Q}(x, dy) \mathcal{Q}^m(y, A) \quad \forall A \in \mathfrak{S}.$$

It is well known that  $Q^m$  is a transition probability on  $\mathcal{S} \times \mathfrak{S}$ .

Asymptotic and non-asymptotic studies on bifurcating Markov chains have shown that the limit law of the Markov chain  $Y$  plays an important role, we refer to Bitseki Penda, Djellout and Guillin [BDG14] and references therein for more details. In the present work, the estimator of the invariant density of the tagged-branch Markov chain will play a crucial role in the analysis of the estimators defined by (1). More precisely, we will see that the denominator converges almost surely to the invariant density of the Markov chain  $Y$  (Proposition 20).

## 2.2 MODEL CONSTRAINTS

The autoregressive functions  $f_0$  and  $f_1$  are devoted to belong to the following class. For  $\gamma \in (0, 1)$  and  $\ell > 0$ , we introduce the class  $\mathcal{F}(\gamma, \ell)$  of continuous functions  $f : \mathcal{S} \rightarrow \mathcal{S}$  such that

$$|f(x)| \leq \gamma|x| + \ell$$

for any  $x \in \mathcal{S}$ .

The two marginal  $G_0$  and  $G_1$  of the noise density are devoted to belong to the following class. For  $r > 0$  and  $\lambda > 2$ , we introduce the class  $\mathcal{G}(r, \lambda)$  of nonnegative continuous functions  $g : \mathcal{S} \rightarrow [0, \infty)$  such that

$$g(x) \leq \frac{r}{1 + |x|^\lambda}$$

for any  $x \in \mathcal{S}$ . When  $(G_0, G_1) \in \mathcal{G}(r, \lambda)^2$  for some  $\lambda > 3$ , we denote the covariance matrix of  $(\varepsilon_0, \varepsilon_1)$ , called noise covariance matrix, by

$$\Gamma = \begin{pmatrix} \sigma_0^2 & \varrho\sigma_0\sigma_1 \\ \varrho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix} \quad (7)$$

with  $\sigma_0, \sigma_1 > 0$  and  $\varrho \in (-1, 1)$ .

It is crucial for our proofs to study the ergodicity of the tagged-branch Markov chain  $Y$ . Geometric ergodicity of nonlinear autoregressive processes has been studied in Bhattacharya and Lee [BL95] (see Theorem 1) and also in An and Huang [AH96] and Cline [Cli07]. The main difference is that we need here a precise control on the ergodicity rate, which should be smaller than  $1/2$ , due to the binary tree structure. We also see, through (6), that the autoregressive function is random in our case. The following crucial assumption will guarantee geometric ergodicity of the tagged-branch Markov chain  $Y$  with an exponential decay rate smaller than  $1/2$  (see Lemma 17). For any  $M > 0$ , we set

$$\delta(M) = \min \left\{ \inf_{|x| \leq M} G_0(x); \inf_{|x| \leq M} G_1(x) \right\}. \quad (8)$$

**Assumption 2.** Set  $M_0 = \ell + \mathbb{E}[|\varepsilon'_1|] < \infty$  where  $\varepsilon'_1$  has density  $(G_0 + G_1)/2$ . There exists  $\eta > 0$  such that

$$\gamma < \frac{1}{2} - \eta$$

and there exists  $M_1 > 2M_0/(1/2 - \eta - \gamma)$  such that

$$2M_1\delta((1 + \gamma)M_1 + \ell) > \frac{1}{2}. \quad (9)$$

The following assumption will guarantee that the invariant density  $\nu$  is positive on some nonempty interval (see Lemma 21). For any  $M > 0$ , we set

$$\eta(M) = \frac{|G_0|_\infty + |G_1|_\infty}{2} \int_{|y|>M} \int_{x \in \mathcal{S}} \frac{r}{1 + |y - \gamma|x| - \ell|^\lambda \wedge |y + \gamma|x| + \ell|^\lambda} dx dy,$$

where for a function  $h : \mathcal{S} \rightarrow \mathbb{R}$ ,  $|h|_\infty$  stands for  $\sup_{x \in \mathcal{S}} |h(x)|$ .

**Assumption 3.** For  $M_2 > 0$  such that  $\eta(M_2) < 1$ , there exists  $M_3 > \ell + \gamma M_2$  such that  $\delta(M_3) > 0$ .

### 2.3 MAIN RESULTS

We need the following property on  $K$ :

**Assumption 4.** The kernel  $K : \mathbb{R} \rightarrow \mathbb{R}$  is bounded with compact support and for some integer  $n_0 \geq 1$ , we have  $\int_{-\infty}^{\infty} x^k K(x) dx = \mathbf{1}_{\{k=0\}}$  for  $k = 1, \dots, n_0$ .

Assumption 4 will enable us to have nice approximation results over smooth functions  $f_0$  and  $f_1$ , described in the following way: for  $\mathcal{X} \subseteq \mathbb{R}$  and  $\beta > 0$ , with  $\beta = \lfloor \beta \rfloor + \{\beta\}$ ,  $0 < \{\beta\} \leq 1$  and  $\lfloor \beta \rfloor$  an integer, let  $\mathcal{H}_\mathcal{X}^\beta$  denote the Hölder space of functions  $h : \mathcal{X} \rightarrow \mathbb{R}$  possessing a derivative of order  $\lfloor \beta \rfloor$  that satisfies

$$|h^{\lfloor \beta \rfloor}(y) - h^{\lfloor \beta \rfloor}(x)| \leq c(h)|x - y|^{\{\beta\}}. \quad (10)$$

The minimal constant  $c(h)$  such that (10) holds defines a semi-norm  $|g|_{\mathcal{H}_\mathcal{X}^\beta}$ . We equip the space  $\mathcal{H}_\mathcal{X}^\beta$  with the norm  $\|h\|_{\mathcal{H}_\mathcal{X}^\beta} = \sup_x |h(x)| + |h|_{\mathcal{H}_\mathcal{X}^\beta}$  and the balls

$$\mathcal{H}_\mathcal{X}^\beta(L) = \{h : \mathcal{X} \rightarrow \mathbb{R}, \|h\|_{\mathcal{H}_\mathcal{X}^\beta} \leq L\}, L > 0.$$

**Theorem 5** (Upper rate of convergence). Let  $\gamma \in (0, 1/2)$  and  $\ell > 0$ , let  $r > 0$  and  $\lambda > 3$ . Specify  $(\widehat{f}_{0,n}, \widehat{f}_{1,n})$  with a kernel  $K$  satisfying Assumption 4 for some  $n_0 > 0$ , with

$$h_n = |\mathbb{T}_n|^{-1/(2\beta+1)}$$

and  $\varpi_n > 0$  such that  $\varpi_n \rightarrow 0$  as  $n \rightarrow \infty$ . For every  $L, L' > 0$  and  $0 < \beta < n_0$ , for every  $G$  such that  $(G_0, G_1) \in (\mathcal{G}(r, \lambda) \cap \mathcal{H}_\mathcal{S}^\beta(L'))^2$  satisfy Assumptions 2 and 3, there exists  $d = d(\gamma, \ell, G_0, G_1) > 0$  such that for every compact interval  $\mathcal{D} \subset [-d, d]$  with non-empty interior and for every  $x$  in the interior of  $\mathcal{D}$ ,

$$\sup_{(f_0, f_1)} \mathbb{E}_\mu \left[ (\widehat{f}_{0,n}(x) - f_0(x))^2 + (\widehat{f}_{1,n}(x) - f_1(x))^2 \right]^{1/2} \lesssim \varpi_n^{-1} |\mathbb{T}_n|^{-\beta/(2\beta+1)}$$

where the supremum is taken among all functions  $(f_0, f_1) \in (\mathcal{F}(\gamma, \ell) \cap \mathcal{H}_\mathcal{D}^\beta(L))^2$ , for any initial probability measure  $\mu(dx)$  on  $\mathcal{S}$  for  $X_0$  such that  $\mu((1 + |\cdot|)^2) < \infty$ . The notation  $\lesssim$  means up to some constant independent of  $n$ .



Some comments are in order:

1) The noise density  $G$  should be at least as regular as  $\beta$ , the regularity of the link functions  $f_0$  and  $f_1$ , in order to obtain the rate  $|\mathbb{T}_n|^{-\beta/(2\beta+1)}$ . Indeed,  $G$  determines the regularity of the mean transition  $y \rightsquigarrow \mathcal{Q}(x, y)$  (see (4)) and thus the regularity of the invariant measure  $\nu$ . Recall now that we built Nadaraya-Watson type estimators: the estimator  $\widehat{f}_{\iota, n}$  is the quotient of an estimator of  $(\nu f_\iota)$  and an estimator of  $\nu$ . Thus the rate of convergence depends not only on the regularity of  $f_\iota$  but also on the regularity of  $\nu$ . A more general result would establish the rate  $|\mathbb{T}_n|^{-\beta'/(2\beta'+1)}$  with  $\beta'$  the minimal regularity between  $f_\iota$  and  $\nu$  (and thus between  $f_\iota$  and  $G$  the noise density).

2) Up to the factor  $\varpi_n^{-1}$ , we obtain the classical rate  $|\mathbb{T}_n|^{-\beta/(2\beta+1)}$  where  $|\mathbb{T}_n|$  is the number of observations. We know it is optimal in a minimax sense in a density estimation framework and we can infer this is optimal in our framework too. To prove it is the purpose of Theorem 6 which follows.

3) The threshold  $\varpi_n$  should be chosen such that it inflates the upper-rate of convergence of a slow factor only. Typically,  $\varpi_n = (\ln n)^{-1}$  is suitable. Looking carefully at the proof, see (27), we actually see that  $\varpi_n \rightarrow 0$  as  $n \rightarrow \infty$  is not necessary. One could choose,  $\varpi_n = \varpi$  with

$$\varpi = \frac{1}{2} \inf_{(f_0, f_1)} \inf_{x \in \mathcal{D}} \nu(x) > 0$$

where the infimum is taken among all  $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$  and where  $\varpi > 0$  is guaranteed by Lemma 21. However, to calibrate in practice  $\varpi_n$  in such a way is not possible since we cannot compute  $\varpi$ .

4) For the sake of simplicity, we have picked a common bandwidth  $h_n$  to define the two estimators, but one can immediately generalize our study for two different bandwidths  $(h_{\iota, n} = |\mathbb{T}_n|^{-\beta_\iota/(2\beta_\iota+1)}, \iota \in \{0, 1\})$  where  $\beta_\iota$  is the Hölder smoothness of  $f_\iota$ .

5) We do not achieve adaptivity in the smoothness of the autoregressive functions since our choice of bandwidth  $h_n$  still depends on  $\beta$ , we refer to the discussion in Section 4.

We complete the previous theorem by

**Theorem 6** (Lower rate of convergence). *Assume the noise density  $G$  is a bivariate Gaussian density. Let  $\mathcal{D} \subset \mathcal{S}$  be a compact interval. For every  $\gamma \in (0, 1)$  and every positive  $\ell, \beta, L$ , there exists  $C > 0$  such that, for every  $x \in \mathcal{D}$ ,*

$$\liminf_{n \rightarrow \infty} \inf_{(\widehat{f}_{0, n}, \widehat{f}_{1, n})} \sup_{(f_0, f_1)} \mathbb{P} \left( |\mathbb{T}_n|^{\beta/(2\beta+1)} (|\widehat{f}_{0, n}(x) - f_0(x)| + |\widehat{f}_{1, n}(x) - f_1(x)|) \geq C \right) > 0,$$

where the supremum is taken among all functions  $(f_0, f_1) \in (\mathcal{F}(\gamma, \ell) \cap \mathcal{H}_{\mathcal{D}}^\beta(L))^2$  and the infimum is taken among all estimators based on  $(X_u, u \in \mathbb{T}_{n+1})$ .

This result obviously implies a lower rate of convergence for the mean quadratic loss at point  $x$ . We see that in a regular case, the Gaussian case, the lower and upper rates match.

### 3 ASYMPTOTIC BEHAVIOUR

#### 3.1 MAIN RESULTS

##### *Asymptotic normality*

The almost-sure convergence of the autoregressive functions estimators is obtained in Proposition 7 for any  $h_n = |\mathbb{T}_n|^{-\alpha}$  with  $\alpha \in (0, 1)$ . Choosing  $\alpha \geq 1/(2\beta + 1)$ , the estimator  $(\widehat{f}_{0,n}(x), \widehat{f}_{1,n}(x))$  recentered by  $(f_0(x), f_1(x))$  and normalised by  $\sqrt{|\mathbb{T}_n|h_n}$  converges in distribution to a Gaussian law. Depending on  $\alpha > 1/(2\beta + 1)$  or not, the limit Gaussian law is centered or not, as we state in Theorems 8 and 9.

**Proposition 7** (Almost sure convergence). *In the same setting as in Theorem 5 with  $h_n = |\mathbb{T}_n|^{-\alpha}$  for  $\alpha \in (0, 1)$ ,*

$$\begin{pmatrix} \widehat{f}_{0,n}(x) \\ \widehat{f}_{1,n}(x) \end{pmatrix} \rightarrow \begin{pmatrix} f_0(x) \\ f_1(x) \end{pmatrix}, \quad \mathbb{P}_\mu - a.s.$$

as  $n \rightarrow \infty$ .

From now on we need to reinforce the assumption on the noise sequence: we require that the noise  $(\varepsilon_0, \varepsilon_1)$  has finite moment of order 4,  $\mathbb{E}[\varepsilon_0^4 + \varepsilon_1^4] < \infty$ , which is guaranteed by  $\lambda > 5$ . We denote by  $|K|_2^2 = \int_{\mathcal{S}} K(x)^2 dx$ .

**Theorem 8** (Asymptotic normality). *In the same setting as in Theorem 5 with  $\lambda > 5$  and  $h_n = |\mathbb{T}_n|^{-\alpha}$  for  $\alpha > 1/(2\beta + 1)$ ,*

$$\sqrt{|\mathbb{T}_n|h_n} \begin{pmatrix} \widehat{f}_{0,n}(x) - f_0(x) \\ \widehat{f}_{1,n}(x) - f_1(x) \end{pmatrix} \xrightarrow{d} \mathcal{N}_2(\mathbf{0}_2, \Sigma_2(x)) \quad \text{with} \quad \Sigma_2(x) = |K|_2^2(\nu(x))^{-1}\Gamma,$$

$\Gamma$  being the noise covariance matrix and  $\mathbf{0}_2 = (0, 0)$ . Moreover, for  $x_1, \dots, x_k$  distinct points in  $\mathcal{D}$ , the sequence

$$\left( \begin{pmatrix} \widehat{f}_{0,n}(x_l) - f_0(x_l) \\ \widehat{f}_{1,n}(x_l) - f_1(x_l) \end{pmatrix}, l = 1, \dots, k \right)$$

is asymptotically independent.

The restriction  $\alpha > 1/(2\beta + 1)$  in Theorem 8 prevents us from choosing  $h_n = |\mathbb{T}_n|^{-1/(2\beta+1)}$ , which is the optimal choice to achieve the minimax rate as we have seen in Theorem 5. The following Theorem remedies to this flaw, but at the cost of an unknown bias. We obtain an explicit expression of this bias for  $\beta$  an integer which depends on the  $\beta$ -th derivatives of the autoregressive functions and the invariant measure of the tagged-branch chain.

**Theorem 9** (Asymptotic normality with bias expression). *In the same setting as in Theorem 5 with  $\lambda > 5$  and  $\beta$  an integer,*

(i) *If  $h_n^\beta \sqrt{|\mathbb{T}_n|h_n} \rightarrow \kappa$  with  $\kappa \in [0, \infty)$  as  $n \rightarrow +\infty$ , then*

$$\sqrt{|\mathbb{T}_n|h_n} \begin{pmatrix} \widehat{f}_{0,n}(x) - f_0(x) \\ \widehat{f}_{1,n}(x) - f_1(x) \end{pmatrix} \xrightarrow{d} \mathcal{N}_2(\kappa \mathbf{m}_2(x), \Sigma_2(x)) \quad \text{with} \quad \Sigma_2(x) = |K|_2^2(\nu(x))^{-1}\Gamma$$

and

$$\mathbf{m}_2(x) = \frac{(-1)^\beta}{\beta! \nu(x)} \int_{\mathcal{S}} y^\beta K(y) dy \left( \frac{(\nu f_0)^\beta(x) - \nu^\beta(x) f_0(x)}{(\nu f_1)^\beta(x) - \nu^\beta(x) f_1(x)} \right).$$

(ii) If  $h_n^\beta \sqrt{|\mathbb{T}_n| h_n} \rightarrow \infty$  as  $n \rightarrow +\infty$ , then

$$h_n^{-\beta} \begin{pmatrix} \widehat{f}_{0,n}(x) - f_0(x) \\ \widehat{f}_{1,n}(x) - f_1(x) \end{pmatrix} \xrightarrow{\mathbb{P}_\mu} \mathbf{m}_2(x).$$

If  $\beta$  is not an integer, we could generalize Theorem 9 but at the cost of introducing fractional derivatives. Note that the definition of this notion is not unique (see [TRB99] or [Use08]). We decided to restrict the parameter  $\beta$  to be an integer in order to avoid here additional technicalities.

### Moderate deviations principle

Statistical estimators are also studied under the angle of large and moderate deviations principles. Large and moderate deviations limit theorems are proved in the independent setting for the kernel density estimator and also for the Nadaraya-Watson estimator (see Louani and Joutard [Lou98, Lou99, Jou06] in the univariate case, see also Mokkadem *et al.* [MPT08] and references therein). We also refer to Mokkadem and Pelletier [MP06] for the study of confidence bands based on the use of moderate deviations principles. The purpose of this section is to establish a moderate deviations principle for the autoregressive functions estimators defined by (1).

Let us introduce the notion of moderate deviations principle in a general setting. Let  $(Z_n)_{n \geq 0}$  be a sequence of random variables with values in a Polish space  $\mathcal{S}$  endowed with its Borel  $\sigma$ -field  $\mathfrak{G}$  and let  $(s_n)_{n \geq 0}$  be a positive sequence that converges to  $+\infty$ . We assume that  $Z_n/s_n$  converges in probability to 0 and that  $Z_n/\sqrt{s_n}$  converges in distribution to a centered Gaussian law. Let  $I : \mathcal{S} \rightarrow \mathbb{R}^+$  be a lower semicontinuous function, that is for all  $c > 0$  the sub-level set  $\{x \in \mathcal{S}, I(x) \leq c\}$  is a closed set. Such a function  $I$  is called a rate function and it is called a good rate function if all the sub-level sets are compact sets. Let  $(a_n)_{n \geq 0}$  be a positive sequence such that  $a_n/s_n$  converges to 0.

**Definition 10** (Moderate deviations principle). *We say that  $Z_n/\sqrt{a_n s_n}$  satisfies a moderate deviations principle on  $\mathcal{S}$  with speed  $a_n$  and the rate function  $I$  if, for any  $A \in \mathfrak{G}$ ,*

$$- \inf_{x \in A^\circ} I(x) \leq \liminf_{n \rightarrow \infty} \frac{1}{a_n} \log \mathbb{P}\left(\frac{Z_n}{\sqrt{a_n s_n}} \in A\right) \leq \limsup_{n \rightarrow \infty} \frac{1}{a_n} \log \mathbb{P}\left(\frac{Z_n}{\sqrt{a_n s_n}} \in A\right) \leq - \inf_{x \in \bar{A}} I(x),$$

where  $A^\circ$  and  $\bar{A}$  denote respectively the interior and the closure of  $A$ .

We will state such a moderate deviations principle for the estimators  $\widehat{f}_0(x)$  and  $\widehat{f}_1(x)$  in a restricted framework. Since the moderate deviations principle requires a strong notion of convergence, the super-exponential convergence (see Definition 23), it is not surprising to ask the process to also satisfy strong assumptions. The autoregressive functions  $f_0$  and  $f_1$  will be restricted to belong to the class  $\mathcal{F}(\gamma, \ell)$  with  $\gamma = 0$ , which means we have

$$\max\{|f_0|_\infty; |f_1|_\infty\} \leq \ell.$$

To guarantee we have geometric ergodicity of the tagged-branch Markov chain  $Y$  uniformly in the initial value with an exponential decay rate smaller than  $1/2$ , we will require the following assumption to replace Assumption 2 (see Lemma 22).

**Assumption 11.** *There exists  $M_4 > \ell$  such that  $2(M_4 - \ell)\delta(M_4) > 1/2$  with  $\delta(\cdot)$  defined by (8).*

We also reinforce the assumption on the kernel with

**Assumption 12.** *The kernel  $K : \mathbb{R} \rightarrow \mathbb{R}$  is such that*

$$\int_{\mathbb{R}} |K^+(x)| dx > \frac{|G_0|_{\infty} + |G_1|_{\infty}}{2\delta(M_3)(1-\eta(M_2))} \int_{\mathbb{R}} |K^-(x)| dx$$

where  $M_2, M_3$  come from Assumption 3 and  $K^+(\cdot) = \max\{K(\cdot); 0\}$ ,  $K^-(\cdot) = \min\{K(\cdot); 0\}$ .

Note that Assumption 12 is valid for any non-negative kernel, in particular for the Gaussian kernel (since  $\int_{\mathbb{R}} |K^-(x)| dx = 0$  for such kernels).

**Theorem 13** (Moderate deviations principle). *In the same setting as in Theorem 5 with  $\gamma = 0$ ,  $\lambda > 5$ ,  $K$  satisfying in addition Assumption 12,  $h_n = |\mathbb{T}_n|^{-\alpha}$  for  $\alpha > 1/(2\beta+1)$  and Assumption 11 replacing Assumption 2, let  $(b_n)_{n \geq 0}$  be a positive sequence such that*

$$(i) \lim_{n \rightarrow \infty} \frac{b_n}{(|\mathbb{T}_n| h_n)^{1/2}} = +\infty, \quad (ii) \lim_{n \rightarrow \infty} \frac{b_n}{|\mathbb{T}_n| h_n} = 0, \quad (iii) \lim_{n \rightarrow \infty} \frac{b_n}{|\mathbb{T}_n| h_n^{1+\beta}} = +\infty.$$

Then the sequence

$$\left( \frac{|\mathbb{T}_n| h_n}{b_n} \begin{pmatrix} \widehat{f}_{0,n}(x) - f_0(x) \\ \widehat{f}_{1,n}(x) - f_1(x) \end{pmatrix}, n \geq 0 \right)$$

satisfies a MDP on  $\mathcal{S}^2$  with speed  $b_n^2/(|\mathbb{T}_n| h_n)$  and good rate function  $J_x : \mathcal{S}^2 \rightarrow \mathbb{R}$  defined by

$$J_x(\mathbf{z}) = (2|K|_2^2)^{-1} \nu(x) \mathbf{z}^t \Gamma^{-1} \mathbf{z}, \quad \mathbf{z} \in \mathcal{S}^2,$$

where  $\mathbf{z}^t$  stands for the transpose of vector  $\mathbf{z}$ .

The contraction principle (see the textbook of Dembo and Zeitouni [DZ98], Chapter 4) enables us to state the following corollary of Theorem 13.

**Corollary 14.** *In the same setting as Theorem 13, for every  $\delta > 0$ ,*

$$\lim_{n \rightarrow +\infty} \frac{|\mathbb{T}_n| h_n}{b_n^2} \log \mathbb{P} \left( \frac{|\mathbb{T}_n| h_n}{b_n} |\widehat{f}_{\iota,n}(x) - f_{\iota}(x)| > \delta \right) = -(2\sigma_{\iota}^2 |K|_2^2)^{-1} \nu(x) \delta^2, \quad \iota \in \{0, 1\}.$$

### 3.2 CONSTRUCTION OF AN ASYMMETRY TEST

Testing in the context of nonparametric regression is a crucial point, especially in applied contexts. The question of no effect in nonparametric regression is early addressed in Eubank and LaRiccia [EL93]. We may also want to compare two regression curves nonparametrically and we refer to Munk [Mun98] and references therein. Specific tools have been developed to compare time series, in particular through autocorrelation functions (see for instance the recent work by Jin [Jin11])

among many others). However the test we propose to study asymmetry follows from the estimation procedure we propose and is more inspired by the regression case.

Let  $x_1, \dots, x_k$  be distinct points in  $\mathcal{S}$ . We are going to build a statistical test that allows us to segregate between hypothesis

$$\mathcal{H}_0 : \forall l \in \{1, \dots, k\}, f_0(x_l) = f_1(x_l) \quad \text{vs.} \quad \mathcal{H}_1 : \exists l \in \{1, \dots, k\}, f_0(x_l) \neq f_1(x_l).$$

In the parametric studies on *E. coli* [Guy07, DM10], these tests are known as detection of cellular aging and they permit to decide if the cell division is symmetric or asymmetric.

Inspired by Bierens [Bie97] we define new estimators in order to both achieve the rate  $|\mathbb{T}_n|^{-\beta/(2\beta+1)}$  in the asymptotic normality theorem and remove the bias. Let  $(\widehat{f}_{\iota,n}^{(a)}(x), \iota \in \{0, 1\})$  be the estimators (1) with bandwidth  $h_n^{(a)} = |\mathbb{T}_n|^{-1/(2\beta+1)}$  and  $(\widehat{f}_{\iota,n}^{(b)}(x), \iota \in \{0, 1\})$  be the estimators (1) with bandwidth  $h_n^{(b)} = |\mathbb{T}_n|^{-\delta/(2\beta+1)}$  for some  $\delta \in (0, 1)$ . We assume that  $\beta$  is an integer so that the results of Theorem 9 are valid. We define

$$\left( \bar{f}_{\iota,n}(x) = (1 - |\mathbb{T}_n|^{-\frac{\beta(1-\delta)}{2\beta+1}})^{-1} (\widehat{f}_{\iota,n}^{(a)}(x) - |\mathbb{T}_n|^{-\frac{\beta(1-\delta)}{2\beta+1}} \widehat{f}_{\iota,n}^{(b)}(x)), \iota \in \{0, 1\} \right), \quad (11)$$

new estimators such that

$$|\mathbb{T}_n|^{-\frac{\beta}{2\beta+1}} \begin{pmatrix} \bar{f}_{0,n}(x) - f_0(x) \\ \bar{f}_{1,n}(x) - f_1(x) \end{pmatrix} \xrightarrow{d} \mathcal{N}_2(\mathbf{0}_2, \Sigma_2(x)) \quad \text{with} \quad \Sigma_2(x) = |K|_2^2 (\nu(x))^{-1} \Gamma. \quad (12)$$

As announced the trick of Bierens [Bie97] enables us to remove the unknown bias while keeping the optimal rate of convergence. We now define a test statistics based on these new estimators by

$$W_n(x_1, \dots, x_k) = \frac{|\mathbb{T}_n|^{\frac{2\beta}{2\beta+1}}}{(\sigma_0^2 + \sigma_1^2 - 2\sigma_0\sigma_1\varrho)|K|_2^2} \sum_{l=1}^k \widehat{\nu}_n(x_l) (\bar{f}_{0,n}(x_l) - \bar{f}_{1,n}(x_l))^2 \quad (13)$$

with  $\widehat{\nu}_n(\cdot) = |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} K_{h_n}(X_u - \cdot)$  where  $h_n = |\mathbb{T}_n|^{-1/(2\beta+1)}$ .

**Theorem 15** (Wald test for asymmetry). *In the same setting as in Theorem 9, let  $x_1, \dots, x_k$  be distinct points in  $\mathcal{D}$ . Then the test statistic  $W_n(x_1, \dots, x_k)$  converges in distribution to the chi-squared distribution with  $k$  degrees of freedom  $\chi^2(k)$ , under  $\mathcal{H}_0$ , and  $\mathbb{P}_\mu$ -almost surely to  $+\infty$ , under  $\mathcal{H}_1$ .*

Note that in (13) we could replace  $\sigma_0^2$ ,  $\sigma_1^2$  and  $\varrho$  by

$$\widehat{\sigma}_{0,n}^2 = |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} (\widehat{\varepsilon}_{u0})^2, \quad \widehat{\sigma}_{1,n}^2 = |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} (\widehat{\varepsilon}_{u1})^2 \quad \text{and} \quad \widehat{\varrho}_n = \widehat{\sigma}_{0,n}^{-1} \widehat{\sigma}_{1,n}^{-1} |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} \widehat{\varepsilon}_{u0} \widehat{\varepsilon}_{u1}$$

with the empirical residuals  $\widehat{\varepsilon}_{uu} = X_{uu} - \widehat{f}_{\iota,n}(X_u)$  for  $u \in \mathbb{T}_n$ . We claim that these estimators are consistent, so that Theorem 15 is still valid with this other test statistics. Proving the convergence in probability of these three quantities would imply some technical calculations and we do not give here more details.

## 4 DISCUSSION

### *Recursive estimators*

For the sake of clarity, we have picked a bandwidth that depends on the size of the whole subtree  $\mathbb{T}_n$ . We could also choose bandwidths that depend on the sizes of the generations. More precisely, we could estimate  $(f_0(x), f_1(x))$  the autoregressive functions at point  $x \in \mathcal{D}$  from the observations  $(X_u, u \in \mathbb{T}_{n+1})$  by

$$\left( \widehat{f}_{\iota, n}(x) = \frac{\sum_{m=0}^n \sum_{u \in \mathbb{G}_m} K_{h_m}(x - X_u) X_{u\iota}}{\sum_{m=0}^n \sum_{u \in \mathbb{G}_m} K_{h_m}(x - X_u)} \right), \iota \in \{0, 1\}$$

with the collection of bandwidths  $(h_m = |\mathbb{G}_m|^{-\alpha}, 0 \leq m \leq n)$  for  $\alpha \in (0, 1)$ . These estimators can be seen as a version of recursive Nadaraya-Watson estimators when the index set has a binary tree structure. We stress that our results also hold for these alternative estimators.

### *Heteroscedasticity*

More generally, given two functions  $\sigma_0, \sigma_1 : \mathcal{S} \rightarrow [0, \infty)$ , we could consider the autoregressive equations

$$X_{u0} = f_0(X_u) + \sigma_0(X_u)\varepsilon_{u0} \quad \text{and} \quad X_{u1} = f_1(X_u) + \sigma_1(X_u)\varepsilon_{u1}$$

with  $\mathbb{E}[\varepsilon_{u0}^2] = \mathbb{E}[\varepsilon_{u1}^2] = 1$  and  $\mathbb{E}[\varepsilon_{u0}\varepsilon_{u1}] = \varrho$  where  $\varrho \in (-1, 1)$ . Assuming  $0 < \inf_{x \in \mathcal{S}} \sigma_\iota(x) \leq \sup_{x \in \mathcal{S}} \sigma_\iota(x) < \infty$  for  $\iota \in \{0, 1\}$ , Theorems 8 and 9 still hold with

$$\Sigma_2(x) = |K|_2^2(\nu(x))^{-1} \begin{pmatrix} \sigma_0^2(x) & \varrho\sigma_0(x)\sigma_1(x) \\ \varrho\sigma_0(x)\sigma_1(x) & \sigma_1^2(x) \end{pmatrix}.$$

The estimation of the variance functions  $\sigma_0$  and  $\sigma_1$  would be in practice of great interest but the theoretical study of such estimators lies here beyond the scope of this work.

### *Adaptive estimation*

Our estimators of the autoregressive functions are not adaptive in the smoothness  $\beta$  of these two functions since the choice of the bandwidth  $h_n = |\mathbb{T}_n|^{-1/(2\beta+1)}$  still depends on  $\beta$ . For classical autoregressive models (*i.e.* non-bifurcating), we refer to the early work by Hoffmann [Hof99]. This has been also investigated by Delouille and van Sachs [DvS05] and they propose there a protocol with design-adapted wavelets. It is well-known that the question of adaptivity boils down to proving deviations inequalities. In our framework, we should prove deviations inequalities for the two empirical means

$$|\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) X_{u\iota} \quad \text{and} \quad |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u).$$

In the restricted class of bounded autoregressive functions  $\mathcal{F}(0, \ell)$ , with compactly supported noise density  $G$ , the results of Chapter II brings such deviations inequalities. Indeed, on the class  $\mathcal{F}(0, \ell)$ ,

we can achieve uniform ergodicity, a key assumption in Chapter II. However, on the more general class  $\mathcal{F}(\gamma, \ell)$  with  $\gamma > 0$ , the problem of deviations inequalities is still open. We recall that on the class  $\mathcal{F}(\gamma, \ell)$  with  $\gamma > 0$ , only  $\mathbb{V}$ -geometric ergodicity with  $\mathbb{V}$  a Lyapunov function is achieved.

### Uniform test

The asymmetry test we have built is based on the choice of a grid of points on  $\mathcal{S}$ . A theoretical result is needed in order to build a uniform test on a interval  $\mathcal{D} \subset \mathcal{S}$ . More precisely, to achieve such a uniform test we should study the asymptotic behaviour of

$$\sup_{x \in \mathcal{D}} \sqrt{|\mathbb{T}_n| h_n} |\hat{f}_{\iota, n}(x) - f_{\iota}(x)|, \quad \iota \in \{0, 1\}.$$

This asymptotic study lies in the scope of the theory of extrema. One can see the study of Liu and Wu [LW10] for autoregressive processes of order 1: an asymptotic Gumbel behaviour is highlighted for the Nadaraya-Watson type estimator of the autoregressive function. One can also see Li [Li14] where the aim is to compare the two autoregressive functions of two autoregressive processes of order 1: a test statistics based on estimated residuals is studied there. Studying the limit distribution of

$$\int_{\mathcal{D}} (\hat{f}_{\iota, n}(x) - f_{\iota}(x))^2 dx, \quad \iota \in \{0, 1\},$$

we could derive an alternative criterion to discriminate between  $f_0 = f_1$  and  $f_0 \neq f_1$ .

### Weaker assumption on the noise

More realistic assumptions on the noise sequence, which represent environmental effects, should be considered. As in Bercu, de Saporta and Gégout-Petit [BdSGP09], we could assume that the noise is such that, for any  $u \in \mathbb{G}_m$ ,

$$\mathbb{E}[\varepsilon_u | \mathcal{F}_{m-1}] = 0$$

where  $\mathcal{F}_{m-1}$  is the sigma-field generated by  $(X_u, u \in \mathbb{T}_{m-1})$ . This is much weaker than assuming  $((\varepsilon_{u0}, \varepsilon_{u1}), u \in \mathbb{T})$  is an independent and identically distributed sequence but the Markovian structure of  $(X_u, u \in \mathbb{T})$  is lost and the techniques we have used in this work are no longer available.

## 5 NUMERICAL IMPLEMENTATION

The goal of this section is to illustrate the theoretical results of the previous sections, in particular the results of Theorem 5 (Upper rate of convergence) and Theorem 15 (Wald test for asymmetry).

### Quality of the estimation procedure.

We pick trial autoregressive functions defined analytically by

$$f_0(x) = x(1/4 + \exp(-x^2)/2) \quad \text{and} \quad f_1(x) = x(1/8 + \exp(-x^2)/2)$$

for  $x \in \mathcal{S} = \mathbb{R}$ . We take a Gaussian noise with  $\sigma_0^2 = \sigma_1^2 = 1$  and  $\varrho = 0.3$ . We simulate  $M = 500$  times a NBAR process up to generation  $n + 1 = 15$ , with root  $X_0 = 1$ . We take a Gaussian kernel  $K(x) = (2\pi)^{-1/2} \exp(-x^2/2)$  and  $h_n = |\mathbb{T}_n|^{-1/5}$  in order to implement estimators given by (1). We did not meet any problem with the denominator in practice and actually set  $\varpi_n = 0$ . We evaluate  $\widehat{f}_{0,n}$  and  $\widehat{f}_{1,n}$  on a regular grid of  $\mathcal{D} = [-3, 3]$  with mesh  $\Delta x = |\mathbb{T}_n|^{-1/2}$ . For each sample we compute the empirical error

$$e_\iota^{(i)} = \frac{\|\widehat{f}_{\iota,n}^{(i)} - f_\iota\|_{\Delta x}}{\|f_\iota\|_{\Delta x}}, \quad i = 1, \dots, M,$$

where  $\|\cdot\|_{\Delta x}$  denotes the discrete norm over the numerical sampling. Table 5 displays the mean-empirical errors together with the empirical standard deviations,

$$\bar{e}_\iota = M^{-1} \sum_{i=1}^M e_\iota^{(i)} \quad \text{and} \quad (M^{-1} \sum_{i=1}^M (e_\iota^{(i)} - \bar{e}_\iota)^2)^{1/2}$$

for  $\iota \in \{0, 1\}$ . The larger  $n$ , the better the reconstruction of  $f_0$  and  $f_1$  as shown in Table 5.

| $n$              | 8             | 9             | 10            | 11            | 12            | 13            | 14            |
|------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| $ \mathbb{T}_n $ | 511           | 1 023         | 2 047         | 4 095         | 8 191         | 16 383        | 32 767        |
| $\bar{e}_0$      | 0.4442        | 0.3417        | 0.2633        | 0.2006        | 0.1517        | 0.1285        | 0.0891        |
| <i>sd.</i>       | <i>0.1509</i> | <i>0.1063</i> | <i>0.0761</i> | <i>0.0558</i> | <i>0.0387</i> | <i>0.0295</i> | <i>0.0209</i> |
| $\bar{e}_1$      | 0.6696        | 0.5141        | 0.4006        | 0.3027        | 0.2356        | 0.1776        | 0.1384        |
| <i>sd.</i>       | <i>0.2482</i> | <i>0.1626</i> | <i>0.1227</i> | <i>0.0831</i> | <i>0.0622</i> | <i>0.0440</i> | <i>0.0326</i> |

Table 5: Mean empirical relative error  $\bar{e}_0$  (resp.  $\bar{e}_1$ ) and its standard deviation computed over  $M = 500$  Monte-Carlo trees, with respect to  $|\mathbb{T}_n|$ , for the autoregressive function  $f_0$  (resp.  $f_1$ ) reconstructed over the interval  $\mathcal{D} = [-3, 3]$  by the estimator  $\widehat{f}_{0,n}$  (resp.  $\widehat{f}_{1,n}$ ).

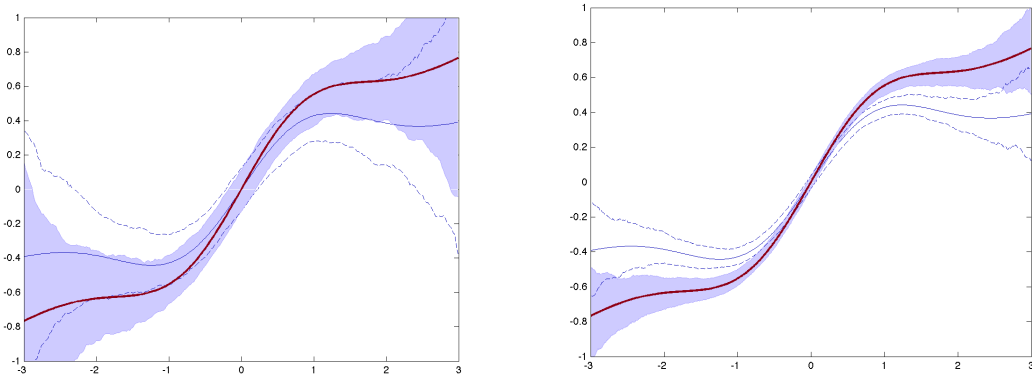


Figure 10: Reconstruction of  $f_0$  over  $\mathcal{D} = [-3, 3]$  with 95%-level confidence bands constructed over  $M = 500$  Monte Carlo trees. In bold red line:  $x \rightsquigarrow f_0(x)$ . In thin blue lines: reconstruction of  $f_1$  with 95%-level confidence bands. Left:  $n = 10$  generations. Right:  $n = 14$  generations.



This is also true at a visual level, as shown on Figure 10 where 95%-level confidence bands are built so that for each point  $x$ , the lower and upper bounds include 95% of the estimators  $(\widehat{f}_{0,n}^{(i)}(x), i = 1 \dots M)$ . As one can see on Figure 10, the reconstruction is good around 0 and deteriorates for large or small  $x$ . Indeed an histogram estimator of the invariant measure  $\nu$  shows that its mass is located around 0 and thus more observations lie in this zone, which ensures a better reconstruction there. The same analysis holds for the reconstruction of  $f_1$ , see the thin blue lines.

The error is close to  $|\mathbb{T}|^{-2/5}$  for both  $\widehat{f}_{0,n}$  and  $\widehat{f}_{1,n}$  as expected: indeed, for a kernel of order  $n_0$ , the bias term in density estimation is of order  $h^{\beta \wedge (n_0+1)}$ . For the smooth  $f_0$ ,  $f_1$  and  $\nu$  we have here, we rather expect for the rate  $|\mathbb{T}_n|^{-(n_0+1)/(2(n_0+1)+1)} = |\mathbb{T}|^{-2/5}$  for the Gaussian kernel with  $n_0 = 1$  that we use here, and this is consistent with what we observe in Figure 11.

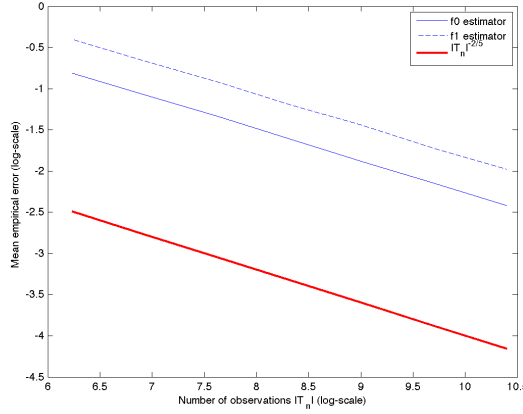


Figure 11: The log-average relative empirical error over  $M = 500$  Monte Carlo trees vs.  $\log(|\mathbb{T}_n|)$  for  $f_0$  (resp.  $f_1$ ) reconstructed over  $\mathcal{D} = [-3, 3]$  by  $\widehat{f}_{0,n}$  (solid blue line) (resp.  $\widehat{f}_{1,n}$  (dashed blue line)) compared to the expected log-rate (solid red line).

### Implementation of the asymmetry test.

We implement now the estimators (11) inspired by [Bie97] in order to compute our test statistics (13). We keep a Gaussian kernel and we pick  $h_n^{(a)} = |\mathbb{T}_n|^{-1/5}$  and  $h_n^{(b)} = |\mathbb{T}_n|^{-1/10}$  (i.e.  $\delta = 1/2$  – and the choice of  $\delta$  proves to have no influence on our numerical results). The numerical study of  $\widehat{f}_{0,n}$  and  $\widehat{f}_{1,n}$  leads to similar results as those of the previous study. For a given grid  $\{x_1, \dots, x_k\}$  of  $\mathcal{D} = [-3, 3]$ , we reject the null hypothesis  $\mathcal{H}_0$  if  $W_n(x_1, \dots, x_k)$  exceeds the 5%-quantile of the chi-squared distribution with  $k$  degrees of freedom and thus obtain a test with asymptotic level 5%. We measure the quality of our test procedure computing the proportion of rejections of the null.

We first implement the two following cases:

$$\text{(Case =)} \quad f_0(x) = f_1(x) = x(1/4 + \exp(-x^2)/2),$$

$$\text{(Case } \neq) \quad f_0(x) = x(1/4 + \exp(-x^2)/2) \quad \text{and} \quad f_1(x) = x(1/8 + \exp(-x^2)/2).$$

The test should reject  $\mathcal{H}_0$  in the second case but not in the first one. The larger  $n$ , the better the test as one can see in Table 6:  $\mathcal{H}_0$  is more and more often rejected for (**Case  $\neq$** ) and less and less often rejected for (**Case =**) as  $n$  increases, which is what we expect. We also observe the influence of the number of points of the grid which enables us to build the test statistics. Three grids of  $\mathcal{D} = [-3, 3]$  are tested with  $k = 13, 25$  and  $61$  points (corresponding to  $\Delta x = 0.5, 0.25$  and  $0.1$  in Table 6). The larger the number of points, the larger the proportion of rejections of  $\mathcal{H}_0$  in both cases. The choice of the number  $k$  of points is delicate but we would recommend to use a low  $k$  to build the test.

| $n$                           |                   | 8     | 9     | 10    | 11    | 12    | 13     | 14     |
|-------------------------------|-------------------|-------|-------|-------|-------|-------|--------|--------|
| $ \mathbb{T}_n $              |                   | 511   | 1 023 | 2 047 | 4 095 | 8 191 | 16 383 | 32 767 |
| <b>Case <math>\neq</math></b> | $\Delta x = 0.5$  | 46.8% | 67.2% | 87.6% | 99.0% | 100%  | 100%   | 100%   |
|                               | $\Delta x = 0.25$ | 59.6% | 77.8% | 92.8% | 99.8% | 100%  | 100%   | 100%   |
|                               | $\Delta x = 0.1$  | 67.8% | 85.4% | 95.6% | 99.8% | 100%  | 100%   | 100%   |
| <b>Case =</b>                 | $\Delta x = 0.5$  | 19.6% | 18.6% | 18.2% | 16.2% | 13.4% | 14.8%  | 12.4%  |
|                               | $\Delta x = 0.25$ | 30.4% | 30.0% | 29.0% | 24.8% | 21.4% | 19.4%  | 19.8%  |
|                               | $\Delta x = 0.1$  | 42.6% | 42.6% | 40.4% | 39.8% | 35%   | 31.6%  | 32.2%  |

Table 6: Proportions of rejections of the null hypothesis  $\mathcal{H}_0 : \{\forall l = 1, \dots, k, f_0(x_l) = f_1(x_l)\}$  for 5% asymptotic level tests over  $M = 500$  Monte-Carlo trees. The test is based on the test statistics  $W_n(x_1, \dots, x_k)$  (13) with the grids  $\{x_l = -3 + (l - 1)\Delta x \leq 3; l \geq 1\}$  for  $\Delta x \in \{0.5; 0.25; 0.1\}$ . (**Case  $\neq$** ): the proportions (power of the test) should be high. (**Case =**): the proportions (type I error) should be low.

The second numerical test aims at studying empirically the power of our test. We keep with the same autoregressive function  $f_0$  for cells of type 0 and parametrize the autoregressive function for cells of type 1 such that it progressively comes closer to  $f_0$ :

$$f_0(x) = x(1/4 + \exp(-x^2)/2) \quad \text{and} \quad f_{1,\tau}(x) = x(\tau + \exp(-x^2)/2)$$

for  $\tau \in [1/8, 1/4]$ . This choice enables us to interpolate between (**Case  $\neq$** ) and (**Case =**).

As  $\tau$  becomes closer to  $1/4$ , *i.e.* as  $f_{1,\tau}$  becomes closer to  $f_0$ , we see the decrease of the proportions of rejections of the null in Figure 12. The steeper the decrease is, the better performs our test. The proportion of rejections of  $\mathcal{H}_0$  is higher than 40% only for  $\tau$  up to 0.1875 for a reasonable number of observations ( $|\mathbb{T}_n| = 2\,047$  on the left in Figure 12). On the right in Figure 12, one can see what become the results for a larger number of observations,  $|\mathbb{T}_n| = 32\,767$ : the performance is good for  $\tau$  up to 0.225, *i.e.* closer to the equality case  $\tau = 1/4$ .

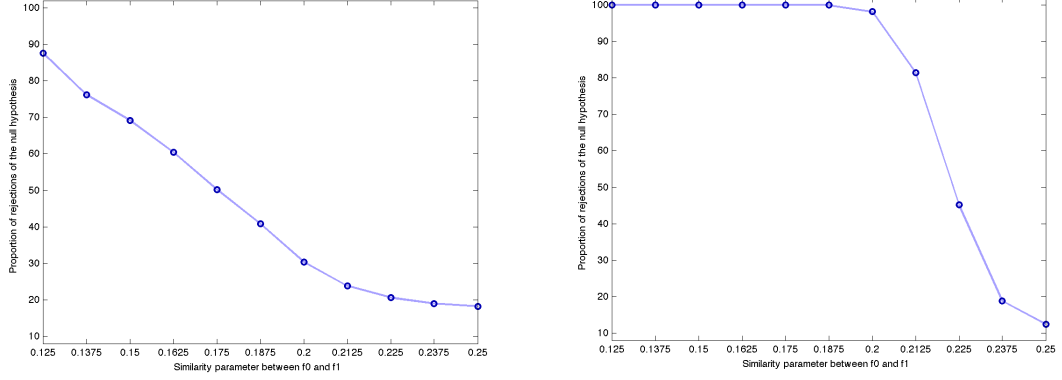


Figure 12: Proportions of rejections of the null hypothesis  $\mathcal{H}_0$  (power of the test):  $\{\forall l \in \{1, \dots, k\}, f_0(x_l) = f_{1,\tau}(x_l)\}$  with respect to  $\tau \in [1/8; 1/4]$  for 5% asymptotic level tests over  $M = 500$  Monte-Carlo trees. The test is based on the test statistics  $W_n(x_1, \dots, x_k)$  (13) with the grid  $\{x_l = -3 + (l-1)\Delta x \leq 3; l \geq 1\}$  for  $\Delta x = 0.5$ . Left:  $n = 10$  generations. Right:  $n = 14$  generations.

## 6 PROOFS

The notation  $\lesssim$  means up to some constant independent of  $n$  and uniform on the class  $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$ .

For a  $\mathfrak{G}$ -measurable function  $h : \mathcal{S} \rightarrow \mathbb{R}$  and a measure  $\mu$  on  $(\mathcal{S}, \mathfrak{G})$  we define  $\mu(h) = \int_{\mathcal{S}} h(x)\mu(dx)$ . For  $K \subset \mathcal{S}$  let

$$|h|_1 = \int_{\mathcal{S}} |h(y)|dy, \quad |h|_2^2 = \int_{\mathcal{S}} h(y)^2 dy, \quad |h|_K = \sup_{y \in K} |h(y)|$$

and  $|h|_{\infty} = |h|_{\mathcal{S}}$ . For a function  $h : \mathcal{S}^2 \rightarrow \mathbb{R}$  and  $K, K' \subset \mathcal{S}$  let

$$|h|_{K,K'} = \sup_{(x,x') \in K \times K'} |h(x,x')|.$$

The following lemma is well-known (see Delmas and Marsalle [DM10], Lemma 2.1 and Guyon [Guy07], Equation (7)) and highlights the role of the tagged-branch Markov chain. We prove it in Appendix for the sake of completeness.

**Lemma 16** (Many-to-one formulae). *Let  $(X_u, u \in \mathbb{T})$  be a bifurcating Markov chain with  $\mathbb{T}$ -transition  $\mathcal{P}$  and let  $\mathcal{Q}$  be defined by (3). Then for  $g : \mathcal{S} \rightarrow \mathbb{R}$  such that  $|g(x)| \leq 1 + |x|$  for any  $x \in \mathcal{S}$ , we have*

$$\mathbb{E}_{\mu} \left[ \sum_{u \in \mathbb{G}_m} g(X_u) \right] = |\mathbb{G}_m| \mathbb{E}_{\mu} [g(Y_m)] = |\mathbb{G}_m| \mu(\mathcal{Q}^m g) \quad (14)$$

and

$$\mathbb{E}_{\mu} \left[ \sum_{\substack{(u,v) \in \mathbb{G}_m^2 \\ u \neq v}} g(X_u)g(X_v) \right] = |\mathbb{G}_m| \sum_{l=1}^m 2^{l-1} \mu \left( \mathcal{Q}^{m-l} (\mathcal{P}(\mathcal{Q}^{l-1} g \otimes \mathcal{Q}^{l-1} g)) \right), \quad (15)$$

for any initial probability measure  $\mu(dx)$  on  $\mathcal{S}$  for  $X_0$  such that  $\mu((1 + |\cdot|)^2) < \infty$ .

## 6.1 PRELIMINARY

Set  $\mathbb{V}(x) = 1 + |x|$  for  $x \in \mathcal{S}$ . It plays the role of the Lyapunov function in the following

**Lemma 17** (Ergodicity). *Let  $\gamma \in (0, 1/2)$  and  $\ell > 0$ , let  $r > 0$  and  $\lambda > 2$ . For every  $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$ , for every  $G$  such that  $(G_0, G_1) \in \mathcal{G}(r, \lambda)^2$  satisfy Assumption 2, there exists a unique probability measure  $\nu$  of the form  $\nu(dx) = \nu(x)dx$  on  $(\mathcal{S}, \mathfrak{S})$ . Moreover, for every  $G$  such that  $(G_0, G_1) \in \mathcal{G}(r, \lambda)^2$  satisfy Assumption 2, there exist a constant  $R > 0$  and  $\rho \in (0, 1/2)$  such that*

$$\sup_{(f_0, f_1)} \sup_{|h| \leq \mathbb{V}} |\mathcal{Q}^m h(x) - \nu(h)| \leq R \mathbb{V}(x) \rho^m, \quad x \in \mathcal{S}, \quad m \geq 0,$$

where the supremum is taken among all functions  $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$  and among all functions  $h : \mathcal{S} \rightarrow \mathbb{R}$  which satisfy  $|h(x)| \leq \mathbb{V}(x)$  for all  $x \in \mathcal{S}$ .

*Proof of Lemma 17.* We shall rely on the results of Hairer and Mattingly [HM11].

*Step 1.* In order to make use of Theorem 1.2 of [HM11] we shall verify their Assumptions 1 and 2. Since  $Y_1 = f_{\iota_1}(Y_0) + \varepsilon'_1$  where  $\iota_1$  is drawn according to the Bernoulli distribution with parameter  $1/2$  and  $\varepsilon'_1$  has density  $(G_0 + G_1)/2$ , we get

$$\mathcal{Q}(|\cdot|)(x) = \mathbb{E}_x[|Y_1|] \leq \mathbb{E}[|f_{\iota_1}(x)|] + \mathbb{E}[|\varepsilon'_1|] \leq \gamma|x| + M_0$$

using  $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$ , with  $M_0 = \ell + \mathbb{E}[|\varepsilon'_1|]$  as defined previously. We have  $\gamma \in (0, 1)$  and  $M_0 \geq 0$ , so that is Assumption 1 in [HM11] (with their  $V(y) = |y|$ ).

Set  $\mathcal{C} = \{x \in \mathcal{S}; |x| \leq M_1\}$  where  $M_1$  comes from Assumption 2. For any  $A \in \mathfrak{S}$  and  $x \in \mathcal{C}$ , using the expression of  $\mathcal{Q}$  given by (4),

$$\mathcal{Q}(x, A) \geq \frac{1}{2} \int_{A \cap \mathcal{C}} G_0(y - f_0(x)) dy + \frac{1}{2} \int_{A \cap \mathcal{C}} G_1(y - f_1(x)) dy.$$

For  $(x, y) \in \mathcal{C}^2$ , we have  $|y - f_\iota(x)| \leq (1 + \gamma)M_1 + \ell$  for  $\iota \in \{0, 1\}$ . Thus

$$\inf_{x \in \mathcal{C}} \mathcal{Q}(x, A) \geq 2M_1 \delta((1 + \gamma)M_1 + \ell) \frac{|A \cap \mathcal{C}|}{|\mathcal{C}|} \quad \forall A \in \mathfrak{S},$$

where  $\delta(\cdot)$  is defined by (8) and  $|A|$  denotes the Lebesgue measure of  $A \in \mathfrak{S}$ . That is Assumption 2 in [HM11] with  $\alpha = 2M_1 \delta((1 + \gamma)M_1 + \ell) > 0$ . The existence and uniqueness of an invariant probability measure  $\nu$  follows from Theorem 1.2 of [HM11]. Moreover  $\nu$  is absolutely continuous with respect to the Lebesgue measure, since  $\mathcal{Q}(x, dy)$  defined by (4) itself is absolutely continuous with respect to the Lebesgue measure. By Assumption 2, for  $M_1$  satisfying (9), there exists some  $\alpha_0 \in (0, 1/2)$  such that

$$\alpha = 2M_1 \delta((1 + \gamma)M_1 + \ell) > 1/2 + \alpha_0. \quad (16)$$

We set  $\beta = \alpha_0/M_0$ . For all  $x \in \mathcal{S}$  we pick  $\mathcal{Q}^m \delta_x$  and  $\nu$  for  $\mu_1$  and  $\mu_2$  in Theorem 1.3 of [HM11] and apply it recursively. We conclude that for any function  $h$  such that  $|h(x)| \leq (1 + \beta|x|)$  for all  $x \in \mathcal{S}$ , for some positive constant  $C$ , we have

$$|\mathcal{Q}^m h(x) - \nu(h)| \leq C \rho^m (1 + \beta|x|)$$

with  $C = 1 + \int_{\mathbb{S}} (1 + \beta|x|)\nu(x)dx < \infty$ .

*Step 2.* A precise control of  $\rho$  with respect to  $\gamma$ ,  $M_0$  and  $\alpha$  is established in [HM11]. Set  $\gamma_0 = \gamma + 2M_0/M_1 + \eta \in (\gamma + 2M_0/M_1, 1)$  where  $\eta$  comes from Assumption 2. Theorem 1.3 of [HM11] states one can take

$$\rho = (1 - (\alpha - \alpha_0)) \vee \left( \frac{2 + M_1\beta\gamma_0}{2 + M_1\beta} \right).$$

Condition (16) gives immediately  $1 - (\alpha - \alpha_0) < 1/2$ . Note that

$$\frac{2 + M_1\beta\gamma_0}{2 + M_1\beta} < \frac{1}{2} \iff M_1 > \frac{(1 + 2\alpha_0)M_0}{1/2 - \gamma - \eta}$$

which is satisfied for  $\alpha_0 \in (0, 1/2)$  with our choice of  $M_1$  (which satisfies (9)). Thus Assumption 2 guarantees we can take  $\rho < 1/2$ .

It immediately follows from Step 1 and Step 2 that, for any function  $h$  such that  $|h| \leq \mathbb{V}$ , for some positive constant  $R$ , we have

$$|\mathcal{Q}^m h(x) - \nu(h)| \leq R\rho^m \mathbb{V}(x)$$

with  $\mathbb{V}(x) = 1 + |x|$ , as asserted. This bound holds uniformly over  $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$  by construction (the uniform choice of  $\rho$  is guaranteed by Step 2 and for a uniform choice of  $C$  in Step 1 recall that  $(G_0, G_1) \in \mathcal{G}(r, \lambda)^2$ ).  $\square$

Step 2 of this proof highlights that Assumption 2 is written to readily obtain  $\rho < 1/2$ . Note that to prove the existence of some  $\rho \in (0, 1)$  in Step 1, one only need the existence of some  $M_1 > 2M_0/(1 - \gamma)$  such that  $2M_1\delta((1 + \gamma)M_1 + \ell) > 0$  with  $\delta(\cdot)$  defined by (8).

## 6.2 ESTIMATION OF THE DENSITY OF THE INVARIANT MEASURE

For  $x \in \mathcal{D}$ , set

$$\widehat{\nu}_n(x) = \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u), \quad (17)$$

a kernel estimator of the density  $\nu$  of the invariant measure of the tagged-branch chain  $Y$  of transition  $\mathcal{Q}$ .

**Proposition 18.** *Let  $\gamma \in (0, 1/2)$  and  $\ell > 0$ , let  $r > 0$  and  $\lambda > 3$ . Specify  $\widehat{\nu}_n$  with a kernel  $K$  satisfying Assumption 4 for some  $n_0 > 0$  and*

$$h_n = |\mathbb{T}_n|^{-1/(2\beta+1)}.$$

*For every  $L' > 0$  and  $0 < \beta < n_0$ , for every  $G$  such that  $(G_0, G_1) \in (\mathcal{G}(r, \lambda) \cap \mathcal{H}_s^\beta(L'))^2$  satisfy Assumptions 2 and 3, for every compact interval  $\mathcal{D} \subset \mathbb{S}$  with non-empty interior and every  $x$  in the interior of  $\mathcal{D}$ ,*

$$\sup_{(f_0, f_1)} \mathbb{E}_\mu \left[ \left( \widehat{\nu}_n(x) - \nu(x) \right)^2 \right] \lesssim |\mathbb{T}_n|^{\frac{-2\beta}{2\beta+1}}$$

*where the supremum is taken among all functions  $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$ , for any initial probability measure  $\mu(dx)$  on  $\mathbb{S}$  for  $X_\emptyset$  such that  $\mu((1 + |\cdot|)^2) < \infty$ .*

*Proof.* The usual bias-variance decomposition can be written here as

$$\begin{aligned} \mathbb{E}\left[\left(\widehat{\nu}_n(x) - \nu(x)\right)^2\right] &= \mathbb{E}\left[\left(\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) - \nu(x)\right)^2\right] \\ &= \mathbb{E}\left[\left(\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) - K_{h_n} \star \nu(x)\right)^2\right] + \left(K_{h_n} \star \nu(x) - \nu(x)\right)^2 \end{aligned}$$

where  $\star$  stands for the convolution. For  $(G_0, G_1) \in \mathcal{H}_s^\beta(L')$ , we have  $\nu \in \mathcal{H}_D^\beta(L'')$  for some  $L'' > 0$ : since  $\nu$  is invariant for  $\mathcal{Q}$ , using (4) which defines  $\mathcal{Q}$ , we can write

$$\nu(y) = \int_{\mathcal{S}} \nu(x) \mathcal{Q}(x, y) dx = \frac{1}{2} \int_{\mathcal{S}} \nu(x) \left( G_0(y - f_0(x)) + G_1(y - f_1(x)) \right) dx,$$

where we immediately see that the regularity of  $\nu$  is inherited from the regularity of  $G_0$  and  $G_1$  the marginals of the noise density  $G$ . By a Taylor expansion up to order  $[\beta]$  (recall that the number  $n_0$  of vanishing moments of  $K$  in Assumption 4 satisfies  $n_0 > \beta$ ), we obtain

$$\left(K_{h_n} \star \nu(x) - \nu(x)\right)^2 \lesssim h_n^{2\beta}, \quad (18)$$

see for instance Proposition 1.2 in Tsybakov [Tsy09]. In addition, we claim that

$$\mathbb{E}\left[\left(\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) - K_{h_n} \star \nu(x)\right)^2\right] \lesssim (|\mathbb{T}_n| h_n)^{-1}. \quad (19)$$

Choosing  $h_n = |\mathbb{T}_n|^{-1/(2\beta+1)}$  brings the announced result. Let us now prove (19) in two steps.

*Step 1. Result over one generation.* We heavily rely on the following controls:

**Lemma 19.** *Let  $F$  be a bounded function with compact support and  $G$  be a locally bounded function. For  $h_n > 0$  and  $x$  in the interior of  $\mathcal{D}$ , we define the function  $H_n : \mathcal{S} \rightarrow \mathbb{R}$  by*

$$H_n(\cdot) = F(h_n^{-1}(x - \cdot))G(\cdot).$$

For  $h_n$  such that  $x - h_n \text{supp}(F) \subset \mathcal{D}$ , we have

$$(i) \quad |\mathcal{Q}H_n|_\infty \leq h_n |F|_1 |G|_{\mathcal{D}} |\mathcal{Q}|_{\mathcal{S}, \mathcal{D}} \quad \text{and} \quad |\nu(H_n)| \leq h_n |F|_1 |G|_{\mathcal{D}} |\nu|_{\mathcal{D}}.$$

(ii) Under Assumption 2, for  $m \geq 1$ ,  $y \in \mathcal{S}$ ,

$$|\mathcal{Q}^m H_n(y) - \nu(H_n)| \lesssim h_n \wedge (\mathbb{V}(y) \rho^m)$$

up to the constant  $\max\{|F|_1 |G|_{\mathcal{D}} (|\mathcal{Q}|_{\mathcal{S}, \mathcal{D}} + |\nu|_{\mathcal{D}}), R |F|_\infty |G|_{\mathcal{D}}\}$ .

The proof of this lemma is postponed to the Appendix. Note that we have  $|\nu|_{\mathcal{D}} \leq |\mathcal{Q}|_{\mathcal{S}, \mathcal{D}} \leq |\mathcal{Q}|_{\mathcal{S}, \mathcal{S}} \leq (|G_0|_\infty + |G_1|_\infty)/2 < \infty$  (recall that  $\nu$  is invariant for  $\mathcal{Q}$  defined by (4) and  $(G_0, G_1) \in \mathcal{G}(r, \lambda)^2$ ). Set

$$H_n(\cdot) = K(h_n^{-1}(x - \cdot)) \quad \text{and} \quad \widetilde{H}_n(\cdot) = H_n(\cdot) - \nu(H_n),$$

with  $h_n$  sufficiently small such that  $x - h_n \text{supp}(K) \subset \mathcal{D}$ . Pick  $m \geq 1$ . On the one hand,

$$\mathbb{E}_\mu \left[ \sum_{u \in \mathbb{G}_m} \tilde{H}_n^2(X_u) \right] = |\mathbb{G}_m| \mu(\mathcal{Q}^m \tilde{H}_n^2) \lesssim |\mathbb{G}_m| h_n$$

relying on the many-to-one formula (14) and Lemma 19(i). Inspired by Doumic, Hoffmann, Krell and Robert [DHKR15] (proof of Proposition 8), set  $l^* = \lfloor |\log h_n| / |\log \rho| \rfloor$ . Since  $\mu(\mathcal{Q}^{m-l}(\mathcal{P}(\mathbb{V} \otimes \mathbb{V}))) < \infty$  (use  $\mathbb{V}(x) = 1 + |x|$  for  $x \in \mathcal{S}$ ,  $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$  and finally  $\mu(\mathbb{V}^2) < \infty$ , one can look at Lemmae 25 and 26 of Guyon [Guy07]), by the many-to-one formula (15),

$$\mathbb{E}_\mu \left[ \sum_{u \neq v \in \mathbb{G}_m} \tilde{H}_n(X_u) \tilde{H}_n(X_v) \right] \lesssim |\mathbb{G}_m| \left( \sum_{l=1}^{l^*} 2^{l-1} h_n^2 + \sum_{l=l^*+1}^m 2^{l-1} \rho^{2(l-1)} \right) \lesssim 2^m h_n,$$

using the first upper-bound given by Lemma 19(ii) for before  $l^*$ , the second one after  $l^*$ , and using  $\rho \in (0, 1/2)$ . Finally, we conclude that

$$\mathbb{E}_\mu \left[ \left( \sum_{u \in \mathbb{G}_m} \tilde{H}_n(X_u) \right)^2 \right] \lesssim |\mathbb{G}_m| h_n.$$

Thus, we have uniformly over  $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$ ,

$$\mathbb{E}_\mu \left[ \left( \frac{1}{|\mathbb{G}_m|} \sum_{u \in \mathbb{G}_m} K_{h_n}(x - X_u) - K_{h_n} \star \nu(x) \right)^2 \right] \lesssim (|\mathbb{G}_m| h_n)^{-1}, \quad m \geq 1. \quad (20)$$

*Step 2. Result over a subtree.* We rely on the previous inequality (20). Decomposing by generation and by the triangle inequality, we obtain

$$\begin{aligned} & \mathbb{E}_\mu \left[ \left( \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) - K_{h_n} \star \nu(x) \right)^2 \right] \\ & \leq \left( \sum_{m=0}^n \frac{|\mathbb{G}_m|}{|\mathbb{T}_n|} \left( \mathbb{E}_\mu \left[ \left( \frac{1}{|\mathbb{G}_m|} \sum_{u \in \mathbb{G}_m} K_{h_n}(x - X_u) - K_{h_n} \star \nu(x) \right)^2 \right] \right)^{1/2} \right)^2 \\ & \lesssim \left( \frac{|\mathbb{G}_0|}{|\mathbb{T}_n|} h_n^{-1} + \sum_{m=1}^n \frac{|\mathbb{G}_m|}{|\mathbb{T}_n|} (|\mathbb{G}_m| h_n)^{-1/2} \right)^2 \lesssim (|\mathbb{T}_n| h_n)^{-1}. \end{aligned}$$

This proves inequality (19) we claimed and the proof is now complete. Note that we have removed the log-term which appears in Proposition 8 of Doumic *et al.* [DHKR15].  $\square$

**Proposition 20.** *In the same setting as in Proposition 18,*

$$\hat{\nu}_n(x) \rightarrow \nu(x), \quad \mathbb{P}_\mu - a.s.$$

as  $n \rightarrow \infty$ .

*Proof.* Write

$$\widehat{\nu}_n(x) - \nu(x) = \left( \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) - K_{h_n} \star \nu(x) \right) + (K_{h_n} \star \nu(x) - \nu(x)).$$

From (18), we deduce that

$$|K_{h_n} \star \nu(x) - \nu(x)| \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (21)$$

Note that we could obtain (21) invoking the result stated by Theorem 2.1.1 of Prakasa Rao [Pra83], result also known as the Bochner Lemma (see section 7.1.2 of Duflo [Duf98]). Using (19) for  $h_n = |\mathbb{T}_n|^{-\alpha}$  with  $\alpha \in (0, 1)$ ,

$$\sum_{n \geq 0} \mathbb{E}_\mu \left[ \left( \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) - K_{h_n} \star \nu(x) \right)^2 \right] < \infty,$$

and by the Borel-Cantelli lemma, we deduce that

$$\left| \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) - K_{h_n} \star \nu(x) \right| \rightarrow 0, \quad \mathbb{P}_\mu - a.s.$$

as  $n \rightarrow \infty$ . Thus  $|\widehat{\nu}_n(x) - \nu(x)| \rightarrow 0$ ,  $\mathbb{P}_\mu - a.s.$  as  $n \rightarrow \infty$ .  $\square$

### 6.3 PROOF OF THEOREM 5

For  $x$  in the interior of  $\mathcal{D}$ , for  $\iota \in \{0, 1\}$ , we plan to use the decomposition

$$\begin{aligned} \widehat{f}_{\iota,n}(x) - f_\iota(x) &= \frac{M_{\iota,n}(x) + L_{\iota,n}(x)}{\widehat{\nu}_n(x) \vee \varpi_n} - \frac{(\nu f_\iota)(x)}{\nu(x)} \\ &= \frac{M_{\iota,n}(x)}{\widehat{\nu}_n(x) \vee \varpi_n} + \frac{L_{\iota,n}(x) - (\nu f_\iota)(x)}{\widehat{\nu}_n(x) \vee \varpi_n} - \frac{\widehat{\nu}_n(x) \vee \varpi_n - \nu(x)}{\widehat{\nu}_n(x) \vee \varpi_n} f_\iota(x) \end{aligned}$$

with

$$M_{\iota,n}(x) = \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) \varepsilon_{u\iota}, \quad (22)$$

$$L_{\iota,n}(x) = \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) f_\iota(X_u). \quad (23)$$

Thus

$$\begin{aligned} \mathbb{E}_\mu \left[ (\widehat{f}_{\iota,n}(x) - f_\iota(x))^2 \right] &\lesssim \varpi_n^{-2} \left( \mathbb{E}_\mu \left[ (M_{\iota,n}(x))^2 \right] + \mathbb{E}_\mu \left[ (L_{\iota,n}(x) - (\nu f_\iota)(x))^2 \right] \right. \\ &\quad \left. + \mathbb{E}_\mu \left[ (\widehat{\nu}_n(x) \vee \varpi_n - \nu(x))^2 \right] \right) \end{aligned} \quad (24)$$

using  $|f_\iota|_{\mathcal{D}} < \infty$ , uniformly over the class  $\mathcal{F}(\gamma, \ell)$  for  $\mathcal{D}$  compact interval. We successively treat the three terms in Steps from 1 to 3.



*Step 1. Term  $M_{\iota,n}(x)$ .* For all  $m \geq 1$  and  $\iota \in \{0, 1\}$  fixed, the sequence  $(\varepsilon_{u\iota}, u \in \mathbb{G}_m)$  is a family independent random variables such that  $\mathbb{E}[\varepsilon_{u\iota}^2] = \sigma_\iota^2$ . Thus

$$\begin{aligned} \mathbb{E}_\mu \left[ \left( \frac{1}{|\mathbb{G}_m|} \sum_{u \in \mathbb{G}_m} K_{h_n}(x - X_u) \varepsilon_{u\iota} \right)^2 \right] &= \frac{\sigma_\iota^2}{|\mathbb{G}_m| h_n^2} \mathbb{E}_\mu \left[ \frac{1}{|\mathbb{G}_m|} \sum_{u \in \mathbb{G}_m} K^2(h_n^{-1}(x - X_u)) \right] \\ &= \frac{\sigma_\iota^2}{|\mathbb{G}_m| h_n^2} \mu \left( \mathcal{Q}^m K^2(h_n^{-1}(x - \cdot)) \right) \lesssim (|\mathbb{G}_m| h_n)^{-1} \end{aligned}$$

by the many-to-one formula (14) and using Lemma 19(i). The result over a subtree follows by the triangle inequality (as in Step 2 of the proof of Proposition 18),

$$\mathbb{E}_\mu \left[ (M_{\iota,n}(x))^2 \right] \lesssim (|\mathbb{T}_n| h_n)^{-1}. \quad (25)$$

*Step 2. Term  $L_{\iota,n}(x)$ .* By usual the bias-variance decomposition,

$$\begin{aligned} \mathbb{E}_\mu \left[ (L_{\iota,n}(x) - (\nu f_\iota)(x))^2 \right] &= \mathbb{E}_\mu \left[ \left( \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) f_\iota(X_u) - K_{h_n} \star (\nu f_\iota)(x) \right)^2 \right] \\ &\quad + (K_{h_n} \star (\nu f_\iota)(x) - (\nu f_\iota)(x))^2. \end{aligned}$$

First, since  $(\nu f_\iota) \in \mathcal{H}_\mathcal{D}^\beta(L'')$  for some constant  $L'' > 0$  and since Assumption 4 is valid with  $n_0 > \beta$ ,

$$(K_{h_n} \star (\nu f_\iota)(x) - (\nu f_\iota)(x))^2 \lesssim h_n^{2\beta}.$$

Secondly, we do the same study as in the proof of Proposition 18 for

$$H_n(\cdot) = K(h_n^{-1}(x - \cdot)) f_\iota(\cdot),$$

relying on Lemma 19 (using  $|f_\iota|_\mathcal{D} < \infty$  uniformly over the class  $\mathcal{F}(\gamma, \ell)$ , with  $\mathcal{D}$  compact interval), with  $h_n$  sufficiently small such that  $x - h_n \text{supp}(K) \subset \mathcal{D}$ . We obtain

$$\mathbb{E}_\mu \left[ \left( \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) f_\iota(X_u) - K_{h_n} \star (\nu f_\iota)(x) \right)^2 \right] \lesssim (|\mathbb{T}_n| h_n)^{-1}.$$

Thus

$$\mathbb{E}_\mu \left[ (L_{\iota,n}(x) - (\nu f_\iota)(x))^2 \right] \lesssim h_n^{2\beta} + (|\mathbb{T}_n| h_n)^{-1}. \quad (26)$$

*Step 3. Denominator  $\widehat{\nu}_n(x) \vee \varpi_n$ .* We prove the following lemma in Appendix.

**Lemma 21.** *Let  $\gamma \in (0, 1/2)$  and  $\ell > 0$ , let  $r > 0$  and  $\lambda > 3$ . For every  $G$  such that  $(G_0, G_1) \in \mathcal{G}(r, \lambda)^2$  satisfy Assumptions 2 and 3, there exists  $d = d(\gamma, \ell, G_0, G_1) > 0$  such that for every  $\mathcal{D} \subset [-d, d]$ ,*

$$\inf_{(f_0, f_1)} \inf_{x \in \mathcal{D}} \nu(x) > 0$$

where the infimum is taken among all functions  $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$ .

Relying first on Lemma 21, we choose  $n$  large enough such that

$$0 < \varpi_n \leq \frac{1}{2} \inf_{(f_0, f_1)} \inf_{x \in \mathcal{D}} \nu(x), \quad (27)$$

and we get

$$\mathbb{E}_\mu \left[ \left( \widehat{\nu}_n(x) \vee \varpi_n - \nu(x) \right)^2 \right] \lesssim \mathbb{E}_\mu \left[ \left( \widehat{\nu}_n(x) - \nu(x) \right)^2 \right] \lesssim h_n^{2\beta} + (|\mathbb{T}_n| h_n)^{-1}, \quad (28)$$

uniformly over  $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$ , using the upper-bound obtained in the proof of Proposition 18 for the second inequality.

Finally, gathering (25), (26) and (28) in (24) and choosing  $h_n = |\mathbb{T}_n|^{-1/(2\beta+1)}$  we obtain the asserted result.

#### 6.4 PROOF OF THEOREM 6

In the following,  $\mathbb{P}_{(f_0, f_1)}^n$  will denote the law on  $\mathcal{S}^{|\mathbb{T}_{n+1}|}$  of the vector  $(X_u, u \in \mathbb{T}_{n+1})$ , NBAR process, in the sense of Definition 1, driven by the autoregressive functions  $f_0$  and  $f_1$  with initial probability measure  $\mu(dx)$  on  $\mathcal{S}$  for  $X_\emptyset$  and with a Gaussian noise *i.e.*

$$G(x, y) = (2\pi(\sigma_0^2\sigma_1^2(1-\rho^2)))^{-1/2} \exp\left(-\frac{\sigma_1^2 x^2 - 2\sigma_0\sigma_1\rho xy + \sigma_0^2 y^2}{2(\sigma_0^2\sigma_1^2(1-\rho^2))}\right), \quad (x, y) \in \mathcal{S}^2,$$

with  $\sigma_0, \sigma_1 > 0$  and  $\rho \in (-1, 1)$ . When  $f_0 = f_1 = f$ , we shorten  $\mathbb{P}_{(f_0, f_1)}^n$  into  $\mathbb{P}_f^n$ . We denote by  $\mathbb{E}_f^n[\cdot]$  the expectation with respect to  $\mathbb{P}_f^n$ .

*Step 1.* Let  $\delta > 0$ . Fix  $f_0 = f_1 = f^*$  with  $f^* \in \mathcal{F}(\gamma, \ell) \cap \mathcal{H}_D^\beta(L - \delta)$  and  $x \in \mathcal{D}$ . Then, for large enough  $n$ , setting  $h_n = |\mathbb{T}_n|^{-1/(2\beta+1)}$ , we construct a perturbation  $(f_{0,n}, f_{1,n})$  of  $(f_0, f_1)$  defined by

$$f_{0,n}(y) = f_{1,n}(y) = f_n^*(y) = f^*(y) + ah_n^\beta K(h_n^{-1}(x - y)), \quad y \in \mathcal{S},$$

for some smooth kernel  $K$  with compact support such that  $K(0) = 1$ , and for some  $a = a_{\delta, K} > 0$  chosen in such a way that  $f_n \in \mathcal{F}(\gamma, \ell) \cap \mathcal{H}_D^\beta(L)$ . Note that at point  $y = x$ ,  $|f_{0,n}(x) - f_0(x)| = |f_{1,n}(x) - f_1(x)| = a_{\delta, K} h_n^\beta = a_{\delta, K} |\mathbb{T}_n|^{-\beta/(2\beta+1)}$ .

*Step 2.* In the sequel, to shorten expressions, we set

$$|(\widehat{f}_{0,n}, \widehat{f}_{1,n}) - f| = |\widehat{f}_{0,n}(x) - f(x)| + |\widehat{f}_{1,n}(x) - f(x)|.$$

For arbitrary estimators  $(\widehat{f}_{0,n}(x), \widehat{f}_{1,n}(x))$  and a constant  $C > 0$ , the maximal risk is bounded below by

$$\begin{aligned} & \max_{f \in \{f^*, f_n^*\}} \mathbb{P}_f^n \left( |\mathbb{T}_n|^{\beta/(2\beta+1)} |(\widehat{f}_{0,n}, \widehat{f}_{1,n}) - f| \geq C \right) \\ & \geq \frac{1}{2} \left( \mathbb{P}_{f^*}^n \left( |\mathbb{T}_n|^{\beta/(2\beta+1)} |(\widehat{f}_{0,n}, \widehat{f}_{1,n}) - f^*| \geq C \right) + \mathbb{P}_{f_n^*}^n \left( |\mathbb{T}_n|^{\beta/(2\beta+1)} |(\widehat{f}_{0,n}, \widehat{f}_{1,n}) - f_n^*| \geq C \right) \right) \\ & \geq \frac{1}{2} \mathbb{E}_{f^*}^n \left[ \mathbf{1}_{\{|\mathbb{T}_n|^{\beta/(2\beta+1)} |(\widehat{f}_{0,n}, \widehat{f}_{1,n}) - f^*| \geq C\}} + \mathbf{1}_{\{|\mathbb{T}_n|^{\beta/(2\beta+1)} |(\widehat{f}_{0,n}, \widehat{f}_{1,n}) - f_n^*| \geq C\}} \right] - \frac{1}{2} \|\mathbb{P}_{f^*}^n - \mathbb{P}_{f_n^*}^n\|_{TV}. \end{aligned}$$

By the triangle inequality, we have

$$|\mathbb{T}_n|^{\beta/(2\beta+1)} \left( |(\widehat{f}_{0,n}, \widehat{f}_{1,n}) - f^*| + |(\widehat{f}_{0,n}, \widehat{f}_{1,n}) - f_n^*| \right) \geq 2|\mathbb{T}_n|^{\beta/(2\beta+1)} |f_n^*(x) - f^*(x)| = 2a_{\delta,K}$$

by Step 1, so if we now take  $C < a_{\delta,K}/4$ , one of the two indicators within the expectation above must be equal to one with full  $\mathbb{P}_{f^*}^n$ -probability. In that case,

$$\max_{f \in \{f^*; f_n^*\}} \mathbb{P}_f^n \left( |\mathbb{T}_n|^{\beta/(2\beta+1)} |(\widehat{f}_{0,n}, \widehat{f}_{1,n}) - f| \geq C \right) \geq \frac{1}{2} (1 - \|\mathbb{P}_{f^*}^n - \mathbb{P}_{f_n^*}^n\|_{TV})$$

and Theorem 6 is thus proved if  $\limsup_{n \rightarrow \infty} \|\mathbb{P}_{f^*}^n - \mathbb{P}_{f_n^*}^n\|_{TV} < 1$ .

*Step 3.* By the Pinsker inequality, we have  $\|\mathbb{P}_{f^*}^n - \mathbb{P}_{f_n^*}^n\|_{TV} \leq \frac{\sqrt{2}}{2} \left( \mathbb{E}_{f^*}^n \left[ \ln \frac{d\mathbb{P}_{f^*}^n}{d\mathbb{P}_{f_n^*}^n} \right] \right)^{1/2}$  and the log-likelihood ratio can be written

$$\begin{aligned} \mathbb{E}_{f^*}^n \left[ \ln \left( \frac{d\mathbb{P}_{f^*}^n}{d\mathbb{P}_{f_n^*}^n} \right) \right] &= \mathbb{E}_{f^*}^n \left[ \sum_{u \in \mathbb{T}_n} \ln \left( \frac{G(X_{u0} - f^*(X_u), X_{u1} - f^*(X_u))}{G(X_{u0} - f_n^*(X_u), X_{u1} - f_n^*(X_u))} \right) \right] \\ &= - \sum_{u \in \mathbb{T}_n} \mathbb{E}_{f^*}^n \left[ \frac{(\sigma_1^2 - \sigma_0\sigma_1\rho)\varepsilon_{u0} + (\sigma_0^2 - \sigma_0\sigma_1\rho)\varepsilon_{u1}}{\sigma_0^2\sigma_1^2(1-\rho^2)} (f_n^* - f^*)(X_u) \right] \\ &\quad + \frac{\sigma_0^2 + \sigma_1^2 - 2\sigma_0\sigma_1\rho}{2\sigma_0^2\sigma_1^2(1-\rho^2)} \sum_{u \in \mathbb{T}_n} \mathbb{E}_{f^*}^n \left[ (f_n^* - f^*)^2(X_u) \right], \end{aligned}$$

since  $G$  is chosen to be the bivariate Gaussian density and, under  $\mathbb{P}_{f^*}^n$ , we know  $X_{u0} = f^*(X_u) + \varepsilon_{u0}$  and  $X_{u1} = f^*(X_u) + \varepsilon_{u1}$ . Recall now that  $X_u$  is independent of  $(\varepsilon_{u0}, \varepsilon_{u1})$  which is centered. Thus

$$\begin{aligned} \mathbb{E}_{f^*}^n \left[ \ln \left( \frac{d\mathbb{P}_{f^*}^n}{d\mathbb{P}_{f_n^*}^n} \right) \right] &= \frac{\sigma_0^2 + \sigma_1^2 - 2\sigma_0\sigma_1\rho}{2\sigma_0^2\sigma_1^2(1-\rho^2)} \sum_{u \in \mathbb{T}_n} \mathbb{E}_{f^*}^n \left[ (f_n^* - f^*)^2(X_u) \right] \\ &= \frac{\sigma_0^2 + \sigma_1^2 - 2\sigma_0\sigma_1\rho}{2\sigma_0^2\sigma_1^2(1-\rho^2)} \sum_{m=0}^n |\mathbb{G}_m| \mu \left( \mathcal{Q}_{f^*}^m \left( (f^* - f_n^*)^2 \right) \right) \end{aligned} \quad (29)$$

using the many-to-one formula (14), with

$$\mathcal{Q}_{f^*}(x, y) = \frac{1}{2} \left( G_0(y - f^*(x)) + G_1(y - f^*(x)) \right),$$

where  $G_0$  and  $G_1$  are the marginals of  $G$ . Since

$$\mathcal{Q}_{f^*} \left( (f^* - f_n^*)^2 \right) (y) = a_{\delta,K}^2 h_n^{2\beta} \int_{\mathcal{S}} K^2(h_n^{-1}(x-z)) \mathcal{Q}_{f^*}(y, z) dz \leq a_{\delta,K}^2 |K|_{\mathcal{S}}^2 |\mathcal{Q}_{f^*}|_{\mathcal{S},\mathcal{S}} h_n^{2\beta+1}$$

where  $|\mathcal{Q}_{f^*}|_{\mathcal{S},\mathcal{S}}$  only depends on  $\sigma_0, \sigma_1$  and  $\rho$ . Gathering this last upper-bound together with (29) and the Pinsker inequality, we finally get, with our choice of  $h_n$ ,

$$\|\mathbb{P}_{f^*}^n - \mathbb{P}_{f_n^*}^n\|_{TV} \lesssim a_{\delta,K}^2$$

and this term can be made arbitrarily small by picking  $a_{K,\delta}$  small enough.

## 6.5 PROOF OF PROPOSITION 7

For the choice  $h_n = |\mathbb{T}_n|^{-\alpha}$  with  $\alpha \in (0, 1)$ , relying successively on (25) and (26), we deduce

$$\sum_{n \geq 0} \mathbb{E}_\mu \left[ (M_{l,n}(x))^2 \right] < \infty \quad \text{and} \quad \sum_{n \geq 0} \mathbb{E}_\mu \left[ (L_{l,n}(x) - (\nu f_l)(x))^2 \right] < \infty.$$

Thus, as  $n \rightarrow \infty$ ,

$$M_{l,n}(x) \rightarrow 0, \quad \mathbb{P}_\mu - a.s. \quad \text{and} \quad L_{l,n}(x) \rightarrow (\nu f_l)(x), \quad \mathbb{P}_\mu - a.s.$$

From Proposition 20 and since  $\varpi_n \rightarrow 0$ ,

$$\widehat{\nu}_n(x) \vee \varpi_n \rightarrow \nu(x), \quad \mathbb{P}_\mu - a.s.$$

We conclude reminding that

$$\widehat{f}_{l,n}(x) = \frac{M_{l,n}(x) + L_{l,n}(x)}{\widehat{\nu}_n(x) \vee \varpi_n}.$$

## 6.6 PROOFS OF THEOREMS 8 AND 9

*Proof of Theorem 8.* Set  $x$  in the interior of  $\mathcal{D}$ . The strategy is to use the following decomposition, which is slightly different from the one used in the proof of Theorem 5,

$$\begin{aligned} \sqrt{|\mathbb{T}_n| h_n} \begin{pmatrix} \widehat{f}_{0,n}(x) - f_0(x) \\ \widehat{f}_{1,n}(x) - f_1(x) \end{pmatrix} &= \frac{1}{\widehat{\nu}_n(x) \vee \varpi_n} \left\{ \sqrt{|\mathbb{T}_n| h_n} \begin{pmatrix} M_{0,n}(x) \\ M_{1,n}(x) \end{pmatrix} \right. \\ &\quad \left. + \sqrt{|\mathbb{T}_n| h_n} \begin{pmatrix} N_{0,n}(x) \\ N_{1,n}(x) \end{pmatrix} + \sqrt{|\mathbb{T}_n| h_n} \begin{pmatrix} R_{0,n}(x) \\ R_{1,n}(x) \end{pmatrix} \right\} \end{aligned}$$

where, for  $\iota \in \{0, 1\}$ ,  $M_{l,n}(x)$  is defined by (22),

$$N_{l,n}(x) = \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) (f_\iota(X_u) - f_\iota(x)), \quad (30)$$

$$R_{l,n}(x) = \left( \widehat{\nu}_n(x) - (\widehat{\nu}_n(x) \vee \varpi_n) \right) f_\iota(x), \quad (31)$$

and  $\widehat{\nu}_n(x)$  is defined by (17). The first part of the decomposition is called main term, the second part negligible term and the third part is a remainder term due to the truncation of the denominator of the estimators. The strategy is the following: prove first that the last two terms goes to zero almost surely and prove a central limit theorem for the main term in a second step.

*Step 1. Negligible and remainder terms,  $N_{l,n}(x)$  and  $R_{l,n}(x)$ .* We use the decomposition  $N_{l,n}(x) = N_{l,n}^{(1)}(x) + N_{l,n}^{(2)}(x)$  where

$$N_{l,n}^{(1)}(x) = \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} \mathbb{E}_\nu [K_{h_n}(x - X_u) (f_\iota(X_u) - f_\iota(x))], \quad (32)$$

$$N_{i,n}^{(2)}(x) = \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} \left( K_{h_n}(x - X_u)(f_i(X_u) - f_i(x)) - \mathbb{E}_\nu [K_{h_n}(x - X_u)(f_i(X_u) - f_i(x))] \right). \quad (33)$$

We claim that

$$\sqrt{|\mathbb{T}_n| h_n} N_{i,n}^{(1)}(x) \longrightarrow 0 \quad \text{and} \quad \sqrt{|\mathbb{T}_n| h_n} N_{i,n}^{(2)}(x) \rightarrow 0, \quad \mathbb{P}_\mu - a.s.$$

as  $n \rightarrow \infty$ .

*Step 1.1.* Set

$$H_n(\cdot) = K(h_n^{-1}(x - \cdot))(f_i(\cdot) - f_i(x)) \quad (34)$$

with  $h_n$  sufficiently small such that  $x - h_n \text{supp}(K) \subset \mathcal{D}$ . By the many-to-one formula (14), after a decomposition of the subtree  $\mathbb{T}_n$  in  $\cup_{m=0}^n \mathbb{G}_m$ ,

$$N_{i,n}^{(1)}(x) = \frac{1}{|\mathbb{T}_n|} \sum_{m=0}^n |\mathbb{G}_m| \mathbb{E}_\nu [K_{h_n}(x - Y_m)(f_i(Y_m) - f_i(x))] = h_n^{-1} \nu(H_n) \quad (35)$$

since  $\nu$  is the invariant measure of the tagged-branch chain  $(Y_m)_{m \geq 0}$  and

$$\begin{aligned} \nu(H_n) &= \int_{\mathcal{S}} K(h_n^{-1}(x - y))(f_i(y) - f_i(x)) \nu(y) dy \\ &= h_n \int_{\mathcal{S}} K(y)(f_i(x - h_n y) - f_i(x)) \nu(x - h_n y) dy \\ &= h_n \int_{\mathcal{S}} K(y) \left( ((\nu f_i)(x - h_n y) - (\nu f_i)(x)) - (\nu(x - h_n y) - \nu(x)) f_i(x) \right) dy. \end{aligned}$$

We now use that both  $(\nu f_i)$  and  $f_i$  have derivatives up to order  $\lfloor \beta \rfloor$ . Also remind that  $K$  is of order  $n_0 > \beta$ . By a Taylor expansion, for some  $\vartheta$  and  $\vartheta' \in (0, 1)$ ,

$$\begin{aligned} \nu(H_n) &= h_n \int_{\mathcal{S}} K(y) \left( \frac{(-h_n y)^{\lfloor \beta \rfloor}}{\lfloor \beta \rfloor!} (\nu f_i)^{\lfloor \beta \rfloor}(x - \vartheta h_n y) - \frac{(-h_n y)^{\lfloor \beta \rfloor}}{\lfloor \beta \rfloor!} \nu^{\lfloor \beta \rfloor}(x - \vartheta' h_n y) f_i(x) \right) dy \quad (36) \\ &= h_n \int_{\mathcal{S}} K(y) \frac{(-h_n y)^{\lfloor \beta \rfloor}}{\lfloor \beta \rfloor!} \left( ((\nu f_i)^{\lfloor \beta \rfloor}(x - \vartheta h_n y) - (\nu f_i)^{\lfloor \beta \rfloor}(x)) \right. \\ &\quad \left. - (\nu^{\lfloor \beta \rfloor}(x - \vartheta' h_n y) - \nu^{\lfloor \beta \rfloor}(x)) f_i(x) \right) dy. \end{aligned}$$

Thus, using  $(\nu f_i) \in \mathcal{H}_{\mathcal{D}}^\beta(L'')$  and  $\nu \in \mathcal{H}_{\mathcal{D}}^\beta(L'')$  for some  $L'' > 0$ ,

$$|\nu(H_n)| \leq h_n \int_{\mathcal{S}} |K(y)| \frac{|h_n y|^{\lfloor \beta \rfloor}}{\lfloor \beta \rfloor!} \left( (L'' |\vartheta h_n y|^{\lfloor \beta \rfloor}) + (L'' |\vartheta' h_n y|^{\lfloor \beta \rfloor}) f_i(x) \right) dy \lesssim h_n^{1+\beta}.$$

Hence, recalling (35),  $N_{i,n}^{(1)}(x) \lesssim h_n^\beta$  and  $\sqrt{|\mathbb{T}_n| h_n} N_{i,n}^{(1)}(x)$  goes to zero when  $n$  goes to infinity choosing  $h_n = |\mathbb{T}_n|^{-\alpha}$  with  $\alpha > 1/(1 + 2\beta)$ .

*Step 1.2.* In the same way we have proved  $|\nu(H_n)| \lesssim h_n^{1+\beta}$ , we prove  $|\mathcal{Q}H_n(y)| \lesssim h_n^{1+\beta}$  using the fact that  $z \rightsquigarrow f_i(z) \mathcal{Q}(y, z)$  and  $z \rightsquigarrow \mathcal{Q}(y, z)$  belong to  $\mathcal{H}_{\mathcal{D}}^\beta(L'')$  for some other  $L'' > 0$  for any fixed  $y \in \mathcal{S}$ . This enables us to reinforce the inequality of Lemma 19(i) and using Lemma 19(ii) we

obtain

$$|\mathcal{Q}^l H_n(y) - \nu(H_n)| \lesssim h_n^{1+\beta} \wedge (\mathbb{V}(y)\rho^l), \quad l \geq 1.$$

It brings the following upper-bound using the same technique as in Step 1 and Step 2 of Proposition 18:

$$\mathbb{E}_\mu [(N_{\iota,n}^{(2)}(x))^2] \lesssim h_n^\beta (|\mathbb{T}_n| h_n)^{-1},$$

which, by the Borel-Cantelli lemma, leads to the  $\mathbb{P}_\mu$ -almost sure convergence of  $\sqrt{|\mathbb{T}_n| h_n} N_{\iota,n}^{(2)}(x)$  to zero, choosing  $h_n = |\mathbb{T}_n|^{-\alpha}$  with  $\alpha > 0$ .

*Step 1.3.* To end the first step of the proof, we prove that the remainder term is such that

$$\sqrt{|\mathbb{T}_n| h_n} R_{\iota,n}(x) \rightarrow 0, \quad \mathbb{P}_\mu - a.s. \quad (37)$$

as  $n \rightarrow \infty$ . Write

$$R_{\iota,n}(x) = (\hat{\nu}_n(x) - \varpi_n) f_\iota(x) \mathbf{1}_{\{\hat{\nu}_n(x) < \varpi_n\}} \quad (38)$$

where  $(\hat{\nu}_n(x) - \varpi_n)$  converges  $\mathbb{P}_\mu$ -almost surely to  $\nu(x)$ . We can easily prove that  $\mathbf{1}_{\{\hat{\nu}_n(x) < \varpi_n\}}$  converges  $\mathbb{P}_\mu$ -almost surely to 0, since  $\hat{\nu}_n(x)$  converges  $\mathbb{P}_\mu$ -almost surely to  $\nu(x)$  and  $\varpi_n \rightarrow 0$ , which means that  $\mathbf{1}_{\{\hat{\nu}_n(x) < \varpi_n\}}$  is null  $\mathbb{P}_\mu$ -almost surely beyond some integer. So

$$\sqrt{|\mathbb{T}_n| h_n} \mathbf{1}_{\{\hat{\nu}_n(x) < \varpi_n\}} = 0, \quad \mathbb{P}_\mu - a.s.$$

beyond some integer and (37) is thus proved.

*Step 2. Main term  $M_{\iota,n}(x)$ .* We will make use of the central limit theorem for martingale triangular arrays (see for instance Duflo [Duf98], Theorem 2.1.9, p 46). We follow Delmas and Marsalle [DM10] (section 4) in order to define the notion of the  $n$  first individuals of  $\mathbb{T}$ . Let  $(\Pi_m^*)_{m \geq 1}$  be independent random variables, where for each  $m$ ,  $\Pi_m^*$  is uniformly distributed over the set of permutations of  $\mathbb{G}_m$ . The collection  $(\Pi_m^*(1), \dots, \Pi_m^*(|\mathbb{G}_m|))$  is a random drawing without replacement of all the elements of  $\mathbb{G}_m$ . For  $k \geq 1$ , set  $\rho_k = \inf\{k' \geq 0, k \leq |\mathbb{T}_{k'}|\}$  (it can be seen as the number of generation to which belongs the  $k$ -th element of  $\mathbb{T}$ ). We finally define a random order on  $\mathbb{T}$  through  $\tilde{\Pi}$  the function from  $\{1, 2, \dots\}$  to  $\mathbb{T}$  such that  $\tilde{\Pi}(1) = \emptyset$  and for  $k \geq 2$ ,  $\tilde{\Pi}(k) = \Pi_{\rho_k}^*(k - |\mathbb{T}_{\rho_k-1}|)$ . We introduce the filtration  $\mathcal{G} = (\mathcal{G}_n, n \geq 0)$  defined by  $\mathbb{G}_0 = \sigma(X_\emptyset)$  and for each  $n \geq 1$ ,

$$\mathcal{G}_n = \sigma\left(\left(X_{\tilde{\Pi}(k)}, X_{(\tilde{\Pi}(k),0)}, X_{(\tilde{\Pi}(k),1)}, 1 \leq k \leq n\right), (\tilde{\Pi}(k), 1 \leq k \leq n+1)\right).$$

For  $n \geq 1$ , we consider the vector of bivariate random variables  $\overline{E}^{(n)}(x) = (\overline{E}_k^{(n)}(x), 1 \leq k \leq |\mathbb{T}_n|)$  defined by

$$\overline{E}_k^{(n)} = \sum_{l=1}^k E_l^{(n)}(x) \quad \text{with} \quad E_l^{(n)}(x) = (|\mathbb{T}_n| h_n)^{-1/2} \begin{pmatrix} K(h_n^{-1}(x - X_{\tilde{\Pi}(l)})) \varepsilon_{(\tilde{\Pi}(l),0)} \\ K(h_n^{-1}(x - X_{\tilde{\Pi}(l)}) \varepsilon_{(\tilde{\Pi}(l),1)} \end{pmatrix}. \quad (39)$$

Notice that  $\overline{E}^{(n)}(x)$  is a square-integrable martingale adapted to  $\mathcal{G} = (\mathcal{G}_n)_{n \geq 0}$ . Then,  $(\overline{E}^{(n)}(x),$

$n \geq 1$ ) is a square-integrable  $\mathcal{G}$ -martingale triangular array whose bracket is given by

$$\langle \bar{E}^{(n)}(x) \rangle_k = \sum_{l=1}^k \mathbb{E} \left[ E_l^{(n)}(x) (E_l^{(n)}(x))^t \middle| \mathcal{G}_{l-1} \right] = \left( \frac{1}{|\mathbb{T}_n| h_n} \sum_{l=1}^k K^2(h_n^{-1}(x - X_{\tilde{\Pi}(l)})) \right) \Gamma$$

where  $\Gamma$  is the noise covariance matrix. We apply Proposition 20 (with  $|K|_2^{-2} K^2$  replacing  $K$  as kernel function) and we obtain

$$\langle \bar{E}^{(n)}(x) \rangle_{|\mathbb{T}_n|} = \left( \frac{1}{|\mathbb{T}_n| h_n} \sum_{u \in \mathbb{T}_n} K^2(h_n^{-1}(x - X_u)) \right) \Gamma \rightarrow |K|_2^2 \nu(x) \Gamma, \quad \mathbb{P}_\mu - a.s. \quad (40)$$

as  $n \rightarrow \infty$ . Condition (A1) of Theorem 2.1.9 of [Duf98] is satisfied, this is exactly (40). Since the bivariate random variables  $((\varepsilon_{u0}, \varepsilon_{u1}), u \in \mathbb{T})$  are independent and identically distributed and since  $\varepsilon_0$  and  $\varepsilon_1$  have finite moment of order four,

$$\begin{aligned} \sum_{k=1}^{|\mathbb{T}_n|} \mathbb{E} \left[ \|\bar{E}_k^{(n)}(x) - \bar{E}_{k-1}^{(n)}(x)\|^4 \middle| \mathcal{G}_{k-1} \right] &= \frac{\mathbb{E}[(\varepsilon_0^2 + \varepsilon_1^2)^2]}{(|\mathbb{T}_n| h_n)^2} \sum_{k=1}^{|\mathbb{T}_n|} K^4(h_n^{-1}(x - X_{\tilde{\Pi}(k)})) \\ &= \frac{\mathbb{E}[(\varepsilon_0^2 + \varepsilon_1^2)^2]}{|\mathbb{T}_n| h_n} \left( \frac{1}{|\mathbb{T}_n| h_n} \sum_{u \in \mathbb{T}_n} K^4(h_n^{-1}(x - X_u)) \right) \end{aligned}$$

where  $\|\cdot\|$  denotes the Euclidian norm for vectors and setting  $\bar{E}_0^{(n)}(x) = 0$ . We apply Proposition 20 (with  $|K^2|_2^{-2} K^4$  replacing  $K$  as kernel function) and we conclude that

$$\sum_{k=1}^{|\mathbb{T}_n|} \mathbb{E} \left[ \|\bar{E}_k^{(n)}(x) - \bar{E}_{k-1}^{(n)}(x)\|^4 \middle| \mathcal{G}_{k-1} \right] \rightarrow 0, \quad \mathbb{P}_\mu - a.s. \quad (41)$$

The Lyapunov condition (41) implies the Lindeberg condition (A2) of Theorem 2.1.9 of [Duf98] (see section 2.1.4, p 47 of [Duf98]). Therefore, by the central limit theorem for martingale triangular arrays,

$$\bar{E}_{|\mathbb{T}_n|}^{(n)}(x) = \sqrt{|\mathbb{T}_n| h_n} \begin{pmatrix} M_{0,n}(x) \\ M_{1,n}(x) \end{pmatrix} \xrightarrow{d} \mathcal{N}_2(\mathbf{0}_2, |K|_2^2 \nu(x) \Gamma).$$

We conclude gathering Step 1 and Step 2, together with

$$\hat{\nu}_n(x) \vee \varpi_n \rightarrow \nu(x), \quad \mathbb{P}_\mu - a.s.$$

and the Slutsky lemma.

*Step 3. Independence.* Let  $x_1 \neq x_2 \in \mathcal{D}$ . We repete Step 2 for

$$\bar{E}_k^{(n)}(x_1, x_2) = \sum_{l=1}^k E_l^{(n)}(x_1, x_2) \quad \text{with} \quad E_l^{(n)}(x_1, x_2) = \begin{pmatrix} E_l^{(n)}(x_1) \\ E_l^{(n)}(x_2) \end{pmatrix}$$

and we are led to

$$\overline{E}_{|\mathbb{T}_n|}^{(n)}(x_1, x_2) = \sqrt{|\mathbb{T}_n|h_n} \begin{pmatrix} M_{0,n}(x_1) \\ M_{1,n}(x_1) \\ M_{0,n}(x_2) \\ M_{1,n}(x_2) \end{pmatrix} \xrightarrow{d} \mathcal{N}_4 \left( \mathbf{0}_4, |K|_2^2 \begin{pmatrix} \nu(x_1)\Gamma & \mathbf{0}_{2,2} \\ \mathbf{0}_{2,2} & \nu(x_2)\Gamma \end{pmatrix} \right)$$

with  $\mathbf{0}_4$  and  $\mathbf{0}_{2,2}$  respectively the null vector of size 4 and the null matrix of size  $2 \times 2$ , which shows asymptotic independence between  $(M_{0,n}(x_1), M_{1,n}(x_1))$  and  $(M_{0,n}(x_2), M_{1,n}(x_2))$  and thus between  $(\widehat{f}_{0,n}(x_1), \widehat{f}_{1,n}(x_1))$  and  $(\widehat{f}_{0,n}(x_2), \widehat{f}_{1,n}(x_2))$  as asserted.  $\square$

*Proof of Theorem 9. Step 1. Case  $\kappa < \infty$ .* We look carefully at the proof of Theorem 8 and see that only Step 1.1 has to be reconsidered. We prove that

$$\sqrt{|\mathbb{T}_n|h_n} N_{i,n}^{(1)}(x) \longrightarrow \kappa \nu(x) \mathbf{m}(x).$$

Indeed for  $\beta$  an integer, using (35) and (36),

$$\sqrt{|\mathbb{T}_n|h_n} N_{i,n}^{(1)}(x) = \frac{(-1)^\beta h_n^\beta \sqrt{|\mathbb{T}_n|h_n}}{\beta!} \int_{\mathcal{S}} y^\beta K(y) ((\nu f_i)^\beta(x - \vartheta h_n y) - \nu^\beta(x - \vartheta h_n y) f_i(x)) dy$$

and we conclude letting  $n$  go to infinity since  $(\nu f_i)^\beta$  and  $\nu^\beta$  are continuous.

*Step 2. Case  $\kappa = \infty$ .* With the same argument we prove that in that case,

$$h_n^{-\beta} N_{i,n}^{(1)}(x) \longrightarrow \nu(x) \mathbf{m}(x)$$

as  $n \rightarrow \infty$ . Looking at Steps 1.2 and 1.3 in the proof of Theorem 8, we obtain

$$h_n^{-\beta} N_{i,n}^{(2)}(x) \longrightarrow 0 \quad \text{and} \quad h_n^{-\beta} R_{i,n}(x) \longrightarrow 0, \quad \mathbb{P}_\mu - a.s.$$

In addition,

$$h_n^{-\beta} M_{i,n}(x) = (h_n^\beta \sqrt{|\mathbb{T}_n|h_n})^{-1} \sqrt{|\mathbb{T}_n|h_n} M_{i,n}(x) \xrightarrow{\mathbb{P}_\mu} 0,$$

since we work in the case  $h_n^\beta \sqrt{|\mathbb{T}_n|h_n} \rightarrow \kappa = \infty$  and  $\sqrt{|\mathbb{T}_n|h_n} M_{i,n}(x)$  is asymptotically Gaussian as proved previously (Step 3 in the proof of Theorem 8).  $\square$

## 6.7 PROOF OF THEOREM 13

### Preliminaries

In the setting of Theorem 13, we achieve uniform ergodicity as stated in the following

**Lemma 22** (Uniform ergodicity). *For every  $(f_0, f_1) \in \mathcal{F}(0, \ell)^2$ , for every  $G$  such that  $(G_0, G_1) \in \mathcal{G}(r, \lambda)^2$  satisfy Assumption 11, there exists a unique probability measure  $\nu$  of the form  $\nu(dx) = \nu(x)dx$  on  $(\mathcal{S}, \mathfrak{S})$ . Moreover, for every  $G$  such that  $(G_0, G_1) \in \mathcal{G}(r, \lambda)^2$  satisfy Assumption 11, there exist a constant  $R > 0$  and  $\rho \in (0, 1/2)$  such that*

$$\sup_{(f_0, f_1)} |\mathcal{Q}^m h(x) - \nu(h)| \leq R |h|_\infty \rho^m,$$



where the supremum is taken among all functions  $(f_0, f_1) \in \mathcal{F}(0, \ell)^2$ , for all  $x \in \mathcal{S}$  and  $m \geq 0$ .

The proof is postponed to the Appendix. Uniform ergodicity achieved by Lemma 22 is here crucial in order to make use of deviations inequalities as obtained in Chapter II. We recall that the ergodicity proved in Lemma 17 was not uniform in the initial point.

We will intensively use the following two concepts: super-exponential convergence and super-exponential equivalence. Let  $(Z_n)_{n \geq 0}$  be a sequence of random variables and  $Z$  a random variable with values in a Polish space  $\mathcal{S}$  endowed with its Borel  $\sigma$ -field  $\mathfrak{S}$  and endowed with  $d$  a metric which makes  $(\mathcal{S}, d)$  complete.

**Definition 23** (Super-exponential convergence). *We say that  $(Z_n)$  converges  $(s_n)$ -super-exponentially fast in probability to  $Z$  and we note  $Z_n \xrightarrow[s_n]{\text{superexp}} Z$  if*

$$\limsup_{n \rightarrow +\infty} \frac{1}{s_n} \log \mathbb{P}(d(Z_n, Z) > \delta) = -\infty$$

for any  $\delta > 0$ .

Let  $(Y_n)_{n \geq 0}$  be another sequence of random variables with values in  $\mathcal{S}$ .

**Definition 24** (Exponential equivalence, see [DZ98]). *We say that  $(Z_n)$  and  $(Y_n)$  are  $(s_n)$ -exponentially equivalent and we note  $Z_n \xrightarrow[s_n]{\text{superexp}} Y_n$  if*

$$\limsup_{n \rightarrow +\infty} \frac{1}{s_n} \log \mathbb{P}(d(Z_n, Y_n) > \delta) = -\infty.$$

for any  $\delta > 0$ .

*Proof of Theorem 13.* Set  $x$  in the interior of  $\mathcal{D}$ . In order to prove the moderate deviations principle, MDP for short, we use the decomposition

$$\begin{aligned} \frac{|\mathbb{T}_n| h_n}{b_n} \begin{pmatrix} \widehat{f}_{0,n}(x) - f_0(x) \\ \widehat{f}_{1,n}(x) - f_1(x) \end{pmatrix} &= \frac{1}{\widehat{v}_n(x) \vee \varpi_n} \left\{ \frac{|\mathbb{T}_n| h_n}{b_n} \begin{pmatrix} M_{0,n}(x) \\ M_{1,n}(x) \end{pmatrix} \right. \\ &\quad \left. + \frac{|\mathbb{T}_n| h_n}{b_n} \begin{pmatrix} N_{0,n}(x) \\ N_{1,n}(x) \end{pmatrix} + \frac{|\mathbb{T}_n| h_n}{b_n} \begin{pmatrix} R_{0,n}(x) \\ R_{1,n}(x) \end{pmatrix} \right\} \end{aligned}$$

where, for  $\iota \in \{0, 1\}$ ,  $M_{\iota,n}(x)$ ,  $N_{\iota,n}(x)$  and  $R_{\iota,n}(x)$  are defined respectively by (22), (30) and (31). The strategy of the proof is the following. After studying the denominator term  $\widehat{v}_n(x) \vee \varpi_n$  in Step 1, we will prove in Step 2 that the last two terms of the decomposition are negligible in the sense of moderate deviations which leads us to

$$\frac{|\mathbb{T}_n| h_n}{b_n} \begin{pmatrix} \widehat{f}_{0,n}(x) - f_0(x) \\ \widehat{f}_{1,n}(x) - f_1(x) \end{pmatrix} \underset{b_n^2 / (|\mathbb{T}_n| h_n)}{\text{superexp}} \frac{1}{\widehat{v}_n(x) \vee \varpi_n} \frac{|\mathbb{T}_n| h_n}{b_n} \begin{pmatrix} M_{0,n}(x) \\ M_{1,n}(x) \end{pmatrix} \quad (42)$$

in the sense of Definition 24. Consequently, this two quantities satisfy the same MDP (see Dembo and Zeitouni [DZ98], Chapter 4) and we prove a MDP for the right-hand side of (42) in Step 3.

Step 1. Denominator  $\widehat{\nu}_n(x) \vee \varpi_n$ . We claim that

$$\widehat{\nu}_n(x) \vee \varpi_n \xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} \nu(x). \quad (43)$$

On the one hand, we have seen that

$$K_{h_n} \star \nu(x) \rightarrow \nu(x)$$

as  $n \rightarrow \infty$  (see (18)). Since the previous sequence is deterministic, we conclude that

$$K_{h_n} \star \nu(x) \xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} \nu(x). \quad (44)$$

On the other hand, from Theorem 4 (ii) of Chapter II,

$$\mathbb{P}_\mu \left( \left| \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) - K_{h_n} \star \nu(x) \right| > \delta \right) \leq 2 \exp \left( \frac{-\delta^2 |\mathbb{T}_n| h_n}{c_1 + c_2 \delta} \right) \quad (45)$$

where  $c_1$  and  $c_2$  are positive constants which depend on  $K$ ,  $R$ ,  $\rho$ ,  $\mathcal{Q}$  but not on  $n$ . Note that Lemma 22 ensures we are in the setting of Chapter II. Applying the log to the two terms of (45), multiplying by  $|\mathbb{T}_n|h_n/b_n^2$  and letting  $n$  go to infinity, we are led to

$$\frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) - K_{h_n} \star \nu(x) \xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} 0 \quad (46)$$

using conditions (i) and (ii) on the sequence  $b_n$ . From the foregoing (44) and (46), we conclude that

$$\widehat{\nu}_n(x) \xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} \nu(x). \quad (47)$$

To reach (43), write the decomposition

$$|\widehat{\nu}_n(x) \vee \varpi_n - \nu(x)| = |\widehat{\nu}_n(x) - \nu(x)| \mathbf{1}_{\{\widehat{\nu}_n(x) \geq \varpi_n\}} + |\varpi_n - \nu(x)| \mathbf{1}_{\{\widehat{\nu}_n(x) < \varpi_n\}},$$

and note that it just remains to prove

$$\mathbf{1}_{\{\widehat{\nu}_n(x) < \varpi_n\}} \xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} 0. \quad (48)$$

We have

$$\mathbb{P}_\mu \left( \mathbf{1}_{\{\widehat{\nu}_n(x) < \varpi_n\}} > \delta \right) \leq \delta^{-1} \mathbb{P}_\mu \left( \widehat{\nu}_n(x) < \varpi_n \right) \leq \delta^{-1} \mathbb{P}_\mu \left( |\widehat{\nu}_n(x) - K_{h_n} \star \nu(x)| > \delta' \right)$$

with  $\delta' = \inf_{(f_0, f_1)} \inf_{x \in \mathcal{D}} K_{h_n} \star \nu(x) - \varpi_n > 0$  for  $n$  large enough under Assumption 12. Using the deviations inequality (45) with  $\delta'$ , we obtain (48). We finally have (43), which ends the first step.

Step 2. Negligible and remainder terms,  $N_{\iota,n}(x)$  and  $R_{\iota,n}(x)$ . We claim that

$$\frac{|\mathbb{T}_n| h_n}{b_n} N_{\iota,n}(x) \xrightarrow[b_n^2/(|\mathbb{T}_n| h_n)]{\text{superexp}} 0. \quad (49)$$

Remind the decomposition  $N_{\iota,n}(x) = N_{\iota,n}^{(1)}(x) + N_{\iota,n}^{(2)}(x)$  with  $N_{\iota,n}^{(1)}(x)$  and  $N_{\iota,n}^{(2)}(x)$  defined by (32) and (33). On the one hand, we have proved that  $|N_{\iota,n}^{(1)}(x)| \lesssim h_n^\beta$  (Step 1.1 in the proof of Theorem 8). Thus

$$\frac{|\mathbb{T}_n| h_n}{b_n} |N_{\iota,n}^{(1)}(x)| \lesssim \frac{|\mathbb{T}_n| h_n^{1+\beta}}{b_n}$$

and since the right-hand side of the previous inequality is deterministic and tends to 0 (condition (iii) on the sequence  $b_n$ ), we conclude that

$$\frac{|\mathbb{T}_n| h_n}{b_n} N_{\iota,n}^{(1)}(x) \xrightarrow[b_n^2/(|\mathbb{T}_n| h_n)]{\text{superexp}} 0,$$

see *e.g.* Worms [Wor01a] for more details. On the other hand, from Theorem 4 (ii) of Chapter II, the following deviations inequality holds

$$\mathbb{P}_\mu(|N_{\iota,n}^{(2)}(x)| > \delta) \leq 2 \exp\left(\frac{-\delta^2 |\mathbb{T}_n| h_n}{c'_1(h_n^\beta + h_n^{2\beta+1} |\mathbb{T}_n|) + c'_2 \delta}\right)$$

where  $c'_1$  and  $c'_2$  are positive constants which depend on  $K, R, \rho, \mathcal{Q}$  but not on  $n$ . Recalling conditions (i) and (ii) on  $b_n$  and  $h_n = |\mathbb{T}_n|^{-\alpha}$  with  $\alpha > 1/(2\beta + 1)$ , it brings

$$\frac{|\mathbb{T}_n| h_n}{b_n} N_{\iota,n}^{(2)}(x) \xrightarrow[b_n^2/(|\mathbb{T}_n| h_n)]{\text{superexp}} 0.$$

Hence (49) is proved. We also have

$$\frac{|\mathbb{T}_n| h_n}{b_n} R_{\iota,n}(x) \xrightarrow[b_n^2/(|\mathbb{T}_n| h_n)]{\text{superexp}} 0,$$

using (38) and recalling (47) and (48). Together with Step 1, it leads us to

$$\frac{1}{\tilde{\nu}_n(x) \vee \varpi_n} \frac{|\mathbb{T}_n| h_n}{b_n} \begin{pmatrix} N_{0,n}(x) \\ N_{1,n}(x) \end{pmatrix} + \frac{1}{\tilde{\nu}_n(x) \vee \varpi_n} \frac{|\mathbb{T}_n| h_n}{b_n} \begin{pmatrix} R_{0,n}(x) \\ R_{1,n}(x) \end{pmatrix} \xrightarrow[b_n^2/(|\mathbb{T}_n| h_n)]{\text{superexp}} 0$$

and finally to (42).

Step 3. Main term  $M_{\iota,n}(x)$ . Consider the  $\mathfrak{G}$ -martingale triangular array  $(\overline{E}_k^{(n)}(x), 1 \leq k \leq n, n \geq 1)$  defined by (39). In order to make use of the MDP for martingale triangular arrays (see [Wor01a, Wor01b, Puh94]), one need to check

$$\langle \overline{E}^{(n)}(x) \rangle_{|\mathbb{T}_n|} \xrightarrow[b_n^2/(|\mathbb{T}_n| h_n)]{\text{superexp}} |K|_2^2 \nu(x) \Gamma, \quad (50)$$

and

$$\sum_{k=1}^{|\mathbb{T}_n|} \mathbb{E} \left[ \|\bar{E}_k^{(n)} - \bar{E}_{k-1}^{(n)}\|^4 \middle| \mathcal{G}_{k-1} \right] \xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} 0. \quad (51)$$

The condition (51) is called exponential Lyapunov condition and it implies exponential Lindeberg condition, we refer to [Wor01a, Wor01b] for more details. Using (40) and (41), one see that to prove (50) and (51), it suffices to show

$$\frac{1}{|\mathbb{T}_n|h_n} \sum_{u \in \mathbb{T}_n} K^2(h_n^{-1}(x - X_u)) \xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} |K|_2^2 \nu(x) \quad (52)$$

and

$$\frac{1}{(|\mathbb{T}_n|h_n)^2} \sum_{u \in \mathbb{T}_n} K^4(h_n^{-1}(x - X_u)) \xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} 0. \quad (53)$$

The same argument as in Step 1 (replacing  $K$  by  $|K|_2^{-2}K^2$  or  $|K^2|_2^{-2}K^4$  in (45)) enables us to prove (52) and (53). Gathering (50) and (51) and using the truncation of the martingale  $(\bar{E}_k^{(n)}(x), 1 \leq k \leq n, n \geq 1)$  as in the proof of Theorem 5.1 of Bitseki Penda and Djellout [BD14], we conclude that

$$\frac{\sqrt{|\mathbb{T}_n|h_n}}{b_n} \bar{E}_{|\mathbb{T}_n|}^{(n)}(x) = \frac{|\mathbb{T}_n|h_n}{b_n} \begin{pmatrix} M_{0,n}(x) \\ M_{1,n}(x) \end{pmatrix}$$

satisfies a MDP on  $\mathcal{S}^2$  with speed  $b_n^2/(|\mathbb{T}_n|h_n)$  and the good rate function defined for all  $\mathbf{z} \in \mathcal{S}^2$  by

$$I_x(\mathbf{z}) = \sup_{\lambda \in \mathbb{R}^2} \left\{ \lambda^t \mathbf{z} - \frac{1}{2} \nu(x) |K|_2^2 \mathbf{z}^t \Gamma \mathbf{z} \right\} = (2|K|_2^2)^{-1} \nu(x)^{-1} \mathbf{z}^t \Gamma^{-1} \mathbf{z}.$$

Finally, using Lemma 4.1 of [Wor00] (which is a consequence of the contraction principle) and Step 1, we conclude that

$$\frac{1}{\bar{\nu}_n(x)^{\vee \varpi_n}} \frac{|\mathbb{T}_n|h_n}{b_n} \begin{pmatrix} M_{0,n}(x) \\ M_{1,n}(x) \end{pmatrix}$$

satisfies a MDP on  $\mathcal{S}^2$  with speed  $b_n^2/(|\mathbb{T}_n|h_n)$  and rate function  $J_x$  defined in Theorem 13. Reminding (42), we get the MDP stated in Theorem 13.  $\square$

## 6.8 PROOF OF THEOREM 15

On the one hand, applying Theorem 9 to the estimator  $\hat{f}_{l,n}^{(a)}$  built with the bandwidth  $h_n^{(a)} = |\mathbb{T}_n|^{-1/(2\beta+1)}$ ,

$$|\mathbb{T}_n|^{-\frac{\beta}{2\beta+1}} A_n(x) \xrightarrow{d} \mathcal{N}_2(\mathbf{m}_2(x), \mathbf{\Sigma}_2(x)) \quad \text{with} \quad A_n(x) = \begin{pmatrix} \hat{f}_{0,n}^{(a)}(x) - f_0(x) \\ \hat{f}_{1,n}^{(a)}(x) - f_1(x) \end{pmatrix}.$$

On the other hand, applying Theorem 9 to the estimator  $\widehat{f}_{\iota,n}^{(b)}$  built with the bandwidth  $h_n^{(b)} = |\mathbb{T}_n|^{-\delta/(2\beta+1)}$  for  $\delta \in (0, 1)$ ,

$$|\mathbb{T}_n|^{-\frac{(1-\delta)\beta}{2\beta+1}} B_n(x) \xrightarrow{\mathbb{P}_\mu} \mathbf{m}_2(x) \quad \text{with} \quad B_n(x) = \begin{pmatrix} \widehat{f}_{0,n}^{(b)}(x) - f_0(x) \\ \widehat{f}_{1,n}^{(b)}(x) - f_1(x) \end{pmatrix}.$$

Combining these two results, we obtain

$$|\mathbb{T}_n|^{-\frac{\beta}{2\beta+1}} \begin{pmatrix} \bar{f}_{0,n}(x) - f_0(x) \\ \bar{f}_{1,n}(x) - f_1(x) \end{pmatrix} = \frac{|\mathbb{T}_n|^{-\frac{\beta}{2\beta+1}} A_n(x) - |\mathbb{T}_n|^{-\frac{(1-\delta)\beta}{2\beta+1}} B_n(x)}{1 - |\mathbb{T}_n|^{-\frac{(1-\delta)\beta}{2\beta+1}}} \xrightarrow{d} \mathcal{N}_2(\mathbf{0}_2, \boldsymbol{\Sigma}_2(x)),$$

as announced in (12). The study of the test statistics  $W_n(x_1, \dots, x_k)$  under  $\mathcal{H}_0$  and  $\mathcal{H}_1$  then follows classical lines. We give here the main argument for  $k = 2$ . By (12) and using in addition the asymptotical independence stated in Theorem 8, we obtain

$$|\mathbb{T}_n|^{-\frac{\beta}{2\beta+1}} \begin{pmatrix} \bar{f}_{0,n}(x_1) - f_0(x_1) \\ \bar{f}_{1,n}(x_1) - f_1(x_1) \\ \bar{f}_{0,n}(x_2) - f_0(x_2) \\ \bar{f}_{1,n}(x_2) - f_1(x_2) \end{pmatrix} \xrightarrow{d} \mathcal{N}_4(\mathbf{0}_4, \boldsymbol{\Sigma}_4(x_1, x_2)) \quad \text{with} \quad \boldsymbol{\Sigma}_4(x_1, x_2) = \begin{pmatrix} \boldsymbol{\Sigma}_2(x_1) & \mathbf{0}_{2,2} \\ \mathbf{0}_{2,2} & \boldsymbol{\Sigma}_2(x_2) \end{pmatrix}.$$

Then, using the Delta-method,

$$|\mathbb{T}_n|^{-\frac{\beta}{2\beta+1}} \begin{pmatrix} (\bar{f}_{0,n}(x_1) - \bar{f}_{1,n}(x_1)) - (f_0(x_1) - f_1(x_1)) \\ (\bar{f}_{0,n}(x_2) - \bar{f}_{1,n}(x_2)) - (f_0(x_2) - f_1(x_2)) \end{pmatrix} \xrightarrow{d} \mathcal{N}_2(\mathbf{0}_2, \boldsymbol{\Sigma}_2(x_1, x_2))$$

with

$$\boldsymbol{\Sigma}_2(x_1, x_2) = |K|_2^2 (\sigma_0^2 + \sigma_1^2 - 2\sigma_0\sigma_1\varrho) \begin{pmatrix} (\nu(x_1))^{-1} & 0 \\ 0 & (\nu(x_2))^{-1} \end{pmatrix},$$

which boils down to, under  $\mathcal{H}_0$ ,

$$\frac{|\mathbb{T}_n|^{-\frac{\beta}{2\beta+1}}}{\sqrt{|K|_2^2 (\sigma_0^2 + \sigma_1^2 - 2\sigma_0\sigma_1\varrho)}} \begin{pmatrix} \nu(x_1)^{1/2} (\bar{f}_{0,n}(x_1) - \bar{f}_{1,n}(x_1)) \\ \nu(x_2)^{1/2} (\bar{f}_{0,n}(x_2) - \bar{f}_{1,n}(x_2)) \end{pmatrix} \xrightarrow{d} \mathcal{N}_2(\mathbf{0}_2, \mathbf{I}_2),$$

with  $\mathbf{I}_2$  the identity matrix of size  $2 \times 2$ . The replacement of  $\nu(\cdot)$  by its estimator  $\widehat{\nu}_n(\cdot)$  is licit by the Slutsky theorem. Thus, under  $\mathcal{H}_0$ ,  $W_n(x_1, x_2) \xrightarrow{d} \chi^2(2)$ , the chi-squared distribution with 2 degrees of freedom. Under  $\mathcal{H}_1$ , we prove that  $W_n(x_1, x_2)$  converges  $\mathbb{P}_\mu$ -almost surely to  $+\infty$  following the same lines and using

$$\sum_{l=1}^2 \widehat{\nu}_n(x_l) (\bar{f}_{0,n}(x_l) - \bar{f}_{1,n}(x_l))^2 \longrightarrow \sum_{l=1}^2 \nu(x_l) (f_0(x_l) - f_1(x_l))^2 \neq 0, \quad \mathbb{P}_\mu - a.s.$$

when  $\mathcal{H}_1$  is valid.

## 7 APPENDIX

### 7.1 PROOF OF LEMMA 16

Let us first prove (i), see Delmas and Marsalle [DM10] (Lemma 2.1). For  $u = (u_1, u_2, \dots, u_m) \in \mathbb{G}_m$ ,

$$\mathbb{E}_\mu [g(X_u)] = \mu(\mathcal{P}_{u_1} \mathcal{P}_{u_2} \dots \mathcal{P}_{u_m}(g)).$$

Then

$$\begin{aligned} \mathbb{E}_\mu \left[ \sum_{u \in \mathbb{G}_m} g(X_u) \right] &= \sum_{\substack{(u_1, \dots, u_m) \\ \in \{0,1\}^m}} \mu(\mathcal{P}_{u_1} \mathcal{P}_{u_2} \dots \mathcal{P}_{u_m}(g)) = \mu \left( \sum_{\substack{(u_1, \dots, u_m) \\ \in \{0,1\}^m}} \mathcal{P}_{u_1} \mathcal{P}_{u_2} \dots \mathcal{P}_{u_m}(g) \right) \\ &= \mu(2^m (\mathcal{P}_0 + \mathcal{P}_1)^m(g)) = |\mathbb{G}_m| \mu(\mathcal{Q}^m g) \end{aligned}$$

since  $\mathcal{Q} = (\mathcal{P}_0 + \mathcal{P}_1)/2$  and  $|\mathbb{G}_m| = 2^m$ . We also know that  $\mu(\mathcal{Q}^m g) = \mathbb{E}_\mu[g(Y_m)]$ .

We now turn to (ii). We refer to Guyon [Guy07] for another strategy of proof (proof of Equation (7)). Some notation first: for  $u = (u_1, \dots, u_m)$  and  $v = (v_1, \dots, v_n)$  in  $\mathbb{T}$ , we write  $uv = (u_1, \dots, u_m, v_1, \dots, v_n)$  for the concatenation. For  $m \geq 0$ , we denote by  $\mathcal{F}_m$  the sigma-field generated by  $(X_u, |u| \leq m)$ .

For  $m \geq 1$ , whenever  $u \neq v \in \mathbb{G}_m$ , there exist  $w \in \mathbb{T}_{m-1}$  together with  $i \neq j \in \{0,1\}$  and  $\tilde{u}, \tilde{v}$  such that  $u = wi\tilde{u}$  and  $v = wj\tilde{v}$ , where we call  $w$  the most recent common ancestor of  $u$  and  $v$ . The main argument uses consecutively a first conditioning by  $\mathcal{F}_{|w|+1}$  which lets  $X_u$  and  $X_v$  conditionally independent and a conditional many-to-one formula of kind (i), a second conditioning by  $\mathcal{F}_{|w|}$  and the definition of the  $\mathbb{T}$ -transition  $\mathcal{P}$  and finally the many-to-one formula (i),

$$\begin{aligned} \mathbb{E}_\mu \left[ \sum_{u \neq v \in \mathbb{G}_m} g(X_u)g(X_v) \right] &= 2 \sum_{l=1}^m \sum_{w \in \mathbb{G}_{m-l}} \mathbb{E}_\mu \left[ \sum_{\substack{\tilde{u} \in \mathbb{G}_{l-1} \\ \tilde{v} \in \mathbb{G}_{l-1}}} g(X_{w0\tilde{u}})g(X_{w1\tilde{v}}) \right] \\ &= 2 \sum_{l=1}^m \sum_{w \in \mathbb{G}_{m-l}} |\mathbb{G}_{l-1}|^2 \mathbb{E}_\mu \left[ \mathcal{Q}^{l-1} g(X_{w0}) \mathcal{Q}^{l-1} g(X_{w1}) \right] \\ &= 2 \sum_{l=1}^m (2^{l-1})^2 \sum_{w \in \mathbb{G}_{m-l}} \mathbb{E}_\mu \left[ \mathcal{P}(\mathcal{Q}^{l-1} g \otimes \mathcal{Q}^{l-1} g)(X_w) \right] \\ &= 2 \sum_{l=1}^m (2^{l-1})^2 |\mathbb{G}_{m-l}| \mathbb{E}_\mu \left[ \mathcal{P}(\mathcal{Q}^{l-1} g \otimes \mathcal{Q}^{l-1} g)(Y_{m-l}) \right] \\ &= 2^m \sum_{l=1}^m 2^{l-1} \mu \left( \mathcal{Q}^{m-l} (\mathcal{P}(\mathcal{Q}^{l-1} g \otimes \mathcal{Q}^{l-1} g)) \right), \end{aligned}$$

as asserted.

## 7.2 PROOF OF LEMMA 19

First,

$$\begin{aligned} |\mathcal{Q}H_n(y)| &\leq \int_{\mathcal{S}} |F(h_n^{-1}(x-z))| |G(z)| \mathcal{Q}(y,z) dz \\ &= h_n \int_{\text{supp}(F)} |F(z)| |G(x-h_n z)| \mathcal{Q}(y, x-h_n z) dz \leq h_n |G|_{\mathcal{D}} |\mathcal{Q}|_{\mathcal{S},\mathcal{D}} |F|_1 \end{aligned}$$

for  $h_n$  such that  $x - h_n \text{supp}(F) \subset \mathcal{D}$  (remind that  $x$  belongs to the interior of  $\mathcal{D}$ ). We prove in the same way the bound on  $\nu(H_n)$ . Hence we have proved (i) and we now turn to (ii). The first bound  $h_n$  obviously comes from (i) and it remains to prove the second bound  $\mathbb{V}(y)\rho^m$ . Under Assumption 2, we apply Proposition 17 to  $g = H_n/|H_n|_{\infty}$  and it brings

$$|\mathcal{Q}^m H_n(y) - \nu(H_n)| \leq R |H_n|_{\infty} \rho^m \mathbb{V}(y).$$

Since  $|H_n|_{\infty} = |F|_{\infty} |G|_{\mathcal{D}}$  for  $h_n$  such that  $x - h_n \text{supp}(F) \subset \mathcal{D}$ , we obtain the announced upper-bound.

## 7.3 PROOF OF LEMMA 21

For every  $|z| \leq d$ ,

$$\nu(z) = \int_{\mathcal{S}} \nu(y) \mathcal{Q}(y,z) dy \geq \inf_{\substack{|y| \leq M_2, \\ |z| \leq d}} \mathcal{Q}(y,z) \int_{|y| \leq M_2} \nu(y) dy.$$

On the one hand,

$$\inf_{\substack{|x| \leq M_2, \\ |y| \leq d}} \mathcal{Q}(x,y) \geq \frac{1}{2} \left\{ \inf_{\substack{|x| \leq M_2, \\ |y| \leq d}} G_0(y-f_0(x)) + \inf_{\substack{|x| \leq M_2, \\ |y| \leq d}} G_1(y-f_1(x)) \right\} \geq \delta(d + (\ell + \gamma M_2)) \geq \delta(M_3) > 0$$

if  $d > 0$  is such that  $d + (\ell + \gamma M_2) \leq M_3$ , which is possible by Assumption 3. On the other hand,

$$\begin{aligned} \int_{|y| > M_2} \nu(y) dy &= \int_{|y| > M_2} \int_{x \in \mathcal{S}} \nu(x) \mathcal{Q}(x,y) dx dy \\ &\leq |\nu|_{\infty} \int_{|y| > M_2} \int_{x \in \mathcal{S}} \frac{1}{2} \left( G_0(y-f_0(x)) + G_1(y-f_1(x)) \right) dx dy \leq \eta(M_2) \end{aligned}$$

since  $(f_0, f_1)$  belongs to  $\mathcal{F}(\gamma, \ell)^2$  and  $(G_0, G_1)$  to  $\mathcal{G}(r, \lambda)^2$ . We know in addition that  $|\nu|_{\infty} \leq |\mathcal{Q}|_{\mathcal{S},\mathcal{S}} \leq (|G_0|_{\infty} + |G_1|_{\infty})/2 < \infty$ . Using  $\eta(M_2) < 1$  brings

$$\int_{|y| \leq M_2} \nu(y) dy > 0.$$

Hence  $\nu(y) \geq \delta(M_3)(1 - \eta(M_2)) > 0$  for any  $|y| \leq d$ . We have uniformity in  $f_0$  and  $f_1$  since  $\delta(M_3)$  and  $\eta(M_2)$  are uniform bounds on the class  $\mathcal{F}(\gamma, \ell)^2$ .

#### 7.4 PROOF OF LEMMA 22

Set  $\mathcal{C} = \{y \in \mathcal{S}; |y| \leq M_4 - \ell\} \neq \emptyset$  since  $M_4 > \ell$  under Assumption 11. We prove that, for all  $y \in \mathcal{C}$ ,

$$\inf_{x \in \mathcal{S}} \mathcal{Q}(x, y) = \frac{1}{2} \left\{ \inf_{x \in \mathcal{S}} G_0(y - f_0(x)) + \inf_{x \in \mathcal{S}} G_1(y - f_1(x)) \right\} \geq \delta(M_4) > 0$$

with  $\delta(\cdot)$  defined by (8), since  $|y - f_i(x)| \leq M_4$  (for  $|f_i|_\infty \leq \ell$  and  $|y| \leq M_4 - \ell$ ). From Douc *et al.* [DMS14], Theorem 6.6 and Lemma 6.10, one can take  $\rho = 1 - |\mathcal{C}|\delta(M_4) = 1 - 2(M_4 - \ell)\delta(M_4)$  and we obtain  $\rho < 1/2$  under Assumption 11.



---

## Références du Chapitre III

---

- [AH96] H. Z. An and F.C. Huang. *The geometric ergodicity of nonlinear autoregressive model*. *Statistica Sinica*, 6 (1996), 943–956.
- [BH99] I. V. Basawa and R. M. Huggins. *Extensions of the bifurcating autoregressive model for cell lineage studies*. *Journal of Applied Probability*, 36 (1999), 1225–1233.
- [BH00] I. V. Basawa and R. M. Huggins. *Inference for the extended bifurcating autoregressive model for cell lineage studies*. *Australian & New Zealand Journal of Statistics*, 42 (2000), 423–432.
- [BZ04] I. V. Basawa and J. Zhou. *Non-Gaussian bifurcating models and quasi-likelihood estimation*. *Journal of Applied Probability*, 41 (2004), 55–64.
- [BZ05] I. V. Basawa and J. Zhou. *Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors*. *Journal of Time Series Analysis*, 26 (2005), 825–842.
- [BB13] B. Bercu and V. Blandin. *A Rademacher-Menchov approach for random coefficient bifurcating autoregressive processes*. *Stochastic Processes and their Applications*, 125 (2015), 1218–1243.
- [BdSGP09] B. Bercu, B. de Saporta and A. Gégout-Petit. *Asymptotic analysis for bifurcating autoregressive processes via a martingale approach*. *Electronic Journal of Probability*, 87 (2009), 2492–2526.
- [BL95] R. Bhattacharya and C. Lee. *On geometric ergodicity of nonlinear autoregressive models*. *Statistics & Probability Letters*, 22 (1995), 311–315.
- [Bie97] H. J. Bierens. *Kernel estimator of regressions functions*. *Advances in econometrics: Fifth world congress*, Vol. 1 (1987).
- [BD14] S. V. Bitseki Penda and H. Djellout. *Deviation inequalities and moderate deviations for estimators of parameters in bifurcating autoregressive models*. *Annales de l’IHP-PS*, 50 (2014), 806–844.
- [BDG14] S. V. Bitseki Penda, H. Djellout and A. Guillin. *Deviation inequalities, moderate deviations and some limit theorems for bifurcating Markov chains with application*. *The Annals of Applied Probability*, 24 (2014), 235–291.
- [BEG15] S. V. Bitseki Penda, M. Escobar-Bach and A. Guillin. *Transportation cost-information and concentration inequalities for bifurcating Markov chains*. arXiv:1501.06693 (2015).
- [Bla13] V. Blandin. *Asymptotic results for random coefficient bifurcating autoregressive processes*. *Statistics* 48 (2014), 1202–1232.
- [Cli07] D. B. H. Cline. *Stability of nonlinear stochastic recursions with application to nonlinear AR-GARCH models*. *Advances in Applied Probability*, 39 (2007), 462–491.
- [CS86] R. Cowan and R. G. Staudte. *The bifurcating autoregressive model in cell lineage studies*. *Biometrics*, 42 (1986), 769–783.
- [DM10] J.-F. Delmas and L. Marsalle. *Detection of cellular aging in Galton-Watson process*. *Stochastic Processes and their Applications*, 12 (2010), 2495–2519.
- [DvS05] V. Delouille and R. von Sachs. *Estimation of nonlinear autoregressive models using design-adapted wavelets*. *Annals of the Institute of Statistical Mathematics*, 57 (2005), 235–253.

- [DZ98] A. Dembo and O. Zeitouni. *Large deviations techniques and applications, second edition*. Second edition, volume 38 of *Applications of Mathematics*, Springer, 1998.
- [dSGPM11] B. de Saporta, A. Gégout-Petit and L. Marsalle. *Parameters estimation for asymmetric bifurcating autoregressive processes with missing data*. *Electronic Journal of Statistics*, 5 (2011), 1313–1353.
- [dSGPM12] B. de Saporta, A. Gégout-Petit and L. Marsalle. *Asymmetry tests for bifurcating autoregressive processes with missing data*. *Statistics & Probability Letters*, 82 (2012), 1439–1444.
- [dSGPM14] B. de Saporta, A. Gégout-Petit and L. Marsalle. *Random coefficients bifurcating autoregressive processes*. *ESAIM: Probability and Statistics*, 18 (2014), 365–399.
- [DMS14] R. Douc, E. Moulines and D. S. Stoffer. *Nonlinear Time Series*. Chapman & Hall/CRC Texts in Statistical Science, 2014.
- [DHKR15] M. Doumic, M. Hoffmann, N. Krell and L. Robert. *Statistical estimation of a growth-fragmentation model observed on a genealogical tree*. *Bernoulli*, 21 (2015), 1760–1799.
- [Duf98] M. Dufflo. *Random Iterative Models*. Springer, Berlin (1998)
- [EL93] R. L. Eubank and V. N. LaRiccia. *Testing for no effect in nonparametric regression*. *Journal of statistical planning and inference*, 36 (1993), 1–14.
- [Guy07] J. Guyon. *Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging*. *The Annals of Applied Probability*, 17 (2007), 1538–1569.
- [HM11] M. Hairer and J. Mattingly. *Yet another look at Harris’ ergodic theorem for Markov chains*. *Seminar on Stochastic Analysis, Random Fields and Applications VI, Progress in Probability*, 63 (2011), 109–117.
- [HV92] W. Härdle and P. Vieu. *Kernel regression smoothing of time series*. *Journal of Time Series Analysis*, 13 (1992), 209–232.
- [Hof99] M. Hoffmann. *On nonparametric estimation in nonlinear AR(1)-models*. *Statistics & Probability Letters*, 44 (1999), 29–45.
- [Jin11] L. Jin. *A data-driven test to compare two or multiple time series*. *Computational Statistics & Data Analysis*, 55 (2011), 2183–2196.
- [Jou06] C. Joutard. *Sharp large deviations in nonparametric estimation*. *Nonparametric Statistics*, 18 (2006), 293–306.
- [Li14] F. Li. *Comparison of autoregressive curves through partial sums of quasi-residuals*. *Contemporary Developments in Statistical Theory, Springer Proceedings in Mathematics & Statistics*, 68 (2014), 239–256.
- [LW10] W. Liu and W. B. Wu. *Simultaneous nonparametric inference for times series*. *The Annals of Statistics*, 38 (2010), 2388–2421.
- [Lou98] D. Louani. *Large deviations limit theorems for the kernel density estimator*. *Scandinavian Journal of Statistics*, 25 (1998), 243–253.
- [Lou99] D. Louani. *Some large deviations limit theorems in conditional nonparametric statistics*. *Statistics: A Journal of Theoretical and Applied Statistics*, 33 (1999), 171–196.
- [MT09] S. Meyn and R. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, 2009.
- [MP06] A. Mokkadem and M. Pelletier. *Confidence bands for densities, logarithmic point of view*. *Alea*, 2 (2006), 231–266.
- [MPT08] A. Mokkadem, M. Pelletier and B. Thiam. *Large and moderate deviations principles for kernel estimators of the multivariate regression*. *Mathematical Methods of Statistics*, 17 (2008), 1–27.
- [Mun98] A. Munk and H. Dette. *Nonparametric comparison of several regression functions: exact and asymptotic theory*. *Annals of Statistics*, 26 (1998), 2339–2368.
- [Nad64] E. A. Nadaraya. *On estimating regression*. *Theory of Probability & Its Applications*, 9 (1964), 141–142.

- 
- [Pra83] B. L. S. Prakasa Rao. *Nonparametric Functional Estimation*. Academic Press, New York (1983)
- [Puh94] A. Puhalskii. *Large deviations of semimartingales via convergence of the predictable characteristics*. Stochastics and Stochastics Reports, 49 (1994), 27–85.
- [Rob83] P. M. Robinson. *Nonparametric estimators for time series*. Journal of Time Series Analysis, 4 (1983), 185–207.
- [TRB99] J. J. Trujillo, M. Rivero and B. Bonilla. *On a Riemann-Liouville generalized Taylor's formula*. Journal of Mathematical Analysis and Applications, 231 (1999), 255–265.
- [Tsy09] A. Tsybakov. *Introduction to Nonparametric Estimation*. Springer series in statistics, Springer-Verlag, New-York, 2009.
- [Use08] D. Usero. *Fractional Taylor series for Caputo fractional derivatives. Construction of numerical schemes*. Preprint (2008)
- [Wat64] G. S. Watson. *Smooth regression analysis*. Sankhyā: The Indian Journal of Statistics, Series A, 26 (1964), 359–372.
- [Wor00] J. Worms. *Principes de déviations modérées pour des martingales et applications statistiques*. Thèse de Doctorat à l'Université Marne-la-Vallée (2000).
- [Wor01a] J. Worms. *Moderate deviations of some dependent variables. I. Martingales*. Mathematical Methods of Statistics, 10 (2001a), 38–72.
- [Wor01b] J. Worms. *Moderate deviations of some dependent variables. II. Some kernel estimators*. Mathematical Methods of Statistics, 10 (2001b), 161–193.

## CHAPITRE IV

---

### Influence de la variabilité individuelle sur le paramètre de Malthus

---

**I**SSUE directement de la biologie, la question suivante est au cœur de ce dernier chapitre : quelle est l'influence de la variabilité individuelle entre cellules sur la croissance globale de la population ? L'objectif général **(C)** *Comparer les évolutions de différentes populations à travers leur vitesse de croissance* (énoncé page 2) est ici abordé spécifiquement sous cet angle.

Des modèles plus riches sont alors considérés dans ce chapitre où la variabilité individuelle est introduite. Les modèles étudiés jusqu'ici viennent être complétés par une variable structurante ad-ditionnelle : un taux de vieillissement pour le modèle en âge du Chapitre I, un taux de croissance pour le modèle en taille comptant parmi les applications du Chapitre II. Il s'agit de comparer, pour un taux de division fixé,

- 1) la croissance d'une **population de référence** composée de cellules vieillissant ou croissant toutes au même taux,
- 2) à la croissance d'une **population dite avec variabilité** où chaque cellule possède son propre taux de vieillissement ou de croissance.

Le paramètre de Malthus, déjà rencontré dans le Chapitre I, mesure la croissance exponentielle de la population et représente donc la quantité adéquate pour cette comparaison. Nous articulons dans ce chapitre outils déterministe et stochastique pour mener à bien une telle comparaison.

## Abstract

---

The aim is to compare the global growth speed of different populations of cells measured by the Malthus parameter. The population of reference is composed of cells all growing at the same growth rate  $\bar{v}$ . The second population is composed of cells each growing at a rate  $v$  drawn according to a non degenerated distribution  $\rho$  with mean  $\bar{v}$ . Analytic answers are provided for a toy model and numerical answers based on stochastic simulations are derived for more realistic models. It appears numerically that in the size-structured model extensively studied, and in both of the size-increment-structured model recently introduced for the study of *E. coli* by Taheri-Araghi *et al.* [TAB+15] and Amir's model [Ami14], the second population grows more slowly than the population of reference. The decrease in the Malthus parameter we measure, around 2% for distributions  $\rho$  with realistic coefficients of variations around 15-20%, is determinant since it controls the *exponential* growth of the whole population. We also explore two ways to model heredity in the transmission of the growth rate and very surprisingly, the previous conclusion can be reversed.

---

## Contents

---

|          |   |            |
|----------|---|------------|
| <b>1</b> | <b>INTRODUCTION</b>   | <b>170</b> |
| 1.1      | Introducing individual variability                          | 170        |
| 1.2      | Estimating the Malthus parameter                            | 174        |
| 1.3      | Main results and outline                                    | 176        |
| <b>2</b> | <b>THE AGE-STRUCTURED MODEL WITH VARIABILITY</b>            | <b>178</b> |
| 2.1      | Description of the model                                    | 178        |
| 2.2      | The Malthus parameter                                       | 181        |
| 2.3      | Influence of variability on the Malthus parameter           | 182        |
| 2.4      | Estimation of the Malthus parameter from one branch         | 187        |
| <b>3</b> | <b>THE SIZE-INCREMENT-STRUCTURED MODEL WITH VARIABILITY</b> | <b>188</b> |
| 3.1      | Description of the model                                    | 188        |
| 3.2      | The Malthus parameter                                       | 189        |
| 3.3      | Influence of variability on the Malthus parameter           | 190        |
| 3.4      | Estimation of the Malthus parameter from one branch         | 200        |
| <b>4</b> | <b>TWO OTHER MODELS OF INTEREST</b>                         | <b>203</b> |
| 4.1      | Size-structured model with variability                      | 203        |
| 4.2      | Model introduced by Amir                                    | 205        |
| <b>5</b> | <b>PERSPECTIVES</b>   | <b>208</b> |
| <b>6</b> | <b>PROOFS</b>   | <b>209</b> |
| 6.1      | Proof of Theorem 3  | 209        |
| 6.2      | Proof of Theorem 4  | 213        |
| 6.3      | Proof of Theorem 7  | 219        |
| <b>7</b> | <b>APPENDIX</b>   | <b>222</b> |
| 7.1      | Proof of Lemma 9  | 222        |
| 7.2      | Proof of Lemma 11   | 223        |
| 7.3      | Complements on the model introduced by Amir                 | 224        |
| <b>8</b> | <b>SUPPLEMENTARY FIGURES AND TABLES</b>                     | <b>225</b> |
|          | <b>REFERENCES</b>   | <b>233</b> |

---

## 1 INTRODUCTION

Recent biological studies draw attention to the question of variability between cells. We refer to the work of Kiviet *et al.* published in 2014 [KNW+14]. A cell in a controlled culture grows at a constant rate  $v > 0$ , but that this rate can differ from one individual to another. The biological question we address in this chapter states as follows. What is the influence of the individual variability in the growth rate on the global growth of the population? The global growth of the population is measured by the Malthus parameter we define thereafter, also called in the literature *fitness*. Even if the variability in the growth rate is small, with a distribution of coefficient of variation around 10%, and even if its influence on the Malthus parameter would be still smaller, since it characterises the exponential global growth of the population such an influence may become determinant. Thus we address two main questions in this work,

- (i) What is the influence of individual variability on the global growth of the population measured by the Malthus parameter?
- (ii) How can we estimate the Malthus parameter?

Question (i) is the seminal question to this work. Question (ii) is subsequent to Question (i) and is motivated by experimental concerns. Indeed, if one wants to measure experimentally the variations of the Malthus parameter for controlled populations, an efficient procedure to estimate this parameter is required. An answer to Question (i) relies on deterministic techniques whereas we introduce a stochastic framework to answer Question (ii).

### 1.1 INTRODUCING INDIVIDUAL VARIABILITY

#### *Deterministic modeling*

Structured models have been successfully used to describe the evolution of a population of cells over the past several years, see Metz and Dickamm [MD86], the textbook of Perthame [Per07] and references therein. We focus in this work on classical structuring variables which are size and age, understood in a broad sense as a physiological age (not necessarily the time elapsed since birth). The quantity  $n(t, x, a)$  of cells of size  $x > 0$  and physiological age  $a \geq 0$  at time  $t \geq 0$  satisfies

$$\begin{cases} \frac{\partial}{\partial t} n(t, x, a) + \frac{\partial}{\partial x} (g_x(x, a)n(t, x, a)) + \frac{\partial}{\partial a} (g_a(x, a)n(t, x, a)) + \gamma(x, a)n(t, x, a) = 0, \\ g_a(x, a = 0)n(t, x, a = 0) = 4 \int_0^\infty \gamma(2x, a)n(t, 2x, a)da, \\ g_x(x = 0, a)n(t, x = 0, a) = 0, \quad n(t = 0, x, a) = n^{\text{in}}(x, a), \end{cases} \quad (1)$$

in a weak sense. The mechanism at work here can be described as a mass balance. The density of cells  $n(t, x, a)$  evolves with two terms of transport which stand for the growth and the aging, and with one term of fragmentation:

- *Transport terms.* Both size and age evolve in a deterministic way, the growth speed and aging speed being respectively  $g_x$  and  $g_a$ . Let  $x_t$  and  $a_t$  be the size and physiological age at time  $t$  of a cell born with size and physiological age  $(x_b, 0) = (x_{t=0}, a_{t=0})$ , then their evolution is given by  $dx_t/dt = g_x(x_t, a_t)$  and  $da_t/dt = g_a(x_t, a_t)$ .
- *Fragmentation terms.* We assume that the division is perfectly symmetric: an individual of size  $x$  divides into two individuals of size  $x/2$  (and age 0). A cell of current size  $x$  and current physiological age  $a$  divides with probability  $\gamma(x, a)dt$  between  $t$  and  $t + dt$ . The boundary condition for  $a = 0$  ensures that the quantity of new cells of size and age  $(x, 0)$  is exactly twice the number of cells of characteristics  $(2x, s)$  for  $s > 0$  that have just divided. The factor 4 is thus the product of a factor 2 arising from the birth of two new cells at each division and another factor 2 coming from the  $2x$  size of dividing cells.

We mention two special cases of interest studied in the literature,

1) The *size-age-structured model* studied in Doumic [Dou07], with  $g_a \equiv 1$  (thus  $a$  is there the time elapsed since birth) and a general  $g_x$ . If  $g_x$  and  $\gamma$  do not depend on size, integrating (1) in  $x$  from 0 to infinity, we obtain the age-structured equation which was the starting point of Chapter I. The age-structured equation is classical and we refer to [Per07] (Section 3.9.1). If  $g_x$  and  $\gamma$  do not depend on age, integrating (1) in  $a$  from 0 to infinity, we obtain the size-structured equation, which is well understood. Perthame and Ryzhik [PR05], Michel [Mic07b] and Doumic and Gabriel [DG10] studied the eigenproblem and proved the existence of a stationary profile (first eigenvector). The exponential convergence, for large times, toward such a stationary profile is studied by Laurençot and Perthame [LP09], Cáceres *et al.* [CCM11], Balagué *et al.* [BCG13], Pakdaman *et al.* [PPS14] and Mischler and Scher [MS13].

2) The *size-increment-structured model* recently proposed in Taheri-Araghi *et al.* [TAB+15] for the study of the bacteria *E. coli* (Supplementary Information, Equations (13)-(14)), with  $g_x(x, a) = g_a(x, a) = x$ , *i.e.* the size grows exponentially, and  $a$ , growing as size  $x$ , represents an increment between the current size  $x$  and the size at birth. The increment of size can be seen as a physiological age, since it goes back to zero after each division, linked to the size: the bigger the cell, the faster the aging. (Note that if the growth is linear,  $g_x \equiv g_a \equiv \bar{v} > 0$ , the increment of size is proportional to the time elapsed since birth.) We also mention the article of Hall, Wake and Gandar of 1991 [HWG91] where an equivalent model is built – the two structuring variables used are there the size and the size at birth.

To take into account variability between cells, we extend the previous framework adding an individual feature – an aging rate or a growth rate – as structuring variable. Let  $\mathcal{V}$  be a compact set of  $(0, \infty)$  where the individual feature takes its values and let  $\rho(v, dv')$  be a Markov kernel with support in  $\mathcal{V} \times \mathcal{V}$  satisfying  $\int_{\mathcal{V}} \rho(v, dv') = 1$  for all  $v \in \mathcal{V}$ . We denote the quantity of cells of size  $x > 0$ , physiological age  $a \geq 0$  and feature  $v \in \mathcal{V}$  at time  $t \geq 0$  by  $n(t, x, a, v)$ . A fairly general





(2). We denote it by  $\lambda_{\gamma,\rho}$  to stress the dependence not only on the division rate  $\gamma$  but also on the variability Markov kernel  $\rho(v, dv')$ . The long-time behaviour of the solution to (2) is expected to be  $n(t, x, a, v) \approx e^{\lambda_{\gamma,\rho} t} N_{\gamma,\rho}(x, a, v)$  with  $N_{\gamma,\rho}$  a stationary profile, where we see that the global growth of the system is governed by  $\lambda_{\gamma,\rho}$ . Let us neglect heredity in the transmission of the individual feature and pick  $\rho(v, dv') = \rho(dv')$  for a non-degenerated distribution  $\rho$  such that  $\int_{\mathcal{V}} v' \rho(dv') = \bar{v}$ . Denote  $\lambda_{\gamma,\rho}$  by  $\lambda_{\gamma,\bar{v}}$  when  $\rho$  is the Dirac mass at point  $\bar{v} > 0$ . Our aim is to compare  $\lambda_{\gamma,\rho}$  to  $\lambda_{\gamma,\bar{v}}$ , *i.e.* to compare the evolution of a population with variability to a population of reference where all cells have the mean characteristic  $\bar{v}$ .

### Related studies

Before going ahead and investigating Question (i) let us describe existing studies on the variations of the Malthus parameter. We sum up below two studies dealing with the size-structured model without variability, particular case of (1). Influence of the asymmetry in the division and influence of the growth speed are respectively treated in Michel [Mic06a] and Calvez, Doumic and Gabriel [CDG12].

We also mention a corpus of articles by Clairambault, Gaubert, Lepoutre, Michel and Perthame [CMP06, CGP07, CGL09, CGL11, GL15] comparing the eigenvalue of a partial differential equation with time dependent periodic coefficients (birth and death rates) to the eigenvalue of the equation with time-averaged coefficients. They mainly focus on the age-structured system for the cell division cycle. Evolution equations with time periodic coefficients require specific techniques (among Floquet's theory).

**Asymmetry of the division.** The influence of the asymmetry in the division was investigated by Michel in [Mic06a] for the model described by

$$\frac{\partial}{\partial t} n(t, x) + \frac{\partial}{\partial x} n(t, x) + B(x)n(t, x) = \frac{1}{\sigma} B\left(\frac{x}{\sigma}\right)n\left(t, \frac{x}{\sigma}\right) + \frac{1}{1-\sigma} B\left(\frac{x}{1-\sigma}\right)n\left(t, \frac{x}{1-\sigma}\right),$$

with  $n(t, x = 0) = 0$ , where the size evolves linearly (at rate  $g_x \equiv 1$ ) and that a cell of size  $x$  divides at a rate  $B(x)$  into two cells of sizes  $\sigma x$  and  $(1 - \sigma)x$ , with a parameter  $\sigma \in (0, 1)$ . The case of reference is the symmetric division case corresponding to  $\sigma = 1/2$ . If we denote  $\lambda_{B,\sigma}$  the Malthus parameter in this model, the aim is to compare  $\lambda_{B,\sigma}$  to  $\lambda_{B,1/2}$ . Depending on the form of the division rate  $B$ , the asymmetry division is beneficial or not at a global level for the growth of the population. Two special cases are highlighted in [Mic06a], **1**) qualitatively, if cells divide at high sizes, then  $\lambda_{B,\sigma} < \lambda_{B,1/2}$  which means that asymmetry slows the global growth of the population, **2**) qualitatively, if cells divides early, at small sizes, then  $\lambda_{B,\sigma} > \lambda_{B,1/2}$  which means that asymmetry creates a global gain speeding up the growth of the population. We refer to Theorems 2.1 and 2.2 of [Mic06a] for precise statements of the assumptions on  $B$ . The method developed in [Mic06a] is an interesting approach to investigate Question (i).

**Influence of the growth rate.** The influence of the growth speed was investigated by Calvez *et al.* in [CDG12] for the model described by

$$\frac{\partial}{\partial t} n(t, x) + \frac{\partial}{\partial x} (\bar{v} x n(t, x)) + B(x)n(t, x) = 4B(2x)n(t, 2x),$$

where the size evolves exponentially at rate  $\bar{v}$  and assuming that division is symmetric for simplicity. On this very simple model, the Malthus parameter is exactly equal to the common individual growth rate  $\bar{v}$ , hence it *seems* that the fastest the cells grow, the fastest the overall cell population grows. This is *not* the case in general: if the growth speed would be  $g(x)$  instead of  $x$ , then increasing  $g$  by a factor may have the effect of diminishing the Malthus parameter. A plausible example: if around infinity  $g(x)$  is equivalent to  $x^\nu$  up to a constant with  $\nu < 1$  and if  $B(x)$  vanishes at infinity, being equivalent to  $x^{-\gamma}$  up to a constant with  $0 < \gamma < 1 - \nu$ , then the Malthus parameter vanishes when inflating  $g$  by a multiplicative factor. This was proved in [CDG12] (see Theorem 1 and Proposition 1 in Appendix 2, see also Figure 2(a)). Note that in the present study, we inflate the growth speed by a factor that depends on each individual.

## 1.2 ESTIMATING THE MALTHUS PARAMETER

### *Stochastic modeling*

As in the previous chapters, let  $\mathbb{T}$  be the infinite binary tree indexed in the following way

$$\mathbb{T} = \bigcup_{m=0}^{\infty} \{0, 1\}^m.$$

We can see the nodes of  $\mathbb{T}$  as representing a given population of cells, where the initial individual is denoted by  $\emptyset$ . Each node  $u \in \mathbb{T}$ , identified with a cell of the population, has a mark

$$(b_u, \zeta_u, \xi_u, \Delta_u, \tau_u),$$

where  $b_u$  is the birth time,  $\zeta_u$  the lifetime,  $\xi_u$  the size at birth,  $\Delta_u$  the physiological age (or increment) at division and  $\tau_u$  the individual feature of cell  $u$ . We introduce  $d_u$  the division time of  $u$ ,  $d_u = b_u + \zeta_u$ . Denote by  $G_x(t; x_b, v)$  the size after a time  $t$  of a cell born with a size  $x_b$  and individual feature  $v$ . If the growth speed  $g_x(x, a, v) = vx$  as in the size-structured or size-increment-structured models then  $G_x(t; x_b, v) = x_b e^{vt}$ . Denote by  $G_a(t; x_b, v)$  the physiological age (or increment) after a time  $t$  of a cell born with a size  $x_b$  and individual feature  $v$ . If the aging speed  $g_a(x, a, v) = v$  as in the age-structured model then  $G_a(t; x_b, v) = vt$  and if  $g_a(x, a, v) = vx$  as in the size-increment-structured model then  $G_a(t; x_b, v) = x_b(e^{vt} - 1)$ . Thus the evolution of the size and the physiological age of  $u$  during its lifetime is governed by respectively

$$\xi_u^t = G_x(t - b_u; \xi_u, \tau_u), \quad \Delta_u^t = G_a(t - b_u; \xi_u, \tau_u), \quad t \in [b_u, d_u].$$

The individual feature of  $u$  is inherited from its parent  $u^-$  and is drawn according to the Markov kernel

$$\rho(v, dv') = \mathbb{P}(\tau_u \in dv' | \tau_{u^-} = v).$$

The individual feature  $\tau_u$  is constant during the lifetime of  $u$ . Division occurs at a rate  $\gamma(x, a, v)dt$ ,

$$\mathbb{P}(\zeta_u \in [t, t + dt) | \zeta_u \geq t, \xi_u = x_b, \tau_u = v) = \gamma(G_x(t; x_b, v), G_a(t; x_b, v), v) dt.$$

When splitting a cell gives birth to two offspring of the same size,

$$\xi_u = \frac{1}{2}G_x(\zeta_{u^-}; \xi_{u^-}, \tau_{u^-})$$

where  $G_x(\zeta_{u^-}; \xi_{u^-}, \tau_{u^-})$  is the size at division of cell  $u^-$ . We have the relation  $\Delta_u = G_a(\zeta_u; \xi_u, \tau_u)$  since  $G_a(\zeta_u; \xi_u, \tau_u)$  is the physiological age after the lifetime  $\zeta_u$ , *i.e.* at division, of cell  $u$ . Then  $((\xi_u, \Delta_u, \tau_u), u \in \mathbb{T})$  is defined as a bifurcating Markov chain (see Guyon [Guy07] and Chapter II for a definition).

This stochastic modeling and the previous deterministic modeling match. Recall that  $b_u$  is the birth time of  $u$ ,  $d_u$  its division time and define the set of living cells at time  $t$ ,

$$\partial\mathcal{T}_t = \{u \in \mathbb{T}; b_u \leq t < d_u\}, \quad (3)$$

keeping on with the notation of Chapter I. Define the measure  $n(t, dx, dv, da)$  as the expectation of the empirical measure at time  $t$  over smooth test functions  $f : \mathcal{S} \rightarrow \mathbb{R}$ ,

$$\iint_{\mathcal{S}} f(x, a, v) n(t, dx, dv, da) = \mathbb{E} \left[ \sum_{u \in \partial\mathcal{T}_t} f(\xi_u^t, \Delta_u^t, \tau_u) \right]$$

where  $(\xi_u^t, \Delta_u^t, \tau_u)$  is the size and physiological age of  $u$  at time  $t$ , together with the individual feature of  $u$ , constant through time. Then  $n(t, dx, dv, da)$  satisfies in a weak sense the partial differential equation (2). We refer to Theorem 1 of [DHKR15] for such a result in the size-structured model with variability. The age-structured model with variability is treated in Section 2 of this chapter (see Theorem 3) and the size-increment-structured model with variability can be treated with the same techniques.

### Continuous time tree versus one branch

The value of the Malthus parameter is encoded in the time evolution of the tree. To build an estimator one can exploit the asymptotic behaviour of empirical means at time  $t$ , as studied in Cloez [Clo11],

$$e^{-\lambda_{\gamma, \rho} t} \sum_{u \in \partial\mathcal{T}_t} f(\xi_u^t, \Delta_u^t, \tau_u) \approx c(f, \gamma, \rho) W_{\gamma, \rho}, \quad (4)$$

when  $t$  is large, in probability, with  $(\xi_u^t, \Delta_u^t)$  defined as the size and physiological age at time  $t$  of cell  $u$ ,  $c(f, \gamma, \rho)$  a constant and  $W_{\gamma, \rho}$  a non-degenerated random variable. Observing  $((\xi_u^t, \Delta_u^t, \tau_u), u \in \partial\mathcal{T}_t)$ , or only a component of it for all living cells, at two different times,  $T/2$  and  $T$  for instance, with  $T$  large enough, one can estimate the Malthus parameter, with a free choice for the smooth test function  $f$ . Common choices are  $f \equiv 1$  or  $f(x, a, v) = x$  where the empirical mean defines the mean number of cells or the mean size of the cells respectively.

It seems more challenging to estimate the Malthus parameter observing one branch of the genealogical tree of the population of cells since we lose the information linked to the time evolution of the population. However one can still estimate it from only one branch. Intuitively, since the Malthus parameter only depends on the division rate  $\gamma$  and the variability Markov kernel  $\rho$ , if

one can reconstruct both  $\gamma$  and  $\rho$  from the observation of only one branch, then the estimation of the Malthus parameter is still possible. The issue is then to propose some procedure achieving statistical efficiency. We discuss this question at length later on, see Sections 2.4 and 3.4.

### 1.3 MAIN RESULTS AND OUTLINE

#### Main results

The three models embedded in (2) we consider in this chapter are summed up in Table 7 below.

|                   | Age-structured     | Size-increment-structured | Size-structured    |
|-------------------|--------------------|---------------------------|--------------------|
| $g_x(x, a, v)$    | Independent on $a$ | $vx$                      | $vx$               |
| $g_a(x, a, v)$    | $v$                | $vx$                      | Independent on $x$ |
| $\gamma(x, a, v)$ | $vB(a)$            | $vxB(a)$                  | $vxB(x)$           |

Table 7: The main models with variability, described by (2). Column 1: choosing the aging speed  $v$  means that the aging is linear at individual rate  $v$ . Columns 2 and 3: choosing the growth speed  $vx$  means that the growth is exponential at individual rate  $v$ .

The choice  $\gamma(x, a, v)dt = g_a(x, a, v)B(a)dt$  or  $g_x(x, a, v)B(x)dt$  means that

$$\gamma(x, a, v)dt = B(a)da \quad \text{or} \quad B(x)dx.$$

We link the unit time division rate  $\gamma$  to a division  $B$  of interest being unit age or unit size. From a stochastic point of view, this specification of  $\gamma$  entails the mutual independence of the physiological ages  $\Delta_u$  in the age-structured model and of the increments of size in the size-increment-structured model, together with their independence of the other quantities of the model: in both models the variables  $\Delta_u$  are drawn according to the density  $f_B(a) = B(a) \exp(-\int_0^a B(s)ds)$ , independently of the aging or growth rate, as highlighted in Equations (23) and (29) below. In the size-structured model, the sizes at birth  $\xi_u$  form a bifurcating Markov chain, independent of the growth rates, as shown in Equation (41) below.

**Question (i).** Our main results are summed up in Table 8. Our analysis begins with the **age-structured model**, for the division rate and the aging and growth speeds given in Column 1 of Table 7, neglecting heredity in the transmission of the aging rate by picking  $\rho(v, dv') = \rho(v')dv'$ . We analytically prove that the Malthus parameter, well-defined by Theorem 4, can increase or decrease when introducing variability in the aging rate, depending on the variations of the density  $f_B$ : it decreases when  $f_B$  is non-increasing and vice versa (Theorem 6). We also compute the perturbation at order two of the Malthus parameter  $\lambda_{\gamma, \rho_\alpha}$  when  $\rho_\alpha$  converges in distribution to a Dirac mass as  $\alpha \rightarrow 0$  (Theorem 7).

This age-structured model still being a toy model for cellular division (at least for *E. coli*, see Robert *et al.* [RHK+14]) we turn to the most recent model which is the **size-increment-structured model** (see Column 2 of Table 7). A numerical study, based on stochastic simulations, is carried out to give preliminary answers. For the unit size division rate  $B$  chosen as a power law with some lag (see Table 8), we infer that the Malthus parameter of a population with variability in the growth rate is lower than the Malthus parameter of a population of reference where all cells

grow at the mean growth rate of the previous population of interest, which writes mathematically with our notation  $\lambda_{\gamma,\rho} < \lambda_{\gamma,\bar{v}}$  for  $\rho(\cdot)$  a probability distribution of mean  $\bar{v}$ . Admittedly this conclusion agrees with biological wisdom but the strength of our methodology is to quantify such a decrease. We evaluate the magnitude of the decrease around 2% when the variability distribution has a realistic coefficient of variation around 15%-20%. Such a decrease is far from being negligible since the Malthus parameter governs the *exponential* growth of the whole population. In addition, we observe the decrease of the Malthus parameter when there is more and more variability in the growth rate. Unexpectedly, introducing heredity in the transmission of the growth rate may have a counter effect and may change this conclusion: in presence of heredity the Malthus parameter may be stable when introducing variability or even larger.

A classical model used so far to model cell division is the **size-structured model** (see Column 3 of Table 7) and our main conclusion, still based on a numerical study using stochastic simulations, is the same as in the previous model.

|               | <b>Age-structured</b>  | <b>Size-increment-structured</b>  | <b>Size-structured</b>                          |
|---------------|--|---|---|
| Division rate | $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$   | $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$  | $B(x) = (x - 1)^2 \mathbf{1}_{\{x \geq 1\}}$    |
| Variability   | In the aging rate  | In the growth rate  | In the growth rate                              |
| Variations    | $\lambda_{\gamma,\rho} \searrow$<br>(Figure 20)  | $\lambda_{\gamma,\rho} \searrow$<br>(Figure 14)   | $\lambda_{\gamma,\rho} \searrow$<br>(Figure 18) |
| Comments      | <i>General result:</i><br>$\lambda_{\gamma,\rho} \searrow$ or $\nearrow$ if $f_B \searrow$ or $\nearrow$<br>(Theorems 6 and 7) | <i>With heredity:</i><br>$\lambda_{\gamma,\rho} \searrow$ or $\rightarrow$ or $\nearrow$<br>(Figures 16 and 17) | -   |

Table 8: Variations of the Malthus parameter compared to the reference value when introducing variability between cells for the main models described in Table 7, for an experimentally realistic unit size division rate  $B$ . Note:  $\lambda_{\gamma,\rho} \searrow$  means that  $\lambda_{\gamma,\rho} < \lambda_{\gamma,\bar{v}}$  for a non-degenerated probability distribution  $\rho(\cdot)$  with mean  $\bar{v}$ , and so on.

**Question (ii).** The estimation of the Malthus parameter can be done observing the population of cells at times  $T/2$  and  $T$  as suggested previously. From an experimental point of view, *the observation of a whole population of cells* can be done up to the 9-th generation, and observing 10 populations is realistic. It leads to a database of around 5 000 cells (aggregating the data).

Experimental protocols developed over the past several years also made possible *the observation of one branch* of the genealogical tree instead of the whole population of cells, see the major experimental study on single cells performed by Wang *el al.* [WRP+10]. One branch can be observed along 150 generations and repeating this experiment 100 times is feasible. Thus accurate measurements can be obtained for a lineage of around 150 000 cells (aggregating the data). One can ask if, alternatively, we could estimate the Malthus parameter from this observation scheme. We expect the estimator built from one branch to be less accurate than the estimator built from a continuous time tree, but it may still be interesting for experimental studies since the number of data is much larger in the first case.

This question is well understood in our toy model, structured by the age and the aging rate (see Section 2.4). As regards the size-increment-structured model with variability in the growth rate, answering Question (ii) observing only one branch is one of the main perspective of this work and we already propose a methodology (see Section 3.4).

### Outline

We fully answer to both questions in the toy model structured by the age and the aging rate (Section 2). This toy model enables us to formalize these two questions and to highlight what is at stake. Then, in Section 3, we turn to a new model, much more realistic for cellular division, structured by the size, the increment of size (which can be seen as a physiological age) and the growth rate. We give preliminary answers to Question (i) for this model based on a numerical study. Our concern in this model is also to design upcoming experiments, and we will see how our numerical study helps. We discuss the size-structured model in Section 4, together with some other recent model introduced by Amir in [Ami14]. We give the main perspectives of this work in Section 5. Finally Section 6 is devoted to the proofs of our results, together with an Appendix in Section 7. Section 8 gives supplementary figures and tables to complete the numerical study of Section 3.

## 2 THE AGE-STRUCTURED MODEL WITH VARIABILITY

### 2.1 DESCRIPTION OF THE MODEL

#### *Deterministic description*

In this section we study the age-structured model with variability embedded in (2). The individual feature  $v$  we introduce represents here an aging rate. The quantity  $n(t, a, v)$  of cells of physiological age  $a$  and aging rate  $v$  at time  $t$  evolves as

$$\begin{cases} \frac{\partial}{\partial t} n(t, a, v) + \frac{\partial}{\partial a} (g(a, v)n(t, a, v)) + \gamma(a, v)n(t, a, v) = 0, & t > 0, a > 0, v \in \mathcal{V}, \\ g(a = 0, v')n(t, a = 0, v') = 2 \iint_{\mathcal{S}} \gamma(a, v)n(t, a, v)\rho(v, dv')dvda, & t > 0, v' \in \mathcal{V}, \\ n(t = 0, a, v) = n^{\text{in}}(a, v), & a \geq 0, v \in \mathcal{V}, \end{cases} \quad (5)$$

which has to be read in the weak sense of measures (we allow us a slight abuse of notation here), see for instance Equation (54) below for the weak formulation. We stress that the state variable  $a$  is not necessarily the time elapsed since birth and that is why we call it *physiological age*. The underlying mechanism to (5) is

- *Transport term*: the physiological age evolves in a deterministic way at a specific rate for each individual. Let  $a_t$  be the age at time  $t$  of a cell of aging rate  $v$ , then its evolution is given by  $da_t/dt = g(a_t, v)$  with  $a_{t=0} = 0$ .
- *Fragmentation term*: a cell of age  $a$  and feature  $v$  divides with probability  $\gamma(a, v)dt$  between  $t$  and  $t + dt$ . It divides into two new cells of age 0 (it could be any other value of refer-

ence). At division, the aging rate is inherited and a new born cell has aging rate  $v'$  with probability  $\rho(v, dv')$  if  $v$  is the aging rate of its mother. The boundary condition guarantees that the quantity of new cells (of age 0) is twice the quantity of the cells that have just divided.

Lebowitz and Rubinow [LR74] and Rotenberg [Rot83] already enriched the age-structured model by an additional feature. In [LR74], the model is structured by the age and the lifetime, called generation time, which is inherited from the mother cell to the daughter cells. The death or disappearance of the cells occur at a constant rate. In [Rot83], the model is structured by a maturity (our physiological age) and by a maturity velocity (our aging rate). Rotenberg's model is further studied by Mischler, Perthame and Ryzhik [MPR02], establishing the existence of a steady state and the long-time behaviour of the solution.

Rotenberg's model is close to the model we are interested in, but different. From a certain viewpoint, the model introduced by Rotenberg [Rot83] and studied in [MPR02] is more general than the model described by (5): there are distinct death and birth rates and maturity velocity can change at any moment during the life of an individual. However, in (5) we allow for a general evolution of the aging rate (through the aging speed  $g$ ) and, most importantly, we allow for heredity of the aging rate (through the Markov kernel  $\rho$ ).

We use the notation  $\mathcal{S} = [0, \infty) \times \mathcal{V}$  for the state space of the physiological age and the aging rate. We require the following two assumptions throughout this section.

**Assumption 1** (Aging speed  $g$  and division rate  $\gamma$ ). *The following conditions are fulfilled:*

- (a) *Both  $(a, v) \rightsquigarrow \gamma(a, v)$  and  $(a, v) \rightsquigarrow g(a, v)$  are continuous.*
- (b) *For any  $(a, v) \in \mathcal{S}$ ,  $g(a, v) > 0$  and  $\sup_{(a,v) \in \mathcal{S}} g(a, v) < \infty$ .*
- (c) *For any  $v \in \mathcal{V}$  there exists  $0 \leq a_{\min}(v) < a_{\max}(v) \leq \infty$  such that  $\gamma(a, v) > 0$  for  $a \in [a_{\min}(v), a_{\max}(v)]$ .*
- (d) *For any  $v \in \mathcal{V}$ , we have  $\int_0^\infty \frac{\gamma(a,v)}{g(a,v)} da = \infty$  and  $\sup_{(a,v) \in \mathcal{S}} \frac{\gamma(a,v)}{g(a,v)} e^{-\int_0^a \frac{\gamma(s,v)}{g(s,v)} ds} < \infty$ .*

**Assumption 2** (Markov kernel  $\rho$ ). *There exists a continuous and bounded  $\rho : \mathcal{V}^2 \rightarrow [0, \infty)$  satisfying  $\int_{\mathcal{V}} \rho(v, v') dv' = 1$  for any  $v \in \mathcal{V}$ , such that  $\rho(v, dv') = \rho(v, v') dv'$ , which satisfies in addition*

$$\inf_{v' \in \mathcal{V}} \int_{\mathcal{V}} \rho(v, v') dv = \varrho > \frac{1}{2}.$$

These two assumptions enable us to define later on the Malthus parameter studying the direct and adjoint eigenproblems (see Section 2.2). In Assumption 1, we require (a) since we look for continuous eigenvectors. The requirement (b) on the aging speed  $g$  authorises the case  $g(a, v) = v$  we are interested in. Note that the nonnegativity of  $g$  means that cells can only age, but not rejuvenate. The boundedness of the aging speed is technical. The condition (c) is needed in our proof for the uniqueness of the eigenlements. Finally note that (d) is not very restrictive and it can be read as follows: for any  $v \in \mathcal{V}$ ,  $a \rightsquigarrow \frac{\gamma(a,v)}{g(a,v)}$  is a hazard rate and the associated density is bounded in  $a$  (and in  $v$ , which belongs to a compact set of  $(0, \infty)$ ). In Assumption 2, the continuity of  $\rho$  is needed since, again, we look for continuous eigenvectors, its boundedness seems technical



but is not a stringent condition. The minorisation condition is used in the study of the adjoint eigenproblem.

Both assumptions are quite strong but they present the advantage to lead to a simple proof. The recent results of Mischler and Scher [MS13] should enable us to weaken it.

### Probabilistic description

To ease the presentation, we set  $g(a, v) = v$ . Such an assumption will be in force in Sections 2.3 and 2.4 and we stress that we adopt it here for the sake of simplicity only. For a cell  $u \in \mathbb{T}$ ,  $\zeta_u$  represents its lifetime,  $\Delta_u$  its physiological age at division and  $\tau_u$  its aging rate. Given  $g(a, v) = v$ , the physiological age at division of  $u$  is given by  $\Delta_u = \tau_u \zeta_u$  and its evolution during the lifetime of  $u$  is governed by

$$\Delta_u^t = \tau_u(t - b_u), \quad t \in [b_u, d_u). \quad (6)$$

The aging rate of  $u$ , given once and for all at birth and constant during its lifetime, is inherited from its parent  $u^-$  and is drawn according to the Markov kernel

$$\rho(v, dv') = \mathbb{P}(\tau_u \in dv' | \tau_{u^-} = v). \quad (7)$$

Cells divide into two cells of physiological age 0 according to the division rate  $\gamma(a, v)dt$ ,

$$\mathbb{P}(\zeta_u \in [t, t + dt) | \zeta_u \geq t, \tau_u = v) = \gamma(vt, v)dt \quad (8)$$

where  $vt = a$  is the physiological age after a time  $t$  for a cell with aging rate  $v$ . Equations (6), (7) and (8) enable us to define  $((\Delta_u, \tau_u), u \in \mathbb{T})$  as a bifurcating Markov chain (see Chapter II, Definition 1).

### Matching deterministic and probabilistic viewpoints

Let  $\mathcal{C}_0^1(\mathcal{S})$  be the set of real-valued test functions which are continuously differentiable in the first variable, with compact support of the form  $[0, a_0] \times \mathcal{V}_0$  for some  $a_0 \in (0, \infty)$  and  $\mathcal{V}_0$  compact subset of  $\mathcal{V}$ . We denote by  $\mathbb{P}_\mu$  the law of the process  $((\Delta_u^t, \tau_u), t \geq 0, u \in \mathbb{T})$  where the root  $(\Delta_\emptyset, \tau_\emptyset)$  has distribution  $\mu$  and  $\mathbb{E}_\mu[\cdot]$  denotes the expectation with respect to  $\mathbb{P}_\mu$ .

**Theorem 3.** *Work under Assumptions 1 and 2. Let  $\mu$  be a probability distribution on  $\mathcal{S}$ . Define the distribution  $n(t, da, dv)$  by*

$$\iint_{\mathcal{S}} f(a, v)n(t, da, dv) = \mathbb{E}_\mu \left[ \sum_{u \in \partial\mathcal{T}_t} f(\Delta_u^t, \tau_u) \right] \quad (9)$$

for every  $f \in \mathcal{C}_0^1(\mathcal{S})$ , where  $\partial\mathcal{T}_t$  is the set of living cells defined by (3),  $\Delta_u^t$  is the physiological age at time  $t$  of cell  $u \in \partial\mathcal{T}_t$  and  $\tau_u$  is the aging rate of  $u$ , constant through time. Then  $n(t, da, dv)$  solves the partial differential equation (5) in a weak sense with initial condition  $n^{\text{in}}(da, dv) = \mu(da, dv)$ .

We stress that the matching between deterministic and probabilistic approaches is far from being an exercise since it can be helpful in the proof of existence and uniqueness of a solution to

(5) in a suitable space, as developed recently for Hawkes processes by Cáceres, Chevallier, Doumic and Reynaud-Bouret [CCDR15].

## 2.2 THE MALTHUS PARAMETER

### *The direct and adjoint eigenproblems*

We now precisely define the Malthus parameter. To that end, let us introduce the direct eigenproblem,

$$\begin{cases} \frac{\partial}{\partial a}(g(a, v)N_{\gamma, \rho}(a, v)) + \gamma(a, v)N_{\gamma, \rho}(a, v) = -\lambda_{\gamma, \rho}N_{\gamma, \rho}(a, v), \\ g(a = 0, v')N_{\gamma, \rho}(a = 0, v') = 2 \iint_{\mathcal{S}} \gamma(a, v)N_{\gamma, \rho}(a, v)\rho(v, v')dvda, \\ N_{\gamma, \rho} \geq 0, \quad \iint_{\mathcal{S}} N_{\gamma, \rho}(a, v)dvda = 1, \end{cases} \quad (10)$$

and the adjoint eigenproblem,

$$\begin{cases} g(a, v)\frac{\partial}{\partial a}\phi_{\gamma, \rho}(a, v) + \gamma(a, v)\left(2 \int_{\mathcal{V}} \phi_{\gamma, \rho}(0, v')\rho(v, v')dv' - \phi_{\gamma, \rho}(a, v)\right) = \lambda_{\gamma, \rho}\phi_{\gamma, \rho}(a, v), \\ \phi_{\gamma, \rho} \geq 0, \quad \iint_{\mathcal{S}} N_{\gamma, \rho}(a, v)\phi_{\gamma, \rho}(a, v)dvda = 1, \end{cases} \quad (11)$$

linked to the evolution equation (5).

Let  $\mathcal{C}_b(\mathcal{V})$  be the set of functions  $f : \mathcal{V} \rightarrow \mathbb{R}$  which are bounded and continuous and let  $\mathcal{C}_b^1(\mathbb{R}^+)$  be the set of functions  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$  which are bounded and continuously differentiable. For  $f : \mathcal{S} \rightarrow [0, \infty)$  the notation  $f \in \mathcal{C}_b^1(\mathbb{R}^+; \mathcal{C}_b(\mathcal{V}))$  means that  $a \rightsquigarrow f(a, v)$  belongs to  $\mathcal{C}_b^1(\mathbb{R}^+)$  for any  $v \in \mathcal{V}$  and that  $v \rightsquigarrow f(a, v)$  belongs to  $\mathcal{C}_b(\mathcal{V})$  for any  $a \in [0, \infty)$ .

**Theorem 4.** *Work under Assumptions 1 and 2. There exists a unique solution  $(\lambda_{\gamma, \rho}, N_{\gamma, \rho}, \phi_{\gamma, \rho})$  to the direct and adjoint eigenproblems (10) and (11) such that  $\lambda_{\gamma, \rho} > 0$ ,  $(gN_{\gamma, \rho}) \in \mathcal{C}_b^1(\mathbb{R}^+; \mathcal{C}_b(\mathcal{V}))$  and  $\phi_{\gamma, \rho} \in \mathcal{C}_b^1(\mathbb{R}^+; \mathcal{C}_b(\mathcal{V}))$ .*

The unique  $\lambda_{\gamma, \rho}$  defined in such a way is what we call the Malthus parameter and let us now briefly expose its key role.

### *Deterministic and stochastic roles of the Malthus parameter*

The behaviour of empirical means at time  $T$  has been investigated by Bansaye *et al.* [BDMT11] and Cloez [Clo11], and we follow their methodology introducing a so-called auxiliary *biased* Markov process  $(\tilde{\chi}_t)_{t \geq 0}$  with infinitesimal generator

$$\mathcal{A}_{\tilde{\gamma}, \tilde{\rho}}f(a, v) = g(a, v)\frac{\partial}{\partial a}f(a, v) + \tilde{\gamma}(a, v)\left(\int_{\mathcal{V}} f(0, v')\tilde{\rho}(v, v')dv' - f(a, v)\right)$$

densely defined on bounded continuous functions, with  $\tilde{\gamma}$  and  $\tilde{\rho}$  biased versions of  $\gamma$  and  $\rho$  (see Lemma 10 below, Section 6.1). For suitable test functions  $f$ , the following approximation holds

for large  $t$ ,

$$\mathbb{E}_\mu \left[ \sum_{u \in \partial \mathcal{T}_t} f(\Delta_u^t, \tau_u) \right] \approx m_\mu e^{\lambda_{\gamma, \rho} t} \iint_{\mathcal{S}} f(a, v) \frac{\nu_{\gamma, \rho}(a, v)}{\phi_{\gamma, \rho}(a, v)} dv da \quad (12)$$

with  $m_\mu = \iint_{\mathcal{S}} \phi_{\gamma, \rho}(a, v) \mu(da, dv)$ ,  $(\lambda_{\gamma, \rho}, \phi_{\gamma, \rho})$  solution to (11) and  $\nu_{\gamma, \rho}$  the invariant measure (assuming it exists) of the biased Markov process  $(\tilde{\chi}_t)_{t \geq 0}$ . Define the distribution  $n(t, \cdot)$  by (9). Then by (12), we heuristically obtain

$$n(t, a, v) \approx m_\mu e^{\lambda_{\gamma, \rho} t} \frac{\nu_{\gamma, \rho}(a, v)}{\phi_{\gamma, \rho}(a, v)} = m_\mu e^{\lambda_{\gamma, \rho} t} N_{\gamma, \rho}(a, v)$$

with  $(\lambda_{\gamma, \rho}, N_{\gamma, \rho})$  solution to (10) if one postulates the relation

$$\nu_{\gamma, \rho} = N_{\gamma, \rho} \phi_{\gamma, \rho}.$$

Using that  $N_{\gamma, \rho}$  and  $\phi_{\gamma, \rho}$  solve respectively (10) and (11), one can check that

$$\int_0^\infty \mathcal{A}_{\tilde{\gamma}, \tilde{\rho}} f(a, v) (N_{\gamma, \rho} \phi_{\gamma, \rho})(a, v) da dv = 0,$$

for any function  $f$  smooth and compactly supported, and this actually characterises the invariant measure. Thus the relation  $\nu_{\gamma, \rho} = N_{\gamma, \rho} \phi_{\gamma, \rho}$  is valid, under appropriate assumptions ensuring existence of these three quantities.

We have formally retrieved a classical deterministic result. Indeed, from a deterministic point of view, the classical long-time behaviour result states as follows: the unique solution  $n(t, a, v)$  to (5) satisfies

$$\iint_{\mathcal{S}} |n(t, a, v) e^{-\lambda_{\gamma, \rho} t} - m^{\text{in}} N_{\gamma, \rho}(a, v)| \phi_{\gamma, \rho}(a, v) dv da \rightarrow 0 \quad \text{as } t \rightarrow \infty, \quad (13)$$

where  $m^{\text{in}} = \iint_{\mathcal{S}} n^{\text{in}}(a, v) \phi_{\gamma, \rho}(a, v) dv da$  (we have  $m^{\text{in}} = m_\mu$  for  $n$  defined by (9)).

For a proof of (13), when the division rate  $\gamma$  is bounded, one can mimic the proof given in Chapter 3 of [Per07] where the renewal equation is studied. Requiring such a strong assumption on the division rate  $\gamma$  leads to a very readable proof, via a fixed point argument for the existence of the time solution and a generalised relative entropy principle for the long-time behaviour. For a stronger result, with milder assumptions on  $\gamma$ , we refer to [MS13] whose results on the age-structured partial differential equation without the state variable  $v$  could be generalised to our case. We expect the rate of convergence in (13) to be given by the spectral gap and to our best knowledge, the more recent and more complete work on this subject is [MS13].

Equations (12) and (13) show that the global growth of the population is driven by the so-called Malthus parameter  $\lambda_{\gamma, \rho}$ , and this quantity is thus the one to consider for our purpose.

### 2.3 INFLUENCE OF VARIABILITY ON THE MALTHUS PARAMETER

In all this section, we neglect heredity in the transmission of the aging rate assuming that

$$\rho(v, dv') = \rho(v') dv' \quad (14)$$

for  $\rho : \mathcal{V} \rightarrow [0, \infty)$  such that  $\int_{\mathcal{V}} \rho(v') dv' = 1$ . This assumption enables us to obtain explicit solutions to the eigenproblem (10)-(11) (see Lemma 12 below, Section 6.2) and our proofs heavily rely on the explicit expressions of the eigenlements.

**Remark 5.** *Let us consider the case of a constant division rate,  $\gamma(a, v) = c > 0$  for any  $(a, v) \in \mathcal{S}$ . Since  $\gamma$  is constant, we expect that variability in the aging rate has no influence on the Malthus parameter. By Lemma 12 below (Section 6.2), one can indeed show that  $\lambda_{\gamma, \rho} = c$  which is independent of  $\rho$ .*

From now on we consider the model described by (5) with the aging speed set to

$$g(a, v) = v \tag{15}$$

which means that the physiological age is proportional to the time elapsed since birth, up to a factor  $v$  which is an individual feature. Let  $0 < a_{\max} \leq \infty$ , for a locally integrable  $B : [0, a_{\max}) \rightarrow [0, \infty)$  such that  $\int_0^{a_{\max}} B(s) ds = \infty$ , we assume that

$$\gamma(a, v) = g(a, v)B(a) = vB(a). \tag{16}$$

One has to see the division rate  $B(a)$  as being unit physiological age. The conversion of  $B(a)da$  to unit time involves the factor  $g(a, v)$  since  $da = g(a, v)dt$ . It leads to the unit time division rate  $\gamma(a, v) = g(a, v)B(a)$  since

$$\gamma(a, v)dt = g(a, v)B(a)dt = B(a)da.$$

Equation (23) in the next section also sheds some light on Specification (16), at a stochastic level.

Setting  $B(a) = 0$  for  $a \geq a_{\max}$  if necessary, define  $f_B(a) = B(a) \exp(-\int_0^a B(s) ds)$  for  $a \geq 0$ .

### Variations

From now on we denote  $\lambda_{\gamma, \rho}$  by  $\lambda_{B, \rho}$  and  $\lambda_{B, \bar{v}}$  stands for  $\lambda_{B, \delta_{\bar{v}}}$  where  $\delta_{\bar{v}}$  is the Dirac mass at point

$$\bar{v} = \int_{\mathcal{V}} v \rho(v) dv. \tag{17}$$

One wants to compare the Malthus parameters  $\lambda_{B, \rho}$  and  $\lambda_{B, \bar{v}}$  respectively solution to

$$2 \iint_{\mathcal{S}} \exp\left(-\frac{\lambda_{B, \rho} a}{v}\right) f_B(a) \rho(v) dv da = 1 \tag{18}$$

and

$$2 \int_0^{\infty} \exp\left(-\frac{\lambda_{B, \bar{v}} a}{\bar{v}}\right) f_B(a) da = 1. \tag{19}$$

**Theorem 6.** *Consider the model described by (5) with specifications (14), (15) and (16). Assuming in addition that  $B : [0, a_{\max}) \rightarrow [0, \infty)$  is differentiable,*

- (i) *for  $a_{\max} \in (0, \infty]$ , if  $B'(a) \leq B(a)^2$  for any  $a \in [0, a_{\max})$  then  $\lambda_{B, \rho} < \lambda_{B, \bar{v}}$ .*
- (ii) *for  $a_{\max} < \infty$ , if  $B$  is such that  $f_B$  is bounded and  $B'(a) \geq B(a)^2$  for any  $a \in [0, a_{\max})$  then  $\lambda_{B, \rho} > \lambda_{B, \bar{v}}$ .*

First note that Theorem 6 is valid for any density  $\rho$  on  $\mathcal{V}$  with mean  $\bar{v}$ . The first derivative of the density  $f_B$  being

$$f'_B(a) = (B'(a) - B(a)^2) \exp\left(-\int_0^a B(s)ds\right),$$

(for  $B$  differentiable), a condition of sign on  $B' - B^2$  is nothing but imposing  $f_B$  to be non-increasing or non-decreasing. The statement (i) of Theorem 6 requires  $f_B$  to be non-increasing whereas the statement (ii) requires  $f_B$  to be non-decreasing.

One canonical example is the following. Assume that  $B$  is constant equal to  $b > 0$  so that  $\gamma(a, v) = vb$  for any  $(a, v) \in \mathcal{S}$ . Then,  $f_B$  is the exponential density of mean  $b^{-1}$ , it is non-increasing and statement (i) of Theorem 6 applies.

The proof of Theorem 6 is short and we give it now. Then we sketch an interpretation of the result with the help of examples.

*Proof of Theorem 6.* Let  $\lambda_{B, \bar{v}}$  be such that

$$H_{\bar{v}}(\lambda_{B, \bar{v}}) = 1 \quad \text{with} \quad H_{\bar{v}}(\lambda) = 2 \int_0^\infty \exp\left(-\frac{\lambda a}{\bar{v}}\right) f_B(a) da,$$

and  $\lambda_{B, \rho}$  be such that

$$H_\rho(\lambda_{B, \rho}) = 1 \quad \text{with} \quad H_\rho(\lambda) = 2 \iint_{\mathcal{S}} \exp\left(-\frac{\lambda a}{v}\right) f_B(a) \rho(v) dv da.$$

We know that **1**) both  $H_{\bar{v}}$  and  $H_\rho$  are continuous on  $[0, \infty)$ , **2**) both  $H_{\bar{v}}$  and  $H_\rho$  are decreasing on  $[0, \infty)$ , **3**) when  $\lambda$  goes to infinity,  $H_{\bar{v}}(\lambda) \rightarrow 0$  and  $H_\rho(\lambda) \rightarrow 0$ , **4**) at point  $\lambda = 0$ ,  $H_{\bar{v}}(\lambda = 0) = H_\rho(\lambda = 0) = 2 > 1$ . Then, by the intermediate value theorem, we know there exists a unique positive  $\lambda_{B, \bar{v}}$  such that  $H_{\bar{v}}(\lambda_{B, \bar{v}}) = 1$  and a unique positive  $\lambda_{B, \rho}$  such that  $H_\rho(\lambda_{B, \rho}) = 1$ . Recall that we want to compare  $\lambda_{B, \rho}$  to  $\lambda_{B, \bar{v}}$ . We claim that **5**) if  $f'_B \leq 0$ , then

$$H_\rho(\lambda) < H_{\bar{v}}(\lambda)$$

for any  $\lambda > 0$ . Together with points **1**) to **4**), this enables us to conclude that  $\lambda_{B, \rho} < \lambda_{B, \bar{v}}$ .

Let us prove **5**). For any  $\lambda > 0$ , introduce

$$h_\lambda : v \in \mathcal{V} \rightsquigarrow 2 \int_0^\infty \exp\left(-\frac{\lambda a}{v}\right) f_B(a) da.$$

Then  $H_{\bar{v}}$  and  $H_\rho$  can be written

$$H_{\bar{v}}(\lambda) = h_\lambda(\bar{v}) \quad \text{and} \quad H_\rho(\lambda) = \int_{\mathcal{V}} h_\lambda(v) \rho(v) dv$$

for  $\lambda > 0$ . We claim that  $h_\lambda$  is strictly concave when  $f'_B \leq 0$ . Then  $h_\lambda(\bar{v}) > \int_{\mathcal{V}} h_\lambda(v) \rho(v) dv$  by Jensen's inequality, since  $\int_{\mathcal{V}} v \rho(v) dv = \bar{v}$ , and **5**) immediately follows.

We now compute the second derivative of  $h_\lambda$ ,

$$h''_\lambda(v) = \frac{1}{v^2} \int_0^\infty \frac{\lambda a}{v} \left( \frac{\lambda a}{v} - 2 \right) e^{-\frac{\lambda a}{v}} f_B(a) da.$$

Integrating by parts, recalling that we choose  $B$  differentiable, we get

$$h''_\lambda(v) = \frac{\lambda}{v^3} \int_0^\infty a^2 e^{-\frac{\lambda a}{v}} f'_B(a) da < 0$$

for any  $v \in \mathcal{V}$  when  $f'_B \leq 0$ . This ends the proof for the case  $f'_B \leq 0$  and the case  $f'_B \geq 0$  is treated in the same way (we need  $f_B$  to be bounded for the integration by parts).  $\square$

To better understand the two statements of Theorem 6, let us consider the following illustration. We choose

$$B(a) = 3(1-a)^{-1} \mathbf{1}_{\{0 \leq a < 1\}} \quad \text{i.e.} \quad f_B(a) = 3(1-a)^2 \mathbf{1}_{\{0 \leq a < 1\}} \quad (20)$$

and

$$B(a) = 3a^2(1-a^3)^{-1} \mathbf{1}_{\{0 \leq a < 1\}} \quad \text{i.e.} \quad f_B(a) = 3a^2 \mathbf{1}_{\{0 \leq a < 1\}}. \quad (21)$$

Let us denote by  $f_{\zeta_u|\tau_u=v}$  the conditional density of the lifetime  $\zeta_u$  conditionally to  $\tau_u = v$  for  $v \in \mathcal{V}$  (one can readily compute it using (8) in both of the cases we have chosen). In the population of reference, all cells age at the common rate  $\bar{v}$  and lifetimes are drawn independently according to  $f_{\zeta_u|\tau_u=\bar{v}}$ . When introducing variability, lifetimes are no longer identically distributed: they are drawn, conditionally on  $v$ , according to  $f_{\zeta_u|\tau_u=v}$ . To simplify, we focus on two subpopulations, the first one composed of cells aging at a *fast* rate  $v$  (compared to  $\bar{v}$ ) and the second one composed of cells aging at a *slow* rate  $v$  (compared to  $\bar{v}$ ). Figure 13 shows  $f_{\zeta_u|\tau_u=1}$  together with  $f_{\zeta_u|\tau_u=v}$  for one *fast* and one *slow* rate  $v$  (compare to  $\bar{v} = 1$ ) in both of the cases described above.

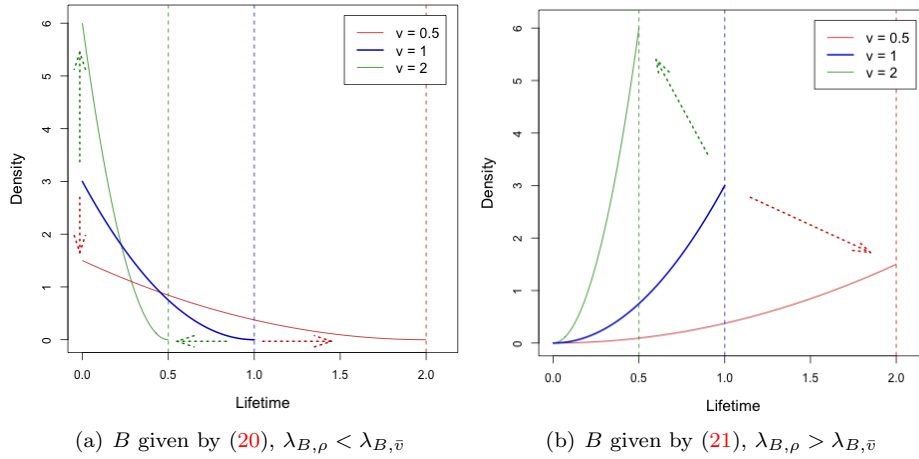


Figure 13: The lifetime density conditionally to the aging rate,  $t \rightsquigarrow f_{\zeta_u|\tau_u=v}(t)$  for different  $v$ . Population without variability:  $f_{\zeta_u|\tau_u=1} = f_B$  (bold blue line). Population with variability composed of slow aging cells,  $f_{\zeta_u|\tau_u=0.5}$  (red line), and fast aging cells,  $f_{\zeta_u|\tau_u=2}$  (green line). (a)  $f_{\zeta_u|\tau_u=v}(t) = 3v(1-vt)^2 \mathbf{1}_{\{0 \leq t < v^{-1}\}}$ . (b)  $f_{\zeta_u|\tau_u=v}(t) = 3v^3 t^2 \mathbf{1}_{\{0 \leq t < v^{-1}\}}$ .

For the choice (20) (see Figure 13(a)),  $f_B$  is non-increasing and thus  $\lambda_{B,\rho} < \lambda_{B,\bar{v}}$  since Theorem 6 (i) applies. The cells of the reference population, with no variability in the aging rate, divide at rather small lifetimes (density in bold blue line). Cells with a *fast* rate  $v$  have the same kind of behaviour as the cells of the reference population and divide at small lifetimes (density in green line). They even divide at smaller lifetimes than the cells of the reference population, which would tend to raise the Malthus parameter, but this effect is proved to be negligible. The cells with a *slow* rate  $v$  behave very differently from the cells of the reference population since they can divide at much larger lifetimes, which tends to decrease the Malthus parameter, and this second effect is dominant. For the choice (21) (see Figure 13(b)),  $f_B$  is non-decreasing and thus  $\lambda_{B,\rho} > \lambda_{B,\bar{v}}$  since Theorem 6 (ii) applies. The cells of the reference population, with no variability in the aging rate, divides at rather large lifetimes (density in bold blue line). Cells with a *slow* rate  $v$  have the same kind of behaviour as the cells of the reference population dividing at large (actually at larger) lifetimes (density in red line) whereas cells with a *fast* rate  $v$  divide at much smaller lifetimes (density in green line). Of the two antagonist effects, the dominant effect is the second one which increase the Malthus parameter. In both cases, it seems that the subpopulation which has a negligible influence is the one which reinforces the behaviour of the cells of the reference population.

### Small perturbations

The density of the aging rates is thought of as a deformation around an average aging rate  $\bar{v}$ . Given a baseline density  $\rho : \mathcal{V} \rightarrow [0, \infty)$  with mean  $\bar{v}$ , we define, for  $\alpha \in (0, 1]$ ,

$$\rho_\alpha(v) = \alpha^{-1} \rho(\alpha^{-1}(v - (1 - \alpha)\bar{v})), \quad v \in \mathcal{V}_\alpha = \alpha\mathcal{V} + (1 - \alpha)\bar{v}, \quad (22)$$

so that  $\int_{\mathcal{V}_\alpha} \rho_\alpha(v) dv = 1$ , the mean value is constant, *i.e.*  $\int_{\mathcal{V}_\alpha} v \rho_\alpha(v) dv = \bar{v}$ . The density  $\rho_\alpha$  converges in distribution to the Dirac mass  $\delta_{\bar{v}}$  as  $\alpha \rightarrow 0$ . Thus we want to investigate the behaviour of the Malthus parameter  $\lambda_{B,\rho_\alpha}$ , which is a perturbation of  $\lambda_{B,\bar{v}}$  for small value of  $\alpha$ .

**Theorem 7.** *Consider the model described by (5) with specifications (15), (16) and  $\rho(v, dv') = \rho_\alpha(v') dv'$ . Then  $\alpha \rightsquigarrow \lambda_{B,\rho_\alpha}$  is twice differentiable and the Malthus parameter  $\lambda_{B,\rho_\alpha}$  defined by (18) satisfies*

$$\lambda_{B,\rho_\alpha} = \lambda_{B,\bar{v}} + \frac{\alpha^2}{2} \frac{d^2 \lambda_{B,\rho_\alpha}}{d\alpha^2} \Big|_{\alpha=0} + o(\alpha^2)$$

with  $\lambda_{B,\bar{v}}$  defined by (19) and

$$\frac{d^2 \lambda_{B,\rho_\alpha}}{d\alpha^2} \Big|_{\alpha=0} = \sigma^2 \left( \int_0^\infty \frac{a}{\bar{v}} e^{-\frac{\lambda_{B,\bar{v}} a}{\bar{v}}} f_B(a) da \right)^{-1} \int_0^\infty \frac{\lambda_{B,\bar{v}} a}{\bar{v}} \left( \frac{\lambda_{B,\bar{v}} a}{\bar{v}} - 2 \right) e^{-\frac{\lambda_{B,\bar{v}} a}{\bar{v}}} f_B(a) da$$

where  $\sigma^2 = \int_{\mathcal{V}} (v - \bar{v})^2 \rho(v) dv$ .

Note that the perturbations play at order 2, but not at order 1 since we prove in Section 6 that  $\frac{d\lambda_{B,\rho_\alpha}}{d\alpha} \Big|_{\alpha=0} = 0$  (see Proposition 14). The amplitude of the perturbation at order 2 depends on the baseline aging rate density  $\rho$  only through its means  $\bar{v}$  and its variance  $\sigma^2$ .

**Remark 8.** *If  $B$  is differentiable, integrating by parts we get*

$$\frac{d^2 \lambda_{B,\rho_\alpha}}{d\alpha^2} \Big|_{\alpha=0} = \lambda_{B,\bar{v}} \sigma^2 \left( \int_0^\infty a e^{-\frac{\lambda_{B,\bar{v}} a}{\bar{v}}} f_B(a) da \right)^{-1} \int_0^\infty a^2 e^{-\frac{\lambda_{B,\bar{v}} a}{\bar{v}}} f'_B(a) da$$

and the variation study of  $\alpha \rightsquigarrow \lambda_{B,\rho_\alpha}$  around  $\alpha = 0$  follows as in the proof of Theorem 6 above.

The class of  $B$  for which Theorem 6 applies is quite restrictive. For applications a monotonous density  $f_B$  seems not realistic since we expect to observe a mode in the density of the physiological ages at division, as for lifetimes. One can see the distribution of lifetimes, also called *doubling times* or *generation times*, in [ONC14] (Figure 1.D), and in [TAB+15] for various medium (Figure S10). Given a more realistic division rate  $B$ , in this simple model, one can numerically solve Equations (18) and (19) in order to compare  $\lambda_{B,\rho}$  to  $\lambda_{B,\bar{v}}$ . A plausible division rate would be of the form  $B(a) = (a - 1)^\beta \mathbf{1}_{\{a \geq 1\}}$ . Unfortunately for such a division rate, Theorem 6 does not apply since  $f_B$  is non-monotonous. Supplementary Figure 20 in Section 8 shows curves  $\alpha \rightsquigarrow \lambda_{B,\alpha}$  for different values of  $\beta$  and we see that all these curves are non-increasing, which means that variability in the aging rate slows down the global growth of the population, for such a  $B$ .

## 2.4 ESTIMATION OF THE MALTHUS PARAMETER FROM ONE BRANCH

We keep on in this section with specifications (14), (15) and (16) and we address Question (ii) when observing one branch. Recalling (8), together with (16), cells divide according to the following mechanism:

$$\mathbb{P}(\zeta_u \in [t, t + dt) | \zeta_u \geq t, \tau_u = v) = vB(vt)dt.$$

In the simple case where the division rate  $B$  is constant equal to  $b > 0$ , the lifetime  $\zeta_u$  conditionally on the aging rate  $\tau_u = v$  for  $v \in \mathcal{V}$  is an exponential random variable with mean  $(vb)^{-1}$ . The joint law of  $(\Delta_u, \tau_u) = (\tau_u \zeta_u, \tau_u)$  is

$$\mathbb{P}((\Delta_u, \tau_u) \in dvda) = f_B(a)\rho(v)dvda. \quad (23)$$

We see that the physiological age  $\Delta_u$  and the aging rate  $\tau_u$  are independent and this is the consequence of the choice (16).

Denote by  $\mathbb{L}_n \subset \mathbb{T}$  one branch of  $n$  cells: for  $(u_1, \dots, u_n) \in \{0, 1\}^n$ ,

$$\mathbb{L}_n = \{v_1 = u_1; v_2 = (u_1, u_2); \dots; v_n = (u_1, u_2, \dots, u_n)\}. \quad (24)$$

We aim at estimating the Malthus parameter  $\lambda_{B,\rho}$  when observing the lifetimes  $(\zeta_u, u \in \mathbb{L}_n)$ . By (18) (see Lemma 12), the Malthus parameter  $\lambda_{B,\rho}$  is solution to

$$\iint_{\mathcal{S}} (2 \exp(-\lambda_{B,\rho} v^{-1} a) - 1) f_B(a)\rho(v)dvda = 0,$$

where  $f_B(a)\rho(v)dvda$  is the joint density of  $(\Delta_u, \tau_u)$  according to (23). Thus we look for  $\widehat{\lambda}_n$  solution to

$$\frac{1}{n} \sum_{u \in \mathbb{L}_n} (2 \exp(-\widehat{\lambda}_n \zeta_u) - 1) = 0,$$

with  $\zeta_u = (\tau_u)^{-1} \Delta_u$  the lifetime of cell  $u$ . What is really interesting in the age model is that we can estimate the Malthus parameter observing only one branch at a parametric rate  $n^{-1/2}$ . Indeed  $(\zeta_u, u \in \mathbb{L}_n)$  is a collection of independent and identically distributed random variables and the estimator  $\widehat{\lambda}_n$  we have defined is nothing but a Z-estimator (see van der Vaart [vdV98], Chapter 5).



### 3 THE SIZE-INCREMENT-STRUCTURED MODEL WITH VARIABILITY

#### 3.1 DESCRIPTION OF THE MODEL

##### *Deterministic description*

We study a model structured by three state variables: the current size  $x$  of the cell, its increment of size  $a = x - x_b$  ( $x_b$  being the size at birth) and its growth rate  $v$ . The model is described deterministically by (2) choosing

$$g_x(x, a, v) = g_a(x, a, v) = g(x, v),$$

one typical choice being  $g(x, v) = vx$ , which means that the growth is exponential. Such an assumption is realistic, we refer to the analysis of single cells growth by Schaechter *et al.* [SWHK62] and by Robert *et al.* [RHK+14] (Figure 2). By the techniques used in the proof of Theorem 3, one can match this deterministic description and the stochastic setting whose description follows.

##### *Fundamental choice of the division rate and probabilistic description*

We specify the division rate as

$$\gamma(x, a, v) = g(x, v)B(a) = vxB(a). \quad (25)$$

It is now time to adopt a stochastic point of view to better understand what are the huge consequences of such a choice. We choose the growth speed as  $vx$  to simplify the presentation but this is not fundamental. For a cell  $u \in \mathbb{T}$ , born at time  $b_u$  and which divides after a lifetime  $\zeta_u$ , let us denote by  $\xi_u$  its size at birth,  $\Delta_u$  its increment of size at division and by  $\tau_u$  its growth rate. The choice  $g(x, v) = vx$  means the growth is exponential, then the size and the increment of size at time  $t$  of cell  $u$  are respectively

$$\xi_u^t = \xi_u e^{\tau_u t}, \quad \Delta_u^t = \xi_u^t - \xi_u = \xi_u (e^{\tau_u t} - 1), \quad t \in [b_u, d_u].$$

The growth rate of one cell is drawn according to the Markov kernel

$$\rho(v, dv') = \mathbb{P}(\tau_u \in dv' | \tau_u = v). \quad (26)$$

With the choice (25), cells divide at the rate  $vxB(a)dt$ ,

$$\mathbb{P}(\zeta_u \in [t, t + dt) | \zeta_u \geq t, \xi_u = x_b, \tau_u = v) = vx_b e^{vt} B(x_b(e^{vt} - 1))dt,$$

where  $x_b e^{vt} = x$  is the size after a time  $t$  of a cell born with size  $x_b$  and growth rate  $v$ , and  $x_b(e^{vt} - 1) = x - x_b = a$  is the size added to the initial size (there are thus respectively the current size and the current increment of size). Cells divide into two equal parts,

$$\xi_u = \frac{1}{2} \xi_u e^{\tau_u - \zeta_u} \quad (27)$$

where  $u^- \in \mathbb{T}$  denotes the parent of  $u$ . By definition the increment of size at division of cell  $u$  is

$$\Delta_u = \xi_u(e^{\tau_u \zeta_u} - 1), \quad (28)$$

and its law is given by

$$\mathbb{P}(\Delta_u \in ds | \xi_u = x_b, \tau_u = v) = \mathbb{P}(x_b(e^{v\zeta_u} - 1) \in ds | \xi_u = x_b, \tau_u = v) = B(s)ds.$$

It means that the increment of size of cell  $u$  is independent of its size at birth and on its growth rate, the increments of size are also independent from one cell to another and identically distributed with common density  $f_B$ . Written shortly,

$$\mathbb{P}(\Delta_u \in da) = B(a) \exp\left(-\int_0^a B(s)ds\right) = f_B(a)da \quad (29)$$

with  $f_B$  density associated to the division rate  $B$  defined previously. Equations (26) to (29) fully describe stochastically the model.

In a nutshell, the choice (25) leads to the fundamental property (29) of independence of the increments of size, mutual independence and independence with respect other quantities of the model. The authors in [TAB+15] show this is experimentally valid: Figure S6 points out the independence between the increment of size and the size at birth of the cell. Figure S8 of [TAB+15] indicates the independence between the increment of size and the growth rate of the cell. Figure S12 in [TAB+15] argue in favor of the independence of the increments of size across generations. This would be interesting to carry out a systematic study based on statistical tests on these three statements.

### 3.2 THE MALTHUS PARAMETER

#### *The direct and adjoint eigenproblems*

A precise definition of the Malthus parameter  $\lambda_{\gamma,\rho}$  is based on the eigenproblem associated to (2). For a general growth speed  $g$  and a general division rate  $\gamma$ , the direct eigenproblem is given by

$$\begin{cases} \frac{\partial}{\partial x}(g(x,v)N_{\gamma,\rho}(x,a,v)) + \frac{\partial}{\partial a}(g(x,v)N_{\gamma,\rho}(x,a,v)) \\ \quad + \gamma(x,a,v)N_{\gamma,\rho}(x,a,v) = -\lambda_{\gamma,\rho}N_{\gamma,\rho}(x,a,v), \\ g(x,v')N_{\gamma,\rho}(x,a=0,v') = 4 \iint_{(0,\infty) \times \mathcal{V}} g(2x,v)\gamma(2x,a,v)N_{\gamma,\rho}(2x,a,v)\rho(v,v')dvda, \\ g(0,v')N_{\gamma,\rho}(x=0,a,v') = 0, \quad N_{\gamma,\rho} \geq 0, \quad \iiint_{\mathcal{S}} N_{\gamma,\rho}(x,a,v)dxdvda = 1, \end{cases} \quad (30)$$

and the adjoint eigenproblem is given by

$$\begin{cases} g(x,v)\frac{\partial}{\partial x}\phi_{\gamma,\rho}(x,a,v) + g(x,v)\frac{\partial}{\partial a}\phi_{\gamma,\rho}(x,a,v) \\ \quad + \gamma(x,a,v)\left(2 \int_{\mathcal{V}} \phi_{\gamma,\rho}\left(\frac{x}{2},0,v'\right)\rho(v,v')dv' - \phi_{\gamma,\rho}(x,a,v)\right) = \lambda_{\gamma,\rho}\phi_{\gamma,\rho}(x,a,v), \\ \phi_{\gamma,\rho} \geq 0, \quad \iiint_{\mathcal{S}} N_{\gamma,\rho}(x,a,v)\phi_{\gamma,\rho}(x,a,v)dxdvda = 1. \end{cases} \quad (31)$$

The existence and uniqueness of a solution  $(N_{\gamma,\rho}, \phi_{\gamma,\rho}, \lambda_{\gamma,\rho})$  to (30) and (31) can be proved in the same spirit as in [Dou07] where the age-size-structured model (*i.e.* the model described by (1) with  $g_a \equiv 1$  and a general  $g_x$ ) is studied. When adding the growth rate as a third structuring variable, the additional technicalities will be the ones we already faced in the proof of Theorem 4.

### *Deterministic and stochastic roles of the Malthus parameter*

The study of the eigenproblem is central since it governs the long-time behaviour of the solution  $n(t, x, a, v)$  to (2). The global growth of the population is driven by the parameter  $\lambda_{\gamma,\rho}$ . For large  $t$ , we expect  $n(t, x, a, v) \approx e^{\lambda_{\gamma,\rho} t} N_{\gamma,\rho}(x, a, v)$ , in a certain sense (for a weighted norm, see (13) for instance), the eigenvector  $N_{\gamma,\rho}$  playing the role of a stationary profile. A stochastic version of this result can also be written. Our model is embedded into the general framework described by Cloez in [Clo11] and thus Approximation (4) is valid in a certain sense (in probability). In view of our initial issue,  $\lambda_{\gamma,\rho}$  is the good quantity to focus on in order to measure the global growth of the population described by (2). The analysis which follows relies entirely on stochastic techniques and in particular on the asymptotic behaviour given by (4).

### 3.3 INFLUENCE OF VARIABILITY ON THE MALTHUS PARAMETER

We give here numerical answers to Question (i) for the size-increment-structured model, described by (2) with  $g_x(x, a, v) = g_a(x, a, v) = vx$  and  $\gamma(x, a, v)$  satisfying (25) (as announced in Table 7, Column 2). To reformulate more precisely Question (i), let us consider the case of reference without variability:

**Lemma 9.** *Consider the model without variability described by (1) with  $g_x(x, a) = g_a(x, a) = \bar{v}x$ ,  $\bar{v} > 0$ . Then the Malthus parameter is given by  $\lambda_{\gamma, \bar{v}} = \bar{v}$ .*

The proof of Lemma 9 is very short, we give in Appendix (Section 7.1). As previously, from now on we denote the Malthus parameter  $\lambda_{\gamma,\rho}$  by  $\lambda_{B,\rho}$  to emphasize we work with Specification (25). Choosing a variability kernel  $\rho(v, dv') = \rho(v')dv'$  with  $\rho : \mathcal{V} \rightarrow [0, \infty)$  a non-degenerated density such that  $\int_{\mathcal{V}} v\rho(v)dv = \bar{v}$ , the general question may state as follows: what can we say about  $\lambda_{B,\rho}$  compared to  $\bar{v}$  the average growth rate, which is the Malthus parameter of a population without variability in the growth rate?

#### *Numerical protocol*

In order to compare  $\lambda_{B,\rho}$  to  $\bar{v}$ , we numerically evaluate  $\lambda_{B,\rho}$  for given division rates  $B$  and growth rate densities  $\rho$ . Our estimates of  $\lambda_{B,\rho}$  are based on trees observed between time 0 and time  $T$ . The theoretical framework legitimating our approach is thus a continuous time setting as in Chapter I.

**Estimation of the Malthus parameter.** Define the sum of all sizes of living cells at time  $t$ ,

$$\mathfrak{M}_t = \sum_{u \in \partial\mathcal{T}_t} \xi_u^t = \sum_{u \in \partial\mathcal{T}_t} \xi_u e^{\tau_u(t-b_u)}, \quad (32)$$

where  $\partial\mathcal{T}_t$  is the set defined by (3). This quantity is called biomass in the biological literature. If we observe  $(\xi_u^{T/2}, u \in \partial\mathcal{T}_{T/2})$  and  $(\xi_u^T, u \in \partial\mathcal{T}_T)$  we can define an estimator of the Malthus

parameter as

$$\widehat{\lambda}_T = \frac{2}{T} \ln \left( \frac{\mathfrak{M}_T}{\mathfrak{M}_{T/2}} \right). \quad (33)$$

Approximation (4) entails

$$\mathfrak{M}_T \approx c_{\gamma,\rho} e^{\lambda_{\gamma,\rho} T} W_{\gamma,\rho},$$

in a certain sense (in probability), as  $T \rightarrow \infty$ , with  $c_{\gamma,\rho} > 0$  a constant that depends only on  $\gamma$  and  $\rho$  and  $W_{\gamma,\rho}$  a non-degenerated random variable. It ensures that the estimator we build is close to the true value for large  $T$ . Note that a large collection of estimators based on (4) can be built, choosing different test functions, and we discuss this later on. We see that the estimator (33) requires the observation of the whole population at two different large times, chosen for simplicity as  $T/2$  and  $T$ . Thus in order to numerically approximate  $\lambda_{B,\rho}$  by (33) for given parameters  $B$  and  $\rho$  one need to simulate continuous time trees up to a large time  $T$ .

**Simulation of a continuous time tree up to time  $T$ .** Let us be given a division rate  $B$  and a density  $\rho$  on  $\mathcal{V}$ . To simulate a full tree up to a time  $T$ , we have to keep the birth and division dates. From (28), the lifetimes in the size-increment-structured model are computed as

$$\zeta_u = \frac{1}{\tau_u} \ln \left( 1 + \frac{\Delta_u}{\xi_u} \right). \quad (34)$$

Recall in addition (26) and (29) giving the distributions of the growth rates and the increments of size. Set  $\xi_\emptyset = x_\emptyset$  and  $\tau_\emptyset = v_\emptyset$ , some given initial values, draw  $\Delta_\emptyset$  according to  $f_B$  and compute its division time  $d_\emptyset = \zeta_\emptyset$  by (34). For a cell  $u \in \mathbb{T}$ , given  $((b_v, \zeta_v, \xi_v, \Delta_v, \tau_v), v \prec u)$ , where the notation  $v \prec u$  means that  $v \neq u$  is an ancestor of  $u$ ,

- (i) If  $d_{u^-} < T$ ,
  - Set its size at birth  $\xi_u = (\xi_{u^-} + \Delta_{u^-})/2$ .
  - Draw its growth rate  $\tau_u$  according to  $\rho(\tau_{u^-}, \cdot)$  and its increment of size  $\Delta_u$  according to  $f_B$ .
  - Compute its lifetime  $\zeta_u$  by (34). Set its birth time  $b_u = d_{u^-}$  and its division time  $d_u = b_u + \zeta_u$ .
- (ii) If  $d_{u^-} \geq T$ , do not simulate the descendant  $u$  of  $u^-$ .

*Main result: the Malthus parameter decreases when introducing variability*

We first specify the division rate  $B$  (recall Specification (25)) and the growth rate Markov kernel  $\rho$ .

**Division rate.** We choose the division rate of the general form

$$B(a) = B_{a_0,\beta}(a) = (a - a_0)^\beta \mathbf{1}_{\{a \geq a_0\}} \quad (35)$$

for  $a_0 \geq 0$  and  $\beta \geq 0$ . This power form in  $\beta$  with some lag  $a_0$  is inspired by experimental data. For instance we refer to [RHK+14], Supplementary Figures S1 and S2, where the division rate is estimated for the age-structured and size-structured models. Adjusting  $\beta$  and  $a_0$  is an easy

way to mimic qualitatively all these curves (similar curves could be drawn for the size-increment-structured model). The estimation is not accurate for large sizes or ages since observations are lacking and we opt not to truncate  $B$  making it constant after some threshold but to let it grow to infinity. In this subsection, we pick  $a_0 = 1$  and  $\beta = 2$ .

**Parametrisation of the variability.** We suppose here that the growth rate of the daughter cell is independent of its parent. Given a non-degenerated baseline density  $\rho : \mathcal{V} \rightarrow [0, \infty)$  with mean  $\bar{v}$ , we define, for  $\alpha \in (0, 1]$ ,

$$\rho_\alpha(v) = \alpha^{-1} \rho(\alpha^{-1}(v - (1 - \alpha)\bar{v})), \quad v \in \mathcal{V}_\alpha = \alpha\mathcal{V} + (1 - \alpha)\bar{v}.$$

We set  $\int_{\mathcal{V}} v \rho(v) dv = \bar{v}$  and  $\int_{\mathcal{V}} (v - \bar{v})^2 \rho(v) dv = \eta^2$ . The coefficient of variation, denoted by CV indexed by the probability distribution considered, is defined as the quotient of the mean and the standard error of this probability distribution. Then  $CV_\rho = \bar{v}/\eta$  and  $CV_{\rho_\alpha} = \alpha CV_\rho$ . Set  $\varphi(x) = (2\pi)^{-1/2} e^{-x^2/2}$  the density of the standard gaussian and  $\Phi(x) = \int_{-\infty}^x \varphi(y) dy$  its cumulative distribution function. We pick for the baseline density  $\rho$ ,

$$\rho(v) = \frac{\varphi(\sigma_\eta^{-1}(v - \bar{v})) \mathbf{1}_{v_{\min} \leq v \leq v_{\max}}}{\Phi(\sigma_\eta^{-1}(v_{\max} - \bar{v})) - \Phi(\sigma_\eta^{-1}(v_{\min} - \bar{v}))} \quad (36)$$

with  $\bar{v} = (v_{\min} + v_{\max})/2$ . That is to say,  $\rho$  is the truncation on  $[v_{\min}, v_{\max}] = \mathcal{V}$  of a Gaussian distribution with mean  $\bar{v}$  and standard deviation  $\sigma_\eta$ . We set  $v_{\min} = 0$ ,  $v_{\max} = 2$  so that  $\bar{v} = 1$  and we choose  $\sigma_\eta = 0.70$  so that  $CV_\rho = 50\%$  (using the known formulae of the moments of a truncated Gaussian distribution). Once  $\rho$  is fixed, it enables us to define the collection  $(\rho_\alpha, \alpha \in (0, 1])$ . For  $\alpha \in (0, 1]$  the support of the density  $\rho_\alpha$  is  $\mathcal{V}_\alpha = [1 - \alpha, 1 + \alpha]$  and  $CV_{\rho_\alpha} = \alpha/2$  for such a choice of  $\rho$ . The larger the coefficient of variation  $CV_{\rho_\alpha}$ , the more variability in the growth rate.

**Estimation of the curve  $CV_{\rho_\alpha} \rightsquigarrow \lambda_{B, \rho_\alpha}$ .** For a given  $\alpha \in (0, 1)$ , we simulate  $M = 50$  continuous time trees up to time  $T$ , picking  $x_\emptyset = 2$ ,  $v_\emptyset = 1$ , with division rate  $B = B_{a_0=1, \beta=2}$  and growth rate density

$$\rho(v, dv') = \rho_\alpha(v') dv'.$$

We obtain a collection

$$(\widehat{\lambda}_{T,m}^{(\alpha)}, m = 1, \dots, M)$$

of  $M$  estimators of the Malthus parameter computed according to (33). We denote the mean of these  $M$  estimators by  $\widehat{\lambda}_T^{(\alpha)}$ . Figure 14 shows the curve

$$CV_{\rho_\alpha} \rightsquigarrow \widehat{\lambda}_T^{(\alpha)},$$

together with a confidence interval for each  $\alpha$ , such that it contains 95% of the  $M$  estimators. The aim is to get confidence intervals of good precision and to that end we choose large times  $T$ . The mean numbers of living individuals at time  $T$  is of magnitude 50 000, see Supplementary Table 9 in Section 8. (Such an amount of data, in a full tree case, would correspond to the observation of

the first 16 generations.) The value of reference is

$$\bar{v} = \int_{\mathcal{V}} v \rho_{\alpha}(v) dv$$

for any  $\alpha \in (0, 1)$ . We first observe that  $\lambda_{B, \rho_{\alpha}}$  is significantly lower than the value of reference  $\bar{v} = 1$ , for  $CV_{\rho_{\alpha}}$  exceeding 10%. In addition, the curve is significantly decreasing as  $CV_{\rho_{\alpha}}$  increases, *i.e.* as there is more and more variability in the growth rate.

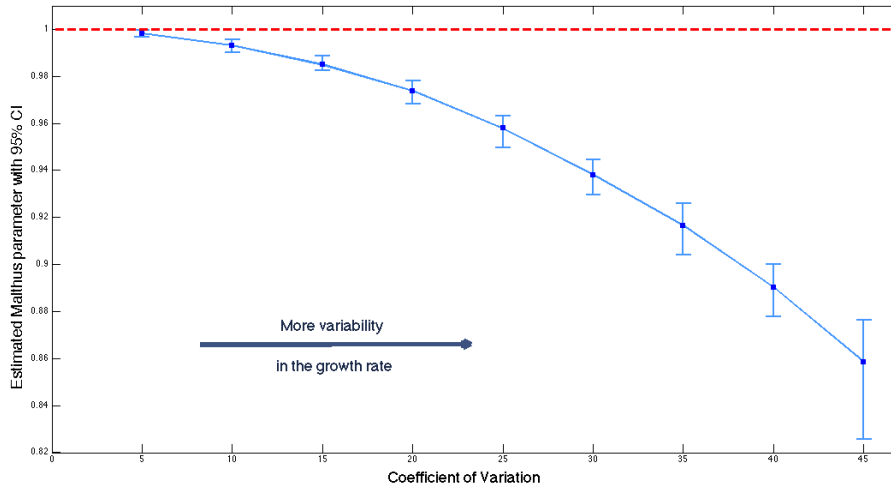


Figure 14: *Size-increment-structured model with variability. Division rate  $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$ . Estimated curve  $CV_{\rho_{\alpha}} \rightsquigarrow \lambda_{B, \rho_{\alpha}}$  using estimator (33) (mean and 95% confidence interval based on  $M = 50$  Monte Carlo continuous time trees). Reference (all cells grow at a rate  $\bar{v} = 1$ ):  $\lambda_{B, \bar{v}} = \bar{v} = 1$ .*

A last comment of importance, as regards upcoming experiments: the estimator of the Malthus parameter (33) we used, based on the biomass, is more precise than another estimator we could build using the number of cells. One, among many others, alternative estimator we could propose is

$$\hat{\lambda}_T = \frac{2}{T} \ln \left( \frac{|\partial \mathcal{T}_T|}{|\partial \mathcal{T}_{T/2}|} \right) \quad (37)$$

where  $|\partial \mathcal{T}_T|$  stands for the cardinality of the set (3) of living particles at time  $T$ . This gives another estimate of the Malthus parameter, still when observing the whole tree at large times  $T/2$  and  $T$ . For a fixed  $T$ , the estimator (33) is better than (37) in the sense that its standard deviation is much smaller (by at least a factor 2 for the lower  $CV_{\rho_{\alpha}}$  as shown in Supplementary Table 10 in Section 8). However as expected the two estimators perform equivalently for large  $T$  as pointed out in Supplementary Figure 21 in Section 8.

### Robustness of the result

The result illustrated by Figure 14 is robust to the change of parameters. The result is robust when changing the division rate  $B$ . In particular, we have explored several couples of parameters  $(a_0, \beta)$  (recall parametrisation (35)). The conclusion remains the same: the Malthus parameter significantly decreases when there is more and more variability in the growth rate. Supplementary Table 11 in Section 8 displays the results for  $a_0 = 1$  and  $\beta = 8$ .

So far we have assumed symmetry in the division. However a slight asymmetry can arise, see Marr *et al.* [MHT66] or Soifer *et al.* [SRBA14] (Figure 8). Asymmetry is defined as the ratio of the size at birth of one daughter cell over the size at division of its parent. The distribution of this ratio shows a mode at 0.50 and has a coefficient of variation of about 4%.

In our numerical study, we amplify asymmetry since we aim at measuring an effect of asymmetry on the curve  $CV_{\rho_\alpha} \rightsquigarrow \lambda_{B, \rho_\alpha}$ . We introduced asymmetry assuming that a cell of size  $x$  divides into two cells of sizes  $ux$  and  $(1-u)x$  with  $u$  uniformly drawn on the interval  $[\varepsilon, 1-\varepsilon]$ , independently of everything else, for  $0 \leq \varepsilon < 1/2$ . A dividing cell does not produce an arbitrary small cell, that is why we chose  $\varepsilon \neq 0$ . However we fix  $\varepsilon$  quite close to 0 for our study. Supplementary Table 12 in Section 8 collects the results for  $\varepsilon = 0.10$ . Once again there is no significant change compared to Figure 14. Thus asymmetry has no significant impact on the Malthus parameter, in presence of variability.

### Toward experiments

The aim here is to be closer to experimental conditions as regards the number of observed cells. The main issue is the following: is the variation of the Malthus parameter significant enough to have a chance to observe it experimentally? The techniques of measurements available enable us to observe at most 9 generations of a population of cells, which means that the population counts at most 500 cells when we stop observing. Repeating 10 times the experiment seems reasonable. Assuming the initial individuals of the 10 populations are generated according to the stationary profile, we can aggregate the data and work as if we observed one population counting approximately 5 000 cells at the final observation time. Thus we want to build confidence intervals for the Malthus parameter when the number of observations is around 5 000, which is realistic for experiments. Note that we do not observe the population of cells up to a fixed time  $T$  any longer. The observation time is instead a random stopping time. However we still use the same estimator (33) as previously.

Figure 15 shows that for a coefficient of variation in the growth rate exceeding 15%, the confidence interval at 95% (built from  $M = 50$  Monte Carlo continuous time trees observed until the random time  $\Pi$  such that  $|\partial\mathcal{T}_\Pi|$  counts between 4 500 and 5 500 individuals) does not overlap the value of reference. We expect the experimental evaluation of the Malthus parameter lie in the 95% confidence intervals we computed.

Our preliminary study based on numerical simulations also enables us to design experiments. Two experimental parameters have to be fixed: the number of cells we observe and the coefficient of variation of the growth rates distribution. As we said, the observation of 5 000 cells is experimentally feasible and a realistic coefficient of variation in the growth rates is around 10%-20%. Our

numerical results (see Figure 15) entail that we can legitimately hope to measure experimentally a significant change in the Malthus parameter observing 5 000 cells for a coefficient of variation around 15% (compared to the value of reference equal to the mean growth rate).

Exploiting the observation of 10 different populations of cells (without aggregating the data) could also be interesting. One could obtain 10 estimates of the Malthus parameter by (33) and build experimental confidence intervals. One could say this approach is model independent since those intervals would be valid for any model described by (2) without any further specification on  $g_x, g_a$  and  $\gamma$ . However the size of each population being small, around 500, we expect that the confidence intervals overlap the value of reference (this guess is based on the same process as previously, we compute 95% confidence intervals, based on  $M = 50$  Monte Carlo continuous time trees observed until the random time  $\Pi$  such that  $|\partial\mathcal{T}_\Pi|$  counts between 450 and 550 individuals, by varying the coefficient of variation in the growth rate between 5% and 25%).

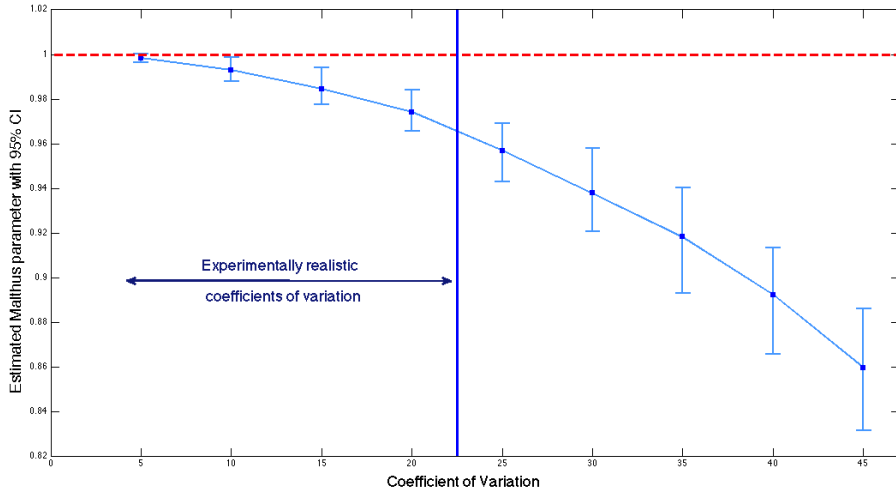


Figure 15: *Size-increment-structured model with variability, experimental conditions.* Division rate  $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$ . Estimated curve  $CV_{\rho_\alpha} \rightsquigarrow \lambda_{B, \rho_\alpha}$  using estimator (33) (mean and 95% confidence interval based on  $M = 50$  Monte Carlo continuous time trees simulated up to the random time  $\Pi$  such that  $4\,500 \leq |\partial\mathcal{T}_\Pi| \leq 5\,500$ ). Reference (all cells grow at a rate  $\bar{v} = 1$ ):  $\lambda_{B, \bar{v}} = \bar{v} = 1$ .

### Extension: heredity

So far we have assumed independence between the growth rates of the cells of the population setting  $\rho(v, dv') = \rho(dv')$ . In this subsection we allow the growth rate to be inherited from one cell to another, which is more realistic in some controlled medium, see Kiviet *et al.* [KNW+14]. For *E. coli*, heredity is *not* negligible in poor medium whereas it is in rich medium. Two ways to take into account heredity are explored, both being biologically plausible.

We first assume that the growth rates  $(\tau_u, u \in \mathbb{T})$  form a bifurcating autoregressive process



(see Chapter III) with linear autoregressive functions. Second, we assume that cells fall into two groups, one with cells growing fast, another one with cells growing slowly (this is why we call this case *bimodal*). Heredity is introduced there assuming the group to which a cell belongs depends on the group of its parent.

**Linear bifurcating autoregressive process.** To take into account heredity in the transmission of the growth rate, we assume the growth rates sequence follows an autoregressive dynamic,

$$\tau_u = (1 - \kappa)\bar{v} + \kappa\tau_{u^-} + \eta_u, \quad u \in \mathbb{T} \setminus \{\emptyset\} \quad (38)$$

with  $\kappa \in (0, 1)$  and  $(\eta_u, u \in \mathbb{T})$  independent and identically distributed with  $\mathbb{E}[\eta_u] = 0$  and with support subset of  $[-(1 - \kappa)(v_{\max} - v_{\min})/2; (1 - \kappa)(v_{\max} - v_{\min})/2]$  so that  $\tau_u \in [v_{\min}, v_{\max}] = \mathcal{V}$ . We set  $\bar{v} = (v_{\min} + v_{\max})/2$ . In addition,  $\eta_u$  is independent of  $(\tau_v, v \preceq u)$  where the notation  $v \preceq u$  means that  $v$  is an ancestor of  $u$ , with possibly  $v = u$ . In other words,  $(\tau_u, u \in \mathbb{T})$  is a symmetric linear autoregressive bifurcating process. The higher the linear coefficient  $\kappa$  is, the stronger heredity is. The case without heredity in the transmission of the growth is given by  $\kappa = 0$ .

Let us denote by  $\eta$  the standard deviation of the noise sequence. At equilibrium (see Chapter III for a precise meaning), we have  $\mathbb{E}[\tau_u] = \bar{v}$  and  $\mathbb{E}[(\tau_u - \bar{v})^2] = \eta^2(1 - \kappa^2)^{-1}$ , thus the coefficient of variation of the equilibrium distribution of the growth rates is

$$CV_\kappa = \frac{\eta(1 - \kappa^2)^{-1/2}}{\bar{v}}.$$

We denote the Malthus parameter by  $\lambda_{B,\kappa}$  in this case and aim at studying  $CV_\kappa \rightsquigarrow \lambda_{B,\kappa}$ .

We still base our study on the simulation of continuous time trees up to a large time (we keep on with the protocole described page 191 drawing  $\tau_u$  conditionally on  $\tau_{u^-}$  complying with (38)). The division rate is still picked as  $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$ . We pick  $\bar{v} = 1$ ,  $v_{\min} = 0.1$  and  $v_{\max} = 1.9$  fixed once and for all. For a given  $\kappa$ , we draw  $\eta_u$  according to a centered Gaussian distribution of variance  $\sigma_\eta^2$  truncated on  $[-6(1 - \kappa)/10; 6(1 - \kappa)/10]$ . We choose values  $\mathbb{E}[\eta_u^2] = \eta^2 = \eta_\kappa^2$  such that  $CV_\kappa$  takes its value in  $\{5\%, 10\%, 15\%, 20\%, 25\%\}$ . For numerical experiments, we pick several  $\kappa$ :  $\kappa = 0$  (no heredity),  $\kappa = 0.25$  and  $\kappa = 0.50$ . On Figure 16, we observe that:

- 1) **For a fixed  $\kappa$ , study of  $CV_\kappa \rightsquigarrow \lambda_{B,\kappa}$ .** In presence of a not too strong heredity ( $\kappa = 0.25$ ), the Malthus parameter  $\lambda_{B,\kappa}$  decreases as the variability in the growth rate measured by  $CV_\kappa$  increases. For a higher level of heredity ( $\kappa = 0.50$ ), the Malthus parameter  $\lambda_{B,\kappa}$  seems constant, close to the mean growth rate  $\bar{v} = 1$ , as  $CV_\kappa$  increases.
- 2) **For a fixed  $CV_\kappa$ , study of  $\kappa \rightsquigarrow \lambda_{B,\kappa}$**  For a coefficient of variation  $CV_\kappa$  fixed between 15% and 25%, we observe that the stronger heredity is, *i.e.* the higher  $\kappa$  is, the higher the Malthus parameter  $\lambda_{B,\kappa}$  is: Figure 16 suggests  $\lambda_{B,\kappa=0} < \lambda_{B,\kappa=0.25} < \lambda_{B,\kappa=0.50}$  and we infer  $\kappa \rightsquigarrow \lambda_{B,\kappa}$  is non-decreasing.
- 3) **Comparison of  $\lambda_{B,\kappa}$  to  $\bar{v}$ .** Introducing variability in the growth rate (with no heredity,  $\kappa = 0$ ) slows down the global growth of the population as we have already seen, we significantly

have  $\lambda_{B,\kappa=0} < \bar{v} = 1$ . However heredity may have a counter effect which fully offset the previous slowing down effect. This phenomenon is observed for  $\kappa$  large enough: Figure 16 shows  $\lambda_{B,\kappa=0.50} \approx \bar{v} = 1$ . One may wonder if for  $\kappa$  higher than 0.50 the global growth of the population can be sped up: one would guess from Figure 16 that  $\lambda_{B,\kappa} > \bar{v} = 1$  for high  $\kappa$  (close to 1). However our numerical tests for  $\kappa = 0.75$  did not show any significant difference from the curve corresponding to  $\kappa = 0.50$ .

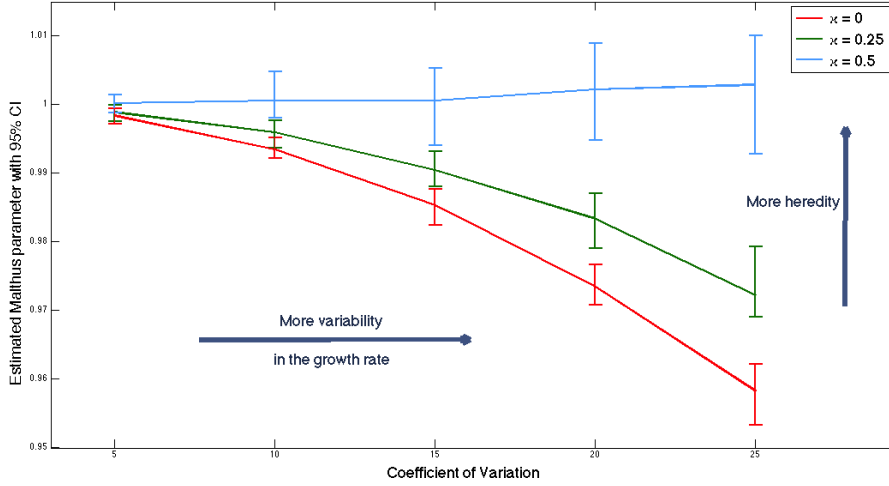


Figure 16: *Size-increment structured model with inherited variability. Division rate  $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$  and the growth rates  $(\tau_u, u \in \mathbb{T})$  satisfy (38). Estimated curve  $CV_\kappa \rightsquigarrow \lambda_{B,\kappa}$  using estimator (33) (mean and 95% confidence interval based on  $M = 50$  Monte Carlo continuous time trees) for different values of  $\kappa$ . Red lower curve:  $\kappa = 0$  (no heredity). Green middle curve:  $\kappa = 0.25$ . Blue upper curve:  $\kappa = 0.50$ .*

**Bimodal case.** An other way to introduce heredity in the transmission of the growth rate is the following. Assume that the growth rate take its values in  $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$  with  $\mathcal{V}_1$  and  $\mathcal{V}_2$  two compact intervals of  $(0, \infty)$ . It means that some cells grow at a fast rate, in the interval  $\mathcal{V}_2$  say, whereas the others grow at a slow rate, in the interval  $\mathcal{V}_1$  say. We specify the model with probabilities of transition  $(p_{ij} > 0, (i, j) \in \{1, 2\}^2)$  such that  $p_{11} + p_{12} = 1$ ,  $p_{21} + p_{22} = 1$  and with two densities  $\rho_1$  and  $\rho_2$  with support  $\mathcal{V}_1$  and  $\mathcal{V}_2$  respectively. The growth rate transmission follows the rules

$$\mathbb{P}(\tau_u \in \mathcal{V}_j | \tau_{u^-} \in \mathcal{V}_i) = p_{ij}, \quad (i, j) \in \{1, 2\}^2, \quad (39)$$

where  $u^-$  is the parent of  $u$ , and

$$\mathbb{P}(\tau_u \in dv | \tau_u \in \mathcal{V}_i) = \rho_i(v)dv, \quad i \in \{1, 2\}. \quad (40)$$

At equilibrium, with a discrete time point of view, cells growing at a slow rate or at a fast rate have proportions respectively given by

$$q_1 = \frac{1 - p_{22}}{(1 - p_{11}) + (1 - p_{22})}, \quad q_2 = \frac{1 - p_{11}}{(1 - p_{11}) + (1 - p_{22})}.$$

Thus at equilibrium, the mean growth rate of the population is

$$\bar{v} = \bar{v}_1 q_1 + \bar{v}_2 q_2$$

where  $\bar{v}_i = \int_{\mathcal{V}_i} v \rho_i(v) dv$  for  $i \in \{1, 2\}$ , and we use this value as reference. We denote the Malthus parameter by  $\lambda_{B, \bar{v}_1, \bar{v}_2, p_{11}, p_{22}}$  and we want to compare it to the value of reference for different parameters  $(\bar{v}_1, \bar{v}_2)$  and  $(p_{11}, p_{22})$ .

We keep on with the same protocol as previously, calculating 95% confidence intervals drawing  $M = 50$  Monte Carlo continuous time trees. For our numerical tests, we pick  $\rho_i$  as a very picky distribution, a Gaussian distribution with mean  $\bar{v}_i$  and standard deviation 0.03 truncated on the interval  $[\bar{v}_i - 0.10, \bar{v}_i + 0.10]$  (thus we have  $\mathcal{V}_1 \cap \mathcal{V}_2 = \emptyset$ ). The mean growth rates of the two groups are chosen as  $\bar{v}_1 = 1 - \varepsilon$  and  $\bar{v}_2 = 1 + \varepsilon$  with  $\varepsilon \in \{0.20, 0.40, 0.60, 0.80\}$ . We use the distance  $|\bar{v}_2 - \bar{v}_1| = 2\varepsilon$  as  $x$ -scale on Figure 17.

For toy example, the transition probabilities are chosen so that **1**) the probability  $p_{11}$  to stay in the slowly growing group is high compared to the probability  $p_{12}$  to become fast growing (*i.e.* innovation among cells is not frequent) and **2**) the probability  $p_{12}$  is smaller than the probability  $p_{21}$  (*i.e.* cells have more chance to degenerate than to increase their growth rate at the next generation). We fix the probability to stay in the slowly growing group  $p_{22} = 0.30$  and test the probability to stay in the fast growing group  $p_{11} = 0.90$  or  $p_{11} = 0.60$ .

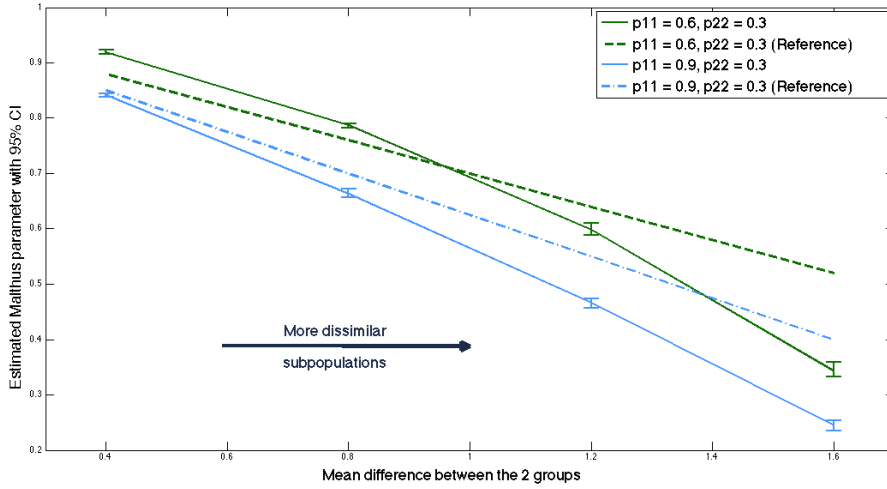


Figure 17: *Size-increment structured model with inherited variability. Division rate  $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$  and the growth rates  $(\tau_u, u \in \mathbb{T})$  satisfy (39) and (40). Estimated curve  $|\bar{v}_2 - \bar{v}_1| \rightsquigarrow \lambda_{B, \bar{v}_1, \bar{v}_2, p_{11}, p_{22}}$  (with  $\bar{v}_i$  the mean growth rate of the subpopulation  $i \in \{1, 2\}$  and  $\bar{v}_1 = 2 - \bar{v}_2$ ) using estimator (33) (mean and 95% confidence interval based on  $M = 50$  Monte Carlo continuous time trees) for different values of  $(p_{11}, p_{22})$ . Green upper curve:  $p_{11} = 0.60$  and  $p_{22} = 0.30$ . Blue lower curve:  $p_{11} = 0.90$  and  $p_{22} = 0.30$ . Reference (all cells grow at the mean growth rate  $\bar{v}$ ):  $\bar{v} = \bar{v}_1 q_1 + \bar{v}_2 q_2$ , in dashed lines for each case.*

Not surprisingly, the higher  $p_{11}$  is, all other parameters remaining equal, the lower the Malthus parameter is, as one can see on Figure 17 (the blue curve is larger than the green one). The more the two groups are different *i.e.* the higher the distance  $|\bar{v}_2 - \bar{v}_1|$  between the mean growth rates of the two groups is, the lower the Malthus parameter is. It is interesting to observe that, for  $p_{11} = 0.60$ , depending on  $\bar{v}_1$  and  $\bar{v}_2$ , the Malthus parameter can be smaller or higher than the value of reference. It means that in some cases the population with variability in the growth rate grows faster than the population of reference composed of cells all growing at the mean growth rate. (As a comparison, the bimodal case without heredity can be studied as well, see Supplementary Figure 22 in Section 8 and Supplementary Table 16. However, for *E. coli*, heredity may not be negligible in the bimodal case.)

Further tests should be implemented to explore the bimodal case. As a first conclusion, we see that the choice of  $p_{11}$  and  $p_{22}$  is determinant. We should also try less spiky distributions  $\rho_1$  and  $\rho_2$  since in practice  $\mathcal{V}_1$  and  $\mathcal{V}_2$  overlap.

Three main conclusions regarding the size-increment-structured model with variability in the growth rate we are studying are in order:

- 1) For a unit size division rate  $B$  experimentally plausible, when there is variability in the growth rate among cells, neglecting heredity at first,
  - (a) The Malthus parameter is lower than the value of reference computed assuming all cells grow at the mean growth rate.
  - (b) In addition the Malthus parameter decreases when there is more and more variability.
  - (c) The variation is of magnitude 2% for experimentally realistic coefficients of variation in the growth rates distribution, around 15-20%.

These conclusions are robust as argued above changing  $B$  or introducing asymmetry in the division.

- 2) When we observe populations of realistic sizes (around 5 000 cells), the decrease of the Malthus parameter still seems significant which augurs well for the experimental part.
- 3) When the individual growth rate is inherited from one cell to another, the Malthus parameter may be equal or even larger than the value of reference computed assuming all cells grow at the equilibrium mean growth rate.

Our conclusion 1)(a) precisely coincides with conventional wisdom in biology. However, our conclusion 3) was perhaps not expected by the biologists. At the light of these conclusions, it seems crucial to measure the strength of heredity in the transmission of the growth rate to decide whether we can neglect its influence or not (see also the perspectives detailed in Section 5).

So far we based our methodology estimating the Malthus parameter from the observation of a continuous-tree (*i.e.* of a whole population until a given time). We stress that our previous conclusion **2**), with experimental concerns, relies on this methodology. Proposing alternative procedures to estimate the Malthus parameter (from alternative observation schemes) seems important to better prepare an experimental part (even if our preliminary conclusion **2**) is positive).

A quite recent and successful protocol is the observation of one branch in the genealogical tree of the population (see some details on page 177) and we now briefly describe a methodology to estimate the Malthus parameter when observing one branch. Further investigating Question (ii) constitutes one of the main perspectives of this work.

### 3.4 ESTIMATION OF THE MALTHUS PARAMETER FROM ONE BRANCH

We keep on with Specification (25) of the division rate,  $\gamma(x, a, v) = g(x, v)B(a)$  for a general growth speed  $g$ . At first, we neglect heredity in the transmission of the growth rate, specifying  $\rho(v, dv') = \rho(v')dv'$ . In the sequel we assume that  $g$  is *known* from previous experimentations whereas  $B$  and  $\rho$  are *unknown*. We observe

$$((\Delta_u, \tau_u), u \in \mathbb{L}_n)$$

with  $\mathbb{L}_n$  a branch of  $n$  cells defined by (24) and aim at reconstructing the Malthus parameter  $\lambda_{B,\rho}$  from these observations.

**One estimation procedure.** Remind the adjoint eigenproblem described by (31) and set

$$\mathcal{A}_{B,\rho}f(x, a, v) = \mathcal{G}f(x, a, v) + \mathcal{F}_{B,\rho}^+f(x, a, v) - \mathcal{F}_B^-f(x, a, v),$$

densely defined on bounded continuous functions, with

$$\begin{aligned} \mathcal{G}f(x, a, v) &= g(x, v)\frac{\partial}{\partial x}f(x, a, v) + g(x, v)\frac{\partial}{\partial a}f(x, a, v), \\ \mathcal{F}_{B,\rho}^+f(x, a, v) &= 2g(x, v)B(a)\int_{\mathcal{V}}f\left(\frac{x}{2}, 0, v'\right)\rho(v, v')dv', \\ \mathcal{F}_B^-f(x, a, v) &= g(x, v)B(a)f(x, a, v). \end{aligned}$$

The Malthus parameter is defined as the dominant eigenvalue of the operator  $\mathcal{A}_{B,\rho}$ . We proceed in two steps: we first approximate the operator  $\mathcal{A}_{B,\rho}$  by a finite-dimensional matrix  $\mathbf{A}_{B,\rho}^N$  using a projection method, the finite-dimensional matrix  $\mathbf{A}_{B,\rho}^N$  is then estimated by  $\widehat{\mathbf{A}}_n^N$ . One can hope the spectrum of the estimated matrix is easy to calculate numerically and calculating its largest eigenvalue provides an estimate for the Malthus parameter. This kind of method was successfully set up in Gobet, Hoffmann and Reiß[GHR04] in another framework. Let us describe more precisely how we carry out these two steps in our case.

*Step 1. Discretisation.* Let us be given a basis  $(w_i, 1 \leq i \leq N)$  of some  $N$ -dimensional approximation space. We define the finite-dimensional operator  $\mathbf{A}_{B,\rho}^N$  by

$$\mathbf{A}_{B,\rho}^N = \left[ \iiint_{\mathcal{S}} (\mathcal{A}_{B,\rho}w_i)(x, a, v)w_j(x, a, v)dxdadv \right]_{1 \leq i, j \leq N},$$

which we decompose into  $\mathbf{A}_{B,\rho}^N = \mathbf{G}^N + \mathbf{F}_{B,\rho}^{+,N} - \mathbf{F}_B^{-,N}$ . The finite-dimensional matrix  $\mathbf{G}^N$ , corresponding to the growth part, whose terms are given by

$$(\mathbf{G}^N)_{i,j} = \iiint_{\mathcal{S}} g(x,v) \left( \frac{\partial}{\partial x} w_i(x,a,v) + \frac{\partial}{\partial a} w_i(x,a,v) \right) w_j(x,a,v) dx da dv,$$

does not depend on the unknown  $(B, \rho)$  and can be computed in a preliminary step, independently of the observed sample. As regards the fragmentation part, the matrix  $\mathbf{F}_{B,\rho}^{+,N}$  has terms

$$(\mathbf{F}_{B,\rho}^{+,N})_{i,j} = \iiint_{\mathcal{S}} (\mathcal{F}_{B,\rho}^+ w_i)(x,a,v) w_j(x,a,v) dx da dv = \iint_{(0,\infty) \times \mathcal{V}} h_{i,j}^+(a,v') B(a) \rho(v') da dv'$$

with

$$h_{i,j}^+(a,v') = 2 \iint_{(0,\infty) \times \mathcal{V}} g(x,v) w_i\left(\frac{x}{2}, 0, v'\right) w_j(x,a,v) dx dv,$$

and the matrix  $\mathbf{F}_B^{-,N}$  has terms

$$(\mathbf{F}_B^{-,N})_{i,j} = \iiint_{\mathcal{S}} (\mathcal{F}_B^- w_i)(x,a,v) w_j(x,a,v) dx da dv = \int_0^\infty h_{i,j}^-(a) B(a) da$$

with

$$h_{i,j}^-(a) = \iint_{(0,\infty) \times \mathcal{V}} g(x,v) w_i(x,a,v) w_j(x,a,v) dx dv.$$

Note that neither  $h_{i,j}^+$  nor  $h_{i,j}^-$  depend on  $B$  or  $\rho$  so that they can be calculated in a preliminary step.

*Step 2. Estimation.* Setting  $f_B(a) = B(a) \exp(-\int_0^a B(s) ds)$  and  $F_B(a) = \int_0^a f_B(s) ds$ , the classical relation

$$B(a) = \frac{f_B(a)}{1 - F_B(a)}$$

holds. For  $h_{i,j}^\pm(a,v')$  being either  $h_{i,j}^+(a,v')$  or  $h_{i,j}^-(a)$ , we write

$$\iint_{(0,\infty) \times \mathcal{V}} h_{i,j}^\pm(a,v') B(a) \rho(v') da dv' = \iint_{(0,\infty) \times \mathcal{V}} \frac{h_{i,j}^\pm(a,v')}{1 - F_B(a)} f_B(a) \rho(v') da dv'.$$

Define  $\widehat{F}_n$  as an estimator of the cumulative distribution function  $F_B$ , converging at the parametric rate  $n^{-1/2}$ . Equation (29) and the assumption  $\rho(v, dv') = \rho(v') dv'$  prove that the couples  $(\Delta_u, \tau_u)$  are independent for  $u \in \mathbb{L}_n$  and distributed as  $f_B(a) \rho(v) da dv$ . Thus we estimate  $(\mathbf{F}_{B,\rho}^{+,N})_{i,j}$  and  $(\mathbf{F}_B^{-,N})_{i,j}$  by respectively

$$(\widehat{\mathbf{F}}_n^{+,N})_{i,j} = \frac{1}{n} \sum_{u \in \mathbb{L}_n} \frac{h_{i,j}^+(\Delta_u, \tau_u)}{1 - \widehat{F}_n(\Delta_u)}, \quad (\widehat{\mathbf{F}}_n^{-,N})_{i,j} = \frac{1}{n} \sum_{u \in \mathbb{L}_n} \frac{h_{i,j}^-(\Delta_u)}{1 - \widehat{F}_n(\Delta_u)},$$

for  $1 \leq i, j \leq N$ . We finally define an estimator of the discretised operator  $\mathbf{A}_{B,\rho}^N$  as

$$\widehat{\mathbf{A}}_n^N = \mathbf{G}^N + \widehat{\mathbf{F}}_n^{+,N} - \widehat{\mathbf{F}}_n^{-,N}.$$

*Statistical error and numerical implementation.* The precise study of this methodology would give a thorough answer to Question (ii) when observing one branch. We sketch here a work approach. The main heuristic idea is that the error in the eigenvalue can be controlled by the error in the operator. We should give a sense to the chain of approximations

$$\mathcal{A}_{B,\rho} \stackrel{(1)}{\approx} \mathbf{A}_{B,\rho}^N \stackrel{(2)}{\approx} \widehat{\mathbf{A}}_n^N,$$

in to order to successively obtain the approximations

$$\lambda_{B,\rho} \stackrel{(3)}{\approx} \lambda_{B,\rho}^N \stackrel{(4)}{\approx} \widehat{\lambda}_n^N.$$

Intuitively, the errors generated by the discretisation step (Approximations (1) and (3)) are related to a usual bias term, which only depends on  $N$ , and the errors generated to the estimation step (Approximations (2) and (4)) are related to a usual variance term, which depends on  $n$  and  $N$ . We expect a balance (governed by the choice of the dimension  $N$  of the approximation space) between the two terms leading to a nonparametric rate depending on the smoothness of the unknown parameters  $(B, \rho)$  and the dimension of the problem.

The difficult point of the whole proof is to prove the existence of a largest eigenvalue  $\lambda_{B,\rho}^N$  of the discretised version  $\mathbf{A}_{B,\rho}^N$  of the operator  $\mathcal{A}_{B,\rho}$ . Indeed, we work in a non-reversible framework and thus the operator  $\mathcal{A}_{B,\rho}$  is *not* self-adjoint and the matrix  $\mathbf{A}_{B,\rho}^N$  is nonsymmetric. As regards Approximation (2), we expect a control of the error as in Proposition 4.14 of [GHR04]. Then, once the existence of  $\lambda_{B,\rho}^N$  is proved, proving that Approximation (2) entails Approximation (4) relies on the spectral approximation theory for finite-dimensional matrices. We stress again that we fall within the spectral approximation theory for *nonsymmetric* matrices. As regards Approximations (1) and (4), we should rely on the general spectral approximation theory (see Chatelin [Cha83] for instance). As explained previously, we expect this estimation procedure based on a single branch of the population to achieve a nonparametric rate for the estimation of the Malthus parameter  $\lambda_{B,\rho}$ , and we expect this rate to be optimal. However remind that in the toy model structured by the age and the aging rate we achieved the parametric rate  $n^{-1/2}$ . A lower bound for such a problem is thus far from being obvious.

As for the numerical implementation, the use of finite element spaces of approximation (see [CL09]) should enable us to easily calculate the matrix  $\mathbf{G}^N$  and to easily implement the functions  $h_{i,j}^+$  and  $h_{i,j}^-$ . The sparse structure of the output matrix, inherited from the properties of the finite elements, would have the huge advantage to ease the calculation of the largest eigenvalue.

*Generalisations.* Taking heredity into account would be possible. One should change the scalar product chosen in Step 1 to proceed to the discretisation and consider

$$\iiint_{\mathcal{S}} (\mathcal{A}_{B,\rho} w_i)(x, a, v) w_j(x, a, v) \nu_\rho(v) dx da dv$$

where  $\nu_\rho$  is the invariant measure of the bifurcating Markov chain ( $\tau_u, u \in \mathbb{T}$ ), we refer to Chapter II for a definition. We should follow [GHR04] introducing the estimated Gram matrix to take into account the fact that  $\nu_\rho$  is unknown. One could consider a general fragmentation, assumed to

be *known*, a cell of size  $x$  giving birth to a cell of size  $y$  with probability  $\kappa(x, y)$  where  $\kappa$  is a fragmentation kernel, nonnegative, such that  $\kappa(x, y) = 0$  for  $y > x$  and  $\int_0^x \kappa(x, y) dy = 1$ ,  $\int_0^x y \kappa(x, y) dy = \frac{x}{2}$ .

**Alternative estimation procedures.** We can alternatively propose two other procedures: we could first estimate the division rate  $B$  and the Markov kernel of the growth rates  $\rho$  (note that such estimators are easily obtained observing one branch) and then apply classical methods *as if we knew*  $B$  and  $\rho$ .

After an estimation of  $B$  and  $\rho$  from the observation of one branch, one could simulate continuous time trees (using the same protocole as described page 191) and use again the estimator defined by (37) based on the biomass. The statistical study of such a procedure remains unclear, but it seems competitive at a computational level.

When knowing  $B$  and  $\rho$ , it is possible to discretise the partial differential equation (2) using a classical finite volume scheme (we refer in particular to [CDG12, BCP08] or more generally to the textbook of Leveque [Lev02]) and to compute in that way the Malthus parameter  $\lambda_{B, \rho}$ . When  $B$  and  $\rho$  are unknown, if we have estimators for them, the same numerical scheme could enable us to estimate  $\lambda_{B, \rho}$ . We should conduct a precise study of how the estimation of  $(B, \rho)$  affects the estimation of  $\lambda_{B, \rho}$ . For the size-structured model (without variability), Perthame and Zubelli [PZ07] (Theorem 3.2) and Doumic and Tine [DT13] (Theorem 1.1) studied the error in the Malthus parameter  $\lambda_B$  when having an error in  $B$ . We should also take into account the error coming from the numerical scheme. A limitation has to be formulated as regards computational efficiency since four dimensions occur (time added to three structuring variables).

The comparison of the previous estimation procedures at a statistical level remains to be investigated. We stress that since we are interested in a three dimensional model (the three structuring variables being the size, the increment of size and the growth rate), achieving computational efficiency is an additional issue. The comparison of the procedures at a computational level remains also to be investigated

## 4 TWO OTHER MODELS OF INTEREST

Our conclusions are robust to the change of model. We discuss here the size-structured model and a recent model introduced by Amir [Ami14]. In both cases, for a reasonable choice of the division rate  $B$  (as a power with some lag) variability in the growth rate slows down the global growth of the population (see Figure 18 and Supplementary Table 18 in Section 8).

### 4.1 SIZE-STRUCTURED MODEL WITH VARIABILITY

We remind that the size-structured model can be derived from the general formalism (2) setting  $g_x(x, a, v) = vx$ ,  $g_a \equiv 1$  and  $\gamma$  independent of  $a$  (see Column 3 of Table 7) and integrating the equation in  $a$  from zero to infinity. We refer to [DG10] and references therein for the study of the eigenproblem. We work with a unit time division rate chosen as

$$\gamma(x, a, v) = vxB(x).$$



For a cell  $u \in \mathbb{T}$ , let  $\zeta_u, \xi_u$  and  $\tau_u$  denote respectively its lifetime, its size at birth and its growth rate. The size of cell  $u$  evolves as  $\xi_u^t = \xi_u \exp(\tau_u(t - b_u))$  for  $t \in [b_u, d_u)$ . Stochastically, cells divide according to the rate  $vx_B(x)dt$ ,

$$\mathbb{P}(\zeta_u \in [t, t + dt) | \zeta_u \geq t, \xi_u = x_b, \tau_u = v) = vx_b e^{vt} B(x_b e^{vt}) dt$$

where  $x_b e^{vt} = x$  is the size after a time  $t$  for a cell born with size  $x_b$  and growth rate  $v$ . Since  $\xi_u = \frac{1}{2} \xi_{u^-} \exp(\tau_{u^-} \zeta_{u^-})$ , where  $u^-$  is the parent of  $u$ , we readily obtain

$$\mathbb{P}(\xi_u \in dy | \xi_{u^-} = x) = 2B(2y) \exp\left(-\int_{x/2}^y 2B(2s) ds\right) \mathbf{1}_{y \geq x/2} dy. \quad (41)$$

Thus  $(\xi_u, u \in \mathbb{T})$  is a bifurcating Markov chain (see Chapter II).

Notice that the model studied in [DHKR15] corresponds to the choice  $\gamma(x, a, v) = B(x)$  (we say that  $B$  is unit time) instead of  $\gamma(x, a, v) = vx_B(x)$  (we say here that  $B$  is unit size). This is fundamentally different. As one can see in (41), with the choice  $\gamma(x, a, v) = vx_B(x)$ , the size at birth of a cell actually does not depend on the growth rate of its parent whereas it would be the case with the choice  $\gamma(x, a, v) = B(x)$  (see [DHKR15], Equation (11)) and this is the main difference.

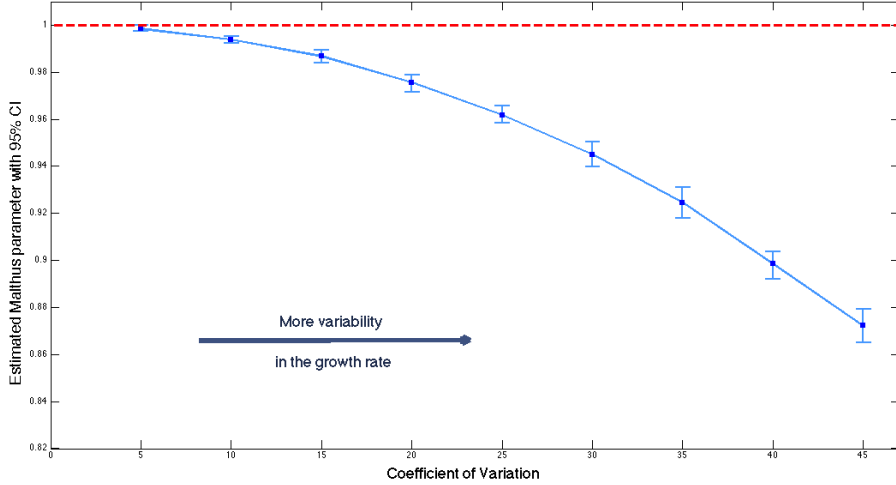


Figure 18: *Size-structured model with variability. Division rate  $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$ . Estimated curve  $CV_{\rho_\alpha} \rightsquigarrow \lambda_{B, \rho_\alpha}$  using estimator (33) (mean and 95% confidence interval based on  $M = 50$  Monte Carlo continuous time trees). Reference (all cells grow at a rate  $\bar{v} = 1$ ):  $\lambda_{B, \bar{v}} = \bar{v} = 1$ .*

With the help of the same protocol as previously, we obtain a numerical answer to Question (i) as a preliminary, see Figure 18 and Supplementary Table 17 in Section 8. For the unit size division rate  $B$  we chose, experimentally realistic, introducing variability in the growth rate diminishes the Malthus parameter controlling the exponential growth of the whole population compared to the

reference value (defined as the Malthus parameter for a population where all cells grow at the mean growth rate  $\bar{v}$ ),  $\lambda_{B,\bar{v}} = \bar{v}$ . This decrease has the same magnitude as previously: about 2% for coefficients of variation around 15%-20%.

As regards Question **(ii)** and more precisely the estimation of the Malthus parameter observing one branch in the genealogical tree of the population of cells, the methodology described in Section 3.4 could be applied here. Since the sizes at birth  $(\xi_u, u \in \mathbb{L}_n)$  form a Markov chain (whereas the increments of size  $(\Delta_u, u \in \mathbb{L}_n)$  were independent), the controls will rely on the ergodicity of this Markov chain (which could be proved using the same kind of techniques as in Doumic *et al.* [DHKR15]).

## 4.2 MODEL INTRODUCED BY AMIR

The model was introduced without variability. Before addressing Question **(i)** we give some hint to better understand this model and in particular to set it with respect to the general size-increment-structured model we have studied. We follow the description of the model given in [Ami14], only adding variability in the growth rate, and then link it to the formalism we have used so far.

The mechanism described by [Ami14] is the following: the cell aims at adding a volume  $\Delta_0 > 0$  to its size at birth  $\xi_u$ . In order to divide precisely at size  $\xi_u + \Delta_0$ , the cell should grow during a time  $\zeta_u^*$  such that

$$\xi_u + \Delta_0 = \xi_u e^{\tau_u \zeta_u^*}$$

if individual  $u$  grows at rate  $\tau_u$  belonging to a state space  $\mathcal{V} = [v_{\min}, v_{\max}]$ . Actually a cell does not live precisely during the lifetime  $\zeta_u^*$  but it lives during a lifetime  $\zeta_u$  which is a noisy version of the target lifetime  $\zeta_u^*$ ,

$$\zeta_u = \zeta_u^* + \eta_u = \frac{1}{\tau_u} \ln \left( 1 + \frac{\Delta_0}{\xi_u} \right) + \eta_u \quad (42)$$

where  $(\eta_u, u \in \mathbb{T})$  is a noise sequence of independent and identically distributed random variables such that  $\mathbb{E}[\eta_u] = 0$  with support  $[-h, h]$  (*e.g.* truncated Gaussian) where  $h > 0$  satisfies  $\exp(v_{\max} h) \leq \sqrt{2}$ . For a given initial size  $\xi_\emptyset = x_\emptyset \leq x_{\max}$  such that

$$\frac{\Delta_0 e^{v_{\max} h}}{2 - e^{v_{\max} h}} \leq x_{\max} \leq \frac{\Delta_0}{e^{v_{\max} h} - 1},$$

the model is well-defined, see Section 7.3. Equation (42) together with (27) describing fragmentation and (26) giving the law of the growth rates completely define the model. The Malthus parameter depends here on the target increment of size  $\Delta_0$ , on the noise distribution  $f_\eta$  and on the distribution of the growth rates. As argued in [Ami14], this model captures well the main observed features as distributions of sizes at birth, sizes at division, ages at division, and correlations between these quantities.

It remains to fix the parameter  $\Delta_0$ . In practice, it should be thought of as the increment of size added by a cell to its initial size in the equilibrium regime,  $\Delta_0 \approx \mathbb{E}[\Delta_u]$ . Since  $\xi_u = (\xi_{u^-} + \Delta_{u^-})/2$  for any  $u \in \mathbb{T}$ , assuming equilibrium is reached so that the sizes at birth  $\xi_u$  and  $\xi_{u^-}$  follow the same law, we deduce  $\mathbb{E}[\Delta_u] = \mathbb{E}[\xi_u]$ . Thus the parameter  $\Delta_0$  satisfies  $\Delta_0 \approx \mathbb{E}[\xi_u]$  and should be fixed in practice as the mean size at birth of cells at equilibrium.

In order to recover the usual description through a partial differential equation as (2), assume that the noise sequence density can be written as

$$\mathbb{P}(\eta_u \in dz) = B(z+h)e^{-\int_0^{z+h} B(y)dy} dz$$

with  $B : [0, \infty) \rightarrow [0, \infty)$  such that  $B(z) = 0$  for  $z > 2h$ . Using (42) we deduce

$$\mathbb{P}(\zeta_u \in [t, t+dt) | \zeta_u \geq t, \xi_u = x_b, \tau_u = v) = B\left(t - \frac{1}{v} \ln\left(1 + \frac{\Delta_0}{x_b}\right) + h\right) dt.$$

Given the time elapsed since birth  $t$  and the size at birth  $x_b$  we can compute the current size  $x$  and the current increment of size  $a$  and conversely: we have

$$x_b = x - a$$

and

$$t = \frac{1}{v} \ln\left(\frac{x}{x-a}\right).$$

Thus

$$B\left(t - \frac{1}{v} \ln\left(1 + \frac{\Delta_0}{x_b}\right) + h\right) = B\left(\frac{1}{v} \ln\left(\frac{x}{x-a+\Delta_0}\right) + h\right) = \gamma(x, a, v).$$

The partial differential equation (2) with this choice of  $\gamma$  describes Amir's model in a deterministic way. Note that we could use a description through different couples of variables instead of size and increment of size, but we want to establish a comparison with the previous model.

The main difference with the size-increment-structured model is the following. The increments  $\Delta_u$  in the size-increment-structured model are independent of the growth rate or of the size at birth, remind (29), whereas it is no longer the case in Amir's model. At first, assume that there is no variability in the growth rate ( $\tau_u = \bar{v}$  for any  $u \in \mathbb{T}$ ). The two models are not equivalent. Heuristically,

$$B\left(t - \frac{1}{\bar{v}} \ln\left(1 + \frac{\Delta_0}{x_b}\right) + h\right) = B\left(\frac{1}{\bar{v}} \ln\left(1 + \frac{a}{x_b}\right) - \frac{1}{\bar{v}} \ln\left(1 + \frac{\Delta_0}{x_b}\right) + h\right) \approx \tilde{B}(a)$$

using the approximation  $x_b \approx \Delta_0$ . The choices  $\gamma(x, a) = \bar{v}xB(a)$  and  $\gamma(x, a) = B(a)$  in the model described by (1) are fundamentally different and lead to very different properties. We obtain the mutual independence of the increments of size in the first case but not in the second one. This huge difference remains adding variability.

This is illustrated on Figure 19 where we see the conditional density of the increments of size conditionally on the size at birth (left) and conditionally on the growth rate (right). Such curves in the size-increment-structured model would all be aligned. We omit the sample autocorrelation plot of the increments of size: we observe that the autocorrelations decrease very fast in both models.

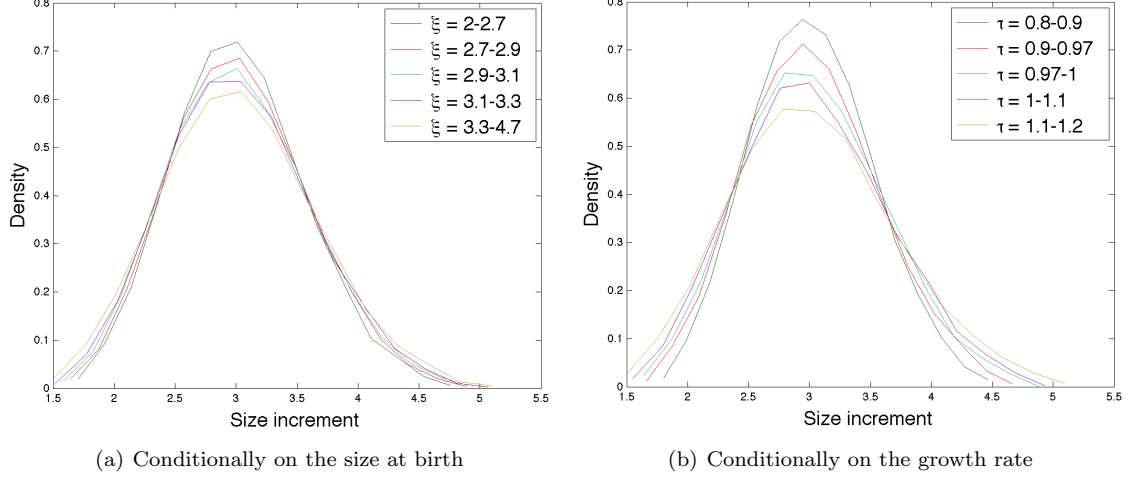


Figure 19: Amir's model [Ami14] with variability. Conditional distributions of the increment of size (estimation by histograms based on a simulated branch of  $n = 100\,000$  cells). Noise distribution: centered Gaussian with standard deviation 0.10 truncated on  $[-h, h]$  for  $h = 0.25$ . Target increment of size:  $\Delta_0 = 3$ . Growth rate distribution: Gaussian with mean  $\bar{v} = 1$  and standard deviation 0.70 truncated on  $[0.80, 1.20]$ .

To put it in a nutshell, still neglecting variability in the growth rate for the sake of presentation,

- 1) **Size-increment-structured model.** This model can be rewritten in the following equivalent way. We assume that the size at division of cell  $u$ ,

$$\xi_u + \Delta_u = \xi_u + \Delta_0 + \varepsilon_u,$$

where  $\Delta_0$  is interpreted as a target increment of size and  $\varepsilon_u$  is a noise. It boils down to  $\Delta_u = \Delta_0 + \varepsilon_u$ , with the  $\varepsilon_u$ 's independent and identically distributed, which enables us to recover the fundamental property of the size-increment-structured model (the  $\Delta_u$ 's independent and identically distributed).

- 2) **Amir's model.** Using (42) (with  $\tau_u = \tau$ , a common growth rate), we derive

$$\xi_u + \Delta_u = (\xi_u + \Delta_0)(1 + \varepsilon_u),$$

where  $\Delta_0$  is the target increment of size, and for some  $\varepsilon_u$ , a noise.

The size at division  $\xi_u + \Delta_u$  in both models is a noisy version of a target size at division  $\xi_u + \Delta_0$ . The noise is additive in the size-increment-structured model whereas it is multiplicative in Amir's model (since the noise is additive for the division time, see (42)). Here lies the fundamental difference, and that is why we derive very different properties. The question is then: can we distin-

guish experimentally the two models, which both seem to capture well the main observed features (distributions of sizes at birth, sizes at division, ages at division, and correlations between these quantities)?

However, as regards Question (i), the same conclusions as in the size-increment-structured model remain valid for Amir’s model. In Supplementary Table 18 in Section 8, we observe a reduction of the Malthus parameter compared to the value of reference (which has to be estimated too here) when introducing variability, and it decreases as there is more and more variability (*i.e.* as the coefficient of variation of the distribution of the growth rate increases).

As a closing remark, we stress once again the similarity between the three models we have studied (size-increment-structured model, size-structured model and Amir’s model, with variability). The main conclusion for these three models as regards the influence of individual variability on the Malthus parameter is the same. When neglecting heredity in the transmission of the growth rate, the Malthus parameter significantly collapse, with a reduction of about 2% when the coefficient of variation of the growth rates distribution is about 15%-20%. Such a variation is not negligible since it characterises the exponential growth of the whole population.

## 5 PERSPECTIVES

The scope of the perspectives is large and includes both theoretical, with analytic and statistical aspects, and experimental issues. At a theoretical level,

- 1) In order to prove existence and uniqueness of the eigenlements in the age-structured and size-increment-structured models with variability, what are minimal assumptions on the growth speed  $g$ , on the division rate  $\gamma$  and on the variability kernel  $\rho$  (based on the general results of Mischler and Scher [MS13] for instance)?

- 2) **Without heredity.** For the *age-structured* model,

- (a) Can we build more general classes of  $B$  (including experimentally more plausible  $B$ ) to discriminate between the two cases  $\lambda_{B,\rho} > \lambda_{B,\bar{v}}$  and  $\lambda_{B,\rho} < \lambda_{B,\bar{v}}$ ?

For the *size-structured* and *size-increment-structured* models with variability,

- (b) Are the two cases,  $\lambda_{B,\rho} > \lambda_{B,\bar{v}}$  and  $\lambda_{B,\rho} < \lambda_{B,\bar{v}}$ , possible? Is there some *plausible*  $B$  such that  $\lambda_{B,\rho} > \lambda_{B,\bar{v}}$ ? Can we build general classes of  $B$  to discriminate between the two cases (specifying a density  $\rho$  if needed, a truncated Gaussian for instance)?
- (c) Can we compute the perturbations of  $\lambda_{B,\rho_\alpha}$  around  $\lambda_{B,\bar{v}}$  for  $\rho_\alpha$  tending to  $\delta_{\bar{v}}$  as  $\alpha \rightarrow 0$  (using the same kind of tools as to prove Theorem 7, see also Michel [Mic06a])?

**3) With heredity.**

- (a) For the *age-structured* model with variability, how does heredity in the transmission of the aging rate influence the results of Theorems 6 and 7?
- (b) For the *size-increment-structured* model with variability, can we analytically shed some light on Figures 16 and 17?

**4) Estimation of the Malthus parameter.** From a statistical point of view, with experimental concerns,

- (a) From the observation of *the whole tree* until a given time, what is the best estimator we could build (based on empirical means), in terms of variance of instance?
- (b) From the observation of *one branch* until a given generation, what is the quality of the different procedures we propose in Section 3.4? What is the best procedure at statistical and computational levels?
- (c) In practice, given the experimental protocols described on page 177, is a procedure of estimation from one branch preferable to a procedure from the whole tree?
- (d) Beyond estimation, how to build confidence intervals and to conduct tests, statistically reliable, on the Malthus parameter?

At an experimental and biological level,

- 1) Based on statistical tests, are the mutual independence of the increments of size and their independence of the growth rates and sizes at birth valid? An answer to this question would enable us to discriminate between the size-increment-structured model and Amir's model.
- 2) Can we experimentally measure the variations of the Malthus parameter in presence of variability in the growth rate?
- 3) How can we explain such an influence of heredity on our answers to the initial question concerning population growth in presence of variability?

## 6 PROOFS

### 6.1 PROOF OF THEOREM 3

We recall that the aim of Theorem 3 is to match stochastic and deterministic approaches in the age-structured model with variability. For given  $\gamma$  and  $\rho$ , let us introduce the operator

$$\mathcal{A}_{\gamma,\rho}f(a, v) = g(a, v) \frac{\partial}{\partial a} f(a, v) + \gamma(a, v) \left( \int_{\mathcal{V}} f(0, v') \rho(v, dv') - f(a, v) \right)$$

densely defined on bounded continuous functions. Our proof relies on the use of a so-called many-to-one formula (see [BDMT11, Clo11]).

**Lemma 10** (Many-to-one formula). *For any nonnegative function  $f : \mathcal{S} \rightarrow [0, \infty)$ , for any probability measure  $\mu(da, dv)$  on  $\mathcal{S} = [0, \infty) \times \mathcal{V}$ ,*

$$\mathbb{E}_\mu \left[ \sum_{u \in \partial \mathcal{T}_t} f(\Delta_u^t, \tau_u) \right] = \mu(\phi_{\gamma, \rho}) e^{\lambda_{\gamma, \rho} t} \mathbb{E}_\mu \left[ ((\phi_{\gamma, \rho})^{-1} f)(\tilde{\chi}_t) \right]$$

where the Markov process  $(\tilde{\chi}_t)_{t \geq 0}$  has for generator  $\mathcal{A}_{\tilde{\gamma}, \tilde{\rho}}$  with

$$\tilde{\gamma}(a, v) = 2 \int_{\mathcal{V}} \phi_{\gamma, \rho}(0, v') \rho(v, dv') \phi_{\gamma, \rho}(a, v)^{-1} \gamma(a, v), \quad (43)$$

$$\tilde{\rho}(v, v') = \left( \int_{\mathcal{V}} \phi_{\gamma, \rho}(0, v') \rho(v, v') dv' \right)^{-1} \phi_{\gamma, \rho}(0, v) \rho(v, v'), \quad (44)$$

and  $\mathbb{P}_\mu(\tilde{\chi}_0 \in dvda) = \mu(da, dv)$ ,  $(\lambda_{\gamma, \rho}, \phi_{\gamma, \rho}) \in (0, \infty) \times \mathcal{C}_b^1(\mathbb{R}^+; \mathcal{C}_b(\mathcal{V}))$  solution to (11),  $\mu(\phi_{\gamma, \rho}) = \iint_{\mathcal{S}} \phi_{\gamma, \rho}(a, v) \mu(da, dv)$ .

For a proof in the case without variability (where all cells age at rate  $g \equiv 1$ ) we refer to Chapter I. The simple proof given there in Appendix can be generalised to the present case with variability. This result is also a particular case of Lemma 3.3 of [Clo11] (see Remark 3.4 of [Clo11] for a general initial measure).

All the proof of Theorem 3 long, we admit the existence of a solution  $(\lambda_{\gamma, \rho}, \phi_{\gamma, \rho}) \in (0, \infty) \times \mathcal{C}_b^1(\mathbb{R}^+; \mathcal{C}_b(\mathcal{V}))$  to (11), with  $\phi_{\gamma, \rho}$  continuous and positive. This result is proved in Section 6.2 (see Theorem 4).

*Proof of Theorem 3.* Let  $t > 0$  be fixed. By Lemma 10, we obtain the new representation

$$\iint_{\mathcal{S}} f(a, v) n(t, da, dv) = \mu(\phi_{\gamma, \rho}) e^{\lambda_{\gamma, \rho} t} \mathbb{E}_\mu \left[ ((\phi_{\gamma, \rho})^{-1} f)(\tilde{\chi}_t) \right]$$

with  $(\tilde{\chi}_t)_{t \geq 0} = ((\mathcal{Z}_t, \mathcal{V}_t))_{t \geq 0}$  a biased process of infinitesimal generator  $\mathcal{A}_{\tilde{\gamma}, \tilde{\rho}}$  with  $\tilde{\gamma}$  and  $\tilde{\rho}$  defined respectively by (43) and (44) and  $\mathbb{P}_\mu(\tilde{\chi}_0 \in dvda) = \mu(da, dv)$ . Let  $N_t = \sum_{s \geq t} \mathbf{1}_{\{\tilde{\chi}_s - \tilde{\chi}_{s-1} < 0\}}$  be the counting process associated to  $(\tilde{\chi}_t)_{t \geq 0}$ . For  $h > 0$ , introduce the difference operator

$$\Delta_h f(t) = h^{-1} (f(t+h) - f(t)).$$

We study the convergence of

$$\Delta_h \left( \iint_{\mathcal{S}} f(a, v) n(t, da, dv) \right) = \mu(\phi_{\gamma, \rho}) (T_0(h) + T_1(h) + T_{\geq 2}(h)) \quad (45)$$

as  $h \rightarrow 0$ , where

$$\begin{aligned} T_0(h) &= \mathbb{E}_\mu \left[ \Delta_h \left( e^{\lambda_{\gamma, \rho} t} ((\phi_{\gamma, \rho})^{-1} f)(\tilde{\chi}_t) \right) \mathbf{1}_{\{N_{t+h} - N_t = 0\}} \right], \\ T_1(h) &= \mathbb{E}_\mu \left[ \Delta_h \left( e^{\lambda_{\gamma, \rho} t} ((\phi_{\gamma, \rho})^{-1} f)(\tilde{\chi}_t) \right) \mathbf{1}_{\{N_{t+h} - N_t = 1\}} \right], \\ T_{\geq 2}(h) &= \mathbb{E}_\mu \left[ \Delta_h \left( e^{\lambda_{\gamma, \rho} t} ((\phi_{\gamma, \rho})^{-1} f)(\tilde{\chi}_t) \right) \mathbf{1}_{\{N_{t+h} - N_t \geq 2\}} \right]. \end{aligned}$$

We successively study the convergence of  $T_0(h)$ ,  $T_1(h)$  and  $T_{\geq 2}(h)$  as  $h \rightarrow 0$  in the three following steps. Along this proof, the notation  $\lesssim$  means up to a constant independent of  $h$  (but that can depend on  $t$ ). We give as a preliminary the following lemma, proved in Section 7.2, useful at different stages of the proof.

**Lemma 11.** *Denote by  $\mathcal{F}_t$  the filtration generated by  $(\tilde{\chi}_s, s \leq t)$ . We have*

- (i)  $\mathbb{P}_\mu(N_{t+h} - N_t = 1 | \mathcal{F}_t) = \tilde{\gamma}(\tilde{\chi}_t)h + h\varepsilon(h)$  with  $|\varepsilon(h)| \leq \bar{\varepsilon}(h)$  for some deterministic  $\bar{\varepsilon}(h) \rightarrow 0$  as  $h \rightarrow 0$ ,
- (ii)  $\mathbb{P}_\mu(N_{t+h} - N_t \geq 2 | \mathcal{F}_t) \lesssim h^2$ .

*Step 1.* To study  $T_0(h)$ , we use the decomposition

$$T_0(h) = \Delta_h(e^{\lambda_{\gamma,\rho}t})\mathbb{E}_\mu\left[\left((\phi_{\gamma,\rho})^{-1}f\right)(\tilde{\chi}_t)\mathbf{1}_{\{N_{t+h}-N_t=0\}}\right] + e^{\lambda_{\gamma,\rho}(t+h)}\mathbb{E}_\mu\left[\Delta_h\left(\left((\phi_{\gamma,\rho})^{-1}f\right)(\tilde{\chi}_t)\right)\mathbf{1}_{\{N_{t+h}-N_t=0\}}\right].$$

Since  $f$  belongs to  $\mathcal{C}_0^1(\mathcal{S})$  and since  $\phi_{\gamma,\rho}$  is continuous and positive (see Remark 13 in the proof of Theorem 4 which follows), we have

$$\left|((\phi_{\gamma,\rho})^{-1}f)(\tilde{\chi}_t)\right| \leq \left(\inf_{\substack{(a,v) \in \\ \text{Supp}(f)}} \phi_{\gamma,\rho}(a,v)\right)^{-1} \sup_{(a,v) \in \mathcal{S}} |f(a,v)|. \quad (46)$$

Then, by dominated convergence, using  $\mathbb{P}_\mu(N_{t+h} - N_t = 0) \rightarrow 1$  as  $h \rightarrow 0$ ,

$$\Delta_h(e^{\lambda_{\gamma,\rho}t})\mathbb{E}_\mu\left[\left((\phi_{\gamma,\rho})^{-1}f\right)(\tilde{\chi}_t)\mathbf{1}_{\{N_{t+h}-N_t=0\}}\right] \rightarrow \lambda_{\gamma,\rho}e^{\lambda_{\gamma,\rho}t}\mathbb{E}_\mu\left[\left((\phi_{\gamma,\rho})^{-1}f\right)(\tilde{\chi}_t)\right] \quad (47)$$

as  $h \rightarrow 0$ .

For the sake of presentation of this step, let us assume  $g(a,v) = v$ . On the event  $\{N_{t+h} - N_t = 0\}$ , the process  $\mathcal{V}_s$  is constant for  $s \in [t, t+h)$ , and  $\mathcal{Z}_{t+h} = \mathcal{Z}_t + \mathcal{V}_t h$  (linear aging at rate  $v$ ). Thus

$$\Delta_h\left(\left((\phi_{\gamma,\rho})^{-1}f\right)(\tilde{\chi}_t)\right) = h^{-1}\left(\left((\phi_{\gamma,\rho})^{-1}f\right)(\mathcal{Z}_t + \mathcal{V}_t h, \mathcal{V}_t) - \left((\phi_{\gamma,\rho})^{-1}f\right)(\mathcal{Z}_t, \mathcal{V}_t)\right).$$

Since

$$\left|\Delta_h\left(\left((\phi_{\gamma,\rho})^{-1}f\right)(\tilde{\chi}_t)\right)\right| \leq v_{\max} \sup_{\substack{(a,v) \in \\ \text{Supp}(f)}} \left|v \frac{\partial}{\partial a} \left(\left(\phi_{\gamma,\rho}\right)^{-1}f\right)(a,v)\right|,$$

by dominated convergence again, we obtain

$$e^{\lambda_{\gamma,\rho}(t+h)}\mathbb{E}_\mu\left[\Delta_h\left(\left((\phi_{\gamma,\rho})^{-1}f\right)(\tilde{\chi}_t)\right)\mathbf{1}_{\{N_{t+h}-N_t=0\}}\right] \rightarrow e^{\lambda_{\gamma,\rho}t}\mathbb{E}_\mu\left[\mathcal{V}_t \frac{\partial}{\partial a} \left(\left(\phi_{\gamma,\rho}\right)^{-1}f\right)(\tilde{\chi}_t)\right] \quad (48)$$

as  $h \rightarrow 0$ .

Gathering (47) and (48) gives the limit of  $T_0(h)$  as  $h \rightarrow 0$ ,

$$T_0(h) \rightarrow \lambda_{\gamma,\rho}e^{\lambda_{\gamma,\rho}t}\mathbb{E}_\mu\left[\left((\phi_{\gamma,\rho})^{-1}f\right)(\tilde{\chi}_t)\right] + e^{\lambda_{\gamma,\rho}t}\mathbb{E}_\mu\left[\mathcal{V}_t \frac{\partial}{\partial a} \left(\left(\phi_{\gamma,\rho}\right)^{-1}f\right)(\tilde{\chi}_t)\right].$$



Using Lemma 10 and the definition of  $n(t, \cdot)$ , we finally get

$$\mu(\phi_{\gamma, \rho})T_0(h) \rightarrow \iint_{\mathcal{S}} \left( \lambda_{\gamma, \rho} f(a, v) + g(a, v) \phi_{\gamma, \rho}(a, v) \frac{\partial}{\partial a} ((\phi_{\gamma, \rho})^{-1} f)(a, v) \right) n(t, da, dv) \quad (49)$$

as  $h \rightarrow 0$ , the extension to a general aging speed  $g$  being straightforward.

*Step 2.* We now turn to  $T_1(h)$  and use the decomposition

$$\begin{aligned} T_1(h) &= \Delta_h(e^{\lambda_{\gamma, \rho} t}) \mathbb{E}_\mu \left[ ((\phi_{\gamma, \rho})^{-1} f)(\tilde{\chi}_t) \mathbf{1}_{\{N_{t+h} - N_t = 1\}} \right] \\ &\quad + e^{\lambda_{\gamma, \rho}(t+h)} h^{-1} \mathbb{E}_\mu \left[ ((\phi_{\gamma, \rho})^{-1} f)(\tilde{\chi}_{t+h}) \mathbf{1}_{\{N_{t+h} - N_t = 1\}} \right] \\ &\quad - e^{\lambda_{\gamma, \rho}(t+h)} h^{-1} \mathbb{E}_\mu \left[ ((\phi_{\gamma, \rho})^{-1} f)(\tilde{\chi}_t) \mathbf{1}_{\{N_{t+h} - N_t = 1\}} \right]. \end{aligned}$$

Since  $\mathbb{P}_\mu(N_{t+h} - N_t = 1) \rightarrow 0$  as  $h \rightarrow 0$ , recalling (46), by dominated convergence, the first term of the decomposition converges to zero. As regards the last term, conditioning with respect to  $\mathcal{F}_t$  and using Lemma 11 (i),

$$\begin{aligned} -h^{-1} \mathbb{E}_\mu \left[ ((\phi_{\gamma, \rho})^{-1} f)(\tilde{\chi}_t) \mathbf{1}_{\{N_{t+h} - N_t = 1\}} \right] &= -\mathbb{E}_\mu \left[ ((\phi_{\gamma, \rho})^{-1} f)(\tilde{\chi}_t) (\tilde{\gamma}(\tilde{\chi}_t) + \varepsilon(h)) \right] \\ &\rightarrow -\mathbb{E}_\mu \left[ ((\phi_{\gamma, \rho})^{-1} f)(\tilde{\chi}_t) \tilde{\gamma}(\tilde{\chi}_t) \right] \quad (50) \end{aligned}$$

as  $h \rightarrow 0$  since  $|\varepsilon(h)| \leq \epsilon(h)$  for some deterministic  $\epsilon(h) \rightarrow 0$  as  $h \rightarrow 0$ , using (46) once again. Introduce the  $(\mathcal{F}_t)_{t \geq 0}$  stopping time

$$\Upsilon_t = \inf\{s > t, N_s - N_t \geq 1\}, \quad (51)$$

being the time of the first jump of the process  $\tilde{\chi}$  after  $t$ . On the event  $\{N_{t+h} - N_t = 1\}$ , we have  $t \leq \Upsilon_t < t + h$  and  $\mathcal{V}_{t+h} = \mathcal{V}_{\Upsilon_t}$ ,  $\mathcal{Z}_{t+h} = \mathcal{V}_{\Upsilon_t}(t + h - \Upsilon_t) = \mathcal{V}_{\Upsilon_t} h + \mathcal{V}_{\Upsilon_t}(t - \Upsilon_t)$ . Then, on the event  $\{N_{t+h} - N_t = 1\}$ ,

$$((\phi_{\gamma, \rho})^{-1} f)(\mathcal{Z}_{t+h}, \mathcal{V}_{t+h}) = ((\phi_{\gamma, \rho})^{-1} f)(\mathcal{V}_{\Upsilon_t} h, \mathcal{V}_{\Upsilon_t}) + h\varepsilon_1(h)$$

with  $|\varepsilon_1(h)| \leq \bar{\varepsilon}_1(h)$  for some deterministic  $\bar{\varepsilon}_1(h)$  such that  $\bar{\varepsilon}_1(h) = O(1)$ . It follows

$$\begin{aligned} h^{-1} \mathbb{E}_\mu \left[ ((\phi_{\gamma, \rho})^{-1} f)(\tilde{\chi}_{t+h}) \mathbf{1}_{\{N_{t+h} - N_t = 1\}} \right] &= h^{-1} \mathbb{E}_\mu \left[ ((\phi_{\gamma, \rho})^{-1} f)(\mathcal{V}_{\Upsilon_t} h, \mathcal{V}_{\Upsilon_t}) \mathbf{1}_{\{N_{t+h} - N_t = 1\}} \right] + \bar{\varepsilon}_2(h) \\ &= h^{-1} \mathbb{E}_\mu \left[ ((\phi_{\gamma, \rho})^{-1} f)(\mathcal{V}_{\Upsilon_t} h, \mathcal{V}_{\Upsilon_t}) \mathbf{1}_{\{N_{t+h} - N_t \geq 1\}} \right] + \bar{\varepsilon}_3(h) \end{aligned}$$

for some deterministic  $\bar{\varepsilon}_2(h), \bar{\varepsilon}_3(h) \rightarrow 0$  as  $h \rightarrow 0$ . We condition with respect to  $\mathcal{F}_t \vee \mathcal{V}_{\Upsilon_t}$  and use that  $\{N_{t+h} - N_t \geq 1\}$  and  $\mathcal{V}_{\Upsilon_t}$  are independent together with Lemma 11 (i) again,

$$h^{-1} \mathbb{E}_\mu \left[ ((\phi_{\gamma, \rho})^{-1} f)(\tilde{\chi}_{t+h}) \mathbf{1}_{\{N_{t+h} - N_t = 1\}} \right] = \mathbb{E}_\mu \left[ ((\phi_{\gamma, \rho})^{-1} f)(\mathcal{V}_{\Upsilon_t} h, \mathcal{V}_{\Upsilon_t}) \tilde{\gamma}(\tilde{\chi}_t) \right] + \bar{\varepsilon}_4(h)$$

for some deterministic  $\bar{\varepsilon}_4(h) \rightarrow 0$  as  $h \rightarrow 0$ . Thus, as  $h \rightarrow 0$ ,

$$\begin{aligned} h^{-1} \mathbb{E}_\mu \left[ ((\phi_{\gamma,\rho})^{-1} f)(\tilde{\chi}_{t+h}) \mathbf{1}_{\{N_{t+h} - N_t = 1\}} \right] \\ \rightarrow \mathbb{E}_\mu \left[ ((\phi_{\gamma,\rho})^{-1} f)(0, \mathcal{V}_{\Upsilon_t}) \tilde{\gamma}(\tilde{\chi}_t) \right] = \mathbb{E}_\mu \left[ \int_{\mathcal{V}} ((\phi_{\gamma,\rho})^{-1} f)(0, v') \tilde{\rho}(\mathcal{V}_t, v') dv' \tilde{\gamma}(\tilde{\chi}_t) \right] \end{aligned} \quad (52)$$

conditioning with respect to  $\mathcal{F}_t$  using that  $\mathcal{V}_{\Upsilon_t}$  is drawn according to the conditional density  $\tilde{\rho}(\mathcal{V}_t, v')$  to rewrite the limit.

Gathering (50) and (52),

$$T_1(h) \rightarrow e^{\lambda_{\gamma,\rho} t} \mathbb{E}_\mu \left[ \int_{\mathcal{V}} ((\phi_{\gamma,\rho})^{-1} f)(0, v') \tilde{\rho}(\mathcal{V}_t, v') dv' \tilde{\gamma}(\tilde{\chi}_t) - ((\phi_{\gamma,\rho})^{-1} f)(\tilde{\chi}_t) \tilde{\gamma}(\tilde{\chi}_t) \right]$$

as  $h \rightarrow 0$ . Using Lemma 10 and the definition of  $n(t, \cdot)$ , we get

$$\begin{aligned} \mu(\phi_{\gamma,\rho}) T_1(h) \rightarrow \iint_{\mathcal{S}} \left( \int_{\mathcal{V}} ((\phi_{\gamma,\rho})^{-1} f)(0, v') \tilde{\rho}(v, v') dv' \tilde{\gamma}(a, v) \phi_{\gamma,\rho}(a, v) \right. \\ \left. - f(a, v) \tilde{\gamma}(a, v) \right) n(t, da, dv). \end{aligned}$$

Recalling the definition (43) of the biased division rate  $\tilde{\gamma}$  and the definition (44) of the biased aging rate Markov kernel  $\tilde{\rho}$ , we conclude

$$\begin{aligned} \mu(\phi_{\gamma,\rho}) T_1(h) \rightarrow \iint_{\mathcal{S}} \left( 2 \int_{\mathcal{V}} f(0, v') \rho(v, dv') \gamma(a, v) \right. \\ \left. - 2f(a, v) \int_{\mathcal{V}} \phi_{\gamma,\rho}(0, v') \rho(v, v') dv' \phi_{\gamma,\rho}(a, v)^{-1} \gamma(a, v) \right) n(t, da, dv). \end{aligned} \quad (53)$$

*Step 3.* As regards  $T_{\geq 2}(h)$ , recalling (46) and by Lemma 11 (ii), we obtain

$$T_{\geq 2}(h) \lesssim h^{-1} \mathbb{P}_\mu(N_{t+h} - N_t \geq 2) \lesssim h.$$

Consequently,  $T_{\geq 2}(h)$  converges to zero as  $h \rightarrow 0$ .

Recalling the decomposition (45), gathering (49) and (53), together with Step 3, and using that  $(\lambda_{\gamma,\rho}, \phi_{\gamma,\rho})$  is solution to (11), we obtain

$$\int_{\mathcal{S}} f(a, v) \frac{\partial}{\partial t} n(t, da, dv) = \int_{\mathcal{S}} \mathcal{A}_{\gamma,\rho} f(a, v) n(t, da, dv), \quad (54)$$

which is the weak formulation of (5). □

## 6.2 PROOF OF THEOREM 4

Theorem 4 concerns the eigenproblem of the age-structured model with variability. As a preliminary, note that neglecting heredity in the transmission of the aging rate enables us to obtain explicit expressions for the eigenvectors  $N_{\gamma,\rho}$  and  $\phi_{\gamma,\rho}$ , and an implicit relation uniquely defining the Malthus parameter  $\lambda_{\gamma,\rho}$ .

**Lemma 12.** *Work under Assumption 1. Assume in addition that*

$$\rho(v, dv') = \rho(v')dv'.$$

*Then the eigenvalue  $\lambda_{\gamma, \rho} > 0$  is uniquely defined by*

$$2 \iint_{\mathcal{S}} \frac{\gamma(a, v)}{g(a, v)} \exp\left(-\int_0^a \frac{\lambda_{\gamma, \rho} + \gamma(s, v)}{g(s, v)} ds\right) \rho(v) dv da = 1.$$

*The unique solution  $N_{\gamma, \rho}$  such that  $(gN_{\gamma, \rho}) \in \mathcal{C}_b^1(\mathbb{R}^+; \mathcal{C}_b(\mathcal{V}))$  to (10) and the unique solution  $\phi_{\gamma, \rho} \in \mathcal{C}_b^1(\mathbb{R}^+; \mathcal{C}_b(\mathcal{V}))$  to (11) are respectively given by, for any  $(a, v) \in \mathcal{S}$ ,*

$$N_{\gamma, \rho}(a, v) = \frac{\kappa \rho(v)}{g(a, v)} \exp\left(-\int_0^a \frac{\lambda_{\gamma, \rho} + \gamma(s, v)}{g(s, v)} ds\right), \quad \phi_{\gamma, \rho}(a, v) = \frac{\kappa' \int_a^\infty \gamma(s, v) N_{\gamma, \rho}(s, v) ds}{g(a, v) N_{\gamma, \rho}(a, v)}$$

*with  $\kappa, \kappa'$  normalizing constants such that  $\iint_{\mathcal{S}} N_{\gamma, \rho} = 1$  and  $\iint_{\mathcal{S}} N_{\gamma, \rho} \phi_{\gamma, \rho} = 1$ .*

For a proof, one can easily check that  $N_{\gamma, \rho}$  and  $\phi_{\gamma, \rho}$  defined in such a way satisfy respectively (10) and (11). We now work in the general case where  $\rho$  is a Markov kernel.

*Proof of Theorem 4.* We first study the direct eigenproblem (10), then we turn to the adjoint eigenproblem (11). At last we prove uniqueness of the eigenelements. The methodology we use is inspired by the one of [Dou07]. In this proof we denote by  $\mathcal{C}_b(\mathcal{V}) = \mathcal{X}$  the Banach space of bounded continuous functions  $f : \mathcal{V} \rightarrow \mathbb{R}$ , equipped with the supremum norm,  $\|f\|_{\mathcal{X}} = \sup_{v \in \mathcal{V}} |f(v)|$ , for  $\mathcal{V}$  compact set of  $(0, \infty)$ .

*Direct eigenproblem.* We split the proof into four steps.

*Step 1.* Since  $N_{\gamma, \rho}$  satisfies (10), we immediately deduce that, for any  $(a, v) \in \mathcal{S}$ ,

$$(gN_{\gamma, \rho})(a, v) = (gN_{\gamma, \rho})(0, v) \exp\left(-\int_0^a \frac{\lambda_{\gamma, \rho} + \gamma(s, v)}{g(s, v)} ds\right). \quad (55)$$

Using the boundary condition leads us to

$$(gN_{\gamma, \rho})(0, v') = 2 \iint_{\mathcal{S}} (gN_{\gamma, \rho})(0, v) e^{-\int_0^a \frac{\lambda_{\gamma, \rho}}{g(s, v)} ds} f_{\gamma/g}(a, v) \rho(v, v') dv da \quad (56)$$

for  $v' \in \mathcal{V}$ , setting

$$f_{\gamma/g}(a, v) = \frac{\gamma(a, v)}{g(a, v)} \exp\left(-\int_0^a \frac{\gamma(s, v)}{g(s, v)} ds\right), \quad (a, v) \in \mathcal{S}. \quad (57)$$

Note that, for every  $v \in \mathcal{V}$ ,  $a \rightsquigarrow f_{\gamma/g}(a, v)$  is a density. Equation (56) leads us to define an operator  $\mathcal{G}_\lambda : \mathcal{X} \rightarrow \mathcal{X}$  by

$$(\mathcal{G}_\lambda f)(v') = 2 \iint_{\mathcal{S}} f(v) e^{-\int_0^a \frac{\lambda}{g(s, v)} ds} f_{\gamma/g}(a, v) \rho(v, v') dv da, \quad v' \in \mathcal{V},$$

for any  $\lambda \geq 0$ . In the next steps, we look for a solution  $(\lambda, f)$  to the equation  $\mathcal{G}_\lambda(f) = f$ . We work on the space  $\mathcal{X}$  of continuous functions since we want to apply the Krein-Rutman theorem

[KT48, DL90] (then the interior of the positive cone is the set of positive functions).

*Step 2.* We introduce a so-called regularized operator, which is strictly positive. For a fixed  $\varepsilon > 0$ , set

$$\rho_\varepsilon(v, v') = \rho(v, v') + |\mathcal{V}|^{-1}\varepsilon, \quad (v, v') \in \mathcal{V}^2, \quad (58)$$

where  $|\mathcal{V}|$  stands for the Lebesgue measure of the compact set  $\mathcal{V} \subset (0, \infty)$  and define the operator  $\mathcal{G}_{\lambda, \varepsilon} : \mathcal{X} \rightarrow \mathcal{X}$  by

$$(\mathcal{G}_{\lambda, \varepsilon} f)(v') = 2 \iint_{\mathcal{S}} f(v) \exp\left(-\int_0^a \frac{\lambda}{g(s, v)} ds\right) f_{\gamma/g}(a, v) \rho_\varepsilon(v, v') dv da, \quad v' \in \mathcal{V},$$

for any  $\lambda \geq 0$ . We claim that  $\mathcal{G}_{\lambda, \varepsilon}$  is **1**) strictly positive on  $\mathcal{X}$  (*i.e.* for any  $f \in \mathcal{X}$  nonnegative and different from the null function,  $(\mathcal{G}_{\lambda, \varepsilon} f)(v') > 0$  for any  $v' \in \mathcal{V}$ ), **2**) a linear mapping from  $\mathcal{X}$  into itself, **3**) continuous and **4**) compact. Thus we are now in position to apply the Krein-Rutman theorem (we use Theorem 6.5 of [Per07]). For any  $\lambda \geq 0$  there exist a unique  $\mu_{\lambda, \varepsilon} > 0$  and a unique positive  $U_{\lambda, \varepsilon} \in \mathcal{X}$  such that

$$\mathcal{G}_{\lambda, \varepsilon}(U_{\lambda, \varepsilon}) = \mu_{\lambda, \varepsilon} U_{\lambda, \varepsilon} \quad (59)$$

and  $\|U_{\lambda, \varepsilon}\|_{\mathcal{X}} = 1$ .

It just remains to prove the four claimed properties. **1**) is precisely achieved thanks to the regularisation  $\rho_\varepsilon$  of  $\rho$  by (58). **2**) The linearity is obvious and for  $f \in \mathcal{X}$ , we have  $\mathcal{G}_{\lambda, \varepsilon}(f) \in \mathcal{X}$  since  $v' \rightsquigarrow \rho_\varepsilon(v, v')$  is continuous and bounded for any  $v \in \mathcal{V}$ . **3**) We even achieve Lipschitz continuity, for any  $(f, g) \in \mathcal{X}^2$ ,

$$\|\mathcal{G}_{\lambda, \varepsilon}(f) - \mathcal{G}_{\lambda, \varepsilon}(g)\|_{\mathcal{X}} \leq 2(|\mathcal{V}|\rho|_{\infty} + \varepsilon)\|f - g\|_{\mathcal{X}}$$

where  $|\rho|_{\infty} = \sup_{(v, v') \in \mathcal{V}^2} \rho(v, v')$ . **4**) We prove that for any  $\lambda \geq 0$  the family  $(\mathcal{G}_{\lambda, \varepsilon}(f), f \in \mathcal{X})$  is equicontinuous. Indeed  $(\mathcal{G}_{\lambda, \varepsilon} f)(v'_1) - (\mathcal{G}_{\lambda, \varepsilon} f)(v'_2)$  is arbitrarily small when  $|v'_1 - v'_2|$  is small enough, uniformly in  $f \in \mathcal{X}$  such that  $\|f\|_{\mathcal{X}} \leq 1$ , since  $v' \rightsquigarrow \rho_\varepsilon(v, v')$  is continuous for any  $v \in \mathcal{V}$ . Therefore by the Ascoli-Arzelà theorem for any  $\lambda \geq 0$  the family  $(\mathcal{G}_{\lambda, \varepsilon}(f), f \in \mathcal{X})$  is compact in  $\mathcal{X}$ .

*Step 3.* We now study the mapping  $\lambda \rightsquigarrow \mu_{\lambda, \varepsilon}$ . Our aim is to prove that there exists a unique  $\lambda_\varepsilon > 0$  such that

$$\mu_{\lambda_\varepsilon, \varepsilon} = 1. \quad (60)$$

To prove so we successively verify that **1**) as  $\lambda$  increases,  $\mu_{\lambda, \varepsilon}$  decreases (strictly), **2**) the mapping  $\lambda \rightsquigarrow \mu_{\lambda, \varepsilon}$  is continuous, **3**) for  $\lambda = 0$ ,  $\mu_{\lambda=0, \varepsilon} = 2(1 + \varepsilon) > 1$ , **4**) as  $\lambda \rightarrow \infty$ ,  $\mu_{\lambda, \varepsilon}$  converges to zero.

To prove **1**), since  $\mu_{\lambda, \varepsilon}$  is the spectral radius of  $\mathcal{G}_{\lambda, \varepsilon}$ , it holds

$$\mu_{\lambda, \varepsilon} = \sup_{f \in \mathcal{X}, \|f\|_{\mathcal{X}}=1} \|\mathcal{G}_{\lambda, \varepsilon} f\|_{\mathcal{X}}. \quad (61)$$

Note that  $\mathcal{G}_{\lambda, \varepsilon}$  itself decreases as  $\lambda$  increases: if  $\lambda_2 > \lambda_1$  then  $\mathcal{G}_{\lambda_2, \varepsilon}(f) < \mathcal{G}_{\lambda_1, \varepsilon}(f)$  for any  $f \in \mathcal{X}$  and we deduce  $\mu_{\lambda_1, \varepsilon} \leq \mu_{\lambda_2, \varepsilon}$  using (61). **2**) Let  $(\lambda_k, k \geq 0)$  be a sequence such that  $\lambda_k \geq 0$  which converges to  $\bar{\lambda} \geq 0$  as  $k \rightarrow \infty$ . Then  $\|\mathcal{G}_{\lambda_k, \varepsilon}(f)\|_{\mathcal{X}}$  converges to  $\|\mathcal{G}_{\bar{\lambda}, \varepsilon}(f)\|_{\mathcal{X}}$  uniformly in  $f \in \mathcal{X}$ ,  $\|f\|_{\mathcal{X}} = 1$ , as  $k \rightarrow \infty$ . By representation (61), we deduce that  $\mu_{\lambda_k, \varepsilon}$  converges to  $\mu_{\bar{\lambda}, \varepsilon}$ . To prove

**3)**, we successively compute,

$$(\mathcal{G}_{\lambda=0,\varepsilon}f)(v') = 2 \iint_{\mathcal{S}} f(v) f_{\gamma/g}(a, v) \rho_{\varepsilon}(v, v') dv da = 2 \int_{\mathcal{V}} f(v) \rho_{\varepsilon}(v, v') dv$$

since  $\int_0^{\infty} f_{\gamma/g}(a, v) da = 1$  for any  $v \in \mathcal{V}$ , and

$$\int_{\mathcal{V}} (\mathcal{G}_{\lambda=0,\varepsilon}f)(v') dv' = 2(1 + \varepsilon) \int_{\mathcal{V}} f(v) dv$$

since  $\int_{\mathcal{V}} \rho_{\varepsilon}(v, v') dv' = 1 + \varepsilon$  for any  $v \in \mathcal{V}$ . Thus, by (59) and choosing  $f = U_{\lambda,\varepsilon}$  in the previous calculus,

$$\int_{\mathcal{V}} (\mathcal{G}_{\lambda=0,\varepsilon}U_{\lambda=0,\varepsilon})(v') dv' = \mu_{\lambda=0,\varepsilon} \int_{\mathcal{V}} U_{\lambda=0,\varepsilon}(v') dv' = 2(1 + \varepsilon) \int_{\mathcal{V}} U_{\lambda=0,\varepsilon}(v) dv.$$

Since  $\int_{\mathcal{V}} U_{\lambda=0,\varepsilon}(v) dv < \infty$  ( $U_{\lambda=0,\varepsilon}$  being bounded and  $\mathcal{V}$  being a compact set), we deduce that  $\mu_{\lambda=0} = 2(1 + \varepsilon)$ . To prove **4)**,

$$\begin{aligned} \int_{\mathcal{V}} (\mathcal{G}_{\lambda,\varepsilon}f)(v') dv' &= 2(1 + \varepsilon) \iint_{\mathcal{S}} f(v) \exp\left(-\int_0^a \frac{\lambda}{g(s, v)} ds\right) f_{\gamma/g}(a, v) dv da \\ &\leq 2(1 + \varepsilon) \frac{|g|_{\infty} |f_{\gamma/g}|_{\infty}}{\lambda} \int_{\mathcal{V}} f(v) dv \end{aligned}$$

using again  $\int_{\mathcal{V}} \rho_{\varepsilon}(v, v') dv' = 1 + \varepsilon$  and Assumption **1** for the upper bound. Then, as previously, by (59) and taking  $f = U_{\lambda,\varepsilon}$  in the previous calculus, we check that

$$\mu_{\lambda,\varepsilon} \leq 2(1 + \varepsilon) \frac{|g|_{\infty} |f_{\gamma/g}|_{\infty}}{\lambda}, \quad (62)$$

which implies that  $\mu_{\lambda,\varepsilon} \rightarrow 0$  as  $\lambda \rightarrow \infty$ .

*Step 4.* In this last step, the aim is to let  $\varepsilon$  tend to zero. Let  $\lambda_{\varepsilon}$  be uniquely defined by (60) and denote by  $U_{\lambda_{\varepsilon},\varepsilon} = U_{\varepsilon} \in \mathcal{X}$  the associated positive eigenvector such that  $\|U_{\varepsilon}\|_{\mathcal{X}} = 1$ . On the one hand, the family,  $(U_{\varepsilon}, 0 < \varepsilon < 1)$  is compact in  $\mathcal{X}$  (recall that  $U_{\varepsilon} = \mathcal{G}_{\lambda_{\varepsilon},\varepsilon}U_{\varepsilon}$  and use again the Ascoli-Arzelà theorem). On the other hand, the family  $(\lambda_{\varepsilon}, 0 < \varepsilon < 1)$  is bounded, recalling (60) and (62). Thus we can extract a subsequence, still denoted by  $(\lambda_{\varepsilon}, U_{\varepsilon})$ , converging to  $(\bar{\lambda}, \bar{U})$  in  $\mathbb{R} \times \mathcal{X}$  with  $\bar{\lambda} \geq 0$  and  $\bar{U} \in \mathcal{X}$  positive such that  $\|\bar{U}\|_{\mathcal{X}} = 1$ . Since

$$U_{\varepsilon}(v) = 2 \iint_{\mathcal{S}} U_{\varepsilon}(v) \exp\left(-\int_0^a \frac{\lambda_{\varepsilon}}{g(s, v)} ds\right) f_{\gamma/g}(a, v) (\rho(v, v') + |\mathcal{V}|^{-1} \varepsilon) dv da,$$

letting  $\varepsilon \rightarrow 0$ , we obtain  $\bar{U} = \mathcal{G}_{\bar{\lambda}}\bar{U}$ , which means that we have found a solution to (56). Now set  $\lambda_{\gamma,\rho} = \bar{\lambda}$  and for any  $(a, v) \in \mathcal{S}$ ,

$$N_{\gamma,\rho}(a, v) = \frac{\kappa \bar{U}(v)}{g(a, v)} \exp\left(-\int_0^a \frac{\bar{\lambda} + \gamma(s, v)}{g(s, v)} ds\right),$$

reminding (55) and  $U_{\gamma,\rho} = gN_{\gamma,\rho}$ , with  $\kappa > 0$  chosen such that  $\iint_{\mathcal{S}} N_{\gamma,\rho} = 1$  (which is possible

since  $\iint_{\mathcal{S}} N_{\gamma,\rho} < \infty$ . □

*Adjoint eigenproblem.* The proof follows the same steps as in the direct eigenproblem.

*Step 1.* Since  $\phi_{\gamma,\rho}$  satisfies (11), one easily checks that

$$\frac{\partial}{\partial a} \left( \phi_{\gamma,\rho}(a, v) e^{-\int_0^a \frac{\lambda_{\gamma,\rho} + \gamma(s,v)}{g(s,v)} ds} \right) = -2f_{\gamma/g}(a, v) e^{-\int_0^a \frac{\lambda_{\gamma,\rho}}{g(s,v)} ds} \int_{\mathcal{V}} \phi_{\gamma,\rho}(0, v') \rho(v, v') dv'$$

with  $f_{\gamma/g}$  defined by (57). Integrating in  $a$  between zero and infinity, we deduce

$$\phi_{\gamma,\rho}(0, v) = 2 \int_0^\infty f_{\gamma/g}(a, v) e^{-\int_0^a \frac{\lambda_{\gamma,\rho}}{g(s,v)} ds} da \int_{\mathcal{V}} \phi_{\gamma,\rho}(0, v') \rho(v, v') dv' \quad (63)$$

and integrating between zero and  $a$ , we deduce

$$\phi_{\gamma,\rho}(a, v) = 2e^{\int_0^a \frac{\lambda_{\gamma,\rho} + \gamma(s,v)}{g(s,v)} ds} \int_a^\infty f_{\gamma/g}(s, v) e^{-\int_0^s \frac{\lambda_{\gamma,\rho}}{g(t,v)} dt} ds \int_{\mathcal{V}} \phi_{\gamma,\rho}(0, v') \rho(v, v') dv'. \quad (64)$$

**Remark 13.** By a *reductio ad absurdum* argument, using (63) and the Markov kernel properties, we prove that for any  $v \in \mathcal{V}$ ,  $\phi_{\gamma,\rho}(0, v) > 0$ . Then, using (64), we deduce that  $\phi_{\gamma,\rho}(a, v) > 0$  for any  $(a, v) \in \mathcal{S}$ .

Equation (63) leads us to define an operator  $\mathcal{G}_\lambda^* : \mathcal{X} \rightarrow \mathcal{X}$  by

$$(\mathcal{G}_\lambda^* f)(v) = 2 \int_0^\infty f_{\gamma/g}(a, v) e^{-\int_0^a \frac{\lambda}{g(s,v)} ds} da \int_{\mathcal{V}} f(v') \rho(v, v') dv', \quad v \in \mathcal{V},$$

for any  $\lambda \geq 0$ . The aim is now to find a solution  $(\lambda, f)$  to the equation  $\mathcal{G}_\lambda^*(f) = f$ .

*Step 2.* For a fixed  $\varepsilon > 0$ , we define a regularized operator  $\mathcal{G}_{\lambda,\varepsilon}^* : \mathcal{X} \rightarrow \mathcal{X}$  by

$$(\mathcal{G}_{\lambda,\varepsilon}^* f)(v) = 2 \int_0^\infty f_{\gamma/g}(a, v) e^{-\int_0^a \frac{\lambda}{g(s,v)} ds} da \int_{\mathcal{V}} f(v') \rho_\varepsilon(v, v') dv', \quad v \in \mathcal{V}, \quad (65)$$

with  $\rho_\varepsilon$  picked as in (58), for any  $\lambda \geq 0$ . With similar arguments as previously, we prove we are in position to apply the Krein-Rutman theorem: for any  $\lambda \geq 0$  there exist a unique  $\mu_{\lambda,\varepsilon} > 0$  and a unique positive  $H_{\lambda,\varepsilon} \in \mathcal{X}$  such that

$$\mathcal{G}_{\lambda,\varepsilon}^*(H_{\lambda,\varepsilon}) = \mu_{\lambda,\varepsilon} H_{\lambda,\varepsilon} \quad (66)$$

and  $\|H_{\lambda,\varepsilon}\|_{\mathcal{X}} = 1$ .

*Step 3.* The study of  $\lambda \rightsquigarrow \mu_{\lambda,\varepsilon}$  consists in proving the same four points as previously. Only the verification of **3**) and **4**) slightly differs. To prove **3**), we successively compute

$$(\mathcal{G}_{\lambda=0,\varepsilon}^* f)(v) = 2 \int_{\mathcal{V}} f(v') \rho_\varepsilon(v, v') dv'$$

and

$$\int_{\mathcal{V}} (\mathcal{G}_{\lambda=0,\varepsilon}^* f)(v) dv = 2 \int_{\mathcal{V}} f(v') \left( \int_{\mathcal{V}} \rho(v, v') dv + \varepsilon \right) dv' \geq 2(\varrho + \varepsilon) \int_{\mathcal{V}} f(v') dv',$$

relying on Assumption 2. Thus, choosing  $f = H_{\lambda,\varepsilon}$  in the previous calculus and using (66), we obtain  $\mu_{\lambda=0,\varepsilon} \geq 2(\varrho + \varepsilon) > 1$  as soon as  $\varrho > 1/2$ . To prove 4), we readily obtain that

$$\mu_{\lambda,\varepsilon} \leq 2(\|\mathcal{V}\|_{\rho|\infty} + \varepsilon) \frac{|g|_{\infty} |f_{\gamma/g}|_{\infty}}{\lambda} \rightarrow 0$$

as  $\lambda \rightarrow 0$ .

*Step 4.* We let  $\varepsilon$  go to zero as previously and we find  $(\bar{\lambda}^*, \bar{H})$  with  $\bar{\lambda}^* \geq 0$  and  $\bar{H}$  nonnegative,  $\|\phi^*\|_{\mathcal{X}} = 1$ , such that  $\mathcal{G}_{\bar{\lambda}^*}^* \bar{H} = \bar{H}$ , which means we have found a solution to (63). Recalling (64), we set

$$\phi_{\gamma,\rho}(a, v) = 2\kappa' e^{\int_0^a \frac{\bar{\lambda}^* + \gamma(s,v)}{g(s,v)} ds} \int_a^{\infty} f_{\gamma/g}(s, v) e^{-\int_0^s \frac{\bar{\lambda}^*}{g(t,v)} dt} ds \int_{\mathcal{V}} \bar{H}(v') \rho(v, v') dv'.$$

and fix  $\kappa' > 0$  such that  $\iint_{\mathcal{S}} N_{\gamma,\rho} \phi_{\gamma,\rho} = 1$ . □

*Uniqueness of the eigenelements.* We successively prove the uniqueness of the eigenvalue, of the direct eigenvector and of the adjoint eigenvector.

*Step 1.* Let  $(\lambda, N)$  be a solution to the direct eigenproblem (10) and  $(\lambda^*, \phi)$  be a solution to the adjoint eigenproblem (11). We first prove that  $\lambda = \lambda^*$ . Indeed,

$$\begin{aligned} \lambda \iint_{\mathcal{S}} N\phi &= \iint_{\mathcal{S}} \left( -\frac{\partial}{\partial a}(gN)(a, v) - \gamma(a, v)N(a, v) \right) \phi(a, v) dvda \\ &= \iint_{\mathcal{S}} N(a, v) \left( g(a, v) \frac{\partial}{\partial a} \phi(a, v) + \gamma(a, v) \left( 2 \int_{\mathcal{V}} \phi(0, v') \rho(v, v') dv' - \phi(a, v) \right) \right) dvda \\ &= \lambda^* \iint_{\mathcal{S}} N\phi \end{aligned}$$

and since  $\iint_{\mathcal{S}} N\phi > 0$  we deduce  $\lambda = \lambda^*$ .

*Step 2.* Let  $(\lambda, N_1)$  and  $(\lambda, N_2)$  be two solutions of the direct eigenproblem (10). We prove that  $N_1 = N_2$ . Following the proof of Proposition 6.3 of [Per07], we prove that  $\tilde{N} = |N_1 - N_2|$  satisfies

$$\iint_{\mathcal{S}} \left( \frac{\partial}{\partial a}(g\tilde{N})(a, v) + \gamma(a, v)\tilde{N}(a, v) \right) \phi(a, v) dvda = 0$$

taking  $\phi$  a solution to (11) as a test function. We deduce that

$$\begin{aligned} 2 \int_{\mathcal{V}} \left( \iint_{\mathcal{S}} \gamma(a, v) |N_1 - N_2|(a, v) \rho(v, v') dvda \right) \phi(0, v') dv' \\ = 2 \int_{\mathcal{V}} \left| \iint_{\mathcal{S}} \gamma(a, v) (N_1 - N_2)(a, v) \rho(v, v') dvda \right| \phi(0, v') dv', \end{aligned}$$

using that  $\phi$  is a solution to (11) and since both  $N_1$  and  $N_2$  satisfy the boundary condition of the eigenproblem (10). Thanks to the fact that  $\phi(0, v') > 0$  for  $v' \in \mathcal{V}$ , we deduce that  $\gamma(a, v)(N_1 - N_2)(a, v)\rho(v, v')$  is of constant sign. Then, integrating in  $v'$ ,  $\gamma(a, v)(N_1 - N_2)(a, v)$  is also of constant sign. Recall that for each rate  $v \in \mathcal{V}$ , the division rate  $\gamma(a, v)$  is positive for  $a$  belonging to some  $[a_{\min}(v), a_{\max}(v)]$ , thus  $(N_1 - N_2)(a, v)$  is of constant sign on  $\{[a_{\min}(v), a_{\max}(v)] \times \{v\}, v \in \mathcal{V}\}$ .

Using (55), we deduce that  $(N_1 - N_2)(0, v)$  is of constant sign on  $\mathcal{V}$  and thus, using (55) again,  $(N_1 - N_2)(a, v)$  is of constant sign on  $\mathcal{S}$ . Since we have  $\iint_{\mathcal{S}} (N_1 - N_2) = 0$ , the conclusion  $N_1 = N_2$  follows. To conclude, Fredholm alternative (see [Bre10]) ensures that uniqueness of a solution to (10) implies uniqueness of a solution to (11).  $\square$

The proof of Theorem 4 is now complete.  $\square$

### 6.3 PROOF OF THEOREM 7

First note that since  $\rho_\alpha$  converges in distribution to the Dirac mass at point  $\bar{v}$  as  $\alpha \rightarrow 0$ , we get the convergence of  $\lambda_{B, \rho_\alpha}$  to  $\lambda_{B, \bar{v}}$  as  $\alpha \rightarrow 0$  using the characterisations (18) and (19).

*Proof of Theorem 7.* For  $\alpha \in [0, 1)$ ,

$$\lambda_{B, \rho_\alpha} = \lambda_{B, \bar{v}} + \alpha \frac{d\lambda_{B, \rho_\alpha}}{d\alpha} \Big|_{\alpha=0} + \frac{\alpha^2}{2} \frac{d^2\lambda_{B, \rho_\alpha}}{d\alpha^2} \Big|_{\alpha=0} + o(\alpha^2)$$

and we use Proposition 14 (or more precisely (72)) and Proposition 15 which follow to get the final result.  $\square$

We aim at computing the first derivative of  $\alpha \rightsquigarrow \lambda_{B, \rho_\alpha}$ .

**Proposition 14** (Age-structured model, first derivative). *For a variability kernel of the form (22), if  $\gamma(a, v) = g(a, v)B(a)$  and  $g(a, v) = v$ ,*

$$\begin{aligned} \frac{d\lambda_{B, \rho_\alpha}}{d\alpha} = & \left( \iint_{\mathcal{S}} \frac{a}{\alpha(v - \bar{v}) + \bar{v}} \exp\left(\frac{-\lambda_{B, \rho_\alpha} a}{\alpha(v - \bar{v}) + \bar{v}}\right) f_B(a) \rho(v) dv da \right)^{-1} \\ & \times \iint_{\mathcal{S}} \frac{\lambda_{B, \rho_\alpha} a (v - \bar{v})}{(\alpha(v - \bar{v}) + \bar{v})^2} \exp\left(\frac{-\lambda_{B, \rho_\alpha} a}{\alpha(v - \bar{v}) + \bar{v}}\right) f_B(a) \rho(v) dv da \end{aligned}$$

for any  $\alpha \in (0, 1]$ .

*Proof.* Introduce the operator

$$\mathcal{A}_{\gamma, \rho} f(a, v) = g(a, v) \frac{\partial}{\partial a} f(a, v) + \gamma(a, v) \left( 2 \int_{\mathcal{V}} f(0, v') \rho(v') dv' - f(a, v) \right) \quad (67)$$

densely defined on bounded continuous functions and let  $\mathcal{A}_{\gamma, \rho}^*$  be its dual operator. The eigenproblem given by (10) and (11) can be written more shortly

$$\mathcal{A}_{\gamma, \rho}^* N_{\gamma, \rho} = \lambda_{\gamma, \rho} N_{\gamma, \rho}, \quad \mathcal{A}_{\gamma, \rho} \phi_{\gamma, \rho} = \lambda_{\gamma, \rho} \phi_{\gamma, \rho},$$

with  $N_{\gamma, \rho} \geq 0$  such that  $\iint_{\mathcal{S}} N_{\gamma, \rho} = 1$  and  $\phi_{\gamma, \rho} \geq 0$  such that  $\iint_{\mathcal{S}} N_{\gamma, \rho} \phi_{\gamma, \rho} = 1$ . In order to ease notation, when no confusion is possible, we abbreviate  $\mathcal{A}_{\gamma, \rho_\alpha}$  by  $\mathcal{A}_\alpha$ ,  $N_{\gamma, \rho_\alpha}$  by  $N_\alpha$ ,  $\lambda_{\gamma, \rho_\alpha}$  by  $\lambda_\alpha$  and so on. We denote the support of  $\rho_\alpha$  by  $\mathcal{V}_\alpha$  and  $[0, \infty) \times \mathcal{V}_\alpha$  by  $\mathcal{S}_\alpha$ .

*Step 1.* Let  $\alpha \in (0, 1]$  be fixed. For  $0 < \varepsilon < \alpha$ , we claim that

$$(\lambda_\alpha - \lambda_{\alpha - \varepsilon}) \iint_{\mathcal{S}_{\alpha - \varepsilon}} \phi_\alpha N_{\alpha - \varepsilon} = \iint_{\mathcal{S}_{\alpha - \varepsilon}} (\mathcal{A}_\alpha \phi_\alpha - \mathcal{A}_{\alpha - \varepsilon} \phi_\alpha) N_{\alpha - \varepsilon}. \quad (68)$$



Indeed, operators  $\mathcal{A}_\alpha$  and  $\mathcal{A}_\alpha^*$  are dual, then

$$\iint_{\mathcal{S}_{\alpha-\varepsilon}} (\mathcal{A}_\alpha \phi_\alpha - \mathcal{A}_{\alpha-\varepsilon} \phi_\alpha) N_{\alpha-\varepsilon} = \iint_{\mathcal{S}_{\alpha-\varepsilon}} (\mathcal{A}_\alpha \phi_\alpha) N_{\alpha-\varepsilon} - \phi_\alpha (\mathcal{A}_{\alpha-\varepsilon}^* N_{\alpha-\varepsilon})$$

which leads to (68) since  $\mathcal{A}_\alpha \phi_\alpha = \lambda_\alpha \phi_\alpha$  and  $\mathcal{A}_{\alpha-\varepsilon}^* N_{\alpha-\varepsilon} = \lambda_{\alpha-\varepsilon} N_{\alpha-\varepsilon}$  (see also [Mic06a], Lemma 3.2 and Equation (3.11)). Using the definition (67) of  $\mathcal{A}_\alpha$ , we get

$$\mathcal{A}_\alpha \phi_\alpha(a, v) - \mathcal{A}_{\alpha-\varepsilon} \phi_\alpha(a, v) = 2\gamma(a, v) \left( \int_{\mathcal{V}_\alpha} \phi_\alpha(0, v') \rho_\alpha(v') dv' - \int_{\mathcal{V}_{\alpha-\varepsilon}} \phi_\alpha(0, v') \rho_{\alpha-\varepsilon}(v') dv' \right),$$

which we insert in (68) to obtain

$$\frac{\lambda_\alpha - \lambda_{\alpha-\varepsilon}}{\varepsilon} = \frac{\iint_{\mathcal{S}_{\alpha-\varepsilon}} 2\gamma(a, v) N_{\alpha-\varepsilon}(a, v) dv da}{\iint_{\mathcal{S}_{\alpha-\varepsilon}} \phi_\alpha(a, v) N_{\alpha-\varepsilon}(a, v) dv da} \times \left( \frac{\int_{\mathcal{V}_\alpha} f(0, v') \rho_\alpha(v') dv' - \int_{\mathcal{V}_{\alpha-\varepsilon}} f(0, v') \rho_{\alpha-\varepsilon}(v') dv'}{\varepsilon} \right) \Big|_{f=\phi_\alpha}$$

and we let  $\varepsilon$  go to zero to study the differentiability on the left of  $\alpha \rightsquigarrow \lambda_\alpha$  at point  $\alpha$ . In the same way we compute  $\varepsilon^{-1}(\lambda_{\alpha+\varepsilon} - \lambda_\alpha)$  to study the differentiability on the right of  $\alpha \rightsquigarrow \lambda_\alpha$  at point  $\alpha$ . Since  $N_{\alpha\pm\varepsilon} \rightarrow N_\alpha$  pointwise when  $\varepsilon \rightarrow 0$  and  $\iint_{\mathcal{S}_\alpha} \phi_\alpha N_\alpha = 1$ , by Lebesgue dominated convergence theorem we get

$$\frac{d\lambda_\alpha}{d\alpha} = \kappa_\alpha \frac{d}{d\alpha} \left( \int_{\mathcal{V}_\alpha} f(0, v') \rho_\alpha(v') dv' \right) \Big|_{f=\phi_\alpha} \quad (69)$$

with  $\iint_{\mathcal{S}_\alpha} 2\gamma(a, v) N_\alpha(a, v) dv da = \kappa_\alpha$  as defined in Lemma 12. Thus,  $\lambda_\alpha$  is differentiable in  $\alpha$  if the derivative of the right-hand side exists.

*Step 2.* The adjoint eigenvector  $\phi_\alpha$  has an explicit expression we exploit now. For  $\beta \in (0, 1]$ , recalling the explicit expression of  $\phi_\beta$  given by Lemma 12,

$$\phi_\beta(0, v) = \kappa'_\beta \int_0^\infty \frac{\gamma(a, v)}{g(a, v)} \exp\left(-\int_0^a \frac{\lambda_\beta + \gamma(s, v)}{g(s, v)} ds\right) da, \quad v \in \mathcal{V}_\beta$$

with  $\kappa'_\beta$  defined in Lemma 12. Thus (69) becomes

$$\frac{d\lambda_\alpha}{d\alpha} = \kappa_\alpha \kappa'_\alpha \int_0^\infty \frac{d}{d\alpha} \left( \int_{\mathcal{V}_\alpha} \frac{\gamma(a, v)}{g(a, v)} \exp\left(-\int_0^a \frac{\lambda_\beta + \gamma(s, v)}{g(s, v)} ds\right) \rho_\alpha(v) dv \right) \Big|_{\beta=\alpha} da.$$

Before going ahead in computations, recall Specifications (15),  $g(a, v) = v$ , and (16),  $\gamma(a, v) = vB(a)$ . The previous equality boils down to

$$\frac{d\lambda_\alpha}{d\alpha} = \bar{\kappa}_\alpha \int_0^\infty \frac{d}{d\alpha} \left( \int_{\mathcal{V}_\alpha} \exp\left(\frac{-\lambda_\beta a}{v}\right) \rho_\alpha(v) dv \right) \Big|_{\beta=\alpha} f_B(a) da \quad (70)$$

with  $\bar{\kappa}_\alpha = \kappa_\alpha \kappa'_\alpha$  equal to

$$\bar{\kappa}_\alpha = \left( \iint_{\mathcal{S}_\alpha} \frac{a}{v} \exp\left(-\frac{\lambda_\alpha a}{v}\right) f_B(a) \rho_\alpha(v) dv da \right)^{-1}, \quad (71)$$

using  $\iint_{\mathcal{S}_\alpha} \phi_\alpha N_\alpha = 1$ .

*Step 3.* In the case of the kernel (22), we can explicitly compute the derivative with respect to  $\alpha$ . After a change of variables (setting  $\alpha^{-1}(v - \bar{v}(1 - \alpha))$  as new variable), (70) reads

$$\frac{d\lambda_\alpha}{d\alpha} = \bar{\kappa}_\alpha \int_0^\infty \frac{d}{d\alpha} \left( \int_{\mathcal{V}} \exp\left(-\frac{\lambda_\beta a}{\alpha(v - \bar{v}) + \bar{v}}\right) \rho(v) dv \right) \Big|_{\beta=\alpha} f_B(a) da.$$

Inverting the derivative in  $\alpha$  and the integral and computing the derivative with respect to  $\alpha$ , we get the announced result, recalling the definition (71) of  $\bar{\kappa}_\alpha$ .  $\square$

As a corollary of Proposition 14, for all division rate  $B$ , when  $\alpha$  goes to zero, the first derivative is null,

$$\frac{d\lambda_{B,\rho_\alpha}}{d\alpha} \Big|_{\alpha=0} = \lim_{\alpha \rightarrow 0} \frac{d\lambda_{B,\rho_\alpha}}{d\alpha} = 0 \quad (72)$$

since  $\int_{\mathcal{V}} (v - \bar{v}) \rho(v) dv = 0$  (we picked a baseline density  $\rho$  with mean  $\bar{v}$ ). So we compute the second derivative when  $\alpha$  converges to zero.

**Proposition 15** (Age-structured model, second derivative at point 0). *For a variability kernel of the form (22), if in addition  $\gamma(a, v) = g(a, v)B(a)$  and  $g(a, v) = v$ ,*

$$\frac{d^2 \lambda_{B,\rho_\alpha}}{d\alpha^2} \Big|_{\alpha=0} = \sigma^2 \left( \int_0^\infty \frac{a}{\bar{v}} e^{-\frac{\lambda_{B,\bar{v}} a}{\bar{v}}} f_B(a) da \right)^{-1} \int_0^\infty \frac{\lambda_{B,\bar{v}} a}{\bar{v}} \left( \frac{\lambda_{B,\bar{v}} a}{\bar{v}} - 2 \right) e^{-\frac{\lambda_{B,\bar{v}} a}{\bar{v}}} f_B(a) da$$

with  $\sigma^2 = \int_{\mathcal{V}} (v - \bar{v})^2 \rho(v) dv$ .

*Proof.* For  $\alpha \in (0, 1]$ , we set

$$\Lambda_1(\alpha, a, u) = \frac{a}{u} \exp\left(-\frac{\lambda_{B,\rho_\alpha} a}{u}\right), \quad \Lambda_2(\alpha, a, u) = \frac{\lambda_{B,\rho_\alpha} a}{u^2} \exp\left(-\frac{\lambda_{B,\rho_\alpha} a}{u}\right),$$

so that Proposition 14 reads

$$\begin{aligned} \frac{d\lambda_{B,\rho_\alpha}}{d\alpha} &= \left( \iint_{\mathcal{S}} \Lambda_1(\alpha, a, \alpha(v - \bar{v}) + \bar{v}) f_B(a) \rho(v) dv da \right)^{-1} \\ &\quad \times \iint_{\mathcal{S}} (v - \bar{v}) \Lambda_2(\alpha, a, \alpha(v - \bar{v}) + \bar{v}) f_B(a) \rho(v) dv da = D_1(\alpha)^{-1} D_2(\alpha) \end{aligned}$$

say. Then,  $\frac{d^2 \lambda_{B,\rho_\alpha}}{d\alpha^2}$  the second derivative of  $\alpha \rightsquigarrow \lambda_{B,\rho_\alpha}$  can be written

$$\begin{aligned} &D_2(\alpha) D_1(\alpha)^{-2} \iint_{\mathcal{S}} \left( \frac{\partial \Lambda_1}{\partial \alpha}(\alpha, a, \alpha(v - \bar{v}) + \bar{v}) + (v - \bar{v}) \frac{\partial \Lambda_1}{\partial u}(\alpha, a, \alpha(v - \bar{v}) + \bar{v}) \right) f_B(a) \rho(v) dv da \\ &+ D_1(\alpha)^{-1} \iint_{\mathcal{S}} \left( (v - \bar{v}) \frac{\partial \Lambda_2}{\partial \alpha}(\alpha, a, \alpha(v - \bar{v}) + \bar{v}) + (v - \bar{v})^2 \frac{\partial \Lambda_2}{\partial u}(\alpha, a, \alpha(v - \bar{v}) + \bar{v}) \right) f_B(a) \rho(v) dv da. \end{aligned}$$

We claim that the first term converges to zero as  $\alpha \rightarrow 0$  and that

$$\frac{\partial \Lambda_2}{\partial \alpha}(\alpha, a, \alpha(v - \bar{v}) + \bar{v}) \rightarrow 0, \quad \text{as } \alpha \rightarrow 0, \quad (73)$$

so that,

$$\frac{d^2 \lambda_{B,\rho_\alpha}}{d\alpha^2} \Big|_{\alpha=0} = \left( \lim_{\alpha \rightarrow 0} D_1(\alpha) \right)^{-1} \iint_{\mathcal{S}} (v - \bar{v})^2 \lim_{\alpha \rightarrow 0} \frac{\partial \Lambda_2}{\partial u}(\alpha, a, \bar{v}) f_B(a) \rho(v) dv da, \quad (74)$$

since  $\lim_{\alpha \rightarrow 0} \frac{\partial \Lambda_2}{\partial u}(\alpha, a, \alpha(v - \bar{v}) + \bar{v}) = \lim_{\alpha \rightarrow 0} \frac{\partial \Lambda_2}{\partial u}(\alpha, a, \bar{v})$  by continuity of  $u \rightsquigarrow \frac{\partial \Lambda_2}{\partial u}(\alpha, a, u)$ .

*Step 1.* We first treat the second term of  $\frac{d^2 \lambda_{B,\rho_\alpha}}{d\alpha^2}$ , with three ingredients. **1)** In order to check (73), let us compute

$$\frac{\partial \Lambda_2}{\partial \alpha}(\alpha, a, u) = \left( \frac{d\lambda_{B,\rho_\alpha}}{d\alpha} a - \frac{\lambda_{B,\rho_\alpha}}{u^2} a \frac{d\lambda_{B,\rho_\alpha}}{d\alpha} a \right) e^{-\frac{\lambda_{B,\rho_\alpha} a}{u}}.$$

As  $\alpha \rightarrow 0$ ,  $\lambda_{B,\rho_\alpha}$  converges to  $\lambda_{B,\bar{v}}$  defined by (19) and  $\frac{d\lambda_{B,\rho_\alpha}}{d\alpha}$  to 0 (recall (72)), thus  $\frac{\partial \Lambda_2}{\partial \alpha}(\alpha, a, u) \rightarrow 0$ . Since  $u \rightsquigarrow \frac{\partial \Lambda_2}{\partial \alpha}(\alpha, a, u)$  is continuous, we deduce (73). **2)** Let us now compute

$$\frac{\partial \Lambda_2}{\partial u}(\alpha, a, u) = \left( \frac{-2\lambda_{B,\rho_\alpha} a}{u^3} + \frac{(\lambda_{B,\rho_\alpha} a)^2}{u^4} \right) e^{-\frac{\lambda_{B,\rho_\alpha} a}{u}},$$

which leads to

$$\lim_{\alpha \rightarrow 0} \frac{\partial \Lambda_2}{\partial u}(\alpha, a, \bar{v}) = \frac{\lambda_{B,\bar{v}} a}{\bar{v}^3} \left( \frac{\lambda_{B,\bar{v}} a}{\bar{v}} - 2 \right) e^{-\frac{\lambda_{B,\bar{v}} a}{\bar{v}}} \quad (75)$$

since  $\lambda_{B,\rho_\alpha}$  converges to  $\lambda_{B,\bar{v}}$  defined by (19) as  $\alpha \rightarrow 0$ . **3)** Since  $\Lambda_1(\alpha, a, \alpha(v - \bar{v}) + \bar{v})$  converges to  $\Lambda_1(0, a, \bar{v})$  as  $\alpha \rightarrow 0$ , we get

$$\lim_{\alpha \rightarrow 0} D_1(\alpha) = \int_0^\infty \frac{a}{\bar{v}} \exp\left(-\frac{\lambda_{B,\bar{v}} a}{\bar{v}}\right) f_B(a) da > 0 \quad (76)$$

using  $\int_{\mathcal{V}} \rho(v) dv = 1$ . Gathering (75) and (76) enables us to compute the right-hand side of (74).

*Step 2.* We now check that the first term of  $\frac{d^2 \lambda_{B,\rho_\alpha}}{d\alpha^2}$  converges to zero as  $\alpha \rightarrow 0$ . One readily checks that

$$\mathbf{1)} \frac{\partial \Lambda_1}{\partial \alpha}(\alpha, a, \alpha(u' - \bar{v}) + \bar{v}) \rightarrow 0, \quad \text{as } \alpha \rightarrow 0, \quad \mathbf{2)} \lim_{\alpha \rightarrow 0} \frac{\partial \Lambda_1}{\partial u}(\alpha, a, \alpha(v - \bar{v}) + \bar{v}) < \infty.$$

We know in addition that **3)**  $\lim_{\alpha \rightarrow 0} D_1(\alpha) > 0$  (see (76)) and that **4)**  $\lim_{\alpha \rightarrow 0} D_2(\alpha) = 0$  since  $\int_{\mathcal{V}} (v - \bar{v}) \rho(v) dv = 0$ . Gathering the four points enables us to conclude.  $\square$

## 7 APPENDIX

### 7.1 PROOF OF LEMMA 9

We focus on the size-increment-structured model without variability described by (1) with  $g_x(x, a) = g_a(x, a) = \bar{v}g(x)$ . The state space is  $\mathcal{S} = [0, \infty) \times [0, \infty)$  for size  $x$  and increment of size  $a$ . The direct eigenproblem associated to (1) is the following. We denote by  $(\lambda_{B,\bar{v}}, N_{\gamma,\bar{v}})$  the eigenpair

solution to

$$\begin{cases} \frac{\partial}{\partial x}(\bar{v}xN_{\gamma,\bar{v}}(x,a)) + \frac{\partial}{\partial a}(\bar{v}xN_{\gamma,\bar{v}}(x,a)) + \gamma(x,a)N_{\gamma,\bar{v}}(x,a) = -\lambda_{B,\bar{v}}N_{\gamma,\bar{v}}(x,a), \\ \bar{v}xN_{\gamma,\bar{v}}(x,a=0) = 4 \int_0^\infty \gamma(2x,a)N_{\gamma,\bar{v}}(2x,a)da, \quad \iint_{\mathcal{S}} N_{\gamma,\bar{v}}(x,a)dxda = 1. \end{cases}$$

Now integrate the evolution equation of the previous system against the weight  $x$  and use the boundary condition to immediately deduce that  $\bar{v} = \lambda_{B,\bar{v}}$ .

**Remark 16.** *It is interesting to note that when  $g(x) = \bar{v}x$ ,  $U_{\gamma,\bar{v}}(x,a) = \bar{v}xN_{\gamma,\bar{v}}(x,a)$  (with  $N_{\gamma,\bar{v}}$  solution to (7.1)) is solution the problem on one branch,*

$$\begin{cases} \frac{\partial}{\partial x}(xU_{\gamma,\bar{v}}(x,a)) + \frac{\partial}{\partial a}(xU_{\gamma,\bar{v}}(x,a)) + \gamma(x,a)U_{\gamma,\bar{v}}(x,a) = 0, \\ xU_{\gamma,\bar{v}}(x,a=0) = \int \gamma(2x,a)N_{\gamma,\bar{v}}(2x,a)da, \quad U_{\gamma,\bar{v}} \geq 0, \quad \int_0^\infty U_{\gamma,\bar{v}}(x)dx = 1. \end{cases}$$

Here we have an explicit link between the problem on one branch and the problem on the tree, which is remarkable.

## 7.2 PROOF OF LEMMA 11

We only give the main steps since it closely follows the proof of Lemma 1 of [DHKR15]. To prove (i), write

$$\mathbb{P}_\mu(N_{t+h} - N_t \geq 1 | \mathcal{F}_t) = \int_0^h \gamma(\mathcal{Z}_t + \mathcal{V}_t s, \mathcal{V}_t) e^{-\int_0^s \gamma(\mathcal{Z}_t + \mathcal{V}_t u, \mathcal{V}_t) du} ds = \gamma(\mathcal{Z}_t, \mathcal{V}_t)h + h\varepsilon(h)$$

with

$$\begin{aligned} \varepsilon(h) &= h^{-1} \gamma(\mathcal{Z}_t, \mathcal{V}_t) \int_0^h (e^{-\int_0^s \gamma(\mathcal{Z}_t + \mathcal{V}_t u, \mathcal{V}_t) du} - 1) ds \\ &\quad + h^{-1} \int_0^h (\gamma(\mathcal{Z}_t + \mathcal{V}_t s, \mathcal{V}_t) - \gamma(\mathcal{Z}_t, \mathcal{V}_t)) e^{-\int_0^s \gamma(\mathcal{Z}_t + \mathcal{V}_t u, \mathcal{V}_t) du} ds \end{aligned}$$

and use  $\gamma(\mathcal{Z}_t, \mathcal{V}_t)$  bounded by  $\sup_{s \in [0,t], v \in \mathcal{V}} \gamma(s, v)$  together with the uniform continuity of  $\gamma(\cdot, \mathcal{V}_t)$ . To prove (ii), for  $\Upsilon_t$  as defined in (51), we have  $\{N_{t+h} - N_t \geq 2\} = \{N_{\Upsilon_t + (t+h - \Upsilon_t)} - N_{\Upsilon_t} \geq 1, \Upsilon_t < t+h\}$ . Thus, conditioning with respect to  $\mathcal{F}_{\Upsilon_t}$ ,

$$\begin{aligned} &\mathbb{P}_\mu(N_{t+h} - N_t \geq 2 | \mathcal{F}_t) \\ &= \mathbb{E}_\mu \left[ \int_0^{t+h-\Upsilon_t} \gamma(\mathcal{Z}_{\Upsilon_t} + \mathcal{V}_{\Upsilon_t} s, \mathcal{V}_{\Upsilon_t}) e^{-\int_0^s \gamma(\mathcal{Z}_{\Upsilon_t} + \mathcal{V}_{\Upsilon_t} u, \mathcal{V}_{\Upsilon_t}) du} ds \mathbf{1}_{\{\Upsilon_t < t+h\}} | \mathcal{F}_t \right] \\ &\lesssim h \mathbb{P}_\mu(\Upsilon_t < t+h | \mathcal{F}_t), \end{aligned}$$

noting that  $\mathbb{P}_\mu$ -almost surely,  $t+h - \Upsilon_t \leq h$ . Noting that  $\mathbb{P}_\mu(\Upsilon_t < t+h | \mathcal{F}_t) = \mathbb{P}_\mu(N_{t+h} - N_t \geq 1 | \mathcal{F}_t)$  and invoking part (i) of the lemma we have just proved give the asserted result.

### 7.3 COMPLEMENTS ON THE MODEL INTRODUCED BY AMIR

Let us check the model is well-defined.

*Step 1.* We first show that if  $\xi_{u^-} \leq x_{\max}$  then  $\xi_u \leq x_{\max}$  for a good choice of parameters  $(\Delta_0, h, v_{\max}, x_{\max})$ . Assume that  $\xi_{u^-} \leq x_{\max}$ , then

$$\xi_u = \frac{1}{2}\xi_{u^-}e^{\tau_u - \zeta_{u^-}} = \frac{1}{2}(\xi_{u^-} + \Delta_0)e^{\tau_u - \eta_{u^-}} \leq \frac{x_{\max} + \Delta_0}{2}e^{v_{\max}h}$$

using (42) and since  $\tau_u \leq v_{\max}$  and  $\eta_u \in [-h, h]$ . To guarantee  $\xi_u \leq x_{\max}$  we impose

$$0 \leq \frac{x_{\max} + \Delta_0}{2}e^{v_{\max}h} \leq x_{\max} \iff x_{\max} \geq \frac{\Delta_0 e^{v_{\max}h}}{2 - e^{v_{\max}h}} \quad \text{and} \quad e^{v_{\max}h} \leq 2. \quad (77)$$

Consequently we should just pick  $\xi_\emptyset \leq x_{\max}$  to have  $\xi_u \leq x_{\max}$  for any cell  $u \in \mathbb{T}$ .

*Step 2.* We should have  $\zeta_u \geq 0$  for any  $u \in \mathbb{T}$ , *i.e.*

$$\eta_u \geq -\frac{1}{\tau_u} \ln \left(1 + \frac{\Delta_0}{\xi_u}\right), \quad u \in \mathbb{T},$$

recalling (42). Assuming that  $\xi_u \in [0, x_{\max}]$  and  $\tau_u \in \mathcal{V} = [v_{\min}, v_{\max}]$ , an upper-bound of the right-hand side is

$$-\frac{1}{v_{\max}} \ln \left(1 + \frac{\Delta_0}{x_{\max}}\right) \geq -\frac{1}{\tau_u} \ln \left(1 + \frac{\Delta_0}{\xi_u}\right),$$

and one should impose

$$0 < h \leq \frac{1}{v_{\max}} \ln \left(1 + \frac{\Delta_0}{x_{\max}}\right) \iff x_{\max} \leq \frac{\Delta_0}{e^{v_{\max}h} - 1} \quad (78)$$

so that

$$\eta_u \geq -h \geq -\frac{1}{v_{\max}} \ln \left(1 + \frac{\Delta_0}{x_{\max}}\right) \geq -\frac{1}{\tau_u} \ln \left(1 + \frac{\Delta_0}{\xi_u}\right)$$

*i.e.*  $\zeta_u \geq 0$ .

Gathering constraints (77) and (78) gives a non-empty solution as soon as  $\exp(v_{\max}h) \leq \sqrt{2}$  and we retrieve the constraint on  $x_{\max}$  (see page 205).

## 8 SUPPLEMENTARY FIGURES AND TABLES

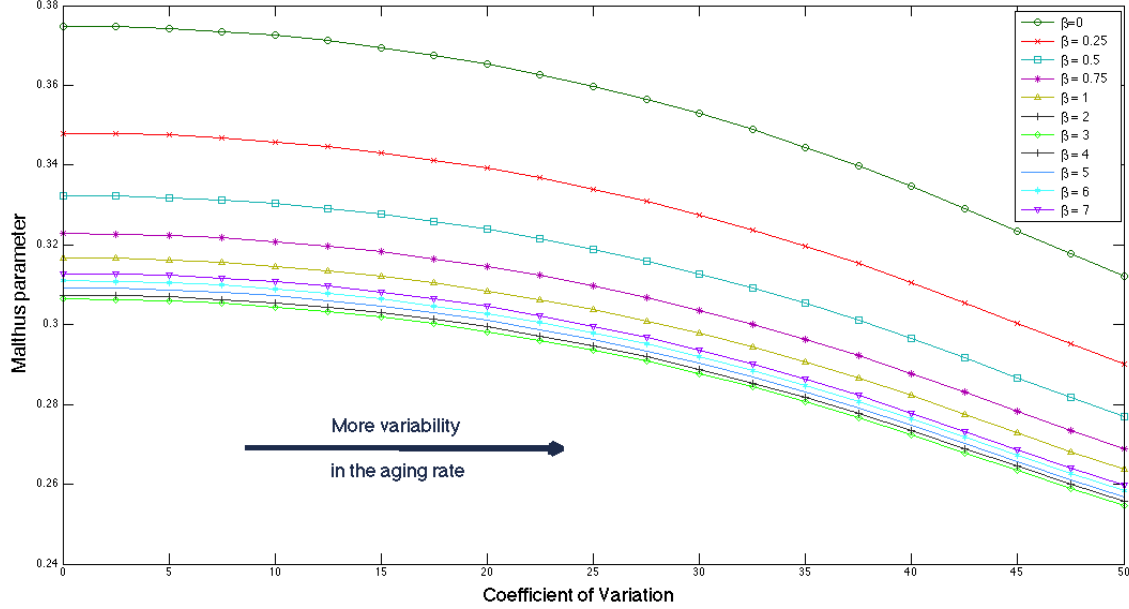


Figure 20: Age-structured model with variability.  $CV_{\rho_\alpha} \rightsquigarrow \lambda_{B,\rho_\alpha}$  defined by (18) for  $\rho_\alpha(v) = \alpha^{-1}\rho(\alpha^{-1}(v - (1 - \alpha)\bar{v}))$  (the baseline density  $\rho$  is a Gaussian density with mean  $\bar{v} = 1$  and standard deviation 0.7 truncated on  $[0, 2]$ ) and different division rates  $B(a) = (a - 1)^\beta \mathbf{1}_{\{a \geq 1\}}$ ,  $\beta \in \{0, 0.25, 0.5, 0.75, 1, 2, \dots, 7\}$ . Reference (all cells age at rate  $\bar{v} = 1$ ): point of abscissa  $\alpha = 0$  and y-coordinate  $\lambda_{B,\bar{v}=1}$  defined by (19).

|                                   | $CV_{\rho_\alpha} = 5\%$               | $CV_{\rho_\alpha} = 10\%$              | $CV_{\rho_\alpha} = 15\%$              | $CV_{\rho_\alpha} = 20\%$             |                                     |
|-----------------------------------|--|--|--|---------------------------------------|-------------------------------------|
| $T$                               | 10.5                                   | 11                                     | 11.25                                  | 11.5                                  |                                     |
| Mean $ \partial \mathcal{J}_T $   | 22 493                                 | 35 588                                 | 39 771                                 | 48 502                                |                                     |
| (Min. $\leq$ $\cdot$ $\leq$ Max.) | (19 877 $\leq$ $\cdot$ $\leq$ 27 198)  | (26 895 $\leq$ $\cdot$ $\leq$ 44 091)  | (27 672 $\leq$ $\cdot$ $\leq$ 52 427)  | (21 549 $\leq$ $\cdot$ $\leq$ 73 785) |                                     |
|                                   | $CV_{\rho_\alpha} = 25\%$              | $CV_{\rho_\alpha} = 30\%$              | $CV_{\rho_\alpha} = 35\%$              | $CV_{\rho_\alpha} = 40\%$             | $CV_{\rho_\alpha} = 45\%$           |
| $T$                               | 12                                     | 12.5                                   | 13                                     | 13.5                                  | 13.75                               |
| Mean $ \partial \mathcal{J}_T $   | 62 903                                 | 80 921                                 | 96 884                                 | 114 857                               | 94 054                              |
| (Min. $\leq$ $\cdot$ $\leq$ Max.) | (26 378 $\leq$ $\cdot$ $\leq$ 116 323) | (19 298 $\leq$ $\cdot$ $\leq$ 147 479) | (24 879 $\leq$ $\cdot$ $\leq$ 177 248) | (7 925 $\leq$ $\cdot$ $\leq$ 356 607) | (465 $\leq$ $\cdot$ $\leq$ 251 256) |

Table 9: Size-increment-structured model with variability. Mean number (minimum, maximum) of living cells at time  $T$  over the  $M = 50$  Monte Carlo continuous time trees simulated up to time  $T$ , for the growth rate densities  $(\rho_\alpha, \alpha \in \{0.1, 0.2, \dots, 0.9\})$  and  $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$ .

|  | $CV_{\rho_\alpha} = 5\%$  | $CV_{\rho_\alpha} = 10\%$ | $CV_{\rho_\alpha} = 15\%$ | $CV_{\rho_\alpha} = 20\%$ |                           |
|--|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Mean of $\hat{\lambda}_T^{(1)}$<br>(sd.) | 0.9982<br>(0.0007)        | 0.9931<br>(0.0014)        | 0.9850<br>(0.0015)        | 0.9738<br>(0.0024)        |                           |
| 95% CI                                   | [0.9967, 0.9996]          | [0.9901, 0.9955]          | [0.9826, 0.9885]          | [0.9684, 0.9782]          |                           |
| Mean of $\hat{\lambda}_T^{(2)}$<br>(sd.) | 0.9883<br>(0.0117)        | 0.9970<br>(0.0059)        | 0.9861<br>(0.0050)        | 0.9745<br>(0.0052)        |                           |
| 95% CI                                   | [0.9718, 1.0123]          | [0.9864, 1.0080]          | [0.9742, 0.9951]          | [0.9628, 0.9848]          |                           |
|  | $CV_{\rho_\alpha} = 25\%$ | $CV_{\rho_\alpha} = 30\%$ | $CV_{\rho_\alpha} = 35\%$ | $CV_{\rho_\alpha} = 40\%$ | $CV_{\rho_\alpha} = 45\%$ |
| 0.9580<br>(0.0031)                       | 0.9381<br>(0.0037)        | 0.9165<br>(0.0047)        | 0.8902<br>(0.0061)        | 0.8585<br>(0.0108)        |                           |
| [0.9497, 0.9631]                         | [0.9297, 0.9446]          | [0.9041, 0.9260]          | [0.8778, 0.8999]          | [0.8256, 0.8764]          |                           |
| 0.9578<br>(0.0058)                       | 0.9374<br>(0.0039)        | 0.9163<br>(0.0062)        | 0.8900<br>(0.0061)        | 0.8599<br>(0.0123)        |                           |
| [0.9467, 0.9676]                         | [0.9283, 0.9447]          | [0.8984, 0.9313]          | [0.8754, 0.8986]          | [0.8282, 0.8960]          |                           |

Table 10: Size-increment-structured model with variability. Division rate  $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$ . Estimation of the Malthus parameter  $\lambda_{B, \rho_\alpha}$  (mean and variance based on  $M = 50$  Monte Carlo continuous time trees) with respect to the coefficient of variation of the growth rate density  $\rho_\alpha$ . Estimation by (33) via the biomass,  $\hat{\lambda}_T^{(1)}$ , and by (37) via the number of cells,  $\hat{\lambda}_T^{(2)}$ .

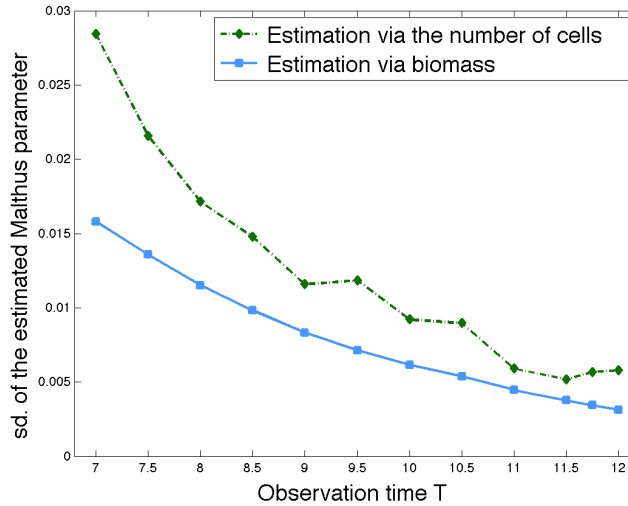


Figure 21: Size-increment-structured model with variability. Standard deviation of two estimators of the Malthus parameter as  $T$  increases (based on  $M = 50$  Monte Carlo continuous time trees simulated up to time  $T$ ), for  $\rho_\alpha = 0.5$  and  $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$ . Blue lower curve: estimation by (33) via the biomass. Green upper curve: estimation by (37) via the number of cells.

|   | $CV_{\rho_\alpha} = 5\%$                        | $CV_{\rho_\alpha} = 10\%$                       | $CV_{\rho_\alpha} = 15\%$                       | $CV_{\rho_\alpha} = 20\%$                       |
|---|---|---|---|---|
| $T$   | 10.5  | 11  | 11.25   | 11.5  |
| Mean $ \partial\mathcal{J}_T $<br>(Min. $\leq$ $\cdot$ $\leq$ Max.) | 23 391<br>(18 944 $\leq$ $\cdot$ $\leq$ 28 357) | 36 620<br>(27 551 $\leq$ $\cdot$ $\leq$ 45 378) | 44 613<br>(29 271 $\leq$ $\cdot$ $\leq$ 56 873) | 46 522<br>(17 455 $\leq$ $\cdot$ $\leq$ 73 855) |
| Mean $\hat{\lambda}_T$<br>(sd.)                                     | 0.9986<br>(0.0006)                              | 0.9933<br>(0.0012)                              | 0.9853<br>(0.0018)                              | 0.9723<br>(0.0024)                              |
| 95% CI  | [0.9977, 1.0000]                                | [0.9904, 0.9952]                                | [0.9818, 0.9880]                                | [0.9679, 0.9769]                                |
| $CV_{\rho_\alpha} = 25\%$   | $CV_{\rho_\alpha} = 30\%$                       | $CV_{\rho_\alpha} = 35\%$                       | $CV_{\rho_\alpha} = 40\%$                       | $CV_{\rho_\alpha} = 45\%$                       |
| 11.75   | 12  | 12.25   | 12.5  | 13  |
| 51 849<br>(27 560 $\leq$ $\cdot$ $\leq$ 81 383)                     | 54 054<br>(26 359 $\leq$ $\cdot$ $\leq$ 85 993) | 50 872<br>(8 316 $\leq$ $\cdot$ $\leq$ 89 667)  | 43 744<br>(1 552 $\leq$ $\cdot$ $\leq$ 97 497)  | 55 450<br>(3 585 $\leq$ $\cdot$ $\leq$ 112 786) |
| 0.9573<br>(0.0028)  | 0.9389<br>(0.0031)                              | 0.9156<br>(0.0046)                              | 0.8864<br>(0.0082)                              | 0.8592<br>(0.0084)                              |
| [0.9516, 0.9623]  | [0.9330, 0.9445]                                | [0.9030, 0.9232]                                | [0.8667, 0.9018]                                | [0.8433, 0.8775]                                |

Table 11: Size-increment-structured model with variability. Division rate  $B(a) = (a - 1)^8 \mathbf{1}_{\{a \geq 1\}}$ . Estimation of the Malthus parameter  $\lambda_{B, \rho_\alpha}$  (mean and 95% confidence interval based on  $M = 50$  Monte Carlo continuous time trees simulated up to time  $T$ ) with respect to the coefficient of variation of the growth rates density  $\rho_\alpha$ .

|   | $CV_{\rho_\alpha} = 5\%$                        | $CV_{\rho_\alpha} = 10\%$                        | $CV_{\rho_\alpha} = 15\%$                       | $CV_{\rho_\alpha} = 20\%$                       |
|---|---|--|---|---|
| $T$   | 10.5  | 11   | 11.25   | 11.5  |
| Mean $ \partial\mathcal{J}_T $<br>(Min. $\leq$ $\cdot$ $\leq$ Max.) | 23 552<br>(20 500 $\leq$ $\cdot$ $\leq$ 26 481) | 36 284<br>(28 956 $\leq$ $\cdot$ $\leq$ 44 799)  | 41 763<br>(29 816 $\leq$ $\cdot$ $\leq$ 59 134) | 46 128<br>(26 790 $\leq$ $\cdot$ $\leq$ 71 492) |
| Mean $\hat{\lambda}_T$<br>(sd.)                                     | 0.9984<br>(0.0007)                              | 0.9937<br>(0.0014)                               | 0.9852<br>(0.0018)                              | 0.9733<br>(0.0022)                              |
| 95% CI  | [0.9972, 0.9999]                                | [0.9913, 0.9964]                                 | [0.9819, 0.9893]                                | [0.9685, 0.9772]                                |
| $CV_{\rho_\alpha} = 25\%$   | $CV_{\rho_\alpha} = 30\%$                       | $CV_{\rho_\alpha} = 35\%$                        | $CV_{\rho_\alpha} = 40\%$                       | $CV_{\rho_\alpha} = 45\%$                       |
| 11.75   | 12  | 12.25  | 12.5  | 13  |
| 52 097<br>(18 007 $\leq$ $\cdot$ $\leq$ 92 617)                     | 51 062<br>(15 933 $\leq$ $\cdot$ $\leq$ 87 886) | 52 465<br>(13 665 $\leq$ $\cdot$ $\leq$ 108 159) | 45 291<br>(3 987 $\leq$ $\cdot$ $\leq$ 91 984)  | 47 440<br>(741 $\leq$ $\cdot$ $\leq$ 112 594)   |
| 0.9580<br>(0.0028)  | 0.9388<br>(0.0035)                              | 0.9143<br>(0.0046)                               | 0.8859<br>(0.0064)                              | 0.8533<br>(0.0069)                              |
| [0.9514, 0.9631]  | [0.9302, 0.9442]                                | [0.9030, 0.9244]                                 | [0.8696, 0.9006]                                | [0.8382, 0.8660]                                |

Table 12: Size-increment-structured model with variability. Division rate  $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$ . Asymmetric division (a cell of size  $x$  splits into two cells of size  $ux$  and  $(1 - u)x$  for  $u$  uniformly drawn on  $[0.1, 0.9]$ ). Estimation of the Malthus parameter  $\lambda_{B, \rho_\alpha}$  (mean and 95% confidence interval based on  $M = 50$  Monte Carlo continuous time trees simulated up to time  $T$ ) with respect to the coefficient of variation of the growth rates density  $\rho_\alpha$ .



|                        | $CV_{\rho_\alpha} = 5\%$  | $CV_{\rho_\alpha} = 10\%$ | $CV_{\rho_\alpha} = 15\%$ | $CV_{\rho_\alpha} = 20\%$ |                           |
|------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Mean $\hat{\lambda}_T$ | 0.9985                    | 0.9930                    | 0.9845                    | 0.9742                    |                           |
| (sd.)                  | (0.0012)                  | (0.0026)                  | (0.0045)                  | (0.0053)                  |                           |
| 95% CI                 | [0.9962, 1.0003]          | [0.9879, 0.9987]          | [0.9776, 0.9939]          | [0.9656, 0.9840]          |                           |
|                        | $CV_{\rho_\alpha} = 25\%$ | $CV_{\rho_\alpha} = 30\%$ | $CV_{\rho_\alpha} = 35\%$ | $CV_{\rho_\alpha} = 40\%$ | $CV_{\rho_\alpha} = 45\%$ |
| 0.9570                 | 0.9379                    | 0.9183                    | 0.8925                    | 0.8599                    |                           |
| (0.0068)               | (0.0096)                  | (0.0107)                  | (0.0113)                  | (0.0140)                  |                           |
| [0.9430, 0.9693]       | [0.9207, 0.9579]          | [0.8930, 0.9404]          | [0.8659, 0.9134]          | [0.8318, 0.8864]          |                           |

Table 13: Size-increment-structured model with variability, experimental conditions. Division rate  $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$ . Estimation of the Malthus parameter  $\lambda_{B, \rho_\alpha}$  (mean and 95% confidence interval based on  $M = 50$  Monte Carlo continuous time trees simulated up to the random time  $\Pi$  such that  $4\,500 \leq |\partial \mathcal{T}_\Pi| \leq 5\,500$ ) with respect to the coefficient of variation of the growth rates density  $\rho_\alpha$ .

|   | $\kappa = 0$                                   | $\kappa = 0.25$                                | $\kappa = 0.50$                                |
|---|--|--|--|
| <b><math>CV_\kappa = 5\%</math></b>                             |  |  |  |
| $T$   | 12   | 12   | 12.5   |
| Mean $ \partial\mathcal{J}_T $<br>(Min. $\leq \cdot \leq$ Max.) | 107 449<br>(99 683 $\leq \cdot \leq$ 117 416)  | 105 431<br>(91137 $\leq \cdot \leq$ 115 549)   | 171 369<br>(151 874 $\leq \cdot \leq$ 204 780) |
| Mean $\hat{\lambda}_T$<br>(sd.)                                 | 0.9984<br>(0.0005)                             | 0.9989<br>(0.0006)                             | 1.0002<br>(0.0006)                             |
| 95% CI  | [0.9972, 0.9994]                               | [0.9976, 0.9999]                               | [0.9988, 1.0014]                               |
| <b><math>CV_\kappa = 10\%</math></b>                            |  |  |  |
| $T$   | 12.25  | 12.25  | 12.5   |
| Mean $ \partial\mathcal{J}_T $<br>(Min. $\leq \cdot \leq$ Max.) | 125 013<br>(105 156 $\leq \cdot \leq$ 145 941) | 128 835<br>(92 082 $\leq \cdot \leq$ 157 695)  | 169 206<br>(124 365 $\leq \cdot \leq$ 226 150) |
| Mean $\hat{\lambda}_T$<br>(sd.)                                 | 0.9935<br>(0.0009)                             | 0.9960<br>(0.0011)                             | 1.0006<br>(0.0016)                             |
| 95% CI  | [0.9922, 0.9952]                               | [0.9937, 0.9977]                               | [0.9981, 1.0048]                               |
| <b><math>CV_\kappa = 15\%</math></b>                            |  |  |  |
| $T$   | 12.5   | 12.5   | 12.5   |
| Mean $ \partial\mathcal{J}_T $<br>(Min. $\leq \cdot \leq$ Max.) | 151 429<br>(110709 $\leq \cdot \leq$ 184 649)  | 158 584<br>(112109 $\leq \cdot \leq$ 202 504)  | 176 701<br>(122377 $\leq \cdot \leq$ 257 817)  |
| Mean $\hat{\lambda}_T$<br>(sd.)                                 | 0.9854<br>(0.0012)                             | 0.9905<br>(0.0014)                             | 1.0006<br>(0.0027)                             |
| 95% CI  | [0.9824, 0.9877]                               | [0.9881, 0.9932]                               | [0.9941, 1.0053]                               |
| <b><math>CV_\kappa = 20\%</math></b>                            |  |  |  |
| $T$   | 12.75  | 12.75  | 12.5   |
| Mean $ \partial\mathcal{J}_T $<br>(Min. $\leq \cdot \leq$ Max.) | 160 789<br>(105 162 $\leq \cdot \leq$ 215 072) | 185 423<br>(107 042 $\leq \cdot \leq$ 264 120) | 184 511<br>(100 129 $\leq \cdot \leq$ 328 265) |
| Mean $\hat{\lambda}_T$<br>(sd.)                                 | 0.9735<br>(0.0014)                             | 0.9834<br>(0.0020)                             | 1.0022<br>(0.0033)                             |
| 95% CI  | [0.9708, 0.9767]                               | [0.9791, 0.9871]                               | [0.9948, 1.0089]                               |
| <b><math>CV_\kappa = 25\%</math></b>                            |  |  |  |
| $T$   | 13   | 13   | 12.5   |
| Mean $ \partial\mathcal{J}_T $<br>(Min. $\leq \cdot \leq$ Max.) | 167 806<br>(86 512 $\leq \cdot \leq$ 225 867)  | 221 010<br>(94 434 $\leq \cdot \leq$ 334 486)  | 174 790<br>(67 461 $\leq \cdot \leq$ 349 999)  |
| Mean $\hat{\lambda}_T$<br>(sd.)                                 | 0.9583<br>(0.0023)                             | 0.9723<br>(0.0023)                             | 1.0029<br>(0.0044)                             |
| 95% CI  | [0.9533, 0.9622]                               | [0.9690, 0.9793]                               | [0.9928, 1.0101]                               |

Table 14: Size-increment-structured model with inherited variability (autoregressive case). Division rate  $B(a) = (a-1)^2 \mathbf{1}_{\{a \geq 1\}}$  and the growth rates  $(\tau_u, u \in \mathbb{T})$  satisfy (38). Estimation of the Malthus parameter  $\lambda_{B,\kappa}$  (mean and 95% confidence interval based on  $M = 50$  Monte Carlo continuous time trees simulated up to time  $T$ ) with respect to  $\kappa$ , for different coefficients of variation  $CV_\kappa$  of the growth rate equilibrium distribution.

| $ \bar{v}_2 - \bar{v}_1 $<br>( $\bar{v}_1, \bar{v}_2$ )          | <b>0.4</b><br>(0.8, 1.2)                        | <b>0.8</b><br>(0.6, 1.4)                       | <b>1.2</b><br>(0.4, 1.6)                       | <b>1.6</b><br>(0.2, 1.8)                      |
|--|---|--|--|---|
| <b><math>p_{11} = 0.90</math> and <math>p_{22} = 0.30</math></b> |   |  |  |   |
| <b><math>T</math></b>  | <i>12</i>                                       | <i>14</i>                                      | <i>19</i>                                      | <i>33</i>                                     |
| <b>Mean</b><br>(Min. $\leq$ $\cdot$ $\leq$ Max.)                 | 21 503<br>(16 203 $\leq$ $\cdot$ $\leq$ 32 857) | 13 025<br>(7 474 $\leq$ $\cdot$ $\leq$ 25 486) | 9 757<br>(4 512 $\leq$ $\cdot$ $\leq$ 19 010)  | 5 479<br>(1 525 $\leq$ $\cdot$ $\leq$ 20 232) |
| <b>Mean <math>\hat{\lambda}_T</math></b><br>(sd.)                | 0.8411<br>(0.0015)                              | 0.6641<br>(0.0043)                             | 0.4662<br>(0.0044)                             | 0.2450<br>(0.0048)                            |
| <b>95% CI</b>  | [0.8386, 0.8441]                                | [0.6563, 0.6723]                               | [0.4562, 0.4743]                               | [0.2350, 0.2536]                              |
| <b>Reference</b>   | 0.85  | 0.70   | 0.55   | 0.40  |
| <b><math>p_{11} = 0.60</math> and <math>p_{22} = 0.30</math></b> |   |  |  |   |
| <b><math>T</math></b>  | <i>12</i>                                       | <i>14</i>                                      | <i>16</i>                                      | <i>21</i>                                     |
| <b>Mean</b><br>(Min. $\leq$ $\cdot$ $\leq$ Max.)                 | 46 589<br>(35 364 $\leq$ $\cdot$ $\leq$ 58 004) | 50 825<br>(30599 $\leq$ $\cdot$ $\leq$ 75 270) | 15 600<br>(6 680 $\leq$ $\cdot$ $\leq$ 35 365) | 2 328<br>(440 $\leq$ $\cdot$ $\leq$ 8 416)    |
| <b>Mean <math>\hat{\lambda}_T</math></b><br>(sd.)                | 0.9190<br>(0.0019)                              | 0.7870<br>(0.0022)                             | 0.5984<br>(0.0058)                             | 0.3430<br>(0.0075)                            |
| <b>95% CI</b>  | [0.9155, 0.9227]                                | [0.7826, 0.7903]                               | [0.5884, 0.6098]                               | [0.3323, 0.3596]                              |
| <b>Reference</b>   | 0.94  | 0.89   | 0.84   | 0.78  |

Table 15: Size-increment-structured model with inherited variability (bimodal case). Division rate  $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$  and the growth rates  $(\tau_u, u \in \mathbb{T})$  satisfy (39) and (40). Estimation of the Malthus parameter  $\lambda_{B, \bar{v}_1, \bar{v}_2, p_{11}, p_{22}}$  (mean and 95% confidence interval based on  $M = 50$  Monte Carlo continuous time trees simulated up to time  $T$ ) with respect to  $|\bar{v}_2 - \bar{v}_1|$ , for different transition probabilities  $(p_{11}, p_{22})$ .

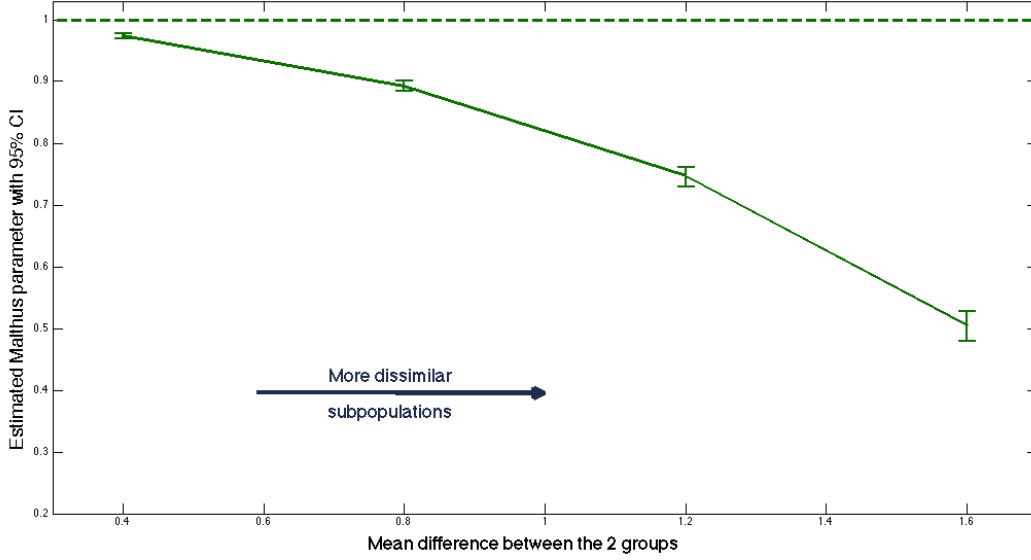


Figure 22: *Size-increment-structured model with bimodal variability (without heredity). Division rate  $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$  and the growth rates  $(\tau_u, u \in \mathbb{T})$  are drawn according to the mixture  $\frac{1}{2}\rho_1 + \frac{1}{2}\rho_2$  where  $\rho_i$  is a Gaussian distribution with mean  $\bar{v}_i$  and standard deviation 0.03 truncated on  $\mathcal{V}_i = [\bar{v}_i - 0.10, \bar{v}_i + 0.10]$ . Estimation of the Malthus parameter  $\lambda_{B, \bar{v}_1, \bar{v}_2}$  (mean and 95% confidence interval based on  $M = 50$  Monte Carlo continuous time trees simulated up to time  $T$ ) with respect to  $|\bar{v}_2 - \bar{v}_1|$*

| $ \bar{v}_2 - \bar{v}_1 $  | <b>0.4</b>                                | <b>0.8</b>                                | <b>1.2</b>                                 | <b>1.6</b>                                |
|--|---|---|--|---|
| $(\bar{v}_1, \bar{v}_2)$   | (0.8, 1.2)                                | (0.6, 1.4)                                | (0.4, 1.6)                                 | (0.2, 1.8)                                |
| $T$  | 11.25                                     | 11.5                                      | 14   | 20  |
| Mean $ \partial \mathcal{J}_T $<br>(Min. $\leq \cdot \leq$ Max.) | 42 042<br>31 144 $\leq \cdot \leq$ 63 748 | 27 248<br>11 979 $\leq \cdot \leq$ 46 040 | 42 566<br>12 414 $\leq \cdot \leq$ 119 911 | 37 202<br>2 813 $\leq \cdot \leq$ 132 317 |
| Mean $\hat{\lambda}_T$<br>(sd.)                                  | 0.9734<br>(0.0025)                        | 0.8920<br>(0.0043)                        | 0.7473<br>(0.0079)                         | 0.5055<br>(0.0113)                        |
| 95% CI   | [0.9694, 0.9781]                          | [0.8835, 0.9011]                          | [0.7298, 0.7607]                           | [0.4791, 0.5272]                          |

Table 16: *Size-increment-structured model with bimodal variability (without heredity). Division rate  $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$  and the growth rates  $(\tau_u, u \in \mathbb{T})$  are drawn according to the mixture  $\frac{1}{2}\rho_1 + \frac{1}{2}\rho_2$  where  $\rho_i$  is a Gaussian distribution with mean  $\bar{v}_i$  and standard deviation 0.03 truncated on  $\mathcal{V}_i = [\bar{v}_i - 0.10, \bar{v}_i + 0.10]$ . Estimation of the Malthus parameter  $\lambda_{B, \bar{v}_1, \bar{v}_2}$  (mean and 95% confidence interval based on  $M = 50$  Monte Carlo continuous time trees simulated up to time  $T$ ) with respect to  $|\bar{v}_2 - \bar{v}_1|$*

| $T$   | $CV_{\rho_\alpha} = 5\%$                  | $CV_{\rho_\alpha} = 10\%$                 | $CV_{\rho_\alpha} = 15\%$                 | $CV_{\rho_\alpha} = 20\%$                 |
|---|---|---|---|---|
| $10.5$  | $11$                                      | $11.25$                                   | $11.5$                                    |   |
| Mean $ \partial\mathcal{J}_T $<br>(Min. $\leq$ $\leq$ Max.) | 46 837<br>(42 358 $\leq$ $\leq$ 52 147)   | 73 100<br>(57 254 $\leq$ $\leq$ 87 282)   | 90 027<br>(53 615 $\leq$ $\leq$ 116 052)  | 98 270<br>(68 946 $\leq$ $\leq$ 128 379)  |
| Mean $\hat{\lambda}_T$<br>(sd.)                             | 0.9985<br>(0.0006)                        | 0.9938<br>(0.0009)                        | 0.9867<br>(0.0014)                        | 0.9757<br>(0.0018)                        |
| 95% CI  | [0.9974, 0.9999]                          | [0.9923, 0.9954]                          | [0.9841, 0.9893]                          | [0.9717, 0.9789]                          |
| $CV_{\rho_\alpha} = 25\%$                                   | $CV_{\rho_\alpha} = 30\%$                 | $CV_{\rho_\alpha} = 35\%$                 | $CV_{\rho_\alpha} = 40\%$                 | $CV_{\rho_\alpha} = 45\%$                 |
| $11.75$   | $12$                                      | $12.25$                                   | $12.5$                                    | $13$                                      |
| 107 305<br>(71 884 $\leq$ $\leq$ 157 032)                   | 120 102<br>(63 409 $\leq$ $\leq$ 200 860) | 104 628<br>(52 116 $\leq$ $\leq$ 172 328) | 117 208<br>(28 171 $\leq$ $\leq$ 192 021) | 114 180<br>(39 238 $\leq$ $\leq$ 238 181) |
| 0.9617<br>(0.0019)  | 0.9450<br>(0.0027)                        | 0.9245<br>(0.0035)                        | 0.8985<br>(0.0030)                        | 0.8722<br>(0.0039)                        |
| [0.9583, 0.9656]  | [0.9397, 0.9505]                          | [0.9178, 0.9312]                          | [0.8920, 0.9036]                          | [0.8650, 0.8794]                          |

Table 17: Size-structured model with variability. Division rate  $B(a) = (a - 1)^2 \mathbf{1}_{\{a \geq 1\}}$ . Estimation of the Malthus parameter  $\lambda_{B, \rho_\alpha}$  (mean and 95% confidence interval based on  $M = 50$  Monte Carlo continuous time trees simulated up to time  $T$ ) with respect to the coefficient of variation of the growth rates density  $\rho_\alpha$ .

| $T$   | $CV_{\rho_\alpha} = 5\%$                | $CV_{\rho_\alpha} = 10\%$               | $CV_{\rho_\alpha} = 15\%$               | $CV_{\rho_\alpha} = 20\%$               |
|---|---|---|---|---|
| $10.5$  | $11$                                    | $11.25$                                 | $11.5$                                  |   |
| Mean $ \partial\mathcal{J}_T $<br>(Min. $\leq$ $\leq$ Max.) | 16 552<br>(15 737 $\leq$ $\leq$ 17 822) | 27 877<br>(19 588 $\leq$ $\leq$ 33 767) | 32 981<br>(21 557 $\leq$ $\leq$ 42 850) | 37 710<br>(20 692 $\leq$ $\leq$ 56 655) |
| Mean $\hat{\lambda}_T$<br>(sd.)                             | 0.9984<br>(0.0007)                      | 0.9935<br>(0.0014)                      | 0.9856<br>(0.0019)                      | 0.9738<br>(0.0029)                      |
| 95% CI  | [0.9968, 0.9997]                        | [0.9908, 0.9961]                        | [0.9817, 0.9892]                        | [0.9680, 0.9799]                        |
| $CV_{\rho_\alpha} = 25\%$                                   | $CV_{\rho_\alpha} = 30\%$               | $CV_{\rho_\alpha} = 35\%$               | $CV_{\rho_\alpha} = 40\%$               | $CV_{\rho_\alpha} = 45\%$               |
| $11.75$   | $12$                                    | $12.25$                                 | $12.5$                                  | $13$                                    |
| 40 786<br>(17 833 $\leq$ $\leq$ 64 689)                     | 37 384<br>(14 053 $\leq$ $\leq$ 65 584) | 37 298<br>(10 954 $\leq$ $\leq$ 71 054) | 34 246<br>(1 028 $\leq$ $\leq$ 95 197)  | 35 683<br>(412 $\leq$ $\leq$ 113 070)   |
| 0.9583<br>(0.0035)  | 0.9382<br>(0.0044)                      | 0.9159<br>(0.0060)                      | 0.8871<br>(0.0065)                      | 0.8594<br>(0.0103)                      |
| [0.9522, 0.9666]  | [0.9289, 0.9453]                        | [0.9047, 0.9290]                        | [0.8718, 0.8999]                        | [0.8301, 0.8786]                        |

Table 18: Amir's model [Ami14] with variability. Noise distribution  $f_\eta$ : centered Gaussian with standard deviation 0.10 truncated on  $[-h, h]$  for  $h = 0.10$ . Target increment of size:  $\Delta_0 = 3$ . Estimation of the Malthus parameter  $\lambda_{\Delta_0, f_\eta, \rho_\alpha}$  (mean and 95% confidence interval based on  $M = 50$  Monte Carlo continuous time trees simulated up to time  $T$ ) with respect to the coefficient of variation of the growth rates density  $\rho_\alpha$ . Reference: the value of reference is not significantly different from the value  $1 = \bar{v}$  (test at level 95% based on  $M = 50$  Monte Carlo continuous time trees simulated up to time  $T = 10$ ).

---

## Références du Chapitre IV

---

- [Ami14] A. Amir. *Cell size regulation in bacteria*. Physical Review Letters, 112 (2014), 208102.
- [BDMT11] V. Bansaye, J.-F. Delmas, L. Marsalle and V. C. Tran. *Limit theorems for Markov processes indexed by continuous time Galton-Watson trees*. The Annals of Applied Probability, 21 (2011), 2263–2314.
- [BCG13] D. Balagué, J. A. Cañizo and P. Gabriel, *Fine asymptotics of profiles and relaxation to equilibrium for growth-fragmentation equations with variable drift rates*. Kinetic Related Models, 6 (2013), 219–243.
- [BCP08] F. Bekkal Brikci, J. Clairambault and B. Perthame. *Analysis of a molecular structured population model with polynomial growth for the cell cycle*. Mathematical and Computer Modelling, 47 (2008), 699–713.
- [Bre10] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010.
- [CCM11] M. J. Cáceres, J. A. Canizo and S. Mischler. *Rate of convergence to an asymptotic profile for the self-similar fragmentation and growth-fragmentation equations*. Journal de mathématiques pures et appliquées, 96 (2011), 334–362.
- [CCDR15] M. J. Cáceres, J. Chevallier, M. Doumic and P. Reynaud-Bouret. *Microscopic approach of a time elapsed neural model*, to appear in Mathematical Models and Methods in Applied Sciences.
- [CDG12] V. Calvez, M. Doumic and P. Gabriel. *Self-similarity in a general aggregation–fragmentation problem. Application to fitness analysis*. Journal de mathématiques pures et appliquées, 98 (2012), 1–27.
- [Cha83] F. Chatelin. *Spectral Approximation of Linear Operator*. Academic Press, New York, 1983.
- [CL09] P. Ciarlet and E. Lunéville. *La méthode des éléments finis: de la théorie à la pratique. I. Concepts généraux*. Les presses de l’ENSTA, 2009.
- [CGL09] J. Clairambault, S. Gaubert and T. Lepoutre. *Comparison of Perron and Floquet eigenvalues in age structured cell division cycle models*. Mathematical Modelling of Natural Phenomena, 4 (2009), 183–209.
- [CGL11] J. Clairambault, S. Gaubert and T. Lepoutre. *Circadian rhythm and cell population growth*. Mathematical and Computer Modelling, 53 (2011), 1558–1567.
- [CGP07] J. Clairambault, S. Gaubert and B. Perthame. *An inequality for the Perron and Floquet eigenvalues of monotone differential systems and age structured equations*. Comptes Rendus Mathématique de l’Académie des Sciences Paris, 345 (2007), 549–554.
- [CMP06] J. Clairambault, P. Michel and B. Perthame. *Circadian rhythm and tumour growth*. Comptes Rendus Mathématique de l’Académie des Sciences Paris, 342 (2006), 17–22.
- [Clo11] B. Cloez. *Limit theorems for some branching measure-valued processes*, hal-00598030 (2011).
- [DL90] R. Dautray and J.-L. Lions. *Mathematical Analysis and Numerical Methods for Science and Technology*. Springer, Chapter VIII, 1990.

- [Dou07] M. Doumic. *Analysis of a population model structured by the cells molecular content*. Mathematical Modelling of Natural Phenomena, 2 (2007), 121–152.
- [DG10] M. Doumic and P. Gabriel. *Eigenelements of a general aggregation-fragmentation model*. Mathematical Models and Methods in Applied Sciences, 20 (2010), 757–783.
- [DT13] M. Doumic and L. M. Tine. *Estimating the Division Rate for the Growth-Fragmentation Equation*. Journal of Mathematical Biology, 67 (2013), 69–103.
- [DHKR15] M. Doumic, M. Hoffmann, N. Krell and L. Robert. *Statistical estimation of a growth-fragmentation model observed on a genealogical tree*. Bernoulli, 21 (2015), 1760–1799.
- [GL15] S. Gaubert and T. Lepoutre. *Discrete limit and monotonicity properties of the Floquet eigenvalue in an age structured cell division cycle model*. Journal of Mathematical Biology, Springer Berlin Heidelberg (2015), 1–41.
- [GHR04] E. Gobet, M. Hoffmann and M. Reiß. *Nonparametric estimation of scalar diffusions based on low frequency data*. The Annals of Statistics, 32 (2004), 2223–2253.
- [Guy07] J. Guyon. *Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging*. The Annals of Applied Probability, 17 (2007), 1538–1569.
- [HWG91] A. J. Hall, G. C. Wake and P. W. Gandar. *Steady size distributions for cells in one-dimensional plant tissues*. Journal of Mathematical Biology, 30 (1991), 101–123.
- [KNW+14] D. J. Kiviet, P. Nghe, N. Walker, S. Boulineau, V. Sunderlikova and S. J. Tans. *Stochasticity of metabolism and growth at the single-cell level*. Nature, 514 (2014), 376–379.
- [KT48] M. G. Krein and M. A. Rutman. *Linear operators leaving invariant a cone in a Banach space*. Uspehi Matem. Nauk (N. S.) (in Russian), 3 (1948), 1–95. English translation: American Mathematical Society Translations, 26 (1950).
- [LP09] P. Laurençot and B. Perthame. *Exponential decay for the growth-fragmentation/cell-division equations*. Communications in Mathematical Sciences, 7 (2009), 503–510.
- [LR74] J. L. Lebowitz and S. I. Rubinow. *A theory for the age and generation time distribution of a microbial population*. Journal of Mathematical Biology, 1 (1974), 17–36.
- [Lev02] R. J. Leveque, *Finite Volume Methods for Hyperbolic Problem*. Cambridge Textbooks in Applied Mathematics, 2002.
- [MHT66] A. G. Marr, R. J. Harvey and W. C. Trentini. *Growth and division of Escherichia coli*. Journal of Bacteriology, 91(1966), 2388–2389.
- [MD86] J. A. J. Metz and O. Dieckmann. *Formulating models for structured populations*. In *The dynamics of physiologically structured populations (Amsterdam, 1983)*, Lecture Notes in Biomathematics, 68 (1986), 78–135.
- [Mic06a] P. Michel. *Optimal proliferation rate in a cell division model*. Mathematical Modelling of Natural Phenomena, 1 (2006), 23–44.
- [Mic07b] P. Michel. *Existence of a solution to the cell division eigenproblem*. Mathematical Models and Methods in Applied Sciences, 16 (2006), 1125–1153.
- [MPR02] S. Mischler, B. Perthame and L. Ryzhik. *Stability in a nonlinear population maturation model*. Mathematical Models and Methods in Applied Sciences, 12 (2002), 1751–1772.
- [MS13] S. Mischler and J. Scher. *Spectral analysis of semigroups and growth-fragmentation equations*. 2013, to appear in Annales de l’Institut Henri Poincaré (C) Analyse Non Linéaire.
- [Mon42] J. Monod. *Recherche sur la croissance des cultures bactériennes*. Thèse, 1942.
- [ONC14] M. Osella, E. Nugent and M. Cosentino Lagomarsino. *Concerted control of Escherichia coli cell division*. PNAS, 111 (2014), 3431–3435.
- [PPS14] K. Pakdaman, B. Perthame and D. Salort. *Adaptation and fatigue model for neuron networks and large time asymptotics in a nonlinear fragmentation equation*. The Journal of Mathematical Neuroscience, (2014), 4–14

- 
- [Per07] B. Perthame. *Transport equations arising in biology*. Birkhäuser Frontiers in mathematics edition, 2007.
- [PR05] B. Perthame and L. Ryzhik. *Exponential decay for the fragmentation or cell-division equation*. Journal of Differential equations, 210 (2005), 155–177.
- [PZ07] B. Perthame and J. Zubelli. *On the inverse problem for a size-structured population model*, 23 (2007), 1037–1052.
- [RHK+14] L. Robert, M. Hoffmann, N. Krell, S. Aymerich, J. Robert and M. Doumic. *Division control in Escherichia Coli is based on a size-sensing rather than a timing mechanism*. BMC Biology, 12 (2014).
- [Rot83] M. Rotenberg. *Transport theory for growing cell populations*. Journal of Theoretical Biology, 103 (1983), 181–199.
- [SWHK62] M. Schaechter, J. P. Williamson, J. R. Hood Jun and A. L. Koch. *Growth, Cell and Nuclear Divisions in some Bacteria*, Microbiology, 29 (1962), 421–434.
- [SRBA14] I. Soifer, L. Robert, N. Barkai and A. Amir. *Single-cell analysis of growth in budding yeast and bacteria reveals a common size regulation strategy*, arXiv:1410.4771 (2014).
- [SMPT05] E. J. Stewart, R. Madden, G. Paul and F. Taddei. *Aging and death in an organism that reproduces by morphologically symmetric division*, PLoS Biology, 3 (2005), 295–300.
- [TAB+15] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola and S. Jun. *Cell-size control and homeostasis in bacteria*. Current Biology, 25 (2015).
- [vdV98] A. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.
- [WRP+10] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright and S. Jun. *Robust growth of Escherichia coli*. Current Biololy, 20 (2010), 1099–1103.



## Analyse statistique des modèles de croissance-fragmentation

**Résumé :** Cette étude théorique est pensée en lien étroit avec un champ d'application : il s'agit de modéliser la croissance d'une population de cellules qui se divisent selon un taux de division inconnu, fonction d'une variable dite structurante – l'âge et la taille des cellules étant les deux exemples paradigmatiques étudiés. Le champ mathématique afférent se situe à l'interface de la statistique des processus, de l'estimation non-paramétrique et de l'analyse des équations aux dérivées partielles. Les trois objectifs de ce travail sont les suivants : reconstruire le taux de division (fonction de l'âge ou de la taille) pour différents schémas d'observation (en temps généalogique ou en temps continu) ; étudier la transmission d'un trait biologique général d'une cellule à une autre et étudier le trait d'une cellule typique ; comparer la croissance de différentes populations de cellules à travers le paramètre de Malthus (après introduction de variabilité dans le taux de croissance par exemple).

**Mots-clés :** Processus de Markov déterministe par morceaux, Processus de branchement, Chaîne de Markov bifurquante, Processus auto-régressif bifurquant, Equation structurée en âge, Equation structurée en taille, Estimation du taux de division, Statistique non-paramétrique, Paramètre de Malthus.

## Statistical analysis of growth-fragmentation models

**Abstract:** This work is concerned with growth-fragmentation models, implemented for investigating the growth of a population of cells which divide according to an unknown splitting rate, depending on a structuring variable – age and size being the two paradigmatic examples. The mathematical framework includes statistics of processes, nonparametric estimations and analysis of partial differential equations. The three objectives of this work are the following: get a nonparametric estimate of the division rate (as a function of age or size) for different observation schemes (genealogical or continuous); to study the transmission of a biological feature from one cell to an other and study the feature of one typical cell; to compare different populations of cells through their Malthus parameter, which governs the global growth (when introducing variability in the growth rate among cells for instance).

**Keywords:** Piecewise deterministic Markov process, Branching process, Bifurcating Markov chain, Bifurcating autoregressive process, Age-structured equation, Size-structured equation, Splitting rate estimation, Nonparametric statistics, Malthus parameter.