



UNIVERSITÉ  
BOURGOGNE  
EUROPE



Institut de Mathématiques  
de Bourgogne - UMR CNRS 5584

ÉCOLE DOCTORALE  
CARNOT - PASTEUR

(ED553)

UNIVERSITÉ BOURGOGNE EUROPE  
Institut de Mathématiques de Bourgogne

ÉCOLE DOCTORALE CARNOT- PASTEUR

## HABILITATION À DIRIGER DES RECHERCHES

Spécialité : Mathématiques appliquées

Mention : Probabilités & Statistiques

Présentée et soutenue par  
Siméon Valère BITSEKI PENDA

### Théorie des chaînes de Markov bifurcantes ou l'histoire d'une compétition entre lois et généalogie

Rapporteurs : KOKONENDJI Celestin Professeur  
LACOUR Claire Professeur  
TRAN Viet-Chi Directeur de recherche

Soutenue publiquement le 26 Novembre 2025, devant le jury composé de :

Examinateurs :	BERCU Bernard	Professeur
	De SAPORTA Benoite	Professeur
	HERRMANN Samuel	Professeur
	KOKONENDJI Celestin	Professeur
	LAMBERT Amaury	Professeur
	LACOUR Claire	Professeur
	LÖCHERBACH Eva	Professeur
	TRAN Viet-Chi	Directeur de recherche
	GUILLIN Arnaud	Professeur



# Remerciements

J'aimerais profiter de cet espace pour remercier Bernard BERCU, Benoite De SAPORTA, Samuel HERRMANN, Celestin KOKONENDJI, Amaury LAMBERT, Claire LACOUR, Eva LÖCHERBACH, Viet-Chi TRAN et Arnaud GUILLIN qui me font l'honneur de participer au jury de soutenance de mon habilitation. J'aimerais en particulier exprimer ma gratitude à Claire, Celestin et Chi qui ont accepté de rapporter ce travail, ainsi qu'à Samuel qui s'est porté garant et a accepté de coordonner le jury.

Je remercie mes collaborateurs : Hacène, Marc, Adélaïde, Mickaël, Angelina, Frédéric, Vincent, Gorgui, Arnaud et Jean-François. Chaque séance de travail avec chacun de vous a toujours été un grand moment de partage. Votre disponibilité à chaque fois est admirable. Je pense avoir reçu de vous plus que ce que j'ai pu donner, je vous en remercie. Je remercie en particulier Jean-François avec qui j'ai travaillé durant toute la période du confinement

À Sylvie et Vincent, qui m'ont donné l'opportunité de travailler une année dans leur équipe. Je vous remercie car ce fut une année très intense où j'ai pu voir les différentes facettes de la recherche. Je remercie en particulier Vincent dont l'aide avait été très précieuse pour mon installation à Massy.

Arnaud, je te remercie pour ton soutien, constant et souvent très discret, qui m'accompagne depuis des années.

À tout ceux qui de près ou de loin ont contribué à mes travaux, je vous remercie.

À Paule, avec qui je partage les hauts et les bas, et en qui je trouve un soutien quotidien ; à nos magnifiques enfants Merici, Raphaële et Archange, je vous dédie ce travail . C'est vous au final qui donnez un sens à tout ceci.



# Liste des notations

$\mathbb{N}$	Ensemble des entiers naturels
$\mathbb{N}^*$	Ensemble des entiers naturels différents de 0
$\mathbb{R}$	Ensemble des nombres réels
$\mathbb{R}^*$	Ensemble des nombres réels différents de 0
$\mathbb{R}_+$	Ensemble des nombres réels positifs
$a \wedge b$	$\min\{a, b\}$ , pour deux réels $a$ et $b$
$a \vee b$	$\max\{a, b\}$ , pour deux réels $a$ et $b$
$a_+$	$\max\{a, 0\}$ , pour un réel $a$
$a_-$	$\min\{a, 0\}$ , pour un réel $a$
$\lfloor a \rfloor$	Partie entière du réel $a$
$\{a\}$	Partie fractionnaire du réel $a$ : $\{a\} = a - \lfloor a \rfloor$
$\mathring{D}$	Intérieur de l'ensemble $D$
$\bar{D}$	Adhérence de l'ensemble $D$
$ A $	Cardinal de $A$ , pour un ensemble fini $A$
$E^p$	$p$ -ième puissance cartésienne de l'ensemble $E$ , où $p \in \mathbb{N}^*$
$\mathcal{E}^{\otimes p}$	$\underbrace{\mathcal{E} \otimes \dots \otimes \mathcal{E}}_{p \text{ fois}}$ est la tribu produit de $p$ exemplaires de la tribu $\mathcal{E}$ , où $p \in \mathbb{N}^*$
$(E^p, \mathcal{E}^{\otimes p})$	Espace mesurable produit de l'espace mesurable $(E, \mathcal{E})$ , où $p \in \mathbb{N}^*$
$\mathcal{B}(E)$	Ensemble des fonctions réelles mesurables définies sur un espace mesurable $(E, \mathcal{E})$
$\mathcal{B}_+(E)$	Sous-ensemble de $\mathcal{B}(E)$ formé de fonctions positives
$\mathcal{B}_b(E)$	Sous-ensemble de $\mathcal{B}(E)$ formé des fonctions bornées
$f \otimes g$	Pour deux fonctions $f, g \in \mathcal{B}(E)$ , $(f \otimes g)(x, y) = f(x)g(y)$
$f \otimes_{\text{sym}} g$	Pour deux fonctions $f, g \in \mathcal{B}(E)$ , $f \otimes_{\text{sym}} g = \frac{1}{2}(f \otimes g + g \otimes f)$
$f \otimes^2$	Pour $f \in \mathcal{B}(E)$ , $f \otimes^2 = f \otimes f$
$\ \bullet\ _\infty$	Norme uniforme (pour $f \in \mathcal{B}(E)$ , $\ f\ _\infty = \sup\{ f(x) , x \in E\}$ )
$\langle \lambda, \bullet \rangle$	Pour une mesure finie $\lambda$ sur $(E, \mathcal{E})$ et $f \in \mathcal{B}(E)$ , $\langle \lambda, f \rangle = \int f(x)d\lambda(x)$
$L^p(\lambda)$	Pour $p \in \mathbb{N}^*$ , ensemble des fonctions $f \in \mathcal{B}(E)$ telles que $\langle \lambda,  f ^p \rangle < \infty$
$\ \bullet\ _{L^2(\lambda)}$	Pour une mesure finie $\lambda$ sur $(E, \mathcal{E})$ et $f \in \mathcal{B}(E)$ , $\ f\ _{L^2(\lambda)} = \langle \lambda, f^2 \rangle^{1/2}$
$\ \bullet\ _p$	Pour $f \in \mathcal{B}(\mathbb{R}^d)$ et $p, d \in \mathbb{N}^*$ , $\ f\ _p = (\int_{\mathbb{R}^d}  f(x) ^p dx)^{1/p}$
$f * g$	Pour deux fonctions intégrables $f, g \in \mathcal{B}(\mathbb{R}^d)$ , $f * g(x) = \int_{\mathbb{R}^d} f(x-t)g(t)dt$
$(\mathbf{x}_{-j}, t)$	Vecteur $\mathbf{x}$ où on a remplacé la $j^{\text{ème}}$ composante $x_j$ par $t$ , pour $\mathbf{x} \in \mathbb{R}^d$ et $t \in \mathbb{R}$

$\mathcal{H}_{\mathcal{X}}^{\beta}$	Espace des fonctions $\beta$ -holdériennes sur $\mathcal{X}$ , où $\beta = (\beta_1, \dots, \beta_d) \in ]0, +\infty[^d$ , $\mathcal{X} \subset \mathbb{R}^d$ et $d \in \mathbb{N}^*$ , i.e., $f \in \mathcal{H}_{\mathcal{X}}^{\beta}$ s'il existe une constante $c(f) > 0$ telle que pour $\mathbf{x} \in \mathbb{R}^d$ , $t \in \mathbb{R}$ et $j \in \{1, \dots, d\}$ ,
	$\left  \frac{\partial^{\lfloor \beta_j \rfloor} f}{\partial x_j^{\lfloor \beta_j \rfloor}}(\mathbf{x}_{-j}, t) - \frac{\partial^{\lfloor \beta_j \rfloor} f}{\partial x_j^{\lfloor \beta_j \rfloor}}(\mathbf{x}) \right  \leq c(f)  x_j - t ^{\{\beta_j\}}$
$ f _{\mathcal{H}_{\mathcal{X}}^{\beta}}$	Pour $f \in \mathcal{H}_{\mathcal{X}}^{\beta}$ , est la plus petite constante $c(f)$ telle que $\forall j \in \{1, \dots, d\}$ , $x \in \mathbb{R}^d$ , $t \in \mathbb{R}$ ,
	$\left  \frac{\partial^{\lfloor \beta_j \rfloor} f}{\partial x_j^{\lfloor \beta_j \rfloor}}(x_{-j}, t) - \frac{\partial^{\lfloor \beta_j \rfloor} f}{\partial x_j^{\lfloor \beta_j \rfloor}}(x) \right  \leq c(f)  x_j - t ^{\{\beta_j\}}$
$\ \bullet\ _{\mathcal{H}_{\mathcal{X}}^{\beta}}$	Norme sur $\mathcal{H}_{\mathcal{X}}^{\beta}$ définie par $\ f\ _{\mathcal{H}_{\mathcal{X}}^{\beta}} = \ f\ _{\infty} +  f _{\mathcal{H}_{\mathcal{X}}^{\beta}}$
$\mathcal{H}_{\mathcal{X}}^{\beta}(L)$	Pour $L > 0$ , $\mathcal{H}_{\mathcal{X}}^{\beta}(L) = \{f \in \mathcal{H}_{\mathcal{X}}^{\beta}; \ f\ _{\mathcal{H}_{\mathcal{X}}^{\beta}} \leq L\}$
$ \mathbf{x} $	Pour un vecteur $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ , $d \in \mathbb{N}^*$ , $ \mathbf{x}  = x_1 \times \dots \times x_d$

Si  $E$  est un espace topologique :

$\mathcal{B}_E$	Tribu borélienne sur $E$
$\mathcal{C}(E)$	Ensemble des fonctions réelles continues sur $E$
$\mathcal{C}_b(E)$	Ensemble des fonctions réelles continues bornées sur $E$
$\mathcal{C}_+(E)$	Ensemble des fonctions réelles continues positives sur $E$
$\nu \ll \mu$	Pour deux mesures de probabilités $\mu$ et $\nu$ sur $E$ , signifie que $\nu$ est absolument continue par rapport à $\mu$

Soit  $(x_n)$  et  $(y_n)$  deux suites de nombres réels.

$x_n = o(1)$	signifie que $x_n \rightarrow 0$ lorsque $n \rightarrow +\infty$
$x_n = o(y_n)$	si $x_n/y_n = o(1)$
$x_n \sim y_n$	si $x_n/y_n \rightarrow 1$ lorsque $n \rightarrow +\infty$
$x_n \lesssim y_n$	s'il existe une constante $c$ , indépendante de $n$ , telle que $x_n \leq cy_n$
$x_n \propto y_n$	s'il existe une constante $c$ , indépendante de $n$ , telle que $x_n = cy_n$
$x_n \asymp y_n$	s'il existe deux constantes $c_1$ et $c_2$ telles que $c_1 y_n \leq x_n \leq c_2 y_n$ lorsque $n \rightarrow +\infty$

$\xrightarrow[n \rightarrow \infty]{\mathbb{P}}$	Convergence en probabilité
$\xrightarrow[n \rightarrow \infty]{(d)}$	Convergence en loi
$\xrightarrow[n \rightarrow \infty]{\text{p.s.}}$	Convergence presque sûre



# Table des matières

<b>1</b>	<b>Introduction générale</b>	<b>5</b>
1.1	Modèle des chaînes de Markov bifurcantes . . . . .	5
1.1.1	L'arbre binaire associé aux chaînes de Markov bifurcantes . . . . .	5
1.1.2	Chaîne de Markov bifurcante . . . . .	6
1.1.3	Quantités utiles à l'études des chaînes de Markov bifurcantes . . . . .	7
1.1.4	Le processus mère-filles associé à une chaîne de Markov bifurcante .	8
1.1.5	Les many-to-one formulae . . . . .	9
1.2	Il était une fois . . . . .	9
1.3	... et aujourd'hui . . . . .	12
1.4	Mes motivations et mes travaux . . . . .	13
1.4.1	Moralité . . . . .	14
1.5	Pour l'avenir . . . . .	15
<b>2</b>	<b>Estimation à noyaux dans les modèles bifurcants autorégressifs non linéaires</b>	<b>16</b>
2.1	Introduction . . . . .	16
2.2	Estimations des fonctions d'autorégression . . . . .	17
2.3	Comportement non-asymptotique des estimateurs $\hat{f}_{\ell,n}$ . . . . .	18
2.3.1	Contraintes sur le modèle . . . . .	18
2.3.2	Bornes de convergence pour les estimateurs $\hat{f}_{\ell,n}$ . . . . .	19
2.4	Normalité asymptotique des estimateurs $\hat{f}_{\ell,n}$ . . . . .	21
2.5	Construction d'un test d'asymétrie . . . . .	23
2.6	Implémentation numérique . . . . .	25
2.6.1	Données simulées . . . . .	25
2.6.2	Données réelles . . . . .	29
2.7	Principe de déviations modérées pour les estimateurs $\hat{f}_{\ell,n}$ . . . . .	30
<b>3</b>	<b>Estimation non paramétrique adaptative de la densité invariante de la chaîne marquée</b>	<b>33</b>
3.1	Cadre général . . . . .	33
3.2	Estimation adaptative de la densité invariante par méthode d'ondelettes .	34
3.2.1	Décomposition en ondelettes et espaces de Besov . . . . .	34

3.2.2	Estimation de la densité invariante $\mu$ par ondelettes . . . . .	35
3.3	Selection locale de la fenêtre dans l'estimation à noyau de la densité invariante	36
3.4	Règle de selection de la fenêtre . . . . .	37
3.5	Quelques pistes pour les preuves . . . . .	38
<b>4</b>	<b>Inégalités de transport et de concentration pour les chaînes de Markov bifurcantes</b>	<b>40</b>
4.1	Inégalité de transport pour les chaînes de Markov bifurcantes . . . . .	41
4.2	Inégalités de concentration gaussiennes pour les chaînes de Markov bifurcantes	43
<b>5</b>	<b>Théorème central limite pour les chaînes de Markov bifurcantes</b>	<b>46</b>
5.1	Cadre général . . . . .	46
5.1.1	Hypothèses pour l'approche ponctuelle . . . . .	47
5.1.2	Hypothèses pour l'approche $L^2$ . . . . .	49
5.1.3	Quelques notations . . . . .	50
5.2	Les principaux résultats . . . . .	51
5.2.1	Cas sous-critique : $2\alpha^2 < 1$ . . . . .	51
5.2.2	Cas critique : $2\alpha^2 = 1$ . . . . .	53
5.2.3	Cas sur-critique : $2\alpha^2 > 1$ . . . . .	54
5.3	Extention au processus mère-filles . . . . .	56
5.3.1	Cas sous critique : $2\alpha^2 < 1$ . . . . .	56
5.3.2	Les cas critique et sur-critique : $2\alpha^2 = 1$ et $2\alpha^2 > 1$ . . . . .	58
5.3.3	Illustration numérique . . . . .	59
5.4	Application à l'étude des estimateurs à noyau de la densité invariante et de la densité de la probabilité de transition . . . . .	63
5.4.1	Définition des estimateurs de $\mu$ et $\mathcal{P}$ . . . . .	64
5.4.2	Principaux résultats . . . . .	65
5.4.3	Sélection de la fenêtre dans le cas particulier $S = \mathbb{R}$ . . . . .	68
<b>6</b>	<b>Principes de déviations modérées pour les chaînes de Markov bifurcantes</b>	<b>70</b>
6.1	Principe de déviation modérées pour les fonctionnelles additives de chaînes de Markov bifurcantes . . . . .	70
6.1.1	Principaux résultats . . . . .	71
6.1.2	Illustration numérique . . . . .	72
6.2	Application : Principe de déviations modérées pour l'estimateur à noyau de la densité invariante . . . . .	73
<b>7</b>	<b>Comportement asymptotique des grandes valeurs du processus bifurcant autorégressif</b>	<b>76</b>
7.1	Cadre général . . . . .	76
7.2	Grandes déviations pour le processus $Y$ . . . . .	78

7.2.1	Grandes déviations et chemin des grandes déviations dans le régime stable strict . . . . .	78
7.2.2	Grandes déviations dans le régime stable faible . . . . .	78
7.3	Densités locales pour le processus BAR . . . . .	79
	<b>Références</b>	<b>88</b>



# Chapitre 1

## Introduction générale

Je présente dans ce mémoire ma contribution à l'élaboration d'une théorie pour les chaînes de Markov bifurcantes. Ce travail, qui a débuté en Octobre 2010 dans le cadre de ma thèse de doctorat, n'a cessé au fil du temps d'aiguiser ma curiosité par la nature des problèmes qu'il soulève. Un phénomène curieux et particulièrement intéressant qui ressort des études est l'existence d'une transition de phase dans le comportement des fonctionnelles de chaînes de Markov bifurcantes. Ce phénomène, qui peut être interprété comme une compétition entre la loi du processus et la taille de la population, s'articule différemment selon le problème que l'on traite. À chaque fois, il se pose alors la question de savoir comment et à quel moment se manifeste cette transition. Ces dernières années, les différentes rencontres que j'ai faites ont permis d'apporter des réponses à un certains nombres de questions, favorisant ainsi des avancées majeures dans la théorie des chaînes de Markov bifurcantes. Avant d'entrer dans le vif du sujet, je commencerai par fixer les idées par une présentation du modèle.

### 1.1 Modèle des chaînes de Markov bifurcantes

#### 1.1.1 L'arbre binaire associé aux chaînes de Markov bifurcantes

On pose  $\mathbb{G}_0 = \{\emptyset\}$  et, pour tout  $m \in \mathbb{N}^*$ ,  $\mathbb{G}_m = \{0, 1\}^m$ , l'ensemble des suites de longueur  $m$  formées de 0 et de 1. On considère l'arbre binaire  $\mathbb{T} = \bigcup_{m=0}^{+\infty} \mathbb{G}_m$ . Pour  $u \in \mathbb{G}_m$ , on pose  $|u| = m$  et on considère les suites  $u0 = (u, 0)$  et  $u1 = (u, 1)$ . L'arbre  $\mathbb{T}$  peut-être vu comme une population (ou plus précisément comme un arbre généalogique) dans laquelle l'individu initial est  $\emptyset$ . Dans cette population, on a deux types d'individus : “0” et “1”. Chaque individu  $u$  donne naissance à deux enfants  $u0$  et  $u1$  de type “0” et “1” respectivement. Ainsi pour tout  $m \in \mathbb{N}$ ,  $\mathbb{G}_m$  peut-être vu comme la  $m$ -ième génération de la population  $\mathbb{T}$ . La génération à laquelle appartient un individu  $u \in \mathbb{T}$  est  $\mathbb{G}_{|u|}$ , où  $|u|$  est la longueur de la suite  $u$ . Pour deux entiers naturels  $n$  et  $p$  tels que  $p \leq n$ ,  $\mathbb{T}_n$  désigne les  $n + 1$  premières générations de la population :  $\mathbb{T}_n = \bigcup_{m=0}^n \mathbb{G}_m$  et  $\mathbb{T}_{p,n}$  désigne la sous-population formée des individus de la génération  $\mathbb{G}_p$  à la génération  $\mathbb{G}_n$  :  $\mathbb{T}_{p,n} = \bigcup_{m=p}^n \mathbb{G}_m$ .

Notons qu'on a en particulier  $\mathbb{T}_n = \mathbb{T}_{0,n}$ .

Pour  $u \in \mathbb{T}$  et  $A \subset \mathbb{T}$ , on pose  $uA = \{uv, v \in A\}$ , où  $uv$  est la concaténation des suites  $u$  et  $v$ , avec la convention  $u\emptyset = \emptyset u = u$ . Soit  $i, j \in \mathbb{T}$ . On écrit  $i \preccurlyeq j$  si  $j \in i\mathbb{T}$ . On désigne par  $i \wedge j$  le plus jeune ancêtre commun de  $i$  et  $j$ , *i.e.*, l'unique individu  $u \in \mathbb{T}$  tel que si  $v \in \mathbb{T}$  et  $v \preccurlyeq i$ ,  $v \preccurlyeq j$  alors  $v \preccurlyeq u$ . On définit également l'ordre lexicographique  $i \leq j$  si soit  $i \preccurlyeq j$  soit  $v0 \preccurlyeq i$  et  $v1 \preccurlyeq j$  pour  $v = i \wedge j$ .

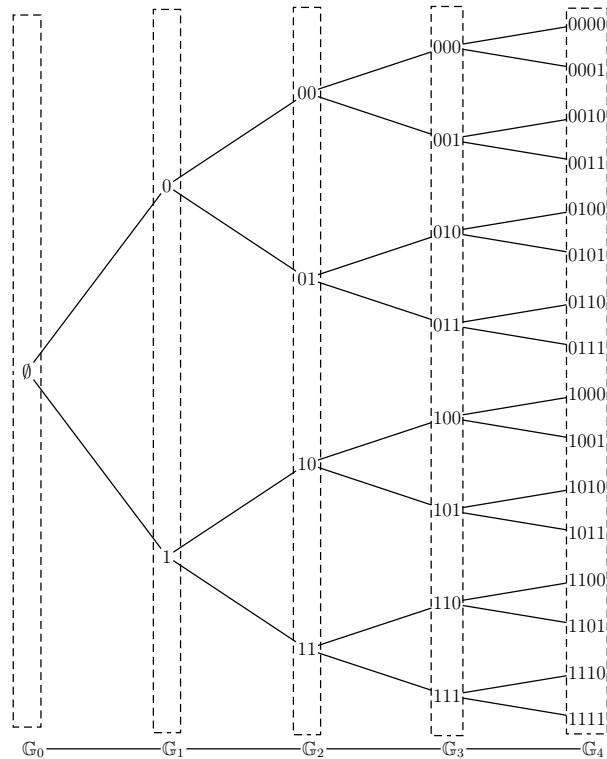


FIGURE 1.1 – L'arbre binaire jusqu'à la quatrième génération

### 1.1.2 Chaîne de Markov bifurcante

Soit  $(S, \mathcal{S})$  un espace mesurable et  $\mathcal{P}$  une probabilité de transition sur  $(S, \mathcal{S}^{\otimes 2})$ , *i.e.*

- $\mathcal{P}(x, \cdot)$  est une mesure de probabilité sur  $(S^2, \mathcal{S}^{\otimes 2})$ ;
- $\mathcal{P}(\cdot, A)$  est une fonction mesurable pour tout  $A \in \mathcal{S}^{\otimes 2}$ .

Soit  $\nu$  une mesure de probabilité sur  $(S, \mathcal{S})$  et  $(\Omega, \mathcal{F}, (\mathcal{F}_m)_{m \in \mathbb{N}}, \mathbb{P})$  un espace de probabilité filtré. La définition suivante a été proposée par Guyon [62].

**Définition 1.1.** *Un processus stochastique  $X = (X_u)_{u \in \mathbb{T}}$ , définie sur  $(\Omega, \mathcal{F}, (\mathcal{F}_m)_{m \in \mathbb{N}}, \mathbb{P})$  et à valeurs dans  $S$ , est une chaîne de Markov bifurcante de mesure initiale  $\nu$  et de probabilité de transition  $\mathcal{P}$  si :*

- $X_\emptyset$  est distribué selon  $\nu$  ;
- $X_u$  est  $\mathcal{F}_{|u|}$ -mesurable pour tout  $u \in \mathbb{T}$  ;
- pour chaque  $m \geq 0$  et pour toute suite  $(g_u)_{u \in \mathbb{G}_m}$  de fonctions mesurables bornées sur  $S^3$ ,

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

où  $\mathcal{P}g(x) = \int_{S^2} g(x, y, z) \mathcal{P}(x, dy dz)$  désigne l'action de  $\mathcal{P}$  sur  $g$ .

**Remarque 1.1.** *Tout au long de ce mémoire,  $\mathbb{P}_\nu$  désignera la loi du processus  $(X_u, u \in \mathbb{T})$  et  $\mathbb{E}_\nu[\cdot]$  désignera l'espérance par rapport à la probabilité  $\mathbb{P}_\nu$ .*

**Remarque 1.2.** *Typiquement, la troisième propriété de la Définition 1.1, que nous appellerons dans la suite “propriété de branchement”, nous dit que conditionnellement à leurs parents, les différentes fratries d'une génération sont indépendantes. Ainsi, même si à nos débuts nous l'avions dit et écrit, voir les chaînes de Markov bifurcantes comme une adaptation des chaînes de Markov sur un arbre n'est certainement pas l'explication la plus efficiente. Il serait plutôt “judicieux” de dire que les chaînes de Markov bifurcantes sont une autre extension des suites de variables aléatoires i.i.d. Je souligne que ce “nouveau” regard sur les chaînes de Markov bifurcantes nous a permis dernièrement d'élaborer des stratégies de preuves plus adaptées qui ont permis de faire évoluer la théorie.*

**Remarque 1.3.** *Le processus  $(X_u, u \in \mathbb{T})$  peut-être vu comme les données de la population  $\mathbb{T}$ . Pour l'individu  $u \in \mathbb{T}$ , la quantité d'intérêt  $X_u$  qui lui est associée est parfois nommée “trait de  $u$ ”. J'utiliserai régulièrement ce terme dans la suite.*

### 1.1.3 Quantités utiles à l'études des chaînes de Markov bifurcantes

Soit  $X = (X_u)_{u \in \mathbb{T}}$  une chaîne de Markov bifurcante à valeurs dans l'espace mesurable  $(S, \mathcal{S})$ , de mesure initiale  $\nu$  et de probabilité de transition  $\mathcal{P}$ . Pour  $u \in \mathbb{T}$ , on désigne par  $X_u^\Delta = (X_u, X_{u0}, X_{u1})$  le triangle mère-filles. Soit  $\mathbb{A} \subset \mathbb{T}$ . On considère les sommes suivantes :

$$\begin{cases} M_{\mathbb{A}}(f) = \sum_{u \in \mathbb{A}} f(X_u) & \text{si } f \in \mathcal{B}(S) \\ M_{\mathbb{A}}(f) = \sum_{u \in \mathbb{A}} f(X_u^\Delta) & \text{si } f \in \mathcal{B}(S^3). \end{cases} \quad (1.1)$$

Dans les études, les cas  $\mathbb{A} = \mathbb{G}_n$  et  $\mathbb{A} = \mathbb{T}_n$ ,  $n \in \mathbb{N}^*$ , sont souvent fréquents. Soulignons que

$$|\mathbb{G}_n| = 2^n \quad \text{et} \quad |\mathbb{T}_n| = 2^{n+1} - 1.$$

On désigne par  $\mathcal{P}_0$  et  $\mathcal{P}_1$  la première et la seconde marginale de  $\mathcal{P}$ , i.e., pour tout  $x \in S$  et  $A \in \mathcal{S}$ ,

$$\mathcal{P}_0(x, A) = \mathcal{P}(x, A \times S) \quad \text{et} \quad \mathcal{P}_1(x, A) = \mathcal{P}(x, S \times A).$$

On pose

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

Alors  $Q$  est une probabilité de transition sur  $(S, \mathcal{S})$ , *i.e.*,

- $\mathcal{Q}(x, \cdot)$  est une mesure de probabilité sur  $(S, \mathcal{S})$  ;
- $\mathcal{Q}(\cdot, A)$  est une fonction mesurable pour tout  $A \in \mathcal{S}$ .

On désigne par  $(Y_m)_{m \in \mathbb{N}}$  la chaîne de Markov de probabilité de transition  $\mathcal{Q}$ . Cette chaîne de Markov correspond aux données d'une lignée aléatoire prise dans la population. Elle peut en effet être définie récursivement comme suit. On se place initialement sur le sommet  $\emptyset$  et on pose  $Y_0 = X_\emptyset$ . Pour tout  $m \in \mathbb{N}$ , si on se trouve sur le sommet  $u \in \mathbb{G}_m$ , alors on choisit l'un des sommet  $u\eta$ ,  $\eta \in \{0, 1\}$ , avec probabilité  $1/2$  et on pose  $Y_{m+1} = X_{u\eta}$ . Dans la suite, nous parlerons de "chaîne marquée" pour désigner cette chaîne de Markov. La chaîne marquée joue un rôle fondamental dans la théorie des chaînes de Markov bifurcantes, en particulier pour l'analyse asymptotique et non-asymptotique des fonctionnelles additives (voir par exemple [23, 62]). Dans [62], Guyon a montré que si la chaîne marquée  $(Y_m)_{m \in \mathbb{N}}$  est ergodique, de mesure invariante  $\mu$ , alors pour une fonction test  $f$  et pour  $\mathbb{A} \in \{\mathbb{G}_n, \mathbb{T}_n\}$ , on a les convergences presque sûres suivantes :

$$\begin{cases} \lim_{n \rightarrow +\infty} |\mathbb{A}|^{-1} M_{\mathbb{A}}(f) = \langle \mu, f \rangle & \text{si } f \text{ est définie sur } S \\ \lim_{n \rightarrow +\infty} |\mathbb{A}|^{-1} M_{\mathbb{A}}(f) = \langle \mu, \mathcal{P}f \rangle & \text{si } f \text{ est définie sur } S^3. \end{cases}$$

On peut ainsi voir que la stabilité d'un modèle de Markov bifurcant est fortement liée à celle de la chaîne marquée associée. En fait, nous dirons qu'un modèle de Markov bifurcant est stable si la chaîne marquée associée est stable.

#### 1.1.4 Le processus mère-filles associé à une chaîne de Markov bifurcante

Considérons une chaîne de Markov bifurcante  $X = (X_u)_{u \in \mathbb{T}}$  comme dans le paragraphe précédent. Pour  $u \in \mathbb{T}$ , nous rappelons le triangle mère-filles  $X_u^\Delta = (X_u, X_{u0}, X_{u1})$ . Alors on vérifie que le processus  $X^\Delta = (X_u^\Delta, u \in \mathbb{T})$  est une chaîne de Markov bifurcante sur  $S^3$  de probabilité de transition  $\mathcal{P}^\Delta$  définie par

$$\mathcal{P}^\Delta(xx_0x_1, dy_0y_1, dz_0z_1) = \delta_{x_0}(dy) \mathcal{P}(y, dy_0, dy_1) \delta_{x_1}(dz) \mathcal{P}(z, dz_0, dz_1),$$

où on pose  $xx_0x_1 = (x, x_0, x_1)$ . La probabilité de transition  $\mathcal{Q}^\Delta$ , de la chaîne marquée associée à  $X^\Delta$ , est définie par

$$\mathcal{Q}^\Delta(xx_0x_1, dy_0y_1) = \frac{1}{2}(\delta_{x_0}(dy) + \delta_{x_1}(dy)) \mathcal{P}(y, dy_0, dy_1).$$

Si on suppose que la chaîne marquée est ergodique et si on désigne par  $\mu$  la mesure de probabilité invariante de sa probabilité de transition  $\mathcal{Q}$ , alors le processus mère-filles est stable et la mesure de probabilité invariante  $\mu^\Delta$  de  $\mathcal{Q}^\Delta$  est définie sur  $S^3$  par

$$\mu^\Delta(dx_0x_1) = \mu(dx) \mathcal{P}(x, dx_0, dx_1). \quad (1.2)$$

**Remarque 1.4.** *L'utilisation par nous du "caractère Markovien" du processus  $X^\Delta$  est très récente (voir par exemple [17]). Le choix de faire apparaître ce processus dans cette introduction vient du fait que, à mon avis, c'est un élément fondamental de la théorie des chaînes de Markov bifurcantes.*

### 1.1.5 Les many-to-one formulae

Ce sont certainement l'un des outils les plus indispensables à l'analyse des chaînes de Markov bifurcantes. En effet, à ma connaissance, excepté les études faisant intervenir les inégalités de concentration, toutes les autres études dans la théorie des chaînes de Markov bifurcantes font appels à ces formules d'une manière ou d'une autre. Grossièrement, l'idée derrière les many-to-one formulae est de dire que, pour une fonction  $f$  définie sur  $S$  et deux entiers naturels  $n$  et  $p$  tels que  $n$  est très grand devant  $p$ , le moment d'ordre  $p$  de  $M_{\mathbb{G}_n}(f)$  est égal à l'espérance de  $|\mathbb{G}_n|^p f(X_{I_1}) \dots f(X_{I_p})$ , où  $I_1, \dots, I_p$  sont  $p$  individus choisis uniformément avec remise dans la génération  $\mathbb{G}_n$ . Ensuite, à l'aide des techniques combinatoires, on est capable d'obtenir des expressions exactes de ces moments. Je renvoie à [23, 62] pour le calcul des moments d'ordre 1 à 4. L'importance des many-to-one formulae réside dans le fait que sous certaines hypothèses, notamment l'ergodicité géométrique de la chaîne marquée, elles permettent de contrôler les moments des fonctionnelles additives de chaînes de Markov bifurcantes (voir par exemple [23]). Ces contrôles, comme nous le verrons dans la suite, sont souvent des outils essentiels pour l'obtention des principaux résultats.

## 1.2 Il était une fois ...

Les modèles de chaînes de Markov bifurcantes ont été introduits dans littérature pour comprendre les mécanismes de la division cellulaire chez *Escherichia Coli* (*E. Coli* en abrégé). *E. Coli* est une bactérie en forme de tige qui se reproduit en se divisant en deux, produisant ainsi deux filles : une qui possède l'extrémité la plus récente de la mère et que l'on appelle fille de type "0" et l'autre qui possède l'extrémité la plus ancienne de la mère et que l'on appelle fille de type "1". L'âge de chaque cellule est alors donné par l'âge de son ancienne extrémité au sens du nombre de divisions depuis lesquelles cette extrémité existe (voir Figure 1.2).

Le premier modèle de chaîne de Markov bifurcante, appelé processus bifurcant autorégressif "linéaire" symétrique (BAR symétrique, en abrégé), a été introduit par Cowan et Staudte [39] dans le but d'analyser les données issues de la division cellulaire de *E. Coli*. Ce modèle est défini comme suit. On tire le trait  $X_\emptyset$ , de la cellule initiale, suivant une mesure de probabilité  $\nu$  sur  $\mathbb{R}$  et pour chaque cellule  $u \in \mathbb{T}$ , les traits  $X_{u0}$  et  $X_{u1}$  de ses filles sont liés au sien par les équations d'autorégression

$$X_{u0} = aX_u + \varepsilon_{u0} \quad \text{et} \quad X_{u1} = aX_u + \varepsilon_{u1}, \quad (1.3)$$

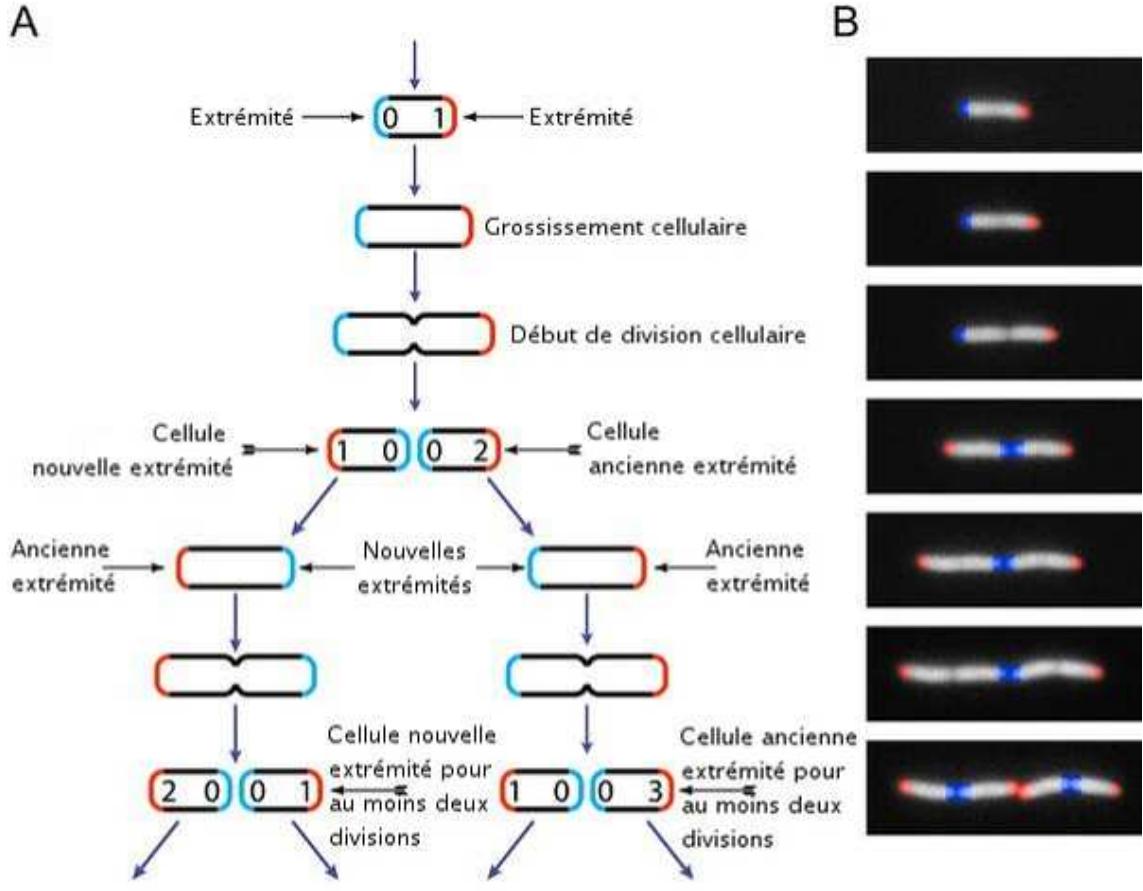


FIGURE 1.2 – Division cellulaire de E.Coli

où  $a \in [-1, 1]$  et  $((\varepsilon_{u0}, \varepsilon_{u1}), u \in \mathbb{T})$  forme une suite de couples aléatoires i.i.d., de loi normale bidimensionnelle de moyenne  $(b, b)$  et de matrice de variance-covariance

$$\Gamma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

avec  $\sigma^2 > 0$  et  $|\rho| < 1$ . Notons que le processus BAR symétrique défini précédemment est une chaîne de Markov bifurcante de mesure initiale  $\nu$  et de probabilité de transition

$$\mathcal{P}(x, dy, dz) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2\sigma^2(1-\rho^2)}g(x, y, z)\right) dy dz,$$

avec

$$g(x, y, z) = (y - ax - b)^2 - 2\rho(y - ax - b)(z - ax - b) + (z - ax - b)^2.$$

Après ces travaux de Cowan et Staudte, le processus BAR a été intensivement étudié et plusieurs extensions ont été proposées. On peut citer entre autre les travaux de Basawa et Zhou [7] où, à notre connaissance, le terme chaîne de Markov bifurcante a été utilisé pour la première fois. Mais, c'est Guyon [62] qui a jeté les bases d'une théorie pour les chaînes de Markov bifurcantes afin d'étudier rigoureusement la question de l'asymétrie dans la division des cellules. En effet, une des questions fondamentales que soulève le mode de reproduction des cellules est celle de savoir si oui ou non la division des cellules est symétrique. Par exemple :

- à la naissance les soeurs ont-elles la même taille ?, le même poids ? ... *etc*
- les nutriments de la mère se partagent-ils de façon égale aux deux soeurs ?
- le cycle cellulaire est-il le même pour deux soeurs ?

Plus généralement, la polarisation des cellules est un processus fondamental dans la biologie cellulaire. L'asymétrie lors de la division cellulaire a été observée et étudiée dans différents contextes, voir par exemple [73] pour les plasmides, [98] pour l'ADN extrachromosomique, [70] pour les mitochondries et [5] pour les parasites. Afin de résoudre le problème de l'asymétrie dans la division des cellules, Guyon & *al* [63] et ensuite Guyon [62] ont introduit une généralisation du modèle de Cowan et Staudte, que nous appelons processus BAR asymétrique définie comme suit. On tire le trait initial  $X_\emptyset$  suivant une mesure de probabilité  $\nu$  sur  $\mathbb{R}$  et pour tout  $u \in \mathbb{T}$  :

$$\begin{cases} X_{u0} = a_0 X_u + b_0 + \varepsilon_{u0} \\ X_{u1} = a_1 X_u + b_1 + \varepsilon_{u1}, \end{cases} \quad (1.4)$$

avec  $a_0, a_1 \in (-1, 1)$ ,  $b_0, b_1 \in \mathbb{R}$  et  $((\varepsilon_{u0}, \varepsilon_{u1}), u \in \mathbb{T})$  une suite de couples aléatoires Gaussiens bidimensionnels  $\mathcal{N}(0, \Gamma)$  indépendants et indépendants de  $X_\emptyset$ , de matrice de variance-covariance

$$\Gamma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

où  $\sigma > 0$  et  $|\rho| < 1$ . Alors, le processus  $X = (X_u, u \in \mathbb{T})$  est une chaîne de Markov bifurcante de mesure initiale  $\nu$  et de probabilité de transition  $\mathcal{P}$  donnée par :

$$\mathcal{P}(x, dy, dz) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2\sigma^2(1-\rho^2)} g(x, y, z)\right) dy dz,$$

avec

$$g(x, y, z) = (y - a_0 x - b_0)^2 - 2\rho(y - a_0 x - b_0)(z - a_1 x - b_1) + (z - a_1 x - b_1)^2.$$

La probabilité de transition de la chaîne marquée est  $\mathcal{Q}$  définie par :

$$\mathcal{Q}(x, dy) = \frac{1}{2\sqrt{2\pi\sigma^2}} \left( e^{-(y-a_0 x-b_0)^2/2\sigma^2} + e^{-(y-a_1 x-b_1)^2/2\sigma^2} \right) dy.$$

En utilisant alors sa théorie sur les chaînes de Markov bifurcantes, Guyon a pu mettre en évidence le fait qu'il existe bien une asymétrie dans la division cellulaire. En particulier, des preuves statistiques ont été apportées sur le fait qu'une cellule qui vieillit a tendance à se multiplier plus lentement qu'une jeune cellule. Les travaux de Guyon ont été généralisés plus tard par Delmas et Marsalle [44] et ensuite par De Saporta, Gegout-Petit et Marsalle [42]. Dans ces travaux, afin de prendre en compte le fait que certains traits ne peuvent-être mesurés, elles (et il) utilisent des arbres de Galton-Watson "binaires" pour modéliser les généalogies. Il a été ainsi introduit le concept de chaînes de Markov bifurcantes sur un arbre de Galton-Watson. Je souligne qu'au cours de ces dernières années, je n'ai pas poussé les recherches vers cette direction. En effet, puisque mon objectif était de comprendre les mécanismes autour des chaînes de Markov bifurcantes, j'ai fait le choix délibéré de me concentrer sur les processus indexés par des arbres binaires réguliers, réservant pour plus tard des extensions aux arbres de Galton-Watson.

À la suite des travaux de Guyon, nous avons continué à enrichir la théorie avec l'étude des inégalités de concentration et du principe de déviations modérées pour les fonctionnelles additives de chaînes de Markov bifurcantes [22, 23, 15]. Nous avons en particulier étudié les inégalités de déviations et le principe de déviations modérées pour les estimateurs des paramètres associés au modèle BAR. Dans ces travaux effectués durant ma thèse de doctorat, nous avons vu apparaître, pour la première fois dans les modèles de Markov bifurcants, un comportement qui s'apparente à "une compétition" et dont l'arbitraire est le taux d'ergodicité géométrique de la chaîne marquée. Ce phénomène nous a semblé d'autant plus curieux qu'il n'apparaissait pas dans les travaux antérieurs, notamment ceux de Guyon, Delmas et Marsalle, De Saporta & *al.* Des extensions du modèle BAR ont été intensivement étudiées dans un cadre non-Markovien [8, 9, 10, 30, 40]. Dans ces travaux, les auteurs ont principalement pris avantage de la structure autoregressive du modèle.

En plus de la question sur l'asymétrie de la division cellulaire, la question de l'estimation du taux de division des cellules a été posée. Cette question a fait l'objet de nombreuses études en statistique. À ce sujet, on peut citer par exemple les travaux de Doumic & *al.* [53] où les données de la population  $\mathbb{T}$  et la théorie des CMBs ont été utilisées pour construire et étudier un estimateur à noyaux pour le taux de division des cellules. Ce travail de Doumic & *al.* est à ma connaissance le premier qui traite de l'estimation non-paramétrique dans les modèles de Markov bifurcants. Nous nous en sommes d'ailleurs beaucoup inspirés ces dernières années.

### 1.3 ... et aujourd'hui

L'engouement autour de la division cellulaire reste toujours grande et les modèles de Markov bifurcants sont très souvent utilisés pour expliquer les mécanismes de cette division. De nouvelles questions émergent, ouvrant ainsi des pistes nouvelles pour la recherche. Sans être exhaustif, je citerai par exemple les travaux de Hughes & *al* [69] où les auteurs s'intéressent aux facteurs qui influencent le cycle cellulaire, *i.e.* le processus conduisant

à la division cellulaire. Ce travail donne des motivations pour l'ouverture d'un nouveau champ dans l'inférence des modèles de Markov bifurcants à savoir "les chaînes de Markov bifurcantes cachées". Je pourrai également citer les travaux de Marzena & *al.* [84] où les auteurs s'intéressent à l'impact de la dynamique du cycle cellulaire sur le comportement à long terme de la population étudiée.

## 1.4 Mes motivations et mes travaux

Plusieurs raisons m'ont convaincu depuis la fin de ma thèse de doctorat de continuer la recherche sur les chaînes de Markov bifurcantes.

Tout d'abord, les différentes études qui avaient été menées depuis les travaux de Cowan et Staudte [39] n'avaient traité que de l'estimation des paramètres réels dans les modèles BAR. De plus, les différentes extensions qui avaient été proposées supposaient toujours une relation "linéaire" entre les traits des parents et ceux des enfants. Il se posa alors la question de considérer des modèles qui pourraient prendre en compte des relations plus complexes entre le trait de la mère et ceux des filles. C'est donc ainsi que nous avons introduit la notion de processus bifurcant autorégressif non-linéaire (NBAR, en abrégé). L'étude de ce modèle a fait l'objet de deux articles publiés en collaboration avec Adélaïde Olivier [27, 28].

Il ressort de la théorie introduite par Guyon [62] que l'étude des chaînes de Markov bifurcantes et par conséquent celle des modèles de Markov bifurcants est fortement liée à la connaissance de la mesure invariante de la chaîne marquée et de la probabilité de transition de la chaîne de Markov bifurcante. Toutefois, il se trouve que ces deux quantités sont inconnues, de telle sorte que leur estimation à partir des données est d'une importance fondamentale. En effet, toute étude, ou presque, sur les modèles de Markov bifurcants requiert leur connaissance. C'est ainsi qu'au cours de ces dernières années, je me suis intéressé à l'inférence statistique pour les modèles de Markov bifurcants. En prenant en compte la structure des données, nous avons développé des méthodes permettant d'estimer les paramètres fonctionnels des modèles de Markov bifurcants [26, 27, 29].

L'étude des estimateurs que nous avons développé a parfois nécessité d'établir des inégalités de concentration pour les fonctionnelles additives de chaînes de Markov bifurcantes. Ce travail a été fait dans [23, 26] sous des conditions d'ergodicité géométrique uniforme de la chaîne marquée. Il se posa alors la question de l'obtention des inégalités de concentration pour les chaînes de Markov bifurcantes sous des conditions plus "faibles". C'est ainsi que dans [24], nous avons étudié les inégalités de transport pour les chaînes de Markov bifurcantes, ce qui nous a permis d'obtenir des inégalités de concentration sous des conditions plus "faibles". Toutefois, les inégalités que nous avons obtenu en utilisant la théorie du transport optimal ne sont que partiellement satisfaisantes. La question de l'obtention des inégalités de concentration pour les chaînes de Markov bifurcantes reste donc ouverte.

Dans [26, 27, 29], nous avions travaillé sous une hypothèse d'ergodicité géométrique

de la chaîne marquée. En plus nous avions supposé que le taux d'ergodicité géométrique est plus petit que  $1/2$ . La question qui se posa par la suite fut naturellement celle de savoir ce qui se passe lorsque le taux d'ergodicité géométrique de la chaîne marquée est plus grand que  $1/2$ . Cette question est d'autant plus pertinente que dans [23], nous avions déjà observé que dans certains cas particuliers, pour le principe de déviations modérées des fonctionnelles additives de chaînes de Markov bifurcantes, la vitesse est fonction du taux d'ergodicité géométrique de la chaîne marquée. Il fut donc question pour nous de comprendre le rôle joué par ce taux dans la détermination de la vitesse de convergence des estimateurs. Pour répondre à cette question, avec Jean-François Delmas, nous choisissons d'étudier premièrement les fluctuations des fonctionnelles additives des chaînes de Markov bifurcantes [20, 19]. Je souligne que les deux travaux précédents sont une avancée majeure dans la théorie des chaînes de Markov bifurcantes. Ceci nous permit alors dans [21] de proposer un théorème central limite pour l'estimateur à noyau de la densité invariante de la chaîne marquée. Plus tard, avec Gorgui Gackou [25], nous complétâmes en partie les travaux de [23] en étudiant le principe de déviations modérées pour les fonctionnelles additives de chaînes de Markov bifurcantes. La compléction venant du fait que dans [25], nous considérons les fonctions qui dépendent d'une seule variable. Je compléta également plus tard les travaux de [21] en étudiant le principe de déviations modérées pour l'estimateur à noyau de la chaîne marquée [16]. J'étendis ensuite dans [17] les résultats de [20] au processus mère-filles, ce qui me permit de faire une étude asymptotique de l'estimateur à noyau de la densité de la transition d'une chaîne de Markov bifurcante [18].

Le comportement en temps long des individus ayant des traits extrêmes, *i.e.* très grands ou très petits, attira également notre attention. C'est ainsi qu'avec Vincent Bansaye [6], nous nous intéressâmes au comportement asymptotique du nombre de cellules d'une génération dont les traits sont très grands ou très petits dans le modèle BAR linéaire Gaussien. Mathématiquement, ceci nous conduisit à l'étude des grandes déviations pour les variables explicatives dans un modèle autorégressif à coefficients aléatoires.

#### 1.4.1 Moralité

Il ressort de tous ces travaux que l'influence du taux d'ergodicité géométrique de la chaîne marquée sur : le comportement asymptotique et non-asymptotique des fonctionnelles de chaînes de Markov bifurcantes, le choix et le comportement des estimateurs des paramètres, n'est pas un artefact dû aux calculs ou à un déficit de techniques de calcul. Cette influence est liée à la structure du modèle qui impose une compétition entre la loi du processus et le taux de reproduction de la population. Cette compétition fait ainsi apparaître une transition de phase autour d'un seuil critique du taux d'ergodicité géométrique de la chaîne marquée. Il ressort aussi que le seuil critique n'est pas universel puisqu'il varie en fonction du problème que l'on traite et des fonctions qui sont utilisées. Toutefois, dans le cadre de l'estimation non-paramétrique, et selon que les inégalités de concentration interviennent ou non dans les études, nous avons vu apparaître régulièrement deux valeurs critiques,  $1/2$  et  $1/\sqrt{2}$ .

## 1.5 Pour l'avenir

L'étude des processus empiriques pour les chaînes de Markov bifurcantes et leurs applications à l'obtention et à l'étude des estimateurs robustes des paramètres dans les modèles de Markov bifurcants est un challenge qui mérite d'être mené. Cette étude est d'autant importante qu'elle facilitera à terme la résolution d'un problème qui me tient particulièrement à cœur : "la construction d'un test d'asymétrie en se basant uniquement sur les estimateurs des transitions marginales  $\mathcal{P}_0$  et  $\mathcal{P}_1$  de la transition  $\mathcal{P}$ ". Je souligne que j'ai commencé en partie l'étude des processus empiriques dans le but de construire et d'étudier les estimateurs des moindres carrés de la densité de transition pour les modèles de Markov bifurcants. Un article est actuellement en cours de rédaction.

Avec Marc Hoffmann et Angelina Roche, nous avons initier une étude qui nous permettra à terme, d'une part d'obtenir, pour toutes les valeurs du taux d'ergodicité géométrique, des bornes minimax pour les estimateurs à noyaux des fonctions d'intérêts dans les modèles de Markov bifurcants. L'autre objectif de cette étude est de construire des estimateurs minimax adaptatifs pour ces fonctions d'intérêts. La particularité ici est que nous considérons désormais le taux d'ergodicité géométrique de la chaîne marquée comme un paramètre de régularité, au même titre que la régularité des fonctions à estimer.

Comme je l'ai précisé précédemment, les travaux de Hughes & *al* [69] nous donne des motivations pour l'étude des "chaînes de Markov bifurcantes cachées". En s'inspirant par exemple des travaux de Lacour [75], je pense qu'il est tout à fait possible de mener à bien cette étude.

Je note que les études et les projets précédents sont très liés au fait que la chaîne marquée est géométriquement ergodique. Une question naturelle qui se pose est celle de savoir ce que deviendraient nos résultats dans les cas où la convergence de la chaîne marquée vers l'équilibre est sous-géométrique. La compétition que nous avons observé dans le cas géométrique reste t-elle toujours valable dans ce cas ? Si oui comment s'exprime t-elle ? Cette question, à mon avis, peut-être le point de départ d'un projet de recherche passionnant.

J'ai en projet plusieurs sujets qui pourraient faire l'objet d'une thèse. Dans l'un des plus récents, il est question de s'intéresser à divers méthodes d'estimation dans les modèles de Markov bifurcants, ainsi que dans les modèles en finance. Ceci me permettra notamment de m'ouvrir sérieusement à l'étude d'autres modèles dans l'avenir.

À présent, la suite de ce mémoire est consacrée à une synthèse de mes travaux. Chaque grand thème fait l'objet d'un chapitre.

## Chapitre 2

# Estimation à noyaux dans les modèles bifurcants autorégressifs non linéaires

Ce chapitre est issu des articles [27, 28] écrit en collaboration avec Adélaïde Olivier. Il était question pour nous de généraliser le modèle de Cowan et Staudte en proposant un modèle prenant en compte des relations plus complexes entre le trait de la mère et ceux des filles.

### 2.1 Introduction

Nous travaillons dans un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ . Pour une mère  $u \in \mathbb{T}$ , nous supposons que le trait  $X_u$  est à valeurs dans  $\mathbb{R}$  muni de sa tribu borélienne  $\mathcal{B}_{\mathbb{R}}$ . Nous supposons que les traits  $X_{u0}$  et  $X_{u1}$  des filles ne dépendent du passé qu'à travers le trait  $X_u$  de la mère via un système d'équations d'autorégression. On a alors la définition suivante.

**Définition 2.1** (Processus NBAR). *Soit  $\nu$  une mesure de probabilité sur  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ ,  $g_\varepsilon$  la densité d'une mesure de probabilité sur  $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}})$  et  $f_0, f_1 : \mathbb{R} \rightarrow \mathbb{R}$  deux fonctions. On dit que  $(X_u, u \in \mathbb{T})$  est un processus bifurcant autorégressif non linéaire (NBAR, en abrégé) de mesure initiale  $\nu$  et de fonctions d'autorégression  $f_0$  et  $f_1$  si*

- $X_\emptyset$  est distribué selon  $\nu$  ;
- Pour chaque  $u \in \mathbb{T}$ ,

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

où  $((\varepsilon_{u0}, \varepsilon_{u1}), u \in \mathbb{T})$  est une suite de couples aléatoires i.i.d. centrés de densité commune  $g_\varepsilon$ .

**Remarque 2.1.** Une observation fondamentale dans les études qui seront menées dans la suite est que le processus  $(X_u, u \in \mathbb{T})$  ainsi défini est une chaîne de Markov bifurcante de mesure initiale  $\nu$  et de probabilité de transition  $\mathcal{P}$  définie par

$$\mathcal{P}(x, dy, dz) = g_\varepsilon(y - f_0(x), z - f_1(x)) dy dz.$$

La chaîne marquée  $(Y_m)_{m \in \mathbb{N}}$  est le processus autorégressif non linéaire défini comme suit :

$$\begin{cases} Y_0 = X_\emptyset \\ Y_m = f_{\iota_m}(Y_{m-1}) + \varepsilon'_m, \quad \forall m \geq 1, \end{cases}$$

où  $(\iota_m)_{m \geq 1}$  est une suite de variables aléatoires i.i.d. de loi de Bernoulli de paramètre 1/2 et  $(\varepsilon'_m)_{m \geq 1}$  est une suite de variables aléatoires i.i.d., indépendantes de  $(X_u, u \in \mathbb{T})$  et distribuées selon une loi ayant pour densité  $(g_0 + g_1)/2$ , où les densités marginales  $g_0$  et  $g_1$  sont définies par :

$$g_0(\cdot) = \int_{\mathbb{R}} g_\varepsilon(\cdot, y) dy \quad \text{et} \quad g_1(\cdot) = \int_{\mathbb{R}} g_\varepsilon(x, \cdot) dx.$$

La probabilité de transition  $\mathcal{Q}$  de la chaîne de Markov  $(Y_m)_{m \in \mathbb{N}}$  est définie explicitement par

$$Q(x, dy) = \frac{1}{2} (g_0(y - f_0(x)) + g_1(y - f_1(x))) dy.$$

## 2.2 Estimations des fonctions d'autorégression

Vu le rôle central que joue les fonctions  $f_0$  et  $f_1$  dans les modèles NBAR, une question naturelle qui se pose est celle de leur estimation à partir des données issues de l'arbre binaire. Pour cela, en nous inspirant des travaux qui avaient été faits pour les modèles de régressions et les séries temporelles autorégressifs, (voir par exemple [67, 85, 96, 104]), nous faisons usage des estimateurs de type Nadaraya-Watson, généralisant ainsi l'usage de cette méthode dans le cadre des modèles autorégressifs indexés par un arbre binaire.

Soit  $D \subset \mathbb{R}$  un intervalle compact. Alors, pour tout  $x \in D$ , nous proposons d'estimer  $(f_0(x), f_1(x))$  à partir des observations  $(X_u, u \in \mathbb{T}_{n+1})$  par

$$\left( \widehat{f}_{\iota, n}(x) = \frac{|\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) X_u}{\left( |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) \right) \vee \varpi_n}, \iota \in \{0, 1\} \right), \quad (2.1)$$

où  $\iota \in \{0, 1\}$ ,  $\varpi_n > 0$  et nous posons  $K_{h_n}(\cdot) = h_n^{-1} K(h_n^{-1} \cdot)$  pour  $h_n > 0$  et le noyau  $K : \mathbb{R} \rightarrow \mathbb{R}$  tel que  $\int_{\mathbb{R}} K = 1$ .

Ce travail a deux objectifs principaux. Premièrement, faire une étude asymptotique et non-asymptotique des estimateurs définis dans (2.1). Ensuite proposer un test qui nous permettra de dire si le phénomène étudié est symétrique ou non. L'étude non-asymptotique comprend principalement le contrôle du risque quadratique au sens minimax. L'étude asymptotique comprend la convergence presque sûre, la normalité asymptotique et le principe de déviations modérées.

## 2.3 Comportement non-asymptotique des estimateurs $\hat{f}_{\iota,n}$

### 2.3.1 Contraintes sur le modèle

Nous travaillerons sous les contraintes suivantes qui assurent la stabilité du modèle.

- Soit  $\gamma \in (0, 1)$  et  $\ell > 0$ . On suppose que les fonctions d'autorégressions  $f_0$  et  $f_1$  appartiennent à la classe  $\mathcal{F}(\gamma, \ell)$  des fonctions continues  $f : \mathbb{R} \rightarrow \mathbb{R}$  telles que

$$|f(x)| \leq \gamma|x| + \ell \quad \forall x \in \mathbb{R}.$$

- Soit  $r > 0$  et  $\lambda > 2$ . On suppose que les densités marginales  $g_0$  et  $g_1$  appartiennent à la classe  $\mathcal{G}(r, \lambda)$  des fonctions positives continues  $g : \mathbb{R} \rightarrow [0, +\infty)$  telles que

$$g(x) \leq \frac{1}{1 + |x|^\lambda} \quad \text{pour } |x| \geq r.$$

**Remarque 2.2.** *Lorsque  $\lambda > 3$ , la matrice de variance-covariance  $\Gamma$  de  $(\varepsilon_0, \varepsilon_1)$ , appelée matrice de covariance du bruit, est bien définie et sera donnée par*

$$\Gamma = \begin{pmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix}, \quad (2.2)$$

avec  $\sigma_0, \sigma_1 > 0$  and  $\rho \in (-1, 1)$ .

- Pour chaque  $m > 0$ , nous posons

$$\delta(m) = \min \left\{ \inf_{|x| \leq m} g_0(x); \inf_{|x| \leq m} g_1(x) \right\} \quad \text{et} \quad M_0 = \ell + \mathbb{E}[|\varepsilon'_1|] < \infty, \quad (2.3)$$

avec  $\varepsilon'_1$  une variable aléatoire distribuée selon une loi de densité  $(g_0 + g_1)/2$ . Alors on suppose qu'il existe  $M_1 > 2M_0/(1 - \gamma)$  tel que  $2M_1\delta((1 + \gamma)M_1 + \ell) > 0$ .

Notons que les trois contraintes ci-dessus sont suffisantes pour montrer l'ergodicité géométrique de la chaîne marquée. On pourra par exemple se référer aux travaux [3, 11, 35] où l'ergodicité géométrique est étudiée pour les processus autorégressifs non linéaires. Toutefois, dans notre cas, nous aurons besoin d'un contrôle plus fin du taux d'ergodicité géométrique de la chaîne marquée qui doit être plus petit que  $1/2$ . Dans ce contexte, les travaux cités précédemment deviennent insuffisants et des conditions supplémentaires sont nécessaires. Nous faisons alors les hypothèses suivantes.

**Hypothèse 2.1.** *Il existe  $\eta > 0$  tel que*

$$\gamma < \frac{1}{2} - \eta$$

*et il existe  $M_1 > 2M_0/(1/2 - \eta - \gamma)$  tel que*

$$2M_1\delta((1 + \gamma)M_1 + \ell) > \frac{1}{2}.$$

Pour chaque  $M > 0$ , posons

$$\eta(M) = \frac{\|g_0\|_\infty + \|g_1\|_\infty}{2} \int_{|y|>M} \int_{x \in \mathbb{R}} \frac{r}{1 + |y - \gamma|x| - \ell|^\lambda \wedge |y + \gamma|x| + \ell|^\lambda} dx dy.$$

**Hypothèse 2.2.** Pour  $M_2 > 0$  tel que  $\eta(M_2) < 1$ , il existe  $M_3 > \ell + \gamma M_2$  tel que  $\delta(M_3) > 0$ .

L'hypothèse 2.1 garantit que la chaîne marquée  $(Y_m)_{m \in \mathbb{N}}$  est géométriquement ergodique et que le taux d'ergodicité géométrique  $\alpha$  est strictement plus petit que  $1/2$ . Plus précisément, nous avons le résultat suivant où la fonction  $V$  est définie sur  $\mathbb{R}$  par  $V(x) = 1 + |x|$ .

**Lemme 2.1** (Ergodicité géométrique de la chaîne marquée). Soit  $\gamma \in (0, 1)$ ,  $\ell > 0$ ,  $r > 0$  et  $\lambda > 2$ . Alors, pour chaque  $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$  et pour chaque  $g_\varepsilon$  tel que  $(g_0, g_1) \in \mathcal{G}(r, \gamma)^2$ , il existe une unique mesure de probabilité  $\mu$  de la forme  $\mu(dx) = \mu(x)dx$  sur  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ , une constante  $R > 0$  et  $\alpha \in (0, 1)$  telles que

$$\sup_{(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2} \sup_{|g| \leq V} |Q^m g(x) - \langle \mu, g \rangle| \leq R V(x) \alpha^m, \quad x \in \mathbb{R}, \quad m \geq 0.$$

De plus, sous l'hypothèse 2.1, on peut choisir  $\alpha \in (0, 1/2)$ .

La preuve du Lemme 2.1 repose essentiellement sur les travaux de Hairer et Mattingly [65]. Nous renvoyons à [27], Lemme 13, pour plus de détails. Quant à l'hypothèse 2.2, elle assure que la densité invariante  $\mu$  est positive sur un intervalle non vide. Plus précisément, nous avons le résultat suivant dont nous renvoyons à [27], Lemme 17, pour plus de détails.

**Lemme 2.2.** Soit  $\gamma \in (0, 1/2)$ ,  $\ell > 0$ ,  $r > 0$  et  $\lambda > 3$ . Pour chaque  $g_\varepsilon$  tel que  $(g_0, g_1) \in \mathcal{G}(r, \lambda)^2$  satisfait les Hypothèses 2.1 et 2.2, il existe  $d = d(\gamma, \ell, r, \lambda, \|g_0\|_\infty, \|g_1\|_\infty) > 0$  tel que pour chaque  $D \subset [-d, d]$ ,

$$\inf_{(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2} \inf_{x \in D} \mu(x) > 0.$$

### 2.3.2 Bornes de convergence pour les estimateurs $\hat{f}_{\ell, n}$

Nous allons à présent donner les résultats principaux de cette section. Avant cela, commençons par compléter le cadre de ces résultats. Nous faisons l'hypothèse suivante sur le noyau.

**Hypothèse 2.3.** Le noyau  $K : \mathbb{R} \rightarrow \mathbb{R}$  est une fonction bornée à support compact et pour un entier naturel  $n_0 \geq 1$ , on a  $\int_{-\infty}^{\infty} x^k K(x) dx = \mathbf{1}_{\{k=0\}}$  pour  $k = 1, \dots, n_0$ .

**Remarque 2.3.** Nous renvoyons à Tsybakov [101], Chapitre 1, pour des exemples et la construction des noyaux vérifiants l'Hypothèse 2.3.

**Théorème 2.1** (Borne supérieure de convergence). *Soit  $\gamma \in (0, 1/2)$ ,  $\ell > 0$ ,  $r > 0$  et  $\lambda > 3$ . On considère les estimateurs  $(\hat{f}_{0,n}, \hat{f}_{1,n})$  définis en (2.1) avec un noyau  $K$  vérifiant l’Hypothèse 2.3 pour un certain  $n_0 > 0$ ,*

$$h_n \propto |\mathbb{T}_n|^{-1/(2\beta+1)}$$

*et  $\varpi_n > 0$  tel que  $\varpi_n \rightarrow 0$  lorsque  $n \rightarrow \infty$ . Pour chaque  $L, L' > 0$  et  $0 < \beta < n_0$ , pour chaque  $g_\varepsilon$  tel que  $(g_0, g_1) \in (\mathcal{G}(r, \lambda) \cap \mathcal{H}_\mathbb{R}^\beta(L'))^2$  satisfont les Hypothèses 2.1 et 2.2, il existe  $d = d(\gamma, \ell, g_0, g_1) > 0$  tel que pour chaque intervalle compact  $D \subset [-d, d]$  tel que  $\mathring{D} \neq \emptyset$ , pour chaque  $x \in \mathring{D}$  et pour toute mesure de probabilité initiale  $\nu(dx)$  sur  $\mathbb{R}$  telle que  $\nu((1 + |\cdot|)^2) < \infty$ ,*

$$\sup_{(f_0, f_1) \in (\mathcal{F}(\gamma, \ell) \cap \mathcal{H}_D^\beta(L))^2} \mathbb{E}_\nu \left[ (\hat{f}_{0,n}(x) - f_0(x))^2 + (\hat{f}_{1,n}(x) - f_1(x))^2 \right]^{1/2} \lesssim \varpi_n^{-1} |\mathbb{T}_n|^{-\beta/(2\beta+1)}.$$

**Remarque 2.4.** En se basant sur le Lemme 2.2, il est possible de choisir  $n$  assez grand tel que le seuil  $\varpi_n$  vérifie

$$0 < \varpi_n \leq \frac{1}{2} \inf_{(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2} \inf_{x \in D} \mu(x).$$

En fait, le choix de  $\varpi_n$  doit être fait de telle sorte qu’il majore la borne supérieure de convergence uniquement d’un facteur lent. Ainsi, dans la pratique, le choix  $\varpi_n = \log(n)^{-1}$  est pertinent.

Principales idées de la preuve (voir [27], paragraphe 6.3, pour plus de détails). soit  $x \in \mathring{D}$ . Pour  $\iota \in \{0, 1\}$ , on a la décomposition suivante :

$$\hat{f}_{\iota,n}(x) - f_\iota(x) = \frac{M_{\iota,n}(x)}{\hat{\mu}_n(x) \vee \varpi_n} + \frac{L_{\iota,n}(x) - (\mu f_\iota)(x)}{\hat{\mu}_n(x) \vee \varpi_n} - \frac{\hat{\mu}_n(x) \vee \varpi_n - \mu(x)}{\hat{\mu}_n(x) \vee \varpi_n} f_\iota(x), \quad (2.4)$$

avec

$$M_{\iota,n}(x) = \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) \varepsilon_{u\iota}, \quad (2.5)$$

$$L_{\iota,n}(x) = \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) f_\iota(X_u),$$

$$\hat{\mu}_n(x) = \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u), \quad (2.6)$$

Ainsi, en utilisant le fait que  $f_\iota$  est bornée sur  $D$ , uniformément sur la classe  $\mathcal{F}(\gamma, \ell)$ , on obtient

$$\begin{aligned} \mathbb{E}_\nu \left[ (\hat{f}_{\iota,n}(x) - f_\iota(x))^2 \right] &\lesssim \varpi_n^{-2} \left( \mathbb{E}_\mu \left[ (M_{\iota,n}(x))^2 \right] + \mathbb{E}_\nu \left[ (L_{\iota,n}(x) - (\mu f_\iota)(x))^2 \right] \right. \\ &\quad \left. + \mathbb{E}_\nu \left[ (\hat{\mu}_n(x) \vee \varpi_n - \mu(x))^2 \right] \right). \end{aligned} \quad (2.7)$$

On termine alors la preuve en contrôlant successivement les trois termes du membre de gauche de (2.7). Pour cela on fait usage des many-to-one formulae et du Lemme 2.2.  $\square$

**Remarque 2.5.** Notons que  $\widehat{\mu}(x)$  défini en (2.6) est l'estimateur de la densité invariante de la chaîne marquée. Dans le contexte du Théorème 2.1, on montre que

$$\sup_{(f_0, f_1) \in (\mathcal{F}(\gamma, \ell) \cap \mathcal{H}_D^\beta(L))^2} \mathbb{E}_\nu \left[ (\widehat{\mu}_n(x) - \mu(x))^2 \right] \lesssim |\mathbb{T}_n|^{\frac{-2\beta}{2\beta+1}}.$$

Nous renvoyons à [27], paragraphe 6.2 pour plus de détails. Nous reviendrons dans les chapitres suivants sur l'étude de cet estimateur dans un cadre plus général.

Le Théorème 2.1 est complété avec le résultat suivant.

**Théorème 2.2** (Borne minimax de convergence). *On suppose que  $g_\varepsilon$  est la densité d'une loi normale bivariée centrée. Soit  $D \subset \mathbb{R}$  un intervalle compact. Pour chaque  $\gamma \in (0, 1)$ ,  $\ell > 0$ ,  $\beta > 0$  et  $L > 0$ , il existe  $C > 0$  tel que, pour chaque  $x \in D$ ,*

$$\liminf_{n \rightarrow \infty} \inf_{(\widehat{f}_{0,n}, \widehat{f}_{1,n})} \sup_{(f_0, f_1)} \mathbb{P} \left( |\mathbb{T}_n|^{\beta/(2\beta+1)} (|\widehat{f}_{0,n}(x) - f_0(x)| + |\widehat{f}_{1,n}(x) - f_1(x)|) \geq C \right) > 0,$$

où l'inf est pris sur l'ensemble de tous les estimateurs basés sur les observations  $(X_u, u \in \mathbb{T}_{n+1})$  et le sup est pris sur l'ensemble de toutes les fonctions  $(f_0, f_1) \in (\mathcal{F}(\gamma, \ell) \cap \mathcal{H}_D^\beta(L))^2$ .

*Idées de la preuve.* La preuve du Théorème 2.2 utilise l'approche basée sur deux hypothèses (voir par exemple [101], paragraphe 2.5), ainsi que les many-to-one formulae. Pour plus détails, nous renvoyons à [27], paragraphe 6.4.  $\square$

## 2.4 Normalité asymptotique des estimateurs $\widehat{f}_{\ell,n}$

Tout d'abord, notons que les estimateurs  $\widehat{f}_{\ell,n}$  sont fortement consistants, comme nous le montre le résultat suivant.

**Proposition 2.1** (Convergence presque sûre). *Dans le même cadre que le Théorème 2.1 avec  $h_n \propto |\mathbb{T}_n|^{-\gamma}$  pour  $\gamma \in (0, 1)$ , on a*

$$\begin{pmatrix} \widehat{f}_{0,n}(x) \\ \widehat{f}_{1,n}(x) \end{pmatrix} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \begin{pmatrix} f_0(x) \\ f_1(x) \end{pmatrix},$$

lorsque  $n \rightarrow +\infty$ .

*Idées de la preuve.* La preuve de ce résultat repose principalement sur l'égalité suivante

$$\widehat{f}_{\ell,n}(x) = \frac{M_{\ell,n}(x) + L_{\ell,n}(x)}{\widehat{\nu}_n(x) \vee \varpi_n},$$

où  $M_{\ell,n}(x)$ ,  $L_{\ell,n}(x)$  et  $\widehat{\nu}_n(x)$  ont été définis en (2.5)-(2.6). Nous renvoyons à [27], paragraphe 6.5 pour plus de détails.  $\square$

Pour l'étude de la normalité asymptotique de nos estimateurs, nous avons besoin des hypothèses supplémentaires sur le bruit. Plus précisément, nous supposons que  $\lambda > 5$ , ce qui assurera que  $\mathbb{E}[\varepsilon_0^4 + \varepsilon_1^4] < \infty$ . Nous avons alors le résultat suivant.

**Théorème 2.3** (Normalité asymptotique sans l'expression du biais). *Dans le même cadre que le Théorème 2.1 avec  $\lambda > 5$  et  $h_n \propto |\mathbb{T}_n|^{-\gamma}$  pour  $\gamma \in (1/(2\beta + 1), 1)$ , on a*

$$\sqrt{|\mathbb{T}_n| h_n} \begin{pmatrix} \widehat{f}_{0,n}(x) - f_0(x) \\ \widehat{f}_{1,n}(x) - f_1(x) \end{pmatrix} \xrightarrow[n \rightarrow \infty]{(d)} \mathcal{N}_2(\mathbf{0}_2, \boldsymbol{\Sigma}_2(x)) \quad \text{avec} \quad \boldsymbol{\Sigma}_2(x) = |K|_2^2 (\mu(x))^{-1} \Gamma,$$

où  $\Gamma$  désigne la matrice de covariance du bruit définie en (2.2) et  $\mathbf{0}_2 = (0, 0)$ . En outre, pour  $k$  points distincts  $x_1, \dots, x_k$  dans  $D$ , les couples

$$\left( \begin{pmatrix} \widehat{f}_{0,n}(x_\ell) - f_0(x_\ell) \\ \widehat{f}_{1,n}(x_\ell) - f_1(x_\ell) \end{pmatrix}, \ell = 1, \dots, k \right)$$

sont asymptotiquement indépendants.

**Remarque 2.6.** *Même si le Théorème 2.3 nous fournit la convergence vers une Gaus-sienne centrée, ce qui est pratique pour les tests d'hypothèses, il souffre du fait que la fenêtre optimale est exclue dans le choix des estimateurs. Le théorème qui suit va résoudre cette “imperfection”, mais le pris à payer est la présence d'un biais.*

**Théorème 2.4** (Normalité asymptotique avec l'expression du biais). *Dans le même cadre que le Théorème 2.1 avec  $\lambda > 5$  et  $\beta$  un entier naturel,*

(i) *Si  $h_n^\beta \sqrt{|\mathbb{T}_n| h_n} \rightarrow \kappa$  avec  $\kappa \in [0, \infty)$  lorsque  $n \rightarrow +\infty$ , alors*

$$\sqrt{|\mathbb{T}_n| h_n} \begin{pmatrix} \widehat{f}_{0,n}(x) - f_0(x) \\ \widehat{f}_{1,n}(x) - f_1(x) \end{pmatrix} \xrightarrow[n \rightarrow \infty]{(d)} \mathcal{N}_2(\kappa \mathbf{m}_2(x), \boldsymbol{\Sigma}_2(x)) \quad \text{avec} \quad \boldsymbol{\Sigma}_2(x) = |K|_2^2 (\mu(x))^{-1} \Gamma$$

et

$$\mathbf{m}_2(x) = \frac{(-1)^\beta}{\beta! \mu(x)} \int_{\mathbb{R}} y^\beta K(y) dy \begin{pmatrix} (\mu f_0)^\beta(x) - \mu^\beta(x) f_0(x) \\ (\mu f_1)^\beta(x) - \mu^\beta(x) f_1(x) \end{pmatrix}.$$

(ii) *Si  $h_n^\beta \sqrt{|\mathbb{T}_n| h_n} \rightarrow +\infty$  lorsque  $n \rightarrow +\infty$ , alors*

$$h_n^{-\beta} \begin{pmatrix} \widehat{f}_{0,n}(x) - f_0(x) \\ \widehat{f}_{1,n}(x) - f_1(x) \end{pmatrix} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mathbf{m}_2(x).$$

**Remarque 2.7.** *Nous soulignons que ce résultat s'étend au cas des  $\beta$  non entiers en utilisant les dérivées fractionnaires. Notons toutefois que la définition de ces dérivées n'est pas unique (voir par exemple [100, 102]). Nous faisons le choix de ne pas utiliser cette notion ici afin d'alléger l'exposé.*

*Idées de preuve des Théorèmes 2.3 et 2.4 (pour plus de détails, voir [27], paragraphe 6.6).*  
Soit  $x \in \mathring{D}$ . La preuve utilise la décomposition suivante :

$$a_n \begin{pmatrix} \widehat{f}_{0,n}(x) - f_0(x) \\ \widehat{f}_{1,n}(x) - f_1(x) \end{pmatrix} = \frac{1}{\widehat{\mu}_n(x) \vee \varpi_n} \left\{ a_n \begin{pmatrix} M_{0,n}(x) \\ M_{1,n}(x) \end{pmatrix} \right. \\ \left. + a_n \begin{pmatrix} N_{0,n}(x) \\ N_{1,n}(x) \end{pmatrix} + a_n \begin{pmatrix} R_{0,n}(x) \\ R_{1,n}(x) \end{pmatrix} \right\}, \quad (2.8)$$

où,  $a_n = \sqrt{|\mathbb{T}_n| h_n}$ , pour  $\iota \in \{0, 1\}$ ,  $M_{\iota,n}(x)$  est défini en (2.5),

$$N_{\iota,n}(x) = \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) (f_\iota(X_u) - f_\iota(x)), \\ R_{\iota,n}(x) = \left( \widehat{\mu}_n(x) - (\widehat{\mu}_n(x) \vee \varpi_n) \right) f_\iota(x),$$

et  $\widehat{\mu}_n(x)$  est défini en (2.6). La stratégie est alors la suivante. On montre premièrement que les deux derniers termes du membre de droite de 2.8 convergent presque sûrement vers 0 (pour le Théorème 2.3) et, pour le Théorème 2.4, que le deuxième terme du membre de droite de (2.8) converge presque sûrement vers  $\kappa \mathbf{m}(x) \mathbf{1}_{\{\kappa < +\infty\}} + \mathbf{m}(x) \mathbf{1}_{\{\kappa = \infty\}}$ . Ceci se fait en utilisant les many-to-one formulae, le développement de Taylor et le Lemme de Borel-Cantelli. Dans un deuxième temps, on montre que

$$\sqrt{|\mathbb{T}_n| h_n} \begin{pmatrix} M_{0,n}(x) \\ M_{1,n}(x) \end{pmatrix}$$

est une martingale par rapport à une filtration engendrée par une transformation du processus mère-filles. On fait alors usage du théorème central limite pour les martingales (voir par exemple [54], paragraphe 2.1.3 et 2.1.4). On obtient la normalité asymptotique voulue en remarquant que  $\widehat{\mu}_n(x) \vee \varpi_n$  converge presque sûrement vers  $\mu(x)$  lorsque  $n$  tend vers l'infini. L'indépendance asymptotique évoquée dans le Théorème 2.3 est une conséquence des propriétés des lois gaussiennes.  $\square$

## 2.5 Construction d'un test d'asymétrie

Nous allons à présent construire un test permettant de décider si oui ou non les fonctions d'autorégression  $f_0$  et  $f_1$  sont égales. Pour cela, nous choisissons de discriminer entre les hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$  définies comme suit. Soit  $x_1, \dots, x_k$  des points distincts dans  $\mathbb{R}$ .

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

Dans les études paramétriques, ce type de test est souvent appelé test du vieillissement cellulaire. En effet, il a été utilisé pour mettre en évidence le fait qu'une cellule qui vieillit a tendance à moins croître et par conséquent se reproduit plus lentement. D'où le nom

de “test du vieillissement cellulaire” qui lui a été attribué, voir par exemple [62]. Afin de construire notre statistique de test, nous avons besoin de construire un estimateur convergeant en loi vers une gaussienne centrée à la vitesse optimale. Pour cela, nous nous inspirons des travaux de Bierens [12]. Soit  $(\widehat{f}_{\iota,n}^{(a)}(x), \iota \in \{0, 1\})$  les estimateurs (2.1) avec la fenêtre  $h_n^{(a)} \propto |\mathbb{T}_n|^{-1/(2\beta+1)}$  et soit  $(\widehat{f}_{\iota,n}^{(b)}(x), \iota \in \{0, 1\})$  les estimateurs (2.1) avec la fenêtre  $h_n^{(b)} \propto |\mathbb{T}_n|^{-\delta/(2\beta+1)}$  pour un certain  $\delta \in (0, 1)$ . Nous définissons

$$\left( \bar{f}_{\iota,n}(x) = \left(1 - |\mathbb{T}_n|^{-\frac{\beta(1-\delta)}{2\beta+1}}\right)^{-1} \left( \widehat{f}_{\iota,n}^{(a)}(x) - |\mathbb{T}_n|^{-\frac{\beta(1-\delta)}{2\beta+1}} \widehat{f}_{\iota,n}^{(b)}(x) \right), \iota \in \{0, 1\} \right). \quad (2.9)$$

Nous avons le résultat suivant.

**Corollaire 2.1.** *Dans le même cadre que le Théorème 2.4,*

$$|\mathbb{T}_n|^{\frac{\beta}{2\beta+1}} \begin{pmatrix} \bar{f}_{0,n}(x) - f_0(x) \\ \bar{f}_{1,n}(x) - f_1(x) \end{pmatrix} \xrightarrow[n \rightarrow \infty]{(d)} \mathcal{N}_2(\mathbf{0}_2, \boldsymbol{\Sigma}_2(x)) \quad \text{avec} \quad \boldsymbol{\Sigma}_2(x) = |K|_2^2 (\mu(x))^{-1} \Gamma,$$

pour chaque  $\delta \in (0, 1)$  dans la définition (2.9) de  $(\bar{f}_{0,n}(x), \bar{f}_{1,n}(x))$ .

Motivé par le Corollaire 2.1, nous utilisons les nouveaux estimateurs  $(\bar{f}_{0,n}(x), \bar{f}_{1,n}(x))$  pour construire la statistique de test  $W_n(x_1, \dots, x_k)$  comme suit.

$$W_n(x_1, \dots, x_k) = \frac{|\mathbb{T}_n|^{\frac{2\beta}{2\beta+1}}}{(\sigma_0^2 + \sigma_1^2 - 2\sigma_0\sigma_1\rho)|K|_2^2} \sum_{\ell=1}^k \widehat{\mu}_n(x_\ell) (\bar{f}_{0,n}(x_\ell) - \bar{f}_{1,n}(x_\ell))^2, \quad (2.10)$$

avec  $\widehat{\mu}_n(\cdot) = |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} K_{h_n}(X_u - \cdot)$ , où  $h_n \propto |\mathbb{T}_n|^{-1/(2\beta+1)}$ . Le résultat qui suit est alors une conséquence des résultats sur la normalité asymptotique que nous avons précédemment établis.

**Théorème 2.5** (Test d’asymétrie de Wald). *Dans le même cadre que le Théorème 2.4, soit  $x_1, \dots, x_k$  des points distincts dans  $D$ . Alors, sous  $\mathcal{H}_0$ , la statistique de test  $W_n(x_1, \dots, x_k)$  converge en loi vers une variable aléatoire du chi-deux à  $k$  degrés de libertés, et sous l’alternative  $\mathcal{H}_1$ , la statistique  $W_n(x_1, \dots, x_k)$  converge presque sûrement vers  $+\infty$ .*

*Idées de preuve du Corollaire 2.1 et du Théorème 2.5.* Le corollaire 2.1 s’obtient en appliquant le Théorème 2.4 aux estimateurs  $\widehat{f}_{\iota,n}^{(a)}$  et  $\widehat{f}_{\iota,n}^{(b)}$ , tandis que le Théorème 2.5 est une conséquence du Corollaire 2.1 et de l’indépendance asymptotique établie au Théorème 2.3. Nous renvoyons à [27], paragraphe 6.7 pour plus détails.  $\square$

**Remarque 2.8.** *Notons que dans (2.10) on peut remplacer  $\sigma_0^2$ ,  $\sigma_1^2$  et  $\rho$  par*

$$\begin{aligned} \widehat{\sigma}_{0,n}^2 &= |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} (\widehat{\varepsilon}_{u0})^2, & \widehat{\sigma}_{1,n}^2 &= |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} (\widehat{\varepsilon}_{u1})^2, \\ \widehat{\rho}_n &= \widehat{\sigma}_{0,n}^{-1} \widehat{\sigma}_{1,n}^{-1} |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} \widehat{\varepsilon}_{u0} \widehat{\varepsilon}_{u1}, \end{aligned} \quad (2.11)$$

avec les résidus empiriques  $\widehat{\varepsilon}_{u\iota} = X_{u\iota} - \widehat{f}_{\iota,n}(X_u)$  pour  $u \in \mathbb{T}_n$ . Nous affirmons que ces estimateurs sont convergents, de telle sorte le Théorème 2.5 reste vrai avec cette nouvelle statistique de test. La preuve de la convergence de ces trois quantités nécessite des calculs assez techniques que nous omettons ici.

## 2.6 Implémentation numérique

Nous allons à présent illustrer les résultats théoriques précédents sur des données simulées et sur des données réelles. Nous illustrerons plus précisément les résultats du Théorème 2.1 et ceux du Théorème 2.5.

### 2.6.1 Données simulées

Afin de vérifier la qualité de notre procédure d'estimation, nous choisissons les fonctions d'autorégression  $f_0$  et  $f_1$  définies pour tout  $x \in \mathbb{R}$  par

$$f_0(x) = x(1/4 + \exp(-x^2)/2) \quad \text{et} \quad f_1(x) = x(1/8 + \exp(-x^2)/2).$$

Nous considérons un bruit gaussien tel que  $\sigma_0 = \sigma_1 = 1$  et  $\rho = 0.3$ . Nous simulons  $M = 500$  échantillons d'un processus NBAR jusqu'à la génération  $n + 1 = 15$ , de quantité initiale  $X_0 = 1$ . Afin d'implémenter les estimateurs donnés en (2.1), nous choisissons le noyau gaussien  $K(x) = (2\pi)^{-1/2} \exp(-x^2/2)$  et la fenêtre  $h_n = |\mathbb{T}_n|^{-1/5}$ . Nous évaluons  $\widehat{f}_{0,n}$  et  $\widehat{f}_{1,n}$  sur une grille régulière de  $D = [-3, 3]$  avec le pas  $\Delta x = |\mathbb{T}_n|^{-1/2}$ . Puisqu'avec ce choix du noyau nous ne rencontrons aucun problème avec le dénominateur, nous posons  $\varpi_n = 0$ . Pour chaque échantillon, nous calculons l'erreur empirique

$$e_\iota^{(i)} = \frac{\|\widehat{f}_{\iota,n}^{(i)} - f_\iota\|_{\Delta x}}{\|f_\iota\|_{\Delta x}}, \quad i = 1, \dots, M, \quad \iota \in \{0, 1\},$$

où  $\|\cdot\|_{\Delta x}$  désigne la norme discrète sur l'espace des échantillons. Le tableau 2.1 présente les erreurs moyennes empiriques et les écarts-types empiriques,  $\bar{e}_\iota = M^{-1} \sum_{i=1}^M e_\iota^{(i)}$  et  $(M^{-1} \sum_{i=1}^M (e_\iota^{(i)} - \bar{e}_\iota)^2)^{1/2}$  pour  $\iota \in \{0, 1\}$ . Comme le montre le tableau 2.1, plus on augmente la taille de l'échantillon, meilleur est la reconstruction de  $f_0$  et  $f_1$ .

La sensibilité de notre méthode à la taille des données peut aussi se voir sur la Figure 2.1, où nous construisons une bande de confiance de 95%, de telle sorte qu'à chaque point, les bornes supérieures et inférieures incluent 95% des valeurs estimées en ce point. On peut voir sur la Figure 2.1 que la reconstruction est bonne dans un voisinage de 0 et se détériore pour les points éloignés de 0. Ceci est dû au fait que la masse de la probabilité invariante de la chaîne marquée est concentrée dans un voisinage de 0, de telle sorte que les valeurs estimées ont plus tendance à tomber dans un voisinage de 0.

D'autre part, comme on peut le voir sur la Figure 2.2, l'erreur pour l'estimation de  $f_0$  et  $f_1$  est proche de  $|\mathbb{T}_n|^{-2/5}$  qui est l'erreur attendue dans ce cas. En effet, les fonctions

$n$	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
$sd.$	0.1509	0.1063	0.0761	0.0558	0.0387	0.0295	0.0209
$\bar{e}_1$	0.6696	0.5141	0.4006	0.3027	0.2356	0.1776	0.1384
$sd.$	0.2482	0.1626	0.1227	0.0831	0.0622	0.0440	0.0326

TABLE 2.1 – [Données simulées] *Erreur moyenne empirique relative  $\bar{e}_0$  (resp.  $\bar{e}_1$ ) et son écart-type calculés sur  $M = 500$  arbres de Monte-Carlo, par rapport à  $|\mathbb{T}_n|$ , pour la fonction d'autorégression  $f_0$  (resp.  $f_1$ ) reconstruite sur l'intervalle  $D = [-3, 3]$  par l'estimateur  $\hat{f}_{0,n}$  (resp.  $\hat{f}_{1,n}$ ).*

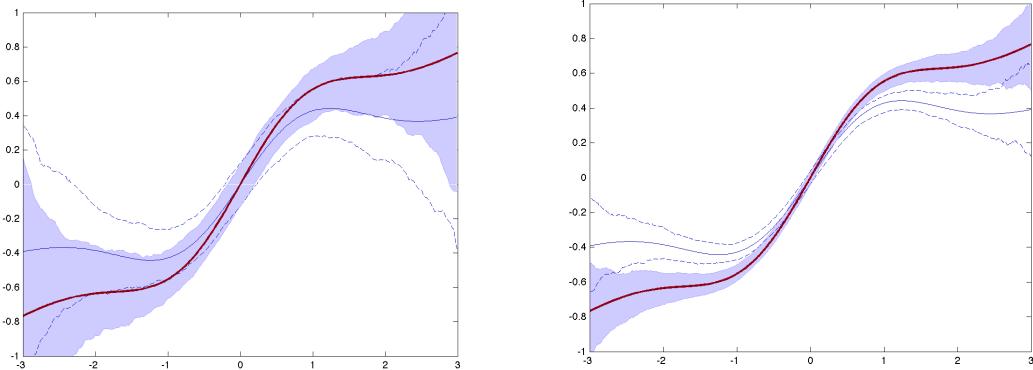


FIGURE 2.1 – [Données simulées] *Reconstruction des fonctions d'autorégression  $f_0$  et  $f_1$  sur  $\mathcal{D} = [-3, 3]$  avec des bandes de confiance de niveau 95% construite sur 500 arbres de Monte Carlo. La courbe rouge en gras représente la reconstruction de la fonction  $x \mapsto f_0(x)$ . La courbe bleue en trait fin est la reconstruction de  $f_1$  avec des bandes de confiance de niveau 95%. À gauche :  $n = 10$  generations. À droite :  $n = 14$  generations.*

d'autorégression sont de classe  $\mathcal{C}^\infty$  et le noyau est d'ordre  $n_0 = 1$ . Ceci nous confirme également qu'on a la bonne vitesse pour la convergence en loi.

Pour l'illustration de notre procédure de test, nous commençons par simuler les estimateurs définis en (2.9) dans le but de calculer la statistiques de test donnée en (2.10). Nous utilisons encore un noyau gaussien et les fenêtres  $h_n^{(a)} = |\mathbb{T}_n|^{-1/5}$  et  $h_n^{(b)} = |\mathbb{T}_n|^{-1/10}$  (i.e.  $\delta = 1/2$ ). Notons que le choix de  $\delta$  n'a eu aucune influence dans nos résultats numériques. Notons que l'étude numérique des estimateurs définies en (2.9) conduit aux mêmes résultats que ceux définies en (2.1). Pour une grille donnée  $\{x_1, \dots, x_k\}$  de l'intervalle  $[-3, 3]$ , on rejette l'hypothèse nulle  $\mathcal{H}_0$  si la statistique de test  $W_n$  dépasse le 95<sup>ème</sup> centile de la loi du chi-deux à  $k$  degrés de libertés. Nous construisons ainsi un test de niveau asymptotique 5%. Afin de mesurer la performance de notre procédure de test, nous

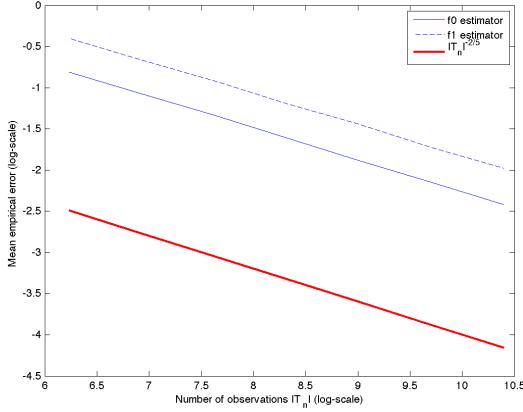


FIGURE 2.2 – [Données simulées] *Tracé du log de l’erreur empirique relative moyenne sur les 500 replications de Monte-Carlo par rapport au log de la taille des données  $\log(|T_n|)$  pour  $f_0$  (resp.  $f_1$ ) reconstruite sur  $D = [-3, 3]$  avec  $\tilde{f}_{0,n}$  (ligne continue bleue) (resp.  $\tilde{f}_{1,n}$  (ligne interrompue bleue)) comparé au log du taux prévue (ligne continue rouge).*

calculons la proportion de rejet dans deux cas :

**(Cas d’égalité)**  $f_0(x) = f_1(x) = x(1/4 + \exp(-x^2)/2)$ ,

**(Cas de différence)**  $f_0(x) = x(1/4 + \exp(-x^2)/2)$  et  $f_1(x) = x(1/8 + \exp(-x^2)/2)$ .

On peut voir dans le Tableau 2.2 que l’hypothèse nulle est de plus en plus rejetée dans le cas de différence et de moins en moins rejetée dans le cas d’égalité lorsque la taille de l’échantillon augmente, ce qui est le résultat que nous espérions. On peut aussi voir l’influence du nombre de points sur notre test. Quand on augmente le nombre de points, on augmente la proportion des rejets dans les deux cas. La moralité de cette étude est que notre test a des bonnes propriétés pour un nombre raisonnable de points.

La deuxième expérience a pour but d’illustrer la puissance empirique de notre test. Nous fixons les fonction  $f_0$  pour les cellules de type 0 et nous paramétrisons la fonction d’autorégression  $f_1$  de telle sorte qu’elle est progressivement proche de  $f_0$  :

$$f_0(x) = x(1/4 + \exp(-x^2)/2) \quad \text{and} \quad f_{1,\tau}(x) = x(\tau + \exp(-x^2)/2)$$

avec le paramètre d’interpolation  $\tau \in [1/8, 1/4]$ . Ce choix nous permet de interpoler entre **(Cas d’égalité)** et **(Cas de différence)**. Les deux courbes de la Figure 2.3 nous donnent la proportion de rejet en fonction du paramètre  $\tau$ . On peut remarquer que quand la fonction  $f_{1,\tau}$  se rapproche de  $f_0$ , *i.e.* lorsque  $\tau$  est proche de  $1/4$ , il y a une décroissance du nombre de rejet de l’hypothèse nulle. Le test est d’autant meilleur que la pente est raide. À gauche de la Figure 2.3, on peut voir que pour  $n = 10$  générations, la proportion des rejets de l’hypothèse nulle est plus grande que 40% seulement pour  $\tau$  plus petit ou

$n$	8	9	10	11	12	13	14
$ \mathbb{T}_n $	511	1 023	2 047	4 095	8 191	16 383	32 767
<b>Cas d'égalité</b>	$\Delta x = 0.5$	46.8%	67.2%	87.6%	99.0%	100%	100%
	$\Delta x = 0.25$	59.6%	77.8%	92.8%	99.8%	100%	100%
	$\Delta x = 0.1$	67.8%	85.4%	95.6%	99.8%	100%	100%
<b>Cas de différence</b>	$\Delta x = 0.5$	19.6%	18.6%	18.2%	16.2%	13.4%	14.8%
	$\Delta x = 0.25$	30.4%	30.0%	29.0%	24.8%	21.4%	19.4%
	$\Delta x = 0.1$	42.6%	42.6%	40.4%	39.8%	35%	31.6%

TABLE 2.2 – [Données simulées] *Proportions de rejet de l'hypothèse nulle  $\mathcal{H}_0 : \{\forall \ell = 1, \dots, k, f_0(x_\ell) = f_1(x_\ell)\}$  pour les tests de niveau asymptotique 5% sur 500 arbres de Monte-Carlo. Le test est basé sur la statistique de test  $W_n(x_1, \dots, x_k)$ , définie en (2.10), avec la grille de points  $\{x_\ell = -3 + (\ell - 1)\Delta x \leq 3; \ell \geq 1\}$  pour  $\Delta x \in \{0.5; 0.25; 0.1\}$  (i.e.  $k = 13, 25$  et 61 points). (Cas d'égalité) : les proportions (puissance du test) doivent être élevées. (Cas de différence) : les proportions (erreur de type I) doivent être basses.*

égal à 0.1875. À droite de la Figure 2.3, on peut voir que pour  $n = 14$  générations, la performance de notre test est bonne pour les valeurs de  $\tau$  jusqu'à 0.225, i.e. proche du **Cas d'égalité**.

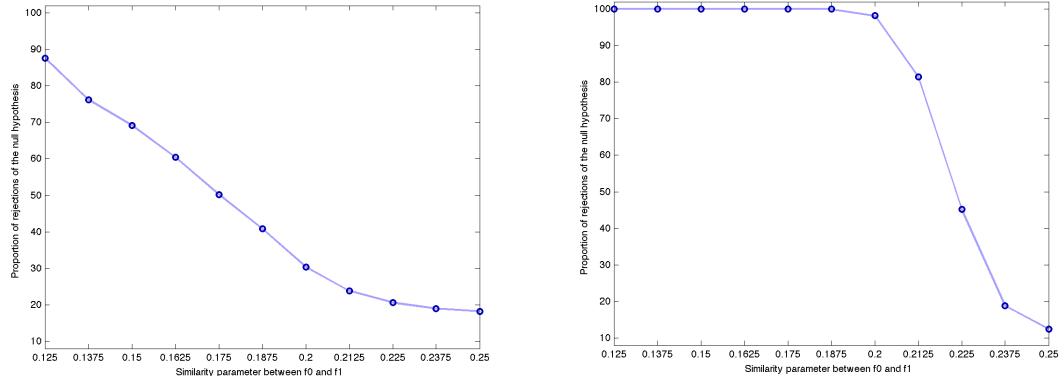


FIGURE 2.3 – [Données simulées] *Proportions de rejet de l'hypothèse nulle  $\mathcal{H}_0$  (puissance du test) :  $\{\forall \ell \in \{1, \dots, k\}, f_0(x_\ell) = f_{1,\tau}(x_\ell)\}$  par rapport à  $\tau \in [1/8; 1/4]$  pour les tests de niveau asymptotique 5% sur 500 arbres de Monte-Carlo. Les tests sont basées sur les statistiques de test  $W_n(x_1, \dots, x_k)$  définies en (2.10), avec la grille de points  $\{x_\ell = -3 + (\ell - 1)\Delta x \leq 3; \ell \geq 1\}$  for  $\Delta x = 0.5$  (i.e.  $k = 13$  points). À gauche :  $n = 10$  generations. À droite :  $n = 14$  generations.*

## 2.6.2 Données réelles

Nous avons également travaillé sur les données réelles, issues des expériences de Stewart *et al.* [99] sur la bactérie *E. coli*. Ces données ont fait l'objet de nombreuses études, l'une des plus récente étant celle Delyon *et al.* [45]. Dans leur expérience, Stewart *et al.* ont observé le taux de croissance plusieurs micro-colonies de *E. coli*. Mais les travaux de Delyon *et al.* [45] ont écarté la possibilité d'agréger les données des différentes micro-colonies. Nous avons donc uniquement travailler sur les échantillons ayant la plus grande taille (entre 443 et 655 cellules). Soulignons qu'afin de prendre en compte les données manquantes, nous avons, comme dans [62], supposé qu'elles se trouvent uniquement à la dernière génération du sous arbre considéré (dans notre cas, la 9<sup>ème</sup> génération). Nous choisissons aussi les échantillons de telle sorte que pour chaque cellule, son type et son taux de croissance, de même que le type et le taux de croissance de ses filles, sont connus. Avec cette approximation, nos résultats théoriques restent valides. Nous avons représenté sur la Figure 2.4a et sur la Figure 2.4b les estimateurs  $(\bar{f}_{0,n}, \bar{f}_{1,n})$  des fonctions d'autorégression, construites avec les échantillons de taille 655 et 446 cellules. Comme on peut l'observer, pour l'échantillon de taille 655, nos courbes sont proches des droites ajustées construites en utilisant les estimateurs de Guyon [62]. Mais pour l'échantillon de taille 446, nous obtenons une relation “complètement” non-linéaire entre le taux de croissance de la mère et ceux des filles.

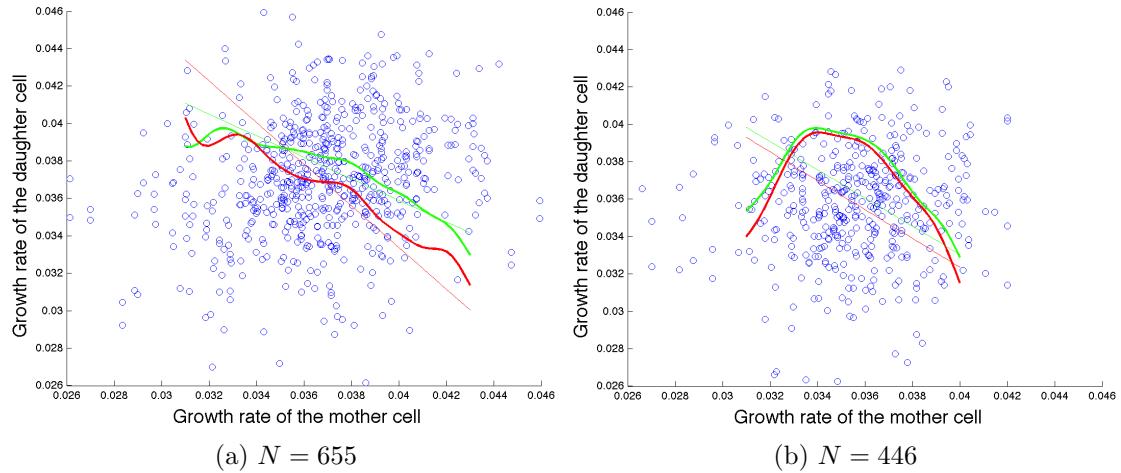


FIGURE 2.4 – [données réels issues de Stewart *et al.* [99]] *Représentation des points  $(X_u, X_{u0})$  et  $(X_u, X_{u1})$  pour  $N$  cellules  $u \in \mathbb{T}_9$ . La courbe verte en gras (resp. la droite verte en trait fin) est la reconstruction de  $x \rightsquigarrow f_0(x)$  avec  $x \rightsquigarrow \bar{f}_{0,n}(x)$  défini en (2.9) (resp. avec l'estimateur linéaire de Guyon [62]). La courbe rouge en gras (resp. la droite rouge en trait fin) est la reconstruction de  $x \rightsquigarrow f_1(x)$  avec  $x \rightsquigarrow \bar{f}_{1,n}(x)$  (resp. avec l'estimateur linéaire de Guyon).*

Nous avons également implémenter notre test statistique (2.10) en utilisant 10 points

équidistants de  $D = [0.0326; 0.0407]$ , où se trouve 80% des données. Nous avons estimé la matrice de covariance du bruit en utilisant (2.11). L'hypothèse nulle est fortement rejeté pour les échantillons ayant la plus grande taille. Ainsi, comme dans [41, 43, 62], nous pouvons conclure qu'il y a une asymétrie dans le processus de transmission du taux de croissance d'une cellule vers ses filles.

## 2.7 Principe de déviations modérées pour les estimateurs $\hat{f}_{t,n}$

Nous commençons par introduire la notion de principe de déviations modérées dans un cadre général. Soit  $(Z_n)_{n \geq 0}$  une suite de variables aléatoires à valeurs dans un espace polonais  $S$  muni de sa tribu borélienne  $\mathcal{B}_S$  et soit  $(s_n)_{n \geq 0}$  une suite de nombres réels positifs qui converge vers  $+\infty$ . On suppose que la suite  $(Z_n/s_n)_{n \geq 0}$  converge en probabilité vers 0 et que la suite  $(Z_n/\sqrt{s_n})_{n \geq 0}$  converge en loi vers une gaussienne centrée. Soit  $I : S \rightarrow \mathbb{R}^+$  une fonction semi-continue inférieurement, *i.e.* pour tout  $c > 0$ , l'ensemble de niveau  $\{x \in S, I(x) \leq c\}$  est un ensemble fermé. La fonction  $I$  est appelée une fonction de taux et elle est appelée une bonne fonction de taux si ses ensembles de niveau sont compacts. Soit  $(b_n)_{n \geq 0}$  une suite de nombres réels positifs telle que  $b_n \rightarrow +\infty$  et  $b_n/\sqrt{s_n} \rightarrow 0$  lorsque  $n$  tend vers  $+\infty$ .

**Définition 2.2** (Principe de déviations modérées). *On dit que  $Z_n/\sqrt{a_n s_n}$  satisfait un principe de déviations modérées dans  $S$  de vitesse  $b_n$  et de fonction de taux  $I$  si l'on a, pour tout  $A \in \mathcal{B}_S$ ,*

$$-\inf_{x \in \mathring{A}} I(x) \leq \liminf_{n \rightarrow \infty} \frac{1}{a_n} \log \mathbb{P}\left(\frac{Z_n}{\sqrt{a_n s_n}} \in A\right) \leq \limsup_{n \rightarrow \infty} \frac{1}{a_n} \log \mathbb{P}\left(\frac{Z_n}{\sqrt{a_n s_n}} \in A\right) \leq -\inf_{x \in \bar{A}} I(x),$$

où  $\mathring{A}$  et  $\bar{A}$  désigne respectivement l'intérieur et l'adhérence de  $A$ .

Pour les suites de variables aléatoires i.i.d et pour les modèles de régression et d'autorégression classiques, des principes de déviations modérées ont été prouvés pour l'estimateur à noyau de la densité et pour l'estimateur de Nadaraya-Watson. Nous renvoyons par exemple au travaux Worms [105, 107] et ceux de Mokkadem et Pelletier [83]. Les deux notions suivantes sont intimement liées au principe de déviations modérées. Soit  $(Z_n)_{n \geq 0}$  et  $(W_n)_{n \geq 0}$  deux suites de variables aléatoires et soit  $Z$  une variable aléatoire à valeurs dans un espace métrique  $(S, d)$ .

**Définition 2.3** (Convergence super-exponentielle). *On dit que  $(Z_n)_{n \geq 0}$  converge  $(b_n^2)$ -super-exponentiellement vite en probabilité vers  $Z$  et on note  $Z_n \xrightarrow[b_n^2]{\text{supereexp}} Z$  si, pour tout  $\delta > 0$ ,*

$$\limsup_{n \rightarrow +\infty} \frac{1}{b_n^2} \log \mathbb{P}(d(Z_n, Z) > \delta) = -\infty.$$

**Définition 2.4** (Equivalence exponentielle, voir [46], Chap 4). *On dit que  $(Z_n)_{n \geq 0}$  et  $(W_n)_{n \geq 0}$  sont  $(b_n^2)_{n \geq 0}$ -exponentiellement équivalentes et on note  $Z_n \underset{b_n^2}{\sim} W_n$  si pour tout  $\delta > 0$ ,*

$$\limsup_{n \rightarrow +\infty} \frac{1}{b_n^2} \log \mathbb{P}(d(Z_n, W_n) > \delta) = -\infty.$$

Puisque le principe de déviations modérées nécessite une notion de convergence plus forte, à savoir la convergence super-exponentielle en probabilité, il est naturel que nous imposions des conditions supplémentaires sur le modèle.

### Contraintes supplémentaires sur le modèle

On suppose donc que les fonctions d'autorégression  $f_0$  et  $f_1$  appartiennent à la classe  $\mathcal{F}(\gamma, \ell)$  avec  $\gamma = 0$ , ce qui signifie que

$$\max\{\|f_0\|_\infty; \|f_1\|_\infty\} \leq \ell.$$

Afin de garantir l'ergodicité géométrique uniforme de la chaîne marquée, Nous remplaçons l'hypothèse 2.1 par la suivante.

**Hypothèse 2.4.** *Il existe  $M_4 > \ell$  tel que  $2(M_4 - \ell)\delta(M_4) > 1/2$ , avec  $\delta(\cdot)$  défini en (2.3).*

Nous faisons également l'hypothèse supplémentaire suivante sur le noyau.

**Hypothèse 2.5.** *Le noyau  $K : \mathbb{R} \rightarrow \mathbb{R}$  est tel que*

$$\frac{\int_{\mathbb{R}} |K_+(x)| dx}{\int_{\mathbb{R}} |K_-(x)| dx} > \frac{\|g_0\|_\infty + \|g_1\|_\infty}{2\delta(M_3)(1 - \eta(M_2))},$$

où  $M_2, M_3$  sont définis dans l'Hypothèse 2.2.

Nous avons alors le résultat suivant.

**Théorème 2.6** (Principe de déviations modérées). *Dans le même cadre que le Théorème 2.1 avec  $\gamma = 0$ ,  $\lambda > 5$ ,  $K$  satisfaisant en plus l'Hypothèse 2.5,  $h_n = |\mathbb{T}_n|^{-\alpha}$  pour  $\alpha > 1/(2\beta + 1)$  et l'Hypothèse 2.1 remplacée par l'Hypothèse 2.4, soit  $(b_n, n \geq 0)$  une suite de nombre réels positifs tels que*

$$(i) \lim_{n \rightarrow \infty} \frac{b_n}{(|\mathbb{T}_n| h_n)^{1/2}} = +\infty, \quad (ii) \lim_{n \rightarrow \infty} \frac{b_n}{|\mathbb{T}_n| h_n} = 0, \quad (iii) \lim_{n \rightarrow \infty} \frac{b_n}{|\mathbb{T}_n| h_n^{1+\beta}} = +\infty.$$

Alors, la suite

$$\left( \frac{|\mathbb{T}_n| h_n}{b_n} \begin{pmatrix} \widehat{f}_{0,n}(x) - f_0(x) \\ \widehat{f}_{1,n}(x) - f_1(x) \end{pmatrix}, n \geq 0 \right)$$

satisfait un principe de déviations modérées dans  $\mathbb{R}^2$  de vitesse  $b_n^2/(|\mathbb{T}_n|h_n)$  et de bonne fonction de taux  $J_x : \mathbb{R}^2 \rightarrow \mathbb{R}$  définie par

$$J_x(\mathbf{z}) = (2|K|_2^2)^{-1} \mu(x) \mathbf{z}^t \Gamma^{-1} \mathbf{z}, \quad \mathbf{z} \in \mathbb{R}^2,$$

où  $\mathbf{z}^t$  désigne la transposée du vecteur  $\mathbf{z}$  et  $\Gamma$  est la matrice de covariance du bruit définie en (2.2).

En particulier, en utilisant le principe de contraction (voir par exemple [46], Chap 4), nous obtenons le corollaire suivant.

**Corollaire 2.2.** *Dans le même cadre que le Théorème 2.6, pour tout  $\delta > 0$ ,*

$$\lim_{n \rightarrow +\infty} \frac{|\mathbb{T}_n|h_n}{b_n^2} \log \mathbb{P}\left(\frac{|\mathbb{T}_n|h_n}{b_n^2} |\widehat{f}_{\iota,n}(x) - f_{\iota}(x)| > \delta\right) = -(2\sigma_{\iota}^2|K|_2^2)^{-1} \mu(x) \delta^2, \quad \iota \in \{0, 1\}.$$

*Principales idées de la preuve du Théorème 2.6.* La preuve utilise la décomposition (2.8) avec  $a_n = (|\mathbb{T}_n|h_n)/b_n$ . La stratégie est la suivante. On montre premièrement que les deux derniers termes du membre de droite de (2.8) convergent  $b_n^2/(|\mathbb{T}_n|h_n)$  super-exponentiellement vite en probabilité vers 0. Ceci nous conduit alors à l'équivalence exponentielle suivante :

$$\frac{|\mathbb{T}_n|h_n}{b_n} \begin{pmatrix} \widehat{f}_{0,n}(x) - f_0(x) \\ \widehat{f}_{1,n}(x) - f_1(x) \end{pmatrix} \xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} \frac{1}{\widehat{\mu}_n(x) \vee \varpi_n} \frac{|\mathbb{T}_n|h_n}{b_n} \begin{pmatrix} M_{0,n}(x) \\ M_{1,n}(x) \end{pmatrix}.$$

On termine ensuite la preuve en montrant que  $\widehat{\mu}_n(x) \vee \varpi_n$  converge  $b_n^2/(|\mathbb{T}_n|h_n)$  super-exponentiellement vite en probabilité vers  $\mu(x)$  et en faisant usage du principe de déviations modérées pour les martingales (voir par exemple [88, 106]). Notons que les convergences super-exponentielles que nous avons évoquées s'établissent à l'aide des inégalités de concentration que nous développerons dans la suite.  $\square$

## Chapitre 3

# Estimation non paramétrique adaptative de la densité invariante de la chaîne marquée

Ce chapitre est issu des articles [26], écrit en collaboration avec Marc Hoffmann et Adélaïde Olivier, et [29], écrit en collaboration avec Angelina Roche. Nous utilisons la transformée en ondelettes et une méthode de type Goldenshluger-Lepski pour l'estimation de la densité invariante de la chaîne marquée. Je fais le choix délibéré de me focaliser uniquement sur l'estimation de la densité invariante de la chaîne marquée. Je renvoie à [26] pour l'estimation de la probabilité de transition  $\mathcal{P}$  (resp.  $\mathcal{Q}$ ) de la chaîne de Markov bifurcante (resp. de la chaîne marquée) par la méthode des ondelettes.

### 3.1 Cadre général

Nous considérons une chaîne de Markov bifurcante  $(X_u, u \in \mathbb{T})$  à valeurs dans  $\mathbb{R}^d$ ,  $d \geq 1$ , de probabilité de transition  $\mathcal{P}$  et de mesure initiale  $\nu$ . Nous supposons que  $\mathcal{P}$  admet une densité, que nous notons également  $\mathcal{P}$ , par rapport à la mesure de Lebesgue. Nous supposons que cette chaîne est stable au sens donné à la fin de la Section 1.1.3. Puisque  $\mathcal{P}$  admet une densité, une conséquence est que la probabilité de transition  $\mathcal{Q}$  et la mesure de probabilité invariante  $\mu$  de la chaîne marquée admettent des densités que nous noterons également  $\mathcal{Q}$  et  $\mu$  respectivement (pour plus de détails, nous renvoyons par exemple à [54], chap 6 et [18], Remarque 1.6). Nous supposerons dans la suite que les densités  $\mathcal{P}$ ,  $\mathcal{Q}$  et  $\mu$  sont uniformément bornées sur  $(\mathbb{R}^d)^3$ ,  $(\mathbb{R}^d)^2$  et  $\mathbb{R}^d$  respectivement. En d'autres termes, nous supposons que  $\max\{\|\mathcal{P}\|_\infty, \|\mathcal{Q}\|_\infty, \|\mu\|_\infty\} < +\infty$ . Nous travaillerons sous l'hypothèse d'ergodicité géométrique uniforme suivante.

**Hypothèse 3.1** (Condition d'ergodicité géométrique uniforme). *Il existe deux constantes*

$\alpha \in (0, 1/2)$  et  $R > 0$  telles que pour toute fonction bornée et  $\mu$ -intégrable  $g$ ,

$$|\mathcal{Q}^m g(x) - \langle \mu, g \rangle| \leq R \|g\|_\infty \alpha^m \quad \forall x \in \mathbb{R}^d \text{ et } m \in \mathbb{N}.$$

### 3.2 Estimation adaptative de la densité invariante par méthode d'ondelettes

Le but ici est de construire un estimateur de la densité invariante  $\mu$  par projection sur une base d'ondelettes.

#### 3.2.1 Décomposition en ondelettes et espaces de Besov

Les bases d'ondelettes ont été documentées dans plusieurs livres (voir par exemple [36]). Nous considérons la base d'ondelettes  $(\psi_\lambda^d)_\lambda$ , adaptées à un domaine compact  $D^d \subset \mathbb{R}^d$ , où le multi-index  $\lambda$  concatène les index de position et le niveau de résolution que nous notons  $|\lambda|$ . On pose  $\Lambda_j = \{\lambda, |\lambda| = j\}$  et  $\Lambda = \cup_{j \geq -1} \Lambda_j$ . Ainsi, pour  $g \in L^\pi(D^d)$ , avec  $\pi \in (0, +\infty]$ , on a la décomposition en ondelettes

$$g = \sum_{j \geq -1} \sum_{\lambda \in \Lambda_j} g_\lambda \psi_\lambda^d = \sum_{\lambda \in \Lambda} g_\lambda \psi_\lambda^d, \quad \text{avec } g_\lambda = \langle g, \psi_\lambda^d \rangle,$$

où nous avons fixé  $j = -1$  afin d'incorporer la partie basse fréquence de la décomposition et  $\langle g, \psi_\lambda^d \rangle = \int g \psi_\lambda^d$  désigne le produit scalaire dans  $L^2(\mathbb{R}^d)$ . Pour  $s > 0$ ,  $g$  appartient à  $B_{\pi, \infty}^s(D^d)$  si la norme suivante est finie :

$$\|g\|_{B_{\pi, \infty}^s(D^d)} = \sup_{j \geq -1} 2^{j(s+d(1/2-1/\pi))} \left( \sum_{\lambda \in \Lambda_j} |\langle g, \psi_\lambda^d \rangle|^\pi \right)^{1/\pi}, \quad (3.1)$$

avec les modifications d'usage si  $\pi = +\infty$ . La relation précise entre cette définition de la norme de Besov et la définition classique peut-être trouvée dans [36]. Étant donnée une base  $(\psi_\lambda^d)_\lambda$ , il existe  $\sigma > 0$  tel que pour  $\pi \geq 1$  et  $s \leq \sigma$ , l'espace de Besov défini par (3.1) coïncide exactement avec la définition usuelle qui utilise les module de continuité pour  $g$ . L'index  $\sigma$  peut-être pris arbitrairement grand. Nous aurons besoin des propriétés additionnelles suivantes sur la base d'ondelettes  $(\psi_\lambda^d)_\lambda$

**Hypothèse 3.2.** Pour  $p \geq 1$ ,

$$\|\psi_\lambda^d\|_{L^p}^p \sim 2^{|\lambda|d(p/2-1)},$$

pour un certain  $\sigma > 0$  et pour tous  $s \leq \sigma$ ,  $j_0 \geq 0$ ,

$$\|g - \sum_{j \leq j_0} \sum_{\lambda \in \Lambda_j} g_\lambda \psi_\lambda^d\|_{L^p} \lesssim 2^{-j_0 s} \|g\|_{B_{p, \infty}^s(D^d)}, \quad (3.2)$$

pour chaque sous ensemble  $\Lambda_0 \subset \Lambda$ ,

$$\int_D \left( \sum_{\lambda \in \Lambda_0} |\psi_\lambda^d(x)|^2 \right)^{p/2} dx \simeq \sum_{\lambda \in \Lambda_0} \|\psi_\lambda^d\|_{L^p}^p. \quad (3.3)$$

Si  $p > 1$ , pour chaque suite  $(u_\lambda)_{\lambda \in \Lambda}$ ,

$$\left\| \left( \sum_{\lambda \in \Lambda} |u_\lambda \psi_\lambda^d|^2 \right)^{1/2} \right\|_{L^p} \simeq \left\| \sum_{\lambda \in \Lambda} u_\lambda \psi_\lambda^d \right\|_{L^p}. \quad (3.4)$$

Le symbole  $\simeq$  signifie une inégalité dans les deux sens, à une constante près dépendant uniquement de  $p$  et  $D$ . La propriété (3.2) reflète que notre définition (3.1) des espaces de Besov correspond à la définition en terme d'approximation linéaire. La propriété (3.4) reflète une propriété de base inconditionnelle (voir par exemple Kerkyacharian et Picard [71], De Vore *et al.* [47]) et (3.3) est appelée inégalité de superconcentration ou propriété de Temlyakov [71]. La formulation (3.3)-(3.4) dans le contexte de l'estimation statistique est postérieure aux articles originaux de Donoho and Johnstone [49, 50] et Donoho *et al.* [52, 51] et est due à Kerkyacharian et Picard [71]. Pour l'existence d'une base d'ondelettes à support compact vérifiant l'hypothèse 3.2, on pourra voir Meyer [82], ou Cohen [36].

### 3.2.2 Estimation de la densité invariante $\mu$ par ondelettes

Nous allons supposer que  $d = 1$  et soit  $D$  un interval compact de  $\mathbb{R}$ . Alors, pour tout  $x \in D$ , on estime la représentation

$$\mu(x) = \sum_{\lambda \in \Lambda} \mu_\lambda \psi_\lambda^1(x), \quad \mu_\lambda = \langle \mu, \psi_\lambda^1 \rangle$$

par

$$\widehat{\mu}_n(x) = \sum_{|\lambda| \leq J} \widehat{\mu}_{\lambda,n} \psi_\lambda^1(x),$$

avec

$$\widehat{\mu}_{\lambda,n} = \mathcal{T}_{\lambda,\eta} \left( \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} \psi_\lambda^1(X_u) \right),$$

et  $\mathcal{T}_{\lambda,\eta}(x) = x \mathbf{1}_{|x| \geq \eta}$  désigne l'opérateur de seuillage dur (avec  $\mathcal{T}_{\lambda,\eta}(x) = x$  pour la partie basse fréquence quand  $\lambda \in \Lambda_{-1}$ ). Ainsi,  $\widehat{\mu}_n$  est spécifié par le niveau de résolution  $J$  et le seuil  $\eta$ . Dans la suite, nous posons

$$\|\mathcal{Q}\|_D = \sup_{x \in \mathbb{R}, y \in D} \mathcal{Q}(x, y).$$

Nous avons le résultat suivant sur la borne supérieure de convergence de notre estimateur.

**Théorème 3.1.** *Sous l'hypothèse 3.1, spécifions  $\hat{\mu}_n$  avec*

$$J = \log_2 \frac{|\mathbb{T}_n|}{\log |\mathbb{T}_n|} \text{ et } \eta = c \sqrt{\log |\mathbb{T}_n| / |\mathbb{T}_n|}$$

*pour un certain  $c > 0$ . Pour tout  $\pi \in (0, \infty]$ ,  $s \in (1/\pi, \sigma]$  et  $p \geq 1$ , pour  $n$  assez grand et  $c$ , on a l'inégalité suivante :*

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

*avec  $\alpha_1(s, p, \pi) = \min \left\{ \frac{s}{2s+1}, \frac{s+1/p-1/\pi}{2s+1-2/\pi} \right\}$ , à une constante près qui dépend continûment de  $s, p, \pi, \|\mu\|_{\mathcal{B}_{\pi, \infty}^s(D)}$ ,  $\alpha$ ,  $R$  et  $\|\mathcal{Q}\|_D$ .*

**Remarque 3.1.** *Notons que notre estimateur est adaptatif dans le sens suivant : pour tous  $s_0 > 0$ ,  $0 < \alpha_0 < 1/2$ ,  $R_0 > 0$  et  $\mathcal{Q}_0 > 0$ , définissons les ensembles  $\mathcal{A}(s_0) = \{(s, \pi), s \geq s_0, s_0 \geq 1/\pi\}$  et*

$$\mathcal{Q}(\alpha_0, R_0, \mathcal{Q}_0) = \{\mathcal{Q} \text{ telle que } \alpha \leq \alpha_0, R \leq R_0, \|\mathcal{Q}\|_D \leq \mathcal{Q}_0\},$$

*où  $\mathcal{Q}$  est prise dans l'ensemble des probabilités de transition satisfaisant l'hypothèse 3.1. Alors, pour tout  $C > 0$ , il existe  $c^* = c^*(\mathcal{D}, p, s_0, \alpha_0, R_0, \mathcal{Q}_0, C)$  tel que  $\hat{\mu}_n$  spécifié avec  $c^*$  satisfait*

$$\sup_n \sup_{(s, \pi) \in \mathcal{A}(s_0)} \sup_{\mu, \mathcal{Q}} \left( \frac{|\mathbb{T}_n|}{\log |\mathbb{T}_n|} \right)^{p\alpha_1(s, p, \pi)} \mathbb{E} [\|\hat{\mu}_n - \mu\|_{L^p(D)}^p] < \infty$$

*où le sup est pris parmi les  $(\mu, \mathcal{Q})$  tels que  $\mu\mathcal{Q} = \mu$  avec  $\mathcal{Q} \in \mathcal{Q}(\alpha_0, R_0, \mathcal{Q}_0)$  et  $\|\mu\|_{\mathcal{B}_{\pi, \infty}^s(D)} \leq C$ . En particulier,  $\hat{\mu}_n$  atteint le taux de convergence (presque) optimal sur les boules de Besov simultanément pour tout  $(s, \pi) \in \mathcal{A}(s_0)$ .*

*Notons également que la question de l'obtention d'une borne minimax reste ouverte pour cet estimateur.*

### 3.3 Sélection locale de la fenêtre dans l'estimation à noyau de la densité invariante

Comme nous l'avons vu au Chapitre 2, la densité invariante  $\mu$  de la chaîne marquée peut-être estimée par

$$\hat{\mu}_{\mathbf{h}}(x) = \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{\mathbf{h}}(x - X_u) \quad \forall x \in \mathbb{R}^d, \quad (3.5)$$

où  $K_{\mathbf{h}}(x) = K(x_1/h_1, \dots, x_d/h_d)/|\mathbf{h}|$  avec  $|\mathbf{h}| = \prod_{j=1}^d h_j$ ,  $K$  est un noyau et  $\mathbf{h} = (h_1, \dots, h_d)^\top \in ]0, +\infty[^d$  est un vecteur de *fenêtres*. Nous avons adapté la définition donnée en 2.6 au cas multidimensionnel. Nous omettons à dessein l'indice  $n$  dans l'écriture du vecteur  $\mathbf{h}$  afin

de ne pas alourdir l'écriture. L'objectif ici est de choisir de façon optimale le vecteur  $\mathbf{h}$ , uniquement à partir des données. Nous utilisons pour cela une règle inspirée des travaux de Goldenshluger et Lepski [59] avec des modifications qui ont été suggérées par Lacour et Massart [76]. Le point de départ de la définition de notre règle de sélection est l'inégalité suivante, qui peut-être vue comme une extension de la décomposition biais-variance. On fait d'abord l'hypothèse suivante sur le noyau.

**Hypothèse 3.3.**

$$\|K\|_\infty < +\infty \quad \text{et} \quad \|K\|_p^p < +\infty, \quad \text{pour } p \in \{1, 2\}.$$

Alors on a :

**Proposition 3.1.** *Sous les hypothèses 3.1 et 3.3, on a*

$$\mathbb{E}[(\widehat{\mu}_{\mathbf{h}}(x) - \mu(x))^2] \leq 2(K_{\mathbf{h}} * \mu(x) - \mu(x))^2 + 2 \frac{C(\mathcal{P}, \mu)}{|\mathbb{T}_n| |\mathbf{h}|}, \quad (3.6)$$

où  $C(P, \nu)$  est une constante qui dépend continûment de  $\mathcal{P}$ ,  $\mu$ ,  $K$  et  $\alpha$ .

Si de plus on suppose que  $\mu \in \mathcal{H}_{\mathbb{R}^d}^{\beta}(L)$ , pour un  $L > 0$ , avec  $\beta = (\beta_1, \dots, \beta_d) \in \mathbb{R}^d$ , et que le noyau  $K$  est d'ordre  $\boldsymbol{\ell} = (\lfloor \beta_1 \rfloor, \dots, \lfloor \beta_d \rfloor)$ , i.e.

$$\int_{\mathbb{R}^d} x_k^j K(x) dx = 0 \quad \forall j \in \{1, \dots, \lfloor \beta_j \rfloor\}, \quad k \in \{1, \dots, d\}, \quad (3.7)$$

alors on peut déduire de la Proposition 3.1 que le meilleur taux de convergence qui peut-être atteint par l'estimateur  $\widehat{\mu}_{\mathbf{h}}$  est  $|\mathbb{T}_n|^{-2\bar{\beta}/(2\bar{\beta}+d)}$ , où  $\bar{\beta} = d/(1/\beta_1 + \dots + 1/\beta_d)$  est la moyenne harmonique de  $\beta$ . Nous avons déjà évoqué ce résultat dans le cas unidimensionnel, voir Remarque 2.5. On peut vérifier que la fenêtre  $\mathbf{h}^* = (h_1^*, \dots, h_d^*)$  telle que  $h_j^* = |\mathbb{T}_n|^{-\bar{\beta}/(\beta_j(2\bar{\beta}+d))}$  permet d'atteindre le meilleur taux de convergence. Cependant, cette valeur de  $\mathbf{h}^*$  dépend du paramètre inconnu  $\mu$  à travers la régularité inconnue  $\beta$  de telle sorte qu'il ne peut-être utilisé en pratique. D'où l'intérêt de l'estimer à partir des données.

### 3.4 Règle de sélection de la fenêtre

Nous considérons une collection  $\mathcal{H}_n \subset [0, +\infty[^d$  de fenêtres possibles. En imitant la décomposition donnée à l'équation 3.6, nous choisissons la fenêtre  $\widehat{\mathbf{h}}$  suivant la règle suivante.

$$\widehat{\mathbf{h}} = \arg \min_{\mathbf{h} \in \mathcal{H}_n} \{A(x, \mathbf{h}) + bV(x, \mathbf{h})\}$$

où

$$A(x, \mathbf{h}) = \max_{\mathbf{h}' \in \mathcal{H}_n} ((\widehat{\mu}_{\mathbf{h}'}(x) - K_{\mathbf{h}} * \widehat{\mu}_{\mathbf{h}'}(x))^2 - aV(x, \mathbf{h}'))_+ \quad \text{avec} \quad b \geq a \geq 1;$$

$$V(x, \mathbf{h}) = C(\mathcal{P}, \mu) \log(|\mathbb{T}_n|) / |\mathbb{T}_n| |\mathbf{h}|.$$

**Remarque 3.2.** Cette procédure dépend de deux constantes  $a$  et  $b$ . Soulignons que le problème de calibration de ces deux constantes est un sujet de recherche en soi (voir par exemple [76]). En dimension 1, nous avons proposé une méthode pour les choisir. En particulier, en s'inspirant des travaux de Arlot et Massart [4], nous avons proposé un algorithme permettant de calibrer le paramètre  $aC(\mathcal{P}, \mu)$ . Nous renvoyons à [29], section 4.1, pour plus de détails.

Pour l'estimateur  $\hat{\mu}_{\hat{\mathbf{h}}}$ , on a l'inégalité oracle suivante.

**Théorème 3.2.** Supposons que la collection  $\mathcal{H}_n$  est telle que  $|\mathcal{H}_n| \leq |\mathbb{T}_n|$  et  $\min_{\mathbf{h} \in \mathcal{H}_n} |\mathbf{h}| \geq \log(|\mathbb{T}_n|)/|\mathbb{T}_n|$ . Alors, sous les hypothèses de la Proposition 3.1, il existe une valeur minimale  $a_{\min} > 0$  indépendante de  $n$ , telle que, pour tout  $a > a_{\min}$ ,

$$\mathbb{E} [(\hat{\mu}_{\hat{\mathbf{h}}}(x) - \mu(x))^2] \leq C_1 \min_{\mathbf{h} \in \mathcal{H}_n} \left\{ \mathcal{B}_{\mathbf{h}}(x) + \frac{\log(|\mathbb{T}_n|)}{|\mathbb{T}_n| |\mathbf{h}|} \right\} + \frac{C_2}{|\mathbb{T}_n|}, \quad (3.8)$$

où  $C_1, C_2 > 0$  ne dépendent ni de  $n$ , ni de  $x$  et

$$\mathcal{B}_{\mathbf{h}}(x) = \max_{\mathbf{h}' \in \mathcal{H}_n} (K_{\mathbf{h}'} * \mu(x) - K_{\mathbf{h}} * K_{\mathbf{h}'} * \mu(x))^2.$$

**Remarque 3.3.** Soulignons que la forme du biais  $\mathcal{B}_{\mathbf{h}}(x)$  dans l'inégalité (3.8) est très similaire à celle obtenue par Rebelles [89]. Comme dans [37], en utilisant l'inégalité de Young, il peut-être remplacé par la borne supérieure  $\|K\|_1 \|\nu - K_{\mathbf{h}} * \nu\|_{\infty}$ .

Une conséquence directe du Théorème 3.2 est que si le noyau  $K$  satisfait (3.7) et si on choisit par exemple

$$\mathcal{H}_n = \left\{ h_{\max} k^{-\gamma}, k = 1, \dots, (|\mathbb{T}_n| h_{\max} / \log(|\mathbb{T}_n|))^{1/\gamma} \right\}^d,$$

avec  $h_{\max} > 0$  et  $\gamma > 1$ , alors

$$\sup_{x \in \mathbb{R}^d} \sup_{\mu \in \mathcal{H}(\beta, L)} \mathbb{E} [(\hat{\mu}_{\hat{\mathbf{h}}}(x) - \mu(x))^2] \lesssim \left( \frac{|\mathbb{T}_n|}{\log(|\mathbb{T}_n|)} \right)^{-2\bar{\beta}/(2\bar{\beta}+d)}.$$

Nous retrouvons ainsi, à une perte logarithmique près, le meilleur taux de convergence obtenu à l'aide de la Proposition 3.1. Soulignons que cette perte logarithmique est due à l'adaptation.

### 3.5 Quelques pistes pour les preuves

L'outil principal pour les preuves des Théorèmes 3.1 et 3.2 est l'inégalité de concentration suivante.

**Théorème 3.3.** *Sous l'hypothèse 3.1, on a*

$$\mathbb{P}\left(\frac{1}{|\mathbb{G}_n|} \sum_{u \in \mathbb{G}_n} g(X_u) - \langle \mu, g \rangle \geq \delta\right) \leq \exp\left(\frac{-|\mathbb{G}_n|\delta^2}{c_1 + c_2\delta}\right),$$

où  $c_1$  et  $c_2$  sont des constantes qui dépendent de  $g$  et des paramètres du modèle.

*Idées de la preuve.* La preuve de ce Théorème 3.3 utilise des arguments similaires à ceux utiliser dans [23]. Toutefois, à la différence des preuves dans [23], nous accordons ici une grande importance au contrôle des variances conditionnelles

$$\mathbb{E}_{X_u} \left[ \left( (\mathcal{Q}^r \tilde{g}(X_{u0}) + \mathcal{Q}^r \tilde{g}(X_{u1}) - 2\mathcal{Q}^{r+1} \tilde{g}(X_u)) \right)^2 \right] \quad \forall u \in \mathbb{T}.$$

Nous renvoyons à [26], Théorème 4.1 ou à [29], Lemme 6.1 pour plus de détails.  $\square$

Une fois l'inégalité de concentration obtenue, le Théorème 3.1 est la conséquence de la théorie générale de l'estimation par seuillage en ondelettes (voir par exemple [71], Corollaire 5.1 et Théorème 6.1). Le Théorème 3.2 quant à lui est la conséquence de la théorie de l'estimation adaptative par la méthode de Goldenshluger-Lepski (voir par exemple [59, 60]).

## Chapitre 4

# Inégalités de transport et de concentration pour les chaînes de Markov bifurcantes

Ce chapitre est issu de l'article [24] écrit en collaboration avec Mikael Escobar-Bach et Arnaud Guillin. L'un des objectifs était de proposer des inégalités de concentration dans un cadre plus général que celui du Théorème 3.3. Pour cela, nous nous appuyons sur la théorie des inégalités de transport qui a été introduite par Marton [77, 78].

Tout d'abord, commençons par faire quelques rappels. Soit  $(E, d)$  un espace métrique et soit  $\mu$  et  $\nu$  deux mesures de probabilités sur  $E$ . Pour  $p \geq 1$ , la  $L^p$ -distance de Wasserstein entre  $\mu$  et  $\nu$  est définie par

$$W_p^d(\nu, \mu) = \inf \left( \int \int d(x, y)^p d\pi(x, y) \right)^{1/p},$$

où l'infimum est pris sur toutes les mesures de probabilités  $\pi$  sur l'espace produit  $E \times E$  ayant pour distributions marginales  $\mu$  et  $\nu$ , appelées couplage de  $(\mu, \nu)$ . Notons que cet infimum est fini dès que  $\mu$  et  $\nu$  ont des moments finis d'ordre  $p$ . En particulier, pour la mesure triviale  $d(x, y) = \mathbf{1}_{x \neq y}$ ,  $2W_1^d(\mu, \nu) = \|\mu - \nu\|_{TV}$  désigne la norme en variation totale de  $\mu - \nu$ . L'information de Kullback ou entropie relative de  $\nu$  par rapport à  $\mu$  est définie par

$$H(\nu|\mu) = \begin{cases} \int \log \frac{d\nu}{d\mu} d\nu, & \text{si } \nu \ll \mu \\ +\infty & \text{sinon.} \end{cases}$$

**Définition 4.1 ( $L^p$ -inégalité de transport).** *On dit qu'une mesure de probabilité  $\mu$  satisfait une  $L^p$ -inégalité de transport  $T_p$  sur  $(E, d)$ , et on note  $\mu \in T_p(C)$ , s'il existe une constante  $C > 0$  telle que pour toute mesure de probabilité  $\nu$ ,*

$$W_p^d(\nu, \mu) \leq \sqrt{2CH(\nu|\mu)}.$$

Les caractérisations suivantes des  $L^1$ -inégalités de transport sont des outils essentiels pour obtenir des inégalités de concentrations ou pour vérifier rapidement qu'une inégalité  $T_1$  est satisfaite.

**Théorème 4.1** ([31]). *La mesure  $\mu$  satisfait la  $L^1$ -inégalité de transport  $T_1(C)$  sur  $(E, d)$  avec une constante  $C > 0$  si et seulement si pour toute fonction Lipschizienne  $F : (E, d) \rightarrow \mathbb{R}$ ,  $F$  est  $\mu$ -intégrable et*

$$\int_E \exp(\lambda(F - \langle \mu, F \rangle)) d\mu \leq \exp\left(\frac{\lambda^2}{2} C \|F\|_{Lip}^2\right), \quad \forall \lambda \in \mathbb{R},$$

où

$$\|F\|_{Lip} = \sup_{x \neq y} \frac{|F(x) - F(y)|}{d(x, y)} < +\infty.$$

En particulier, on a l'inégalité de concentration

$$\mu(F - \langle \mu, F \rangle \leq -t) \vee \mu(F - \langle \mu, F \rangle \geq t) \leq \exp\left(-\frac{t^2}{2C\|F\|_{Lip}^2}\right) \quad \forall t \in \mathbb{R}.$$

**Théorème 4.2** ([48, 61]).  *$\mu$  satisfait une  $L^1$ -inégalité de transport  $T_1(C)$  sur  $(E, d)$  si et seulement si il existe  $\delta > 0$  et  $x_0 \in E$  telle que*

$$\mu\left(e^{\delta d^2(x, x_0)}\right) < \infty,$$

où la constante  $C$  peut être explicite.

## 4.1 Inégalité de transport pour les chaînes de Markov bifurcantes

Dans toute la suite,  $(X_u, u \in \mathbb{T})$  est une chaîne de Markov bifurcante à valeurs dans un espace métrique  $(S, d)$ , de probabilité de transition  $\mathcal{P}$  et de mesure initiale  $\nu$ . Nous désignons par  $\mathcal{P}_n$  la loi du  $|\mathbb{T}_n|$ -échantillon  $(X_u, u \in \mathbb{T}_n)$ . Pour  $p \geq 1$  et  $N \geq 1$ , nous considérons la métrique  $d_{\ell_p}$  définie sur  $S^N$  par

$$d_{\ell_p}(x, y) = \left( \sum_{i=1}^N d(x_i, y_i)^p \right)^{1/p} \quad \forall x, y \in S^N.$$

L'hypothèse suivante sera appelée  $H_p(C)$ .

**Hypothèse 4.1** ( $H_p(C)$ ). *Il existe  $q > 0$  tel que :*

- (a)  $\nu \in T_p(C)$  ;
- (b)  $\mathcal{P}(x, \cdot, \cdot) \in T_p(C)$  sur  $(S^2, d_{\ell_p})$ ,  $\forall x \in S$  ;

$$(c) \quad W_p^{d_{\ell_p}}(\mathcal{P}(x, \cdot, \cdot), \mathcal{P}(\tilde{x}, \cdot, \cdot)) \leq q d(x, \tilde{x}), \quad \forall x, \tilde{x} \in S.$$

Dans la suite,  $q$  sera appelé taux de contraction de la probabilité de transition  $\mathcal{P}$ .

**Remarque 4.1.** Soulignons que sous  $H_p(C)$ , les marginales  $\mathcal{P}_0$  et  $\mathcal{P}_1$  satisfont également  $T_p(C)$ . Aussi, pour  $q > 0$  associé à  $H_p(C)$ , ils existent  $q_0, q_1 \in (0, q)$  tels que, pour  $\iota \in \{0, 1\}$  et  $\forall x, \tilde{x} \in S$ ,

$$W_p^d(\mathcal{P}_\iota(x, \cdot), \mathcal{P}_\iota(\tilde{x}, \cdot)) \leq q_\iota d(x, \tilde{x}).$$

De plus, si  $\mathcal{P}$  admet la décomposition  $\mathcal{P}(x, dy, dz) = \mathcal{P}_0(x, dy)\mathcal{P}_1(x, dz)$ , alors  $q \leq (q_0^p + q_1^p)^{1/p}$ .

Nous avons le résultat suivant.

**Théorème 4.3.** On suppose  $H_p(C)$  satisfaite pour  $1 \leq p \leq 2$  et  $C > 0$ . Alors  $\mathcal{P}_n \in T_p(C_n)$  où

$$C_n \propto \begin{cases} C \frac{|\mathbb{T}_n|^{2/p-1}}{(1-q)^2} & \text{si } q < 1 \\ C \exp\left(2 - \frac{2}{p}\right) |\mathbb{T}_n|^{2/p+1} & \text{si } q = 1 \\ C |\mathbb{T}_n| \left(\frac{\exp(p-1)q^{p|\mathbb{T}_n|}}{q^p-1}\right)^{2/p} & \text{si } q > 1. \end{cases}$$

Quelques idées de la preuve du Théorème 4.3. La clé pour la preuve de ce Théorème réside dans la construction d'un couplage, “à la Marton” (voir par exemple [79, 80, 81], ou encore [48], section 2.4), qui prenne en compte la structure binaire du processus. Ensuite, le reste de la preuve est une conséquence de l'additivité de l'entropie (voir par exemple [103], Lemme 22.8) et de l'analyse. Nous renvoyons à [24], preuve du Théorème 2.3, pour plus de détails.  $\square$

**Remarque 4.2.** Dans le Théorème 4.3, on voit apparaître deux régimes avec une transition de phase autour de  $q = 1$ . Lorsque  $q < 1$ , on a un comportement classique (comme pour les suites de variables aléatoires i.i.d.) : nous disons alors que  $T_p(C_n)$  est conduit par la loi du processus. Lorsque  $q \geq 1$ , le comportement n'est plus habituel. Pour  $q > 1$ , on observe que la taille de l'échantillon joue un rôle significatif dans l'inégalité de transport à travers le terme  $q^{p|\mathbb{T}_n|}$  : nous disons alors que  $T_p(C_n)$  est conduit par la généalogie. On observe donc ici une compétition entre la loi du processus et la généalogie avec pour arbitre le taux de contraction en distance de Wasserstein de la probabilité de transition  $\mathcal{P}$ . Ce genre de comportement avait déjà été observé dans [23] avec comme arbitre de la compétition le taux d'ergodicité géométrique de la chaîne marquée.

**Remarque 4.3.** Comme conséquence directe des Théorèmes 4.3, 4.1 et du Lemme 2.1 de [48], nous avons l'inégalité de concentration suivante. Soit  $f$  une fonction réelle Lipschitzienne définie sur  $(S, d)$ . Alors sous  $H_p(C)$ , pour  $1 \leq p \leq 2$  et pour tout  $t > 0$ , on a, pour  $\mathbb{A}_n \in \{\mathbb{G}_n, \mathbb{T}_n\}$ ,

$$\mathbb{P}(|\mathbb{A}_n|^{-1} |M_{\mathbb{A}_n}(f) - \mathbb{E}[M_{\mathbb{A}_n}(f)]| > t) \leq 2 \exp\left(-\frac{t^2 |\mathbb{A}_n|^{2/p}}{2C_n \|f\|_{Lip}^2}\right), \quad (4.1)$$

où  $C_n$  est défini dans le Théorème 4.3. La même inégalité reste valable si  $f$  est définie sur  $(S^3, d_{\ell_1})$ , avec  $4C_n$  à la place de  $2C_n$ .

Notons que dans le cas  $p = 1$ , l'inégalité de concentration 4.1 est pertinente seulement dans le cas  $q < 1$ . Ce cas correspond en particulier à une stabilité des chaînes de Markov marginales, i.e. les chaînes de Markov de probabilités de transitions  $\mathcal{P}_0$  et  $\mathcal{P}_1$ .

## 4.2 Inégalités de concentration gaussiennes pour les chaînes de Markov bifurcantes

L'inégalité 4.1 pose deux problèmes. Le premier c'est qu'elle est établie pour des fonctionnelles centrées par rapport à la loi de l'échantillon alors que, pour les applications statistiques par exemple, ce qui est intéressant c'est un "centrage" par rapport à la loi invariante. L'autre problème que pose cette inégalité c'est qu'elle impose aux chaînes marginales d'être stables. Mais notons que la stabilité d'une chaîne de Markov bifurcante n'impose pas que les deux chaînes de Markov marginales soient stables. Le but de cette section est donc d'étendre l'inégalité 4.1 afin de prendre en compte au mieux ces deux "préoccupations". Rappelons de la Remarque 4.1 que sous  $H_1(C)$ , on peut associer, à un taux de contraction  $q > 0$  de  $\mathcal{P}$ , des taux de contraction  $q_0$  et  $q_1$  des marginales  $\mathcal{P}_0$  et  $\mathcal{P}_1$ . Nous faisons l'hypothèse suivante.

**Hypothèse 4.2.** *On suppose  $H_1(C)$  satisfaite avec un taux de contraction  $q > 0$  de  $\mathcal{P}$ . Soit  $q_0, q_1 > 0$  les taux de contraction associés aux marginales  $\mathcal{P}_0$  et  $\mathcal{P}_1$  respectivement. On suppose  $q_0 + q_1 < 2$ .*

**Remarque 4.4.** *Sous l'Hypothèse 4.2 et en utilisant la convexité de  $W_1$  (voir par exemple [103]), nous avons*

$$W_1^d(\mathcal{Q}(x, \cdot), \mathcal{Q}(\tilde{x}, \cdot)) \leq \frac{q_0 + q_1}{2} d(x, \tilde{x}), \quad \forall x, \tilde{x},$$

assurant ainsi une contraction stricte de  $\mathcal{Q}$  et par conséquent une convergence exponentielle vers la probabilité invariante  $\mu$  en distance de wassertein (dès que  $\mu$  admet un moment d'ordre 1). Plus précisément, on a

$$W_1^d(\mathcal{Q}^n(x, \cdot), \mu) \leq \left( \frac{q_0 + q_1}{2} \right)^n \int d(x, y) \mu(dy).$$

Dans la suite, nous posons  $\gamma = (q_0 + q_1)/2$ , le taux de convergence vers l'équilibre de la chaîne marquée en distance de Wasserstein  $W_1$ . Soit  $\mathbb{A}_n \in \{\mathbb{G}_n, \mathbb{T}_n\}$ . En utilisant alors les many-to-one formulae et la définition duale de la distance de Wasserstein  $W_1$ , on montre que

$$|\mathbb{E}(|\mathbb{A}_n|^{-1} M_{\mathbb{A}_n}(f)) - \langle \mu, f \rangle| \leq c_n \quad \text{où} \quad c_n = \begin{cases} c\gamma^{n+1} & \text{si } \gamma \neq 1/2 \\ c\frac{n}{2^{n+1}} & \text{si } \gamma = 1/2, \end{cases}$$

et  $c$  est une constante universelle. Ainsi, pour tout  $\delta > c_n$  et  $\mathbb{A}_n \in \{\mathbb{G}_n, \mathbb{T}_n\}$ , nous avons

$$\mathbb{P}(|\mathbb{A}_n|^{-1} M_{\mathbb{A}_n}(f) - \langle \mu, f \rangle > \delta) \leq \mathbb{P}(|\mathbb{A}_n|^{-1} (M_{\mathbb{A}_n}(f) - \mathbb{E}[M_{\mathbb{A}_n}(f)]) > \delta - c_n). \quad (4.2)$$

Nous avons alors le résultat suivant.

**Théorème 4.4.** *Soit  $f$  une fonction lipschitzienne sur  $S$  ou sur  $S^3$ . Sous l'hypothèse 4.2, il existe  $n_0 \in \mathbb{N}$  et une constante positive  $c$  dépendante de  $C$ ,  $q$  et  $\gamma$  tels que  $\forall n > n_0$  et  $\forall \delta > 0$ ,*

$$\mathbb{P}(|\mathbb{A}_n|^{-1} M_{\mathbb{A}_n}(f - \langle \mu, f \rangle) > \delta) \leq \begin{cases} \exp\left(-\frac{(\delta - c_n)^2 |\mathbb{A}_n|}{c \|f\|_{Lip}^2}\right) & \text{si } \gamma^2 < 1/2 \\ \exp\left(-\frac{(\delta - c_n)^2 |\mathbb{A}_n|}{c \|f\|_{Lip}^2 n}\right) & \text{si } \gamma^2 = 1/2 \\ \exp\left(-\frac{(\delta - c_n)^2 |\mathbb{A}_n|}{c \|f\|_{Lip}^2 (2\gamma^2)^n}\right) & \text{si } 1/2 < \gamma^2 < 1, \end{cases}$$

**Remarque 4.5.** *Notons que le Théorème 4.4 généralise, dans un certain sens, le Théorème 3.1 de [23] (le taux d'ergodicité géométrique est remplacé ici par le taux de convergence en distance de Wasserstein  $\gamma$ ). Une fois de plus nous voyons apparaître deux régimes avec une transition de phase autour de  $\gamma = \sqrt{2}/2$ . Comme nous l'avons dit précédemment, ceci matérialise la compétition qui existe entre la loi du processus et la taille de la population considérée.*

*Quelques pistes pour la preuve du Théorème 4.4.* La preuve utilise la remarque suivante. Si une variable aléatoire  $X$  satisfait, pour  $\kappa > 0$ , la propriété dite de concentration Gaussienne :

$$\mathbb{E}[\exp(t(X - \mathbb{E}[X]))] \leq \exp(\kappa t^2/2) \quad \forall t \in \mathbb{R},$$

alors,  $\forall r > 0$ , on

$$\mathbb{P}(X - \mathbb{E}[X] \geq r) \leq e^{-\frac{r^2}{2\kappa}}.$$

Ainsi, grâce à (4.2), la preuve du Théorème 4.4 revient à montrer une propriété de concentration Gaussienne pour  $|\mathbb{A}_n|^{-1} (M_{\mathbb{A}_n}(f) - \mathbb{E}[M_{\mathbb{A}_n}(f)])$ . Cette propriété de concentration Gaussienne s'obtient à l'aide de la propriété de branchement (*i.e.* la troisième propriété de la Définition 1.1) et d'une utilisation récursive du Lemme 2.1 de [48] et du Théorème 4.1. Nous renvoyons à [24], preuve des Propositions 3.3, 3.5 et des Corollaires 3.4 et 3.6, pour plus de détails.  $\square$

**Remarque 4.6.** *Un exemple où peut-être appliquée la théorie précédente est le processus NBAR (voir Définition 2.1). En effet, avec les notations de la Définition 2.1, on montre que si :*

1. *les fonctions  $f_0$  et  $f_1$  sont Lipschitziennes,*
2.  *$(\varepsilon_u)_{u \in \mathbb{T}}$  est une suite de variables aléatoires i.i.d. gaussiennes ou bornées,*

alors l'Hypothèse 4.2 est satisfaite avec  $q_0$ ,  $q_1$  et  $q$  qui dépendent uniquement de  $\|f_0\|_{Lip}$  et  $\|f_1\|_{Lip}$ . Alors, en utilisant la décomposition (2.4) et en appliquant le Théorème 4.4, on est capable de donner des bornes exponentielles pour les probabilités  $\mathbb{P}(|\hat{f}_{\iota,n}(x) - f_{\iota}(x)| > \delta)$ ,  $\delta > 0$  et  $\iota \in \{0, 1\}$ , où les  $\hat{f}_{\iota,n}(x)$  sont les estimateurs définis en (2.1). Toutefois, un pareil contrôle reste insuffisant puisque l'idéal serait de pouvoir contrôler les probabilités  $\mathbb{P}(\sup_{x \in D} |\hat{f}_{\iota,n}(x) - f_{\iota}(x)| > \delta)$ . Cette dernière probabilité est difficile à traiter. Mais pour l'avenir, des pistes que nous pourrions exploiter sont données dans les travaux de Bolley, Guillin et Villani [32] et ceux de Fournier et Guillin [58].

## Chapitre 5

# Théorème central limite pour les chaînes de Markov bifurcantes

Ce chapitre est issu des articles [19, 20, 21] écrit en collaboration avec Jean-françois Delmas et des articles [17, 18]. Nous prouvons un théorème central limite pour les fonctionnelles additives de chaînes de Markov bifurcantes. Nous prouvons ainsi l'existence deux régimes avec une transition de phase. Ceci confirme alors que la compétition qui existe entre la loi du processus, représentée par le taux d'ergodicité de la chaîne marquée, et la taille de la population étudiée n'est pas une simple vue de l'esprit. Ce phénomène est intimement lié à la structure du modèle. Nous appliquons ensuite le résultat précédent à l'obtention d'un théorème central limite pour les estimateurs à noyau de la densité invariante de la chaîne marquée et la densité de la probabilité de transition d'une chaîne de Markov bifurcante. Même s'il semble curieux que les différents régimes disparaissent pour ces estimateurs, nous verrons qu'en fait il n'en est rien : le problème se déplace juste ailleurs. Les résultats obtenus (et les techniques utilisées) dans ce chapitre sont une avancée majeure dans la théorie des chaînes de Markov bifurcantes à plusieurs titres. En premier, ils complètent merveilleusement bien les résultats de Guyon [62]. Ensuite, ils donnent des pistes nouvelles pour une extension des résultats obtenus aux Chapitres 1 et 2. Ils ouvrent également la voie pour l'exploration d'autres techniques d'estimation non paramétrique pour les modèles de Markov bifurcants.

### 5.1 Cadre général

Nous considérons une chaîne de Markov bifurcante  $(X_u, u \in \mathbb{T})$  à valeurs dans un espace mesurable  $(S, \mathcal{S})$ , de probabilité de transition  $\mathcal{P}$  et de mesure initiale  $\nu$ . Nous considérons deux cadres pour les hypothèses : le cadre ponctuelle et le cadre  $L^2$ .

### 5.1.1 Hypothèses pour l'approche ponctuelle

Pour un ensemble  $F \subset \mathcal{B}(S)$  de fonctions à valeurs réelles, nous écrivons  $F^2 = \{f^2; f \in F\}$ ,  $F \otimes F = \{f_0 \otimes f_1; f_0, f_1 \in F\}$ , et, pour une probabilité de transition  $P$ ,  $P(F) = \{Pf; f \in F\}$  à condition que  $P$  agisse sur  $F$ . L'hypothèse suivante, tirée de [62], établit les propriétés de l'espace des fonctions.

**Hypothèse 5.1.** *Soit  $F \subset \mathcal{B}(S)$  un ensemble de fonctions à valeurs réelles tel que :*

- (i)  *$F$  est un sous espace vectoriel qui contient les constantes ;*
- (ii)  *$F^2 \subset F$  ;*
- (iii)  *$F \subset L^1(\nu)$  ;*
- (iv)  *$F \otimes F \subset L^1(\mathcal{P}(x, \cdot))$  pour tout  $x \in S$ , et  $\mathcal{P}(F \otimes F) \subset F$ .*

La condition (iv) implique que  $\mathcal{P}_0(F) \subset F$ ,  $\mathcal{P}_1(F) \subset F$  et  $\mathcal{Q}(F) \subset F$ , où  $\mathcal{P}_0$ ,  $\mathcal{P}_1$  et  $\mathcal{Q}$  désignent respectivement la première et la deuxième marginale de  $\mathcal{P}$ , et la probabilité de transition de la chaîne marquée. Notons que si  $f \in F$ , même si  $|f|$  n'appartient pas à  $F$ , en utilisant les conditions (i) et (ii), nous obtenons, avec  $g = (1 + f^2)/2$ , que  $|f| \leq g$  et  $g \in F$ . Typiquement, si  $(S, d)$  est un espace métrique, l'ensemble  $\mathcal{C}_b(S)$  des fonctions continues bornées sur  $S$  et l'ensemble des fonctions réelles de classe  $\mathcal{C}^1$  dont les dérivées sont à croissance polynomiale vérifient l'Hypothèse 5.1.

En suivant toujours [62], nous considérons les propriétés d'ergodicité ponctuelles suivantes pour  $\mathcal{Q}$ .

**Hypothèse 5.2.** *Il existe une mesure de probabilité  $\mu$  sur  $(S, \mathcal{S})$  telle que  $F \subset L^1(\mu)$  et pour tout  $f \in F$ , on a la convergence ponctuelle  $\lim_{n \rightarrow \infty} \mathcal{Q}^n f = \langle \mu, f \rangle$  et il existe  $g \in F$  avec :*

$$|\mathcal{Q}^n(f)| \leq g \quad \forall n \in \mathbb{N}. \quad (5.1)$$

**Hypothèse 5.3.** *Il existe une mesure de probabilité  $\mu$  sur  $(S, \mathcal{S})$  telle que  $F \subset L^1(\mu)$ , et  $\alpha \in (0, 1)$  tel que pour tout  $f \in F$ , il existe  $g \in F$  tel que :*

$$|\mathcal{Q}^n f - \langle \mu, f \rangle| \leq \alpha^n g \quad \forall n \in \mathbb{N}. \quad (5.2)$$

Une suite  $\mathfrak{f} = (f_\ell, \ell \in \mathbb{N})$  d'éléments de  $F$  satisfait uniformément (5.1) et (5.2) s'il existe  $g \in F$  telle que :

$$|\mathcal{Q}^n(f_\ell)| \leq g \quad \text{and} \quad |\mathcal{Q}^n f_\ell - \langle \mu, f_\ell \rangle| \leq \alpha^n g \quad \forall n, \ell \in \mathbb{N}. \quad (5.3)$$

Ceci entraîne en particulier que  $|f_\ell| \leq g$  et  $|\langle \mu, f_\ell \rangle| \leq \langle \mu, g \rangle$ . Notons que (5.3) est par exemple vérifié si  $\mathfrak{f}$  a un nombre fini d'éléments satisfaisant (5.1) et (5.2).

**Exemple 5.1.** *Soit  $(S, d)$  un espace métrique, et  $Y$  une chaîne de Markov uniformément géométriquement ergodique i.e. il existe  $\alpha \in (0, 1)$  et une constante finie  $C$  telle que pour tout  $x \in S$  :*

$$\|\mathcal{Q}^n(x, \cdot) - \mu\|_{TV} \leq C\alpha^n, \quad (5.4)$$

où, pour une mesure signée finie  $\pi$  sur  $(S, \mathcal{B}_S)$ , sa norme en variation totale est définie par  $\|\pi\|_{TV} = \sup_{f \in \mathcal{B}(S), \|f\|_\infty \leq 1} |\langle \pi, f \rangle|$ . Alors, en prenant  $F = \mathcal{C}_b(S)$ , on obtient que les propriétés (i-iii) de l’Hypothèse 5.1 et l’Hypothèse 5.3 sont satisfaites. En particulier, (5.4) implique que (5.2) est satisfaite avec  $g = C \|f\|_\infty$ .

Nous considérons la propriété, plus forte, suivante basée sur le second trou spectral de l’opérateur  $\mathcal{Q}$ .

**Hypothèse 5.4.** *Il existe une mesure de probabilité  $\mu$  sur  $(S, \mathcal{S})$  telle que  $F \subset L^1(\mu)$ ,  $\alpha \in (0, 1)$ , un ensemble d’indices non vides  $J$ , des valeurs propres complexes distinctes  $\{\alpha_j, j \in J\}$  de l’opérateur  $\mathcal{Q}$  avec  $|\alpha_j| = \alpha$ , des projections complexes non nulles  $\{\mathcal{R}_j, j \in J\}$  définies sur  $\mathbb{C}F$ , le  $\mathbb{C}$ -espace vectoriel engendré par  $F$ , tel que  $\mathcal{R}_j \circ \mathcal{R}_{j'} = \mathcal{R}_{j'} \circ \mathcal{R}_j = 0$  pour tout  $j \neq j'$  (de sorte  $\sum_{j \in J} \mathcal{R}_j$  est aussi une projection définie sur  $\mathbb{C}F$ ) et une suite de nombres positifs  $(\beta_n, n \in \mathbb{N})$  convergeant vers 0, tels que pour tout  $f \in F$ , il existe  $g \in F$  et, avec  $\theta_j = \alpha_j/\alpha$  :*

$$\left| \mathcal{Q}^n(f) - \langle \mu, f \rangle - \alpha^n \sum_{j \in J} \theta_j^n \mathcal{R}_j(f) \right| \leq \beta_n \alpha^n g \quad \forall n \in \mathbb{N}. \quad (5.5)$$

Sans perte de généralité, nous supposerons que la suite  $(\beta_n, n \in \mathbb{N})$  dans l’Hypothèse 5.4 est décroissante et bornée supérieurement par 1.

**Remarque 5.1.** *Dans [62], seuls les Hypothèses 5.1 et 5.2 sont considérées. Si  $F$  contient un ensemble  $A$  de fonctions bornées qui séparent (dans le sens où deux mesures de probabilités qui coïncident sur  $A$  sont égales), alors les Hypothèses 5.1 et 5.2 impliquent en particulier que  $\mu$  est l’unique mesure invariante de  $\mathcal{Q}$ . Notons que l’Hypothèse d’ergodicité géométrique 5.3 implique l’Hypothèse 5.2, et que l’Hypothèse 5.4 implique l’Hypothèse 5.3 (avec le même  $\alpha$  mais une fonction  $g$  qui peut-être différente).*

**Exemple 5.2.** *Nous considérons le modèle BAR défini en (1.3) où nous supposons que  $(\varepsilon_v, v \in \mathbb{T})$  est une suite de variables aléatoires i.i.d. centrées de loi normale  $\mathcal{N}(0, \sigma^2)$  avec  $\sigma > 0$ . Soit  $G$  une variable aléatoire de loi normale standard  $\mathcal{N}(0, 1)$ . Alors la probabilité de transition  $\mathcal{Q}$  de la chaîne marquée vérifie*

$$\mathcal{Q}^n f(x) = \mathbb{E} \left[ f \left( a^n x + \sqrt{1 - a^{2n}} \sigma_a G \right) \right], \quad (5.6)$$

où  $\sigma_a = \sigma(1 - a^2)^{-1/2}$ . La probabilité de transition  $\mathcal{Q}$  admet une unique mesure de probabilité invariante  $\mu$ , qui est la Gaussienne  $\mathcal{N}(0, \sigma_a^2)$ . L’opérateur  $\mathcal{Q}$  (sur  $L^2(\mu)$ ) est un opérateur intégral de Hilbert-Schmidt symétrique dont les valeurs propres, de multiplicité algébrique 1, sont données par  $\sigma_p(\mathcal{Q}) = (a^n, n \in \mathbb{N})$  et les fonctions propres correspondantes  $(\bar{g}_n(x), n \in \mathbb{N})$  sont définies pour  $n \in \mathbb{N}$  par  $\bar{g}_n(x) = g_n(\sigma_a^{-1} x)$ , où  $g_n$  est le polynôme de Hermite de degré  $n$ . En particulier, on a  $\bar{g}_0 = 1$  et  $\bar{g}_1(x) = \sigma_a^{-1} x$ . Soit  $\mathcal{R}$  la

projection orthogonale sur l'espace vectoriel engendré par  $\bar{g}_1$ , i.e.  $\mathcal{R}f = \langle \mu, f\bar{g}_1 \rangle \bar{g}_1$  ou de manière équivalente, pour  $x \in \mathbb{R}$  :

$$\mathcal{R}f(x) = \sigma_a^{-1} x \mathbb{E}[Gf(\sigma_a G)]. \quad (5.7)$$

Nous considérons  $F$ , l'ensemble des fonctions  $f \in \mathcal{C}^2(\mathbb{R})$  telles que  $f, f'$  et  $f''$  sont à croissance au plus polynomial. Supposons que  $F \subset L^1(\nu)$ . Alors,  $F$  satisfait l'Hypothèse 5.1. On a aussi que  $F \subset L^1(\mu)$ . On vérifie également que l'Hypothèse 5.4 est satisfaite avec  $J = \{j_0\}$ ,  $\alpha_{j_0} = \alpha = |a|$ ,  $\beta_n = |a|^n$  et  $\mathcal{R}_{j_0} = \mathcal{R}$ . De même, les Hypothèses 5.2 et 5.3 sont satisfaites.

### 5.1.2 Hypothèses pour l'approche $L^2$

**Hypothèse 5.5.** Il existe une mesure de probabilité invariante  $\mu$  pour la probabilité de transition  $\mathcal{Q}$ .

(i) Il existe une constante  $M$  telle que pour tout  $f, g, h \in L^2(\mu)$  :

$$\begin{aligned} \|\mathcal{P}(\mathcal{Q}f \otimes_{\text{sym}} \mathcal{Q}g)\|_{L^2(\mu)} &\leq M \|f\|_{L^2(\mu)} \|g\|_{L^2(\mu)}, \\ \|\mathcal{P}(\mathcal{P}(\mathcal{Q}f \otimes_{\text{sym}} \mathcal{Q}g) \otimes_{\text{sym}} \mathcal{Q}h)\|_{L^2(\mu)} &\leq M \|f\|_{L^2(\mu)} \|g\|_{L^2(\mu)} \|h\|_{L^2(\mu)}, \\ \|\mathcal{P}(f \otimes_{\text{sym}} \mathcal{Q}g)\|_{L^2(\mu)} &\leq M \|f\|_{L^2(\mu)} \|g\|_{L^2(\mu)}. \end{aligned}$$

(ii) Il existe  $k_0 \in \mathbb{N}$ , tel que la mesure de probabilité  $\nu \mathcal{Q}^{k_0}$  a une densité bornée, noté  $\nu_0$ , par rapport à  $\mu$ . C'est-à-dire :

$$\nu \mathcal{Q}^{k_0}(dy) = \nu_0(y) \mu(dy) \quad \text{et} \quad \|\nu_0\|_\infty < +\infty.$$

La propriété suivante garantit que  $\mathcal{Q}$  converge exponentiellement dans  $L^2(\mu)$ .

**Hypothèse 5.6.** La probabilité de transition  $\mathcal{Q}$  a une unique mesure de probabilité invariante  $\mu$ . De plus, il existe  $\alpha \in (0, 1)$  et  $M$  fini tel que pour tout  $f \in L^2(\mu)$  :

$$\|\mathcal{Q}^n f - \langle \mu, f \rangle\|_{L^2(\mu)} \leq M \alpha^n \|f\|_{L^2(\mu)} \quad \text{pour tout } n \in \mathbb{N}.$$

L'hypothèse suivante est une version plus forte de l'Hypothèse 5.6 utilisant le second trou spectral.

**Hypothèse 5.7.** La probabilité de transition  $\mathcal{Q}$  a une unique mesure de probabilité invariante  $\mu$ , et il existe  $\alpha \in (0, 1)$ , un ensemble d'indices  $J$  non vide, des valeurs propres complexes distinctes  $\{\alpha_j, j \in J\}$  de l'opérateur  $\mathcal{Q}$  avec  $|\alpha_j| = \alpha$ , des projections complexes non nulles  $\{\mathcal{R}_j, j \in J\}$  définies sur  $\mathbb{C}L^2(\mu)$ , le  $\mathbb{C}$ -espace vectoriel engendré par  $L^2(\mu)$ , telles que  $\mathcal{R}_j \circ \mathcal{R}_{j'} = \mathcal{R}_{j'} \circ \mathcal{R}_j = 0$  pour tout  $j \neq j'$  (de sorte que  $\sum_{j \in J} \mathcal{R}_j$  est aussi une projection définie sur  $\mathbb{C}L^2(\mu)$ ) et une suite de nombres réels positifs  $(\beta_n, n \in \mathbb{N})$  convergente vers 0 telle que pour tout  $f \in L^2(\mu)$ , avec  $\theta_j = \alpha_j/\alpha$  :

$$\|\mathcal{Q}^n f - \langle \mu, f \rangle - \alpha^n \sum_{j \in J} \theta_j^n \mathcal{R}_j(f)\|_{L^2(\mu)} \leq \beta_n \alpha^n \|f\|_{L^2(\mu)} \quad \forall n \in \mathbb{N}. \quad (5.8)$$

### Remarque 5.2.

1. Comme dans le cadre ponctuelle, on montre que le modèle BAR défini en (1.3) vérifie les Hypothèses 5.5, 5.6 et 5.7.
2. Soulignons que les Hypothèses 5.6 et 5.7 de l'approche  $L^2$  sont les équivalentes des Hypothèses 5.3 et 5.4 de l'approche ponctuelle. L'hypothèse structurelle 5.1 de l'approche ponctuelle est remplacée dans l'approche  $L^2$  par l'Hypothèse 5.5.

**Remarque 5.3.** Dans les Hypothèses 5.4 et 5.7, si les projections de  $f$  sur les espaces propres associés aux valeurs propres  $\alpha_j$  sont nulles, i.e.  $\mathcal{R}_j(f) = 0$  pour tout  $j \in J$ , alors dans ce cas, on utilise le troisième trou spectral et ainsi de suite. En fait l'idée est, à partir de la deuxième famille de valeurs propres en terme de module, de chercher la première pour laquelle l'une des projections de  $f$  sur l'un des espaces propres associés est non nulle. Notons que c'est une idée qui a déjà été développée dans le cas des processus de Markov branchants à temps continu et des superprocessus par Ren, Song et Zhang [90, 91, 92, 93, 94]. Leurs résultats reposent sur des hypothèses fortes sur le générateur infinitésimal du processus. Dans notre cas, cela reviendrait à supposer que la transition  $\mathcal{Q}$  est un opérateur compact de  $L^2$  et par conséquent que son spectre est entièrement connu. Une pareille hypothèse est évidemment plus forte que les Hypothèses 5.4 et 5.7.

#### 5.1.3 Quelques notations

Nous terminons cette section avec quelques notations. Pour  $f \in L^1(\mu)$ , nous posons

$$\tilde{f} = f - \langle \mu, f \rangle$$

Soit  $\mathbf{f} = (f_\ell, \ell \in \mathbb{N})$  une suite d'éléments de  $L^1(\mu)$ . Nous posons

$$N_{n,\emptyset}(\mathbf{f}) = |\mathbb{G}_n|^{-1/2} \sum_{\ell=0}^n M_{\mathbb{G}_{n-\ell}}(\tilde{f}_\ell). \quad (5.9)$$

**Remarque 5.4.** Nous considérerons en particulier les suites  $\mathbf{f}$  suivantes. Soit  $f \in L^1(\mu)$ . Si  $f_0 = f$  et  $f_\ell = 0$  pour tout  $\ell \in \mathbb{N}^*$ , nous écrirons

$$\mathbf{f}_0 = (f, 0, \dots), \quad (5.10)$$

et nous obtenons :

$$N_{n,\emptyset}(\mathbf{f}_0) = |\mathbb{G}_n|^{-1/2} M_{\mathbb{G}_n}(\tilde{f}).$$

Si  $f_\ell = f$  pour tout  $\ell \in \mathbb{N}$ , alors on écrirons

$$\mathbf{f} = (f, f, \dots), \quad (5.11)$$

et nous obtenons, puisque  $|\mathbb{T}_n| = 2^{n+1} - 1$  et  $|\mathbb{G}_n| = 2^n$  :

$$N_{n,\emptyset}(\mathbf{f}) = |\mathbb{G}_n|^{-1/2} M_{\mathbb{T}_n}(\tilde{f}) = \sqrt{2 - 2^{-n}} |\mathbb{T}_n|^{-1/2} M_{\mathbb{T}_n}(\tilde{f}).$$

Ainsi, à partir des résultats pour  $N_{n,\emptyset}(\mathbf{f})$ , nous pourrons facilement déduire des résultats pour les sommes  $M_{\mathbb{T}_n}(f)$  et  $M_{\mathbb{G}_n}(f)$ .

## 5.2 Les principaux résultats

À présent que nous avons fixé le cadre, nous pouvons énoncer les principaux résultats pour les fonctions appartenant à  $\mathcal{B}(S)$ . Nous rappelons les sommes définies en (1.1). Il existe trois cas, fonction du paramètre d'ergodicité géométrique  $\alpha$ .

### 5.2.1 Cas sous-critique : $2\alpha^2 < 1$

Pour une suite  $\mathbf{f} = (f_\ell, \ell \in \mathbb{N})$  de  $\mathcal{B}(S)$ , lorsqu'elles sont bien définies, nous considérerons les quantités suivantes.

$$\Sigma^{\text{sub}}(\mathbf{f}) = \Sigma_1^{\text{sub}}(\mathbf{f}) + 2\Sigma_2^{\text{sub}}(\mathbf{f}), \quad (5.12)$$

où :

$$\begin{aligned} \Sigma_1^{\text{sub}}(\mathbf{f}) &= \sum_{\ell \geq 0} 2^{-\ell} \langle \mu, \tilde{f}_\ell^2 \rangle + \sum_{\ell \geq 0, k \geq 0} 2^{k-\ell} \langle \mu, \mathcal{P}((\mathcal{Q}^k \tilde{f}_\ell) \otimes^2) \rangle, \\ \Sigma_2^{\text{sub}}(\mathbf{f}) &= \sum_{0 \leq \ell < k < \infty} 2^{-\ell} \langle \mu, \tilde{f}_k \mathcal{Q}^{k-\ell} \tilde{f}_\ell \rangle + \sum_{\substack{0 \leq \ell < k < \infty \\ r \geq 0}} 2^{r-\ell} \langle \mu, \mathcal{P}(\mathcal{Q}^r \tilde{f}_k \otimes_{\text{sym}} \mathcal{Q}^{k-\ell+r} \tilde{f}_\ell) \rangle. \end{aligned}$$

Alors, nous avons le résultat suivant.

**Théorème 5.1.** *Soit  $X$  une chaîne de Markov bifurcante de probabilité de transition  $\mathcal{P}$  et de distribution initiale  $\nu$ . Soit  $\mathbf{f} = (f_\ell, \ell \in \mathbb{N})$  une suite de  $\mathcal{B}(S)$ .*

- *Dans le cadre ponctuel, on suppose que les Hypothèses 5.1 et 5.3 sont satisfaites avec  $\alpha \in (0, 1/\sqrt{2})$  et que la suite  $\mathbf{f}$  est une suite de  $F$  satisfaisant uniformément l'Hypothèse 5.3, i.e. (5.3) pour un  $g \in F$ .*
- *Dans le cadre  $L^2$ , on suppose que les Hypothèses 5.5 et 5.6 sont satisfaites avec  $\alpha \in (0, 1/\sqrt{2})$  et que la suite  $\mathbf{f}$  est bornée dans  $L^4(\mu)$ , i.e.  $\sup_{\ell \in \mathbb{N}} \|f_\ell\|_{L^4(\mu)} < +\infty$ .*

*Alors on a la convergence en loi suivante :*

$$N_{n,\emptyset}(\mathbf{f}) \xrightarrow[n \rightarrow +\infty]{(d)} G,$$

où  $G$  est une variable aléatoire Gaussienne centrée de variance  $\Sigma^{\text{sub}}(\mathbf{f})$  définie en (5.12), qui est bien définie et finie.

En considérant les suites définies en (5.10) et (5.11), nous obtenons aisément le corollaire suivant.

**Corollaire 5.1.** *Soit  $f \in \mathcal{B}(S)$  et  $X$  une chaîne de Markov bifurcante de probabilité de transition  $\mathcal{P}$  et de distribution initiale  $\nu$  telles que, avec  $\alpha \in (0, 1/\sqrt{2})$ , soit les Hypothèses 5.1 et 5.3 sont satisfaites et  $f \in F$ , ou alors les Hypothèses 5.5 et 5.6 sont satisfaites et  $f \in L^4(\mu)$ . Alors, on a les convergences en loi suivantes :*

$$|\mathbb{G}_n|^{-1/2} M_{\mathbb{G}_n}(\tilde{f}) \xrightarrow[n \rightarrow +\infty]{(d)} G_1 \quad \text{et} \quad |\mathbb{T}_n|^{-1/2} M_{\mathbb{T}_n}(\tilde{f}) \xrightarrow[n \rightarrow +\infty]{(d)} G_2,$$

où  $G_1$  et  $G_2$  sont des variables aléatoires Gaussiennes centrées de variances respectives  $\Sigma_{\mathbb{G}}^{\text{sub}}(f)$  et  $\Sigma_{\mathbb{T}}^{\text{sub}}(f)$  définies par

$$\Sigma_{\mathbb{G}}^{\text{sub}}(f) = \langle \mu, \tilde{f}^2 \rangle + \sum_{k \geq 0} 2^k \langle \mu, \mathcal{P} \left( \mathcal{Q}^k \tilde{f} \otimes^2 \right) \rangle \quad \text{et} \quad \Sigma_{\mathbb{T}}^{\text{sub}}(f) = \Sigma_{\mathbb{G}}^{\text{sub}}(f) + 2 \Sigma_{\mathbb{T},2}^{\text{sub}}(f), \quad (5.13)$$

avec

$$\Sigma_{\mathbb{T},2}^{\text{sub}}(f) = \sum_{k \geq 1} \langle \mu, \tilde{f} \mathcal{Q}^k \tilde{f} \rangle + \sum_{\substack{k \geq 1 \\ r \geq 0}} 2^r \langle \mu, \mathcal{P} \left( \mathcal{Q}^r \tilde{f} \otimes_{\text{sym}} \mathcal{Q}^{r+k} \tilde{f} \right) \rangle.$$

*Aperçu rapide de la preuve du Théorème 5.1* (voir [20], Section 4, et [19], Section 5, pour plus de détails). Soit  $(p_n, n \in \mathbb{N})$  une suite décroissante d'éléments de  $\mathbb{N}^*$  telle que, pour tout  $\lambda > 0$  :

$$p_n < n, \quad \lim_{n \rightarrow \infty} p_n/n = 1 \quad \text{et} \quad \lim_{n \rightarrow \infty} n - p_n - \lambda \log(n) = +\infty.$$

Rappelons l'ordre lexicographique sur  $\mathbb{T}$ , “ $\leq$ ”, défini au Chapitre 1. Pour  $i \in \mathbb{T}$ , nous définissons la  $\sigma$ -algèbre

$$\mathcal{F}_i = \{X_u; u \in \mathbb{T} \text{ tel que } u \leq i\}.$$

Par construction, on  $\mathcal{F}_i \subset \mathcal{F}_j$  dès que  $i \leq j$ . Nous définissons pour  $n \in \mathbb{N}$ ,  $i \in \mathbb{G}_{n-p_n}$  et  $\mathfrak{f} = (f_\ell, \ell \in \mathbb{N})$  les incrément de martingale

$$\Delta_{n,i}(\mathfrak{f}) = N_{n,i}(\mathfrak{f}) - \mathbb{E}[N_{n,i}(\mathfrak{f}) | \mathcal{F}_i] \quad \text{et} \quad \Delta_n(\mathfrak{f}) = \sum_{i \in \mathbb{G}_{n-p_n}} \Delta_{n,i}(\mathfrak{f}), \quad (5.14)$$

où

$$N_{n,i}(\mathfrak{f}) = |\mathbb{G}_n|^{-1/2} \sum_{\ell=0}^{p_n} M_{i\mathbb{G}_{p_n-\ell}}(\tilde{f}_\ell).$$

Alors, en utilisant la propriété de branchement, on obtient la décomposition

$$N_{n,\emptyset}(\mathfrak{f}) = \Delta_n(\mathfrak{f}) + R_0(n) + R_1(n), \quad (5.15)$$

où  $\Delta_n(\mathfrak{f})$  est défini en (5.14),

$$R_0(n) = |\mathbb{G}_n|^{-1/2} \sum_{k=0}^{n-p_n-1} M_{\mathbb{G}_k}(\tilde{f}_{n-k}) \quad \text{et} \quad R_1(n) = \sum_{i \in \mathbb{G}_{n-p_n}} \mathbb{E}[N_{n,i}(\mathfrak{f}) | \mathcal{F}_i]. \quad (5.16)$$

On montre premièrement que  $R_0(n)$  et  $R_1(n)$ , définis en (5.16), convergent en probabilité vers 0. Ensuite, on termine la preuve du Théorème 5.1 en montrant un théorème central limite pour  $\Delta_n(\mathfrak{f})$ . Pour cela, on utilise le théorème central limite pour les martingales. Les preuves des convergences en probabilités utilisent de façon intensive les many-to-one formulae et le contrôle des moments d'ordre 1, 2 et 4.  $\square$

### 5.2.2 Cas critique : $2\alpha^2 = 1$

Soit  $J$  l'ensemble des indices mentionné dans les Hypothèses 5.4 et 5.7. Pour  $j \in J$ , nous désignons par  $\mathcal{R}_j$  la projection sur l'espace propre associé à  $\alpha_j$  avec  $\alpha_j = \theta_j \alpha$ ,  $|\theta_j| = 1$ . Puisque  $\mathcal{Q}$  est un opérateur réel, on obtient que si  $\alpha_j$  est une valeur propre non réelle, alors son conjugué  $\bar{\alpha}_j$  est également une valeur propre. Nous désignerons par  $\bar{\mathcal{R}}_j$  la projection associée à  $\bar{\alpha}_j$ . Pour toute fonction  $f \in \mathcal{B}(S)$ , si ces quantités sont bien définies, nous posons

$$\hat{f} = \tilde{f} - \sum_{j \in J} \mathcal{R}_j(f) \quad \text{avec} \quad \tilde{f} = f - \langle \mu, f \rangle.$$

Pour une suite  $\mathbf{f} = (f_\ell, \ell \in \mathbb{N})$  de  $\mathcal{B}(S)$ , quand elles sont bien définies, nous considérons les quantités :

$$\Sigma^{\text{crit}}(\mathbf{f}) = \Sigma_1^{\text{crit}}(\mathbf{f}) + 2\Sigma_2^{\text{crit}}(\mathbf{f}), \quad (5.17)$$

où :

$$\begin{aligned} \Sigma_1^{\text{crit}}(\mathbf{f}) &= \sum_{k \geq 0} 2^{-k} \langle \mu, \mathcal{P}f_{k,k}^* \rangle = \sum_{k \geq 0} 2^{-k} \sum_{j \in J} \langle \mu, \mathcal{P}(\mathcal{R}_j(f_k) \otimes_{\text{sym}} \bar{\mathcal{R}}_j(f_k)) \rangle, \\ \Sigma_2^{\text{crit}}(\mathbf{f}) &= \sum_{0 \leq \ell < k} 2^{-(k+\ell)/2} \langle \mu, \mathcal{P}f_{k,\ell}^* \rangle, \end{aligned}$$

avec, pour  $k, \ell \in \mathbb{N}$  :

$$f_{k,\ell}^* = \sum_{j \in J} \theta_j^{\ell-k} \mathcal{R}_j(f_k) \otimes_{\text{sym}} \bar{\mathcal{R}}_j(f_\ell).$$

**Remarque 5.5.** Nous dirons qu'une suite  $\mathbf{f} = (f_\ell, \ell \in \mathbb{N})$  de  $F$  satisfait uniformément l'Hypothèse 5.4 s'il existe  $g \in F$  tel que :

$$|\mathcal{Q}^n(f_\ell)| \leq g, \quad |\mathcal{Q}^n(\tilde{f}_\ell)| \leq \alpha^n g \quad \text{et} \quad |\mathcal{Q}^n(\hat{f}_\ell)| \leq \beta_n \alpha^n g \quad \forall n, \ell \in \mathbb{N}. \quad (5.18)$$

**Théorème 5.2.** Soit  $X$  une chaîne de Markov bifurcante de probabilité de transition  $\mathcal{P}$  et de distribution initiale  $\nu$ . Soit  $\mathbf{f} = (f_\ell, \ell \in \mathbb{N})$  une suite de  $\mathcal{B}(S)$ .

- Dans le cadre ponctuel, on suppose que les Hypothèses 5.1 et 5.4 sont satisfaites avec  $\alpha = 1/\sqrt{2}$  et que la suite  $\mathbf{f}$  est une suite de  $F$  satisfaisant uniformément l'Hypothèse 5.4, i.e. (5.18) pour un  $g \in F$ .
- Dans le cadre  $L^2$ , on suppose que les Hypothèses 5.5 (avec  $k_0 \in \mathbb{N}$ ), 5.6 et 5.7 sont satisfaites avec  $\alpha = 1/\sqrt{2}$  et que la suite  $\mathbf{f}$  est bornée dans  $L^4(\mu)$ , i.e.  $\sup_{\ell \in \mathbb{N}} \|f_\ell\|_{L^4(\mu)} < +\infty$ .

Alors on a la convergence en loi suivante :

$$n^{-1/2} N_{n,\emptyset}(\mathbf{f}) \xrightarrow[n \rightarrow +\infty]{(d)} G,$$

où  $G$  est une variable aléatoire Gaussienne centrée de variance  $\Sigma^{\text{crit}}(\mathbf{f})$  définie en (5.17), qui est bien définie et finie.

Les grandes lignes de la preuve du Théorème 5.2 sont les mêmes que celles du Théorème 5.1. Nous renvoyons à [20], Section 5 pour plus de détails. Pour une fonction  $f \in \mathcal{B}(S)$ , nous posons

$$\Sigma_{\mathbb{G}}^{\text{crit}}(f) = \sum_{j \in J} \langle \mu, \mathcal{P}(\mathcal{R}_j(f) \otimes_{\text{sym}} \overline{\mathcal{R}}_j(f)) \rangle \quad \text{et} \quad \Sigma_{\mathbb{T}}^{\text{crit}}(f) = \Sigma_{\mathbb{G}}^{\text{crit}}(f) + 2\Sigma_{\mathbb{T},2}^{\text{crit}}(f), \quad (5.19)$$

où

$$\Sigma_{\mathbb{T},2}^{\text{crit}}(f) = \sum_{j \in J} \frac{1}{\sqrt{2}\theta_j - 1} \langle \mu, \mathcal{P}(\mathcal{R}_j(f) \otimes_{\text{sym}} \overline{\mathcal{R}}_j(f)) \rangle.$$

Alors, en considérant les suites définies en (5.10) et (5.11), nous obtenons aisément le corollaire suivant.

**Corollaire 5.2.** *Soit  $f \in \mathcal{B}(S)$  et  $X$  une chaîne de Markov bifurcante de probabilité de transition  $\mathcal{P}$  et de distribution initiale  $\nu$  telles que, avec  $\alpha = 1/\sqrt{2}$ , soit les Hypothèses 5.1 et 5.4 sont satisfaites et  $f \in F$ , soit les Hypothèses 5.5 (avec  $k_0 \in \mathbb{N}$ ), 5.6 et 5.7 sont satisfaites et  $f \in L^4(\mu)$ . Alors, on a les convergences en loi suivantes :*

$$(n|\mathbb{G}_n|)^{-1/2} M_{\mathbb{G}_n}(\tilde{f}) \xrightarrow[n \rightarrow \infty]{(d)} G_1, \quad \text{et} \quad (n|\mathbb{T}_n|)^{-1/2} M_{\mathbb{T}_n}(\tilde{f}) \xrightarrow[n \rightarrow \infty]{(d)} G_2,$$

où  $G_1$  et  $G_2$  sont des variables aléatoires Gaussiennes centrées de variances respectives  $\Sigma_{\mathbb{G}}^{\text{crit}}(f)$  et  $\Sigma_{\mathbb{T}}^{\text{crit}}(f)$  qui sont bien définies et finies.

### 5.2.3 Cas sur-critique : $2\alpha^2 > 1$

Nous allons travailler dans le cadre suivant.

- Pour l'approche ponctuelle, nous faisons les Hypothèses 5.1 et 5.4 avec  $\alpha \in (1/\sqrt{2}, 1)$  dans (5.5).
- Pour l'approche  $L^2$ , nous faisons l'Hypothèse 5.5 (ii) et l'Hypothèse 5.7 avec  $\alpha \in (1/\sqrt{2}, 1)$  dans (5.8).

Nous considérons la filtration  $\mathcal{H} = (\mathcal{H}_n, n \in \mathbb{N})$  définie par  $\mathcal{H}_n = \sigma(X_i, i \in \mathbb{T}_n)$ . Nous rappelons l'ensemble  $J$  et les projections  $\mathcal{R}_j$  définies dans les Hypothèses 5.4 et 5.7. Nous avons le lemme suivant.

**Lemme 5.1.** *Pour tout  $j \in J$  et pour  $f \in F$  ou  $f \in L^2(\mu)$  (selon qu'on est dans l'approche ponctuelle ou l'approche  $L^2$ ), la suite  $M_j(f) = (M_{n,j}(f), n \in \mathbb{N})$ , avec*

$$M_{n,j}(f) = (2\alpha_j)^{-n} M_{\mathbb{G}_n}(\mathcal{R}_j(f)),$$

*est une  $\mathcal{H}$ -martingale qui converge presque sûrement et dans  $L^2$  vers une variable aléatoire que nous notons  $M_{\infty,j}(f)$ .*

*Idées de la preuve.* La preuve du Lemme 5.1 utilise la décomposition suivante :

$$M_{\mathbb{G}_n}(f) = \sum_{u \in \mathbb{G}_{n-1}} (f(X_{u0}) + f(X_{u1})), \quad \forall f \in \mathcal{B}(S),$$

la propriété de branchement, les many-to-one formulae et le théorème de convergence des martingales.  $\square$

**Théorème 5.3.** *Soit  $X$  une chaîne de Markov bifurcante de probabilité de transition  $\mathcal{P}$  et de distribution initiale  $\nu$ . Soit  $\mathbf{f} = (f_\ell, \ell \in \mathbb{N})$  une suite de  $\mathcal{B}(S)$  telle que :*

- *dans le cadre ponctuel, la suite  $\mathbf{f}$  est une suite de  $F$  satisfaisant uniformément l'Hypothèse 5.4 ;*
- *dans le cadre  $L^2$ , la suite  $\mathbf{f}$  est bornée dans  $L^8(\mu)$ .*

*Alors, on a la convergence en probabilité suivante*

$$(2\alpha^2)^{-n/2} N_{n,\emptyset}(\mathbf{f}) - \sum_{\ell \in \mathbb{N}} (2\alpha)^{-\ell} \sum_{j \in J} \theta_j^{n-\ell} M_{\infty,j}(f_\ell) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

*Idées principales de la preuve (pour plus de détails, voir [20], Section 6).* Soit  $(\hat{p}_n, n \in \mathbb{N})$  une suite de nombres pairs telle que (pour  $n \geq 3$ ) :

$$\frac{5n}{6} < \hat{p}_n < n, \quad \lim_{n \rightarrow \infty} (n - \hat{p}_n) = \infty \quad \text{et} \quad \lim_{n \rightarrow \infty} \alpha^{-(n-\hat{p}_n)} \beta_{\hat{p}_n/2} = 0.$$

La preuve repose principalement sur la décomposition suivante

$$N_{n,\emptyset}(\mathbf{f}) = R_0(n) + R_4(n) + T_n^{(1)}(\mathbf{f}) + T_n^{(2)}(\mathbf{f}),$$

où, avec les notations (5.15) et (5.16),

$$\begin{aligned} R_0(n) &= |\mathbb{G}_n|^{-1/2} \sum_{k=0}^{n-\hat{p}_n-1} M_{\mathbb{G}_k}(\tilde{f}_{n-k}), \\ T_n^{(1)}(\mathbf{f}) &= |\mathbb{G}_n|^{-1/2} \sum_{i \in \mathbb{G}_{n-\hat{p}_n}} \sum_{\ell=0}^{\hat{p}_n} 2^{\hat{p}_n-\ell} \left( \mathcal{Q}^{\hat{p}_n-\ell}(\hat{f}_\ell) \right) (X_i), \\ T_n^{(2)}(\mathbf{f}) &= |\mathbb{G}_n|^{-1/2} \sum_{i \in \mathbb{G}_{n-\hat{p}_n}} \sum_{\ell=0}^{\hat{p}_n} 2^{\hat{p}_n-\ell} \alpha^{\hat{p}_n-\ell} \sum_{j \in J} \theta_j^{\hat{p}_n-\ell} \mathcal{R}_j(f_\ell)(X_i), \\ R_4(n) &= \sum_{i \in \mathbb{G}_{n-\hat{p}_n}} (N_{n,i}(\mathbf{f}) - \mathbb{E}[N_{n,i}(\mathbf{f}) | \mathcal{H}_{n-\hat{p}_n}]). \end{aligned}$$

On montre premièrement que  $R_0(n)$ ,  $R_4(n)$  et  $T_n^{(1)}(\mathbf{f})$  convergent en probabilité vers 0. On termine ensuite la preuve en montrant que

$$(2\alpha^2)^{-n/2} T_n^{(2)}(\mathbf{f}) - \sum_{\ell \in \mathbb{N}} (2\alpha)^{-\ell} \sum_{j \in J} \theta_j^{n-\ell} M_{\infty,j}(f_\ell) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0.$$

La preuve de ces convergences utilisent principalement le contrôle des moments d'ordre 1 et 2 à travers les many-to-one formulae.  $\square$

Comme conséquence direct du Théorème 5.3, nous avons le résultat suivant.

**Corollaire 5.3.** *Soit  $f \in F$  ou  $f \in L^8(\mu)$ , selon qu'on travaille avec l'approche ponctuelle ou l'approche  $L^2$ , et soit  $X$  une chaîne de Markov bifurcante de probabilité de transition  $\mathcal{P}$  et de distribution initiale  $\nu$ . On suppose que  $\alpha$  est l'unique valeur propre de module  $\alpha$  (et par conséquent,  $J$  est réduit à un singleton  $\{j_0\}$ ). Alors on a*

$$(2\alpha)^{-n} M_{\mathbb{G}_n}(\tilde{f}) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} M_{\infty, j_0}(f) \quad \text{and} \quad (2\alpha)^{-n} M_{\mathbb{T}_n}(\tilde{f}) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \frac{2\alpha}{2\alpha - 1} M_{\infty, j_0}(f),$$

où  $M_{\infty, j_0}(f)$  est la variable aléatoire définie dans le Lemme 5.1.

**Remarque 5.6.** *Dans le cas sur-critique, nous n'avons pas un théorème central limite au sens classique. Ceci est dû au fait que le centrage et la normalisation ne sont pas adaptés dans ce cas. Obtenir le “bon” théorème central limite dans le cas sur-critique fait partie de nos objectifs pour l'avenir.*

### 5.3 Extention au processus mère-filles

Soit  $X$  une chaîne de Markov bifurcante de probabilité de transition  $\mathcal{P}$  et de distribution initiale  $\nu$ . Dans cette section, nous nous plaçons exclusivement dans le cadre ponctuel. En particulier, nous rappelons que  $\mu$  est la mesure de probabilité invariante de  $\mathcal{Q}$ , la probabilité de transition de la chaîne de Markov bifurcante  $X$ . Nous considérons le processus mère-filles  $X^\Delta = (X_u^\Delta, u \in \mathbb{T})$  défini à la Section 1.1.4. En se rappelant de (1.1) et des notations de la Section 1.1.4, pour une suite  $\mathfrak{f} = (f_\ell, \ell \in \mathbb{N})$  d'éléments de  $\mathcal{B}(S^3)$ , nous posons

$$N_{n, \emptyset}(\mathfrak{f}) = |\mathbb{G}_n|^{-1/2} \sum_{\ell=0}^n M_{\mathbb{G}_{n-\ell}}(\tilde{f}_\ell), \quad (5.20)$$

où nous avons posé  $\tilde{f}_\ell = f_\ell - \langle \mu^\Delta, f_\ell \rangle = f_\ell - \langle \mu, \mathcal{P}f_\ell \rangle$ . Alors, en utilisant le caractère “Markovien bifurcant” de  $X^\Delta$  et les résultats de la Section précédente, nous sommes capables d'établir les résultats suivants.

#### 5.3.1 Cas sous critique : $2\alpha^2 < 1$

**Théorème 5.4.** *On suppose les Hypothèses 5.1 et 5.3 satisfaites avec  $\alpha \in (0, 1/\sqrt{2})$ . Pour toute suite  $\mathfrak{f} = (f_\ell, \ell \in \mathbb{N})$  d'éléments de  $\mathcal{B}(S^3)$  telle que  $(\mathcal{P}f_\ell, \ell \in \mathbb{N})$ ,  $(\mathcal{P}f_\ell^2, \ell \in \mathbb{N})$  et  $(\mathcal{P}f_\ell^4, \ell \in \mathbb{N})$  existent et sont des suites d'éléments de  $F$  qui satisfont (5.3) pour un  $g \in F$ , on a la convergence en loi suivante*

$$N_{n, \emptyset}(\mathfrak{f}) \xrightarrow[n \rightarrow +\infty]{(d)} G,$$

où  $G$  est une variable aléatoire gaussienne centrée de variance  $\Sigma^{\Delta, \text{sub}}(\mathfrak{f})$  définie par

$$\Sigma^{\Delta, \text{sub}}(\mathfrak{f}) = \Sigma_1^{\Delta, \text{sub}}(\mathfrak{f}) + 2 \Sigma_2^{\Delta, \text{sub}}(\mathfrak{f}),$$

avec  $\Sigma_1^{\Delta, \text{sub}}(\mathfrak{f})$  et  $\Sigma_2^{\Delta, \text{sub}}(\mathfrak{f})$  définies par :

$$\Sigma_1^{\Delta, \text{sub}}(\mathfrak{f}) = \sum_{\ell=0}^{+\infty} 2^{-\ell} \langle \mu, \mathcal{P}((\tilde{f}_\ell)^2) \rangle + \sum_{\ell \geq 0, k \geq 0} 2^{k-\ell} \langle \mu, \mathcal{P}(\mathcal{Q}^k(\mathcal{P}\tilde{f}_\ell) \otimes \mathcal{Q}^k(\mathcal{P}\tilde{f}_\ell)) \rangle,$$

et

$$\Sigma_2^{\Delta, \text{sub}}(\mathfrak{f}) = \Sigma_{2,1}^{\Delta, \text{sub}}(\mathfrak{f}) + \Sigma_{2,2}^{\Delta, \text{sub}}(\mathfrak{f}),$$

avec

$$\Sigma_{2,1}^{\Delta, \text{sub}}(\mathfrak{f}) = \sum_{0 \leq \ell < k} 2^{-\ell-1} \langle \mu, \mathcal{P}(\tilde{f}_k(\mathcal{Q}^{k-\ell-1}\mathcal{P}(\tilde{f}_\ell) \oplus \mathcal{Q}^{k-\ell-1}\mathcal{P}(\tilde{f}_\ell))) \rangle,$$

$$\Sigma_{2,2}^{\Delta, \text{sub}}(\mathfrak{f}) = \sum_{\substack{0 \leq \ell < k \\ r \geq 0}} 2^{r-\ell} \langle \mu, \mathcal{P}(\mathcal{Q}^r(\mathcal{P}(\tilde{f}_k)) \otimes_{\text{sym}} \mathcal{Q}^{r+k-\ell}(\mathcal{P}(\tilde{f}_\ell))) \rangle.$$

*Idée principale de la preuve.* On applique le Théorème 5.1 à la chaîne de Markov bifurcante  $X^\Delta$ .  $\square$

### Remarque 5.7.

1. Une conséquence du Théorème 5.4 est que si  $f \in \mathcal{B}(S^3)$  est telle que  $\mathcal{P}(f)$ ,  $\mathcal{P}(f^2)$  et  $\mathcal{P}(f^4)$  existent et appartiennent à  $F$ , alors  $M_{\mathbb{G}_n}(f - \mathcal{P}f)$  et  $M_{\mathbb{G}_n}(\mathcal{P}f)$  sont asymptotiquement indépendantes, ce qui n'est pas le cas pour  $M_{\mathbb{T}_n}(f - \mathcal{P}f)$  et  $M_{\mathbb{T}_n}(\mathcal{P}f)$ , et pour  $M_{\mathbb{G}_n}(f - \mathcal{P}f)$  et  $M_{\mathbb{G}_n}(f)$ .
2. Dans le cas spécial où la suite  $(f_\ell, \ell \in \mathbb{N})$  est telle que  $\mathcal{P}f_\ell = 0$  pour tout  $\ell \in \mathbb{N}$ , on a la convergence en loi suivante pour tout  $\alpha \in (0, 1)$  :

$$N_{n,\emptyset}(\mathfrak{f}) \xrightarrow[n \rightarrow \infty]{(d)} G, \quad (5.21)$$

où  $G$  est une variable aléatoire gaussienne centrée de variance  $\Sigma(\mathfrak{f})$  définie par

$$\Sigma(\mathfrak{f}) = \sum_{\ell=0}^{\infty} 2^{-\ell} \langle \mu, \mathcal{P}f_\ell^2 \rangle.$$

On peut observer dans ce cas spécial que la variance asymptotique ne dépend pas du taux d'ergodicité géométrique de la chaîne marquée. Ceci à pour conséquence la disparition des trois régimes de convergence de la Section précédente. Nous généralisons ainsi les résultats de Guyon [62] et les résultats de Delmas et Marsalle [44]. On retrouve en particulier que si  $f_1, \dots, f_d$  sont des éléments de  $\mathcal{B}(S^3)$  tels que  $\mathcal{P}f_\ell = 0$  et  $\mathcal{P}(f_\ell^2) \in F$  pour tout  $\ell \in \{1, \dots, d\}$ , alors les variables aléatoires  $|\mathbb{G}_n|^{-1/2} M_{\mathbb{G}_{n-\ell-1}}(f_\ell)$ ,  $\ell \in \{1, \dots, d\}$ , sont asymptotiquement indépendantes.

### 5.3.2 Les cas critique et sur-critique : $2\alpha^2 = 1$ et $2\alpha^2 > 1$ .

Dans le cas où les fonctions ne sont pas centrées par rapport à  $\mathcal{P}$ , on voit apparaître de nouveau les différents régimes de convergence. Soit  $\mathfrak{f} = \{f_\ell, \ell \in \mathbb{N}\}$  une suite de  $\mathcal{B}(S^3)$ . Nous rappelons l'ensemble  $J$ , les nombres complexes  $\theta_j$  et les projections  $\mathcal{R}_j$ ,  $j \in J$ , définies dans l'Hypothèse 5.4. Pour  $k, \ell \in \mathbb{N}$ , nous introduisons la notation suivante :

$$\mathcal{P}f_{k,\ell}^* = \sum_{j \in J} \theta_j^{\ell-k} \mathcal{R}_j(\mathcal{P}f_k) \otimes_{\text{sym}} \overline{\mathcal{R}}_j(\mathcal{P}f_\ell).$$

Alors nous avons le résultat de convergence qui suit.

**Théorème 5.5.** *Soit  $\mathfrak{f} = (f_\ell, \ell \in \mathbb{N})$  une suite d'éléments de  $\mathcal{B}(S^3)$  telle que les suites  $(\mathcal{P}f_\ell, \ell \in \mathbb{N})$ ,  $(\mathcal{P}f_\ell^2, \ell \in \mathbb{N})$  et  $(\mathcal{P}f_\ell^4, \ell \in \mathbb{N})$  sont des suites d'éléments de  $F$  qui satisfont (5.18) pour un  $g \in F$ .*

1. *Si les Hypothèses 5.1 et 5.4 sont satisfaites avec  $\alpha = 1/\sqrt{2}$ , alors on a*

$$n^{-1/2} N_{n,\emptyset}(\mathfrak{f}) \xrightarrow[n \rightarrow \infty]{(d)} G,$$

où  $G$  est une variable aléatoire réelle gaussienne centrée de variance  $\Sigma^{\Delta, \text{crit}}(\mathfrak{f})$  définie par

$$\Sigma^{\Delta, \text{crit}}(\mathfrak{f}) = \Sigma_1^{\Delta, \text{crit}}(\mathfrak{f}) + 2 \Sigma_2^{\Delta, \text{crit}}(\mathfrak{f}),$$

où

$$\Sigma_1^{\Delta, \text{crit}}(\mathfrak{f}) = \sum_{\ell=0}^{+\infty} 2^{-\ell} \langle \mu, \mathcal{P}(\mathcal{P}f_{\ell,\ell}^*) \rangle \quad \text{et} \quad \Sigma_2^{\Delta, \text{crit}}(\mathfrak{f}^\Delta) = \sum_{0 \leq \ell < k} 2^{-(k+\ell)/2} \langle \mu, \mathcal{P}(\mathcal{P}f_{k,\ell}^*) \rangle.$$

2. *Si les Hypothèses 5.1 et 5.4 sont satisfaites avec  $\alpha > 1/\sqrt{2}$ , alors on a*

$$(2\alpha^2)^{-n/2} N_{n,\emptyset}(\mathfrak{f}) - \sum_{\ell \in \mathbb{N}} (2\alpha)^{-\ell} \sum_{j \in J} \theta_j^{n-\ell} M_{\infty,j}(\mathcal{P}(f_\ell)) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

où les  $M_{\infty,j}(\mathcal{P}(f_\ell))$  sont définies dans le Lemme 5.1.

*Idées principales de la preuve.* La preuve utilise la décomposition suivante. Soit  $(s_n, n \in \mathbb{N})$  une suite de nombres réels convergeant vers 0 :

$$s_n N_{n,\emptyset}(\mathfrak{f}) = s_n |\mathbb{G}_n|^{-1/2} \sum_{\ell=0}^n M_{\mathbb{G}_{n-\ell}}(f_\ell - \mathcal{P}f_\ell) + s_n |\mathbb{G}_n|^{-1/2} \sum_{\ell=0}^n M_{\mathbb{G}_{n-\ell}}(\mathcal{P}f_\ell - \langle \mu, \mathcal{P}f_\ell \rangle),$$

Le premier terme du membre de droite converge en probabilité vers 0 grâce à (5.21) et le second terme permet d'obtenir le résultat du Théorème 5.5 grâce aux Théorèmes 5.2 et 5.3.  $\square$

**Remarque 5.8.** *Comme pour les théorèmes 5.2 et 5.3, la vitesse de convergence dans le Théorème 5.5 est plus rapide que celles des résultats donnés dans [44, 62]. En effet, comme nous allons l'observer dans les illustrations numériques, dans le cas sur-critique, cette vitesse est une fonction croissante du taux d'ergodicité géométrique. Une fois de plus, pour une fonction  $f \in \mathcal{B}(S^3)$  telle que  $\mathcal{P}(f) \neq 0$ , si toutes les projections de  $\mathcal{P}f$  sur les sous espaces propres associés aux valeurs propres  $\{\alpha_j, j \in J\}$  (voir Hypothèse 5.4) sont nulles, alors on fait des projections dans la famille des sous espaces propres suivantes, i.e. les sous espaces propres associés à la famille des valeurs propres dont le module suit directement celui des  $\alpha_j$ .*

### 5.3.3 Illustration numérique

Le but de cette Section est d'illustrer l'impact du taux d'ergodicité géométrique de la chaîne marquée sur la convergence en loi des fonctionnelles de chaînes de Markov bifurcantes. Nous considérons le processus BAR symétrique défini en (1.3) avec  $a = \alpha$ . Nous rappelons que  $\alpha$  est une valeur propre de multiplicité 1 et l'expression de la projection orthogonale sur le sous espace propre associé à  $\alpha$  est donnée dans (5.7). Afin d'illustrer l'effet du taux d'ergodicité géométrique sur les fluctuations, nous allons représenter, pour  $\mathbb{A}_n \in \{\mathbb{G}_n, \mathbb{T}_n\}$  et pour une fonction  $f$ , la pente  $b_{\alpha,n}$ , de la droite de régression de  $\log(\text{Var}(|\mathbb{A}_n|^{-1}M_{\mathbb{A}_n}(f)))$  en  $\log(|\mathbb{A}_n|)$ , en fonction du taux d'ergodicité géométrique  $\alpha$ . Pour les processus classiques (par exemple les variables aléatoires i.i.d.), les points sont situés autour de la droite horizontale  $y = -1$ . Dans notre cas et pour  $n$  assez grand, en conséquence des Corollaires 5.1, 5.2 et 5.3, la pente théorique espérée est  $b_{\alpha,n} \sim h_1(\alpha) = \log(\alpha^2 \vee 2^{-1})/\log(2)$  si la projection de  $f$  sur le sous espace propre associé à  $\alpha$  est non nulle. Par contre, dans le cas où la fonction  $f$  appartient au noyau de l'opérateur  $\mathcal{Q}$  définie en (5.6), on observe que cette pente n'est plus la bonne : il faut alors projeter  $f$  dans le sous espace propre qui suit.

Nous allons considérer les fonctions  $f$  définies sur  $\mathbb{R}$  par  $f(x) = x^p$ , pour  $p \in \{1, \dots, 4\}$  et les fonctions  $f_1$  et  $f_2$  définies sur  $\mathbb{R}^3$  par  $f_1(x, y, z) = x^2y$  et  $f_2(x, y, z) = xz$ . Comme on peut l'observer, la pente  $b_{\alpha,n}$  est une fonction croissante de  $\alpha$ . On peut aussi observer que les courbes présentent deux tendances avec une transition de phase :

- autour de  $\alpha = 1/\sqrt{2}$  pour  $f(x) = x^p$  (avec  $p \in \{1, 3\}$ ) et pour  $f_1$  ;
- autour de  $\alpha^2 = 1/\sqrt{2}$  pour  $f(x) = x^p$  (avec  $\{p \in \{2, 4\}\}$ ) et pour  $f_2$ .

Les différentes courbes ont des comportements standards pour  $\alpha < 1/\sqrt{2}$  (dans le cas  $f(x) = x^p$  (avec  $p \in \{1, 3\}$ ) et  $f_1$ ) ou pour  $\alpha^2 < 1/\sqrt{2}$  (dans le cas  $f(x) = x^p$  (avec  $\{p \in \{2, 4\}\}$ ) et  $f_2$ ) et un comportement non standard au-delà des seuils  $1/\sqrt{2}$  ou  $1/2^{1/4}$  suivant les cas. Pour le cas  $p \in \{2, 4\}$  et  $f_2$ , la bonne normalisation est  $(2\alpha^4)^{-n/2}$ . Ceci est dû au fait que pour ce cas, nous projetons les fonctions dans le sous espace propre associé à la valeur propre  $\alpha^2$ , puisque les projections de ces fonctions dans le deuxième sous espace propre sont nulles, ce qui intuitivement permet de dire que la pente théorique est la fonction  $h_2$  définie par  $h_2(\alpha) = \log(\alpha^4 \vee 2^{-1})/\log(2)$ .

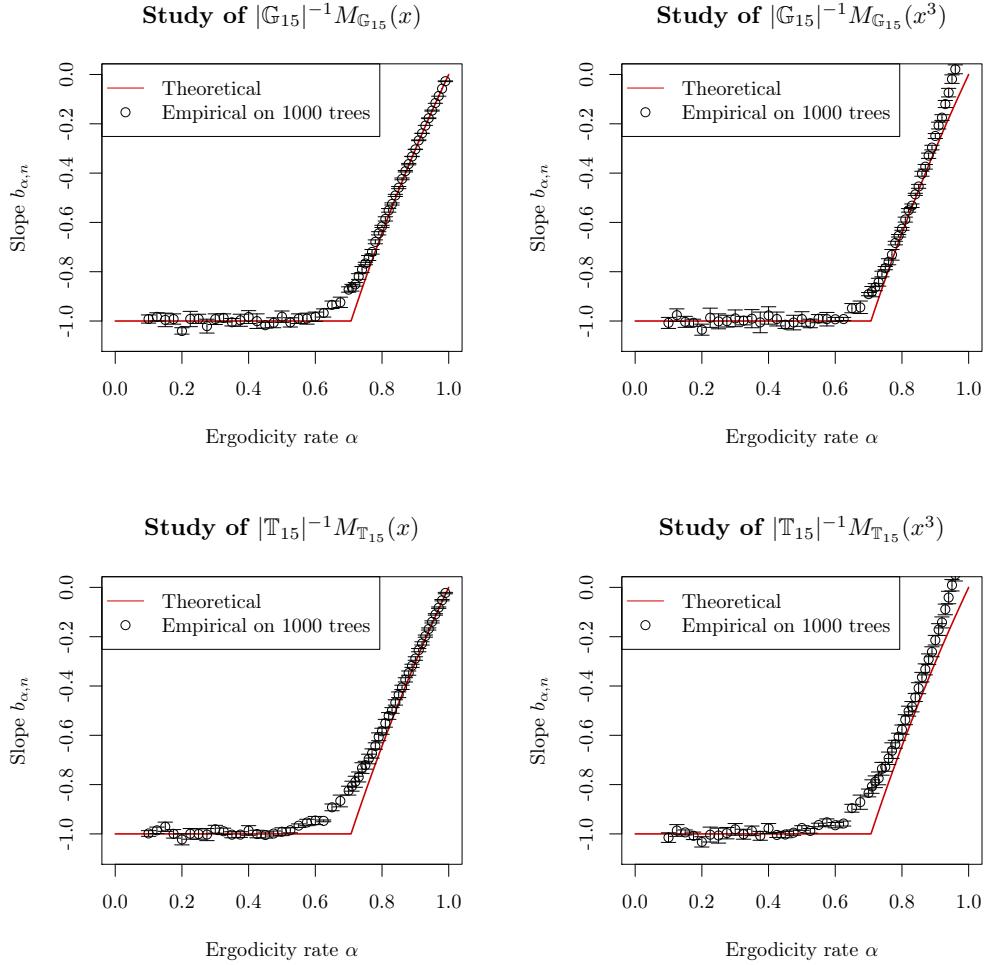


FIGURE 5.1 – Courbe de la pente  $b_{\alpha,n}$  de la droite de régression de  $\log(\text{Var}(|\mathbb{A}_n|^{-1} M_{\mathbb{A}_n}(f)))$  en  $\log(|\mathbb{A}_n|)$  en fonction du taux d'ergodicité géométrique  $\alpha$ , pour  $n = 15$ ,  $\mathbb{A}_n \in \{\mathbb{G}_n, \mathbb{T}_n\}$  et  $f(x) = x^p$  avec  $p \in \{1, 3\}$ . Dans ce cas, on a  $\mathcal{R}(f) \neq 0$ , où  $\mathcal{R}$  est le projecteur défini en (5.7). On peut voir que la courbe empirique (en noir) est proche de la courbe théorique attendue (en rouge) de la fonction  $h_1(\alpha) = \log(\alpha^2 \vee 2^{-1}) / \log(2)$  pour  $\alpha \in (0, 1)$ .

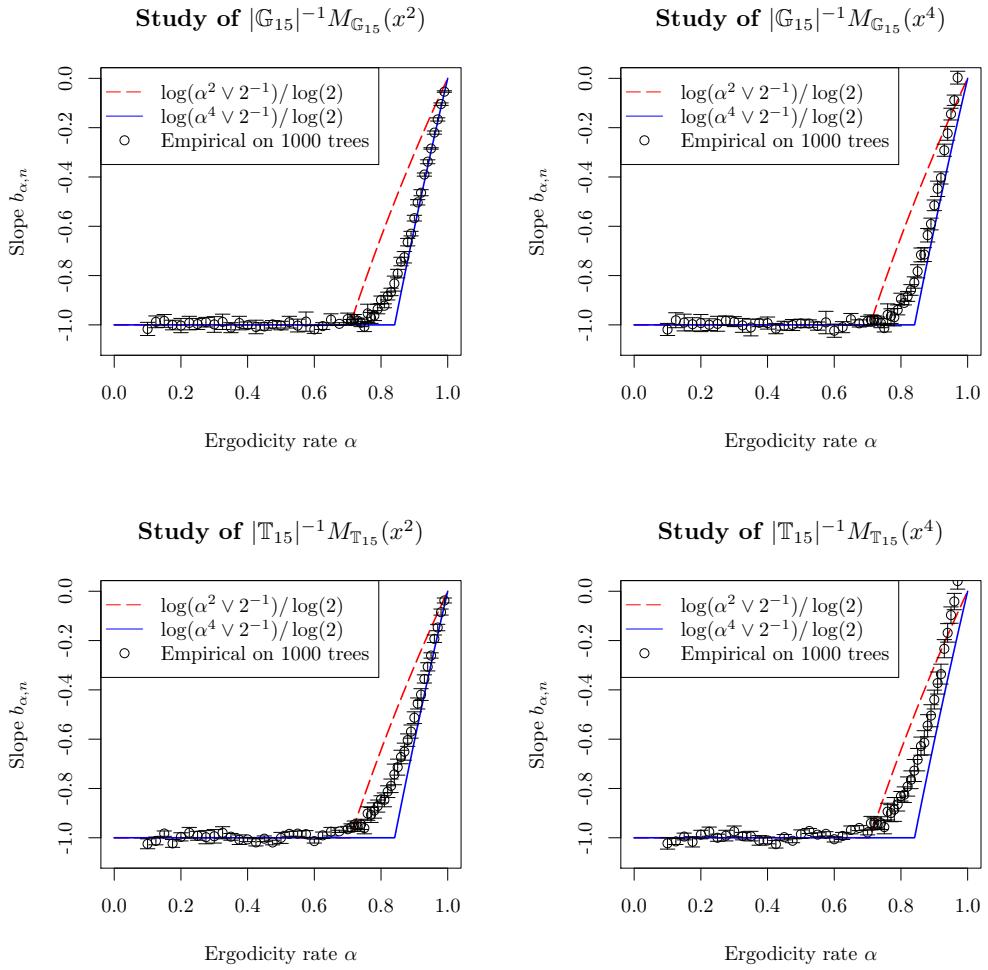


FIGURE 5.2 – Courbe de la pente  $b_{\alpha,n}$  de la droite de régression de  $\log(\text{Var}(|\mathbb{A}_n|^{-1}M_{\mathbb{A}_n}(f)))$  en  $\log(|\mathbb{A}_n|)$  en fonction du taux d'ergodicité géométrique  $\alpha$ , pour  $n = 15$ ,  $\mathbb{A}_n \in \{\mathbb{G}_n, \mathbb{T}_n\}$  et  $f(x) = x^p$  avec  $p \in \{2, 4\}$ . Dans ce cas,  $\mathcal{R}(f) = 0$ , où  $\mathcal{R}$  est le projecteur défini en (5.7). On peut observer que la courbe empirique (en noir) n'est plus proche de la courbe de la fonction  $h_1(\alpha) = \log(\alpha^2 \vee 2^{-1}) / \log(2)$  pour  $2\alpha^2 > 1$  (ligne pointillée en rouge) ; par contre elle proche de la courbe (en bleue) de la fonction  $h_2(\alpha) = \log(\alpha^4 \vee 2^{-1}) / \log(2)$  pour  $\alpha \in (0, 1)$ .

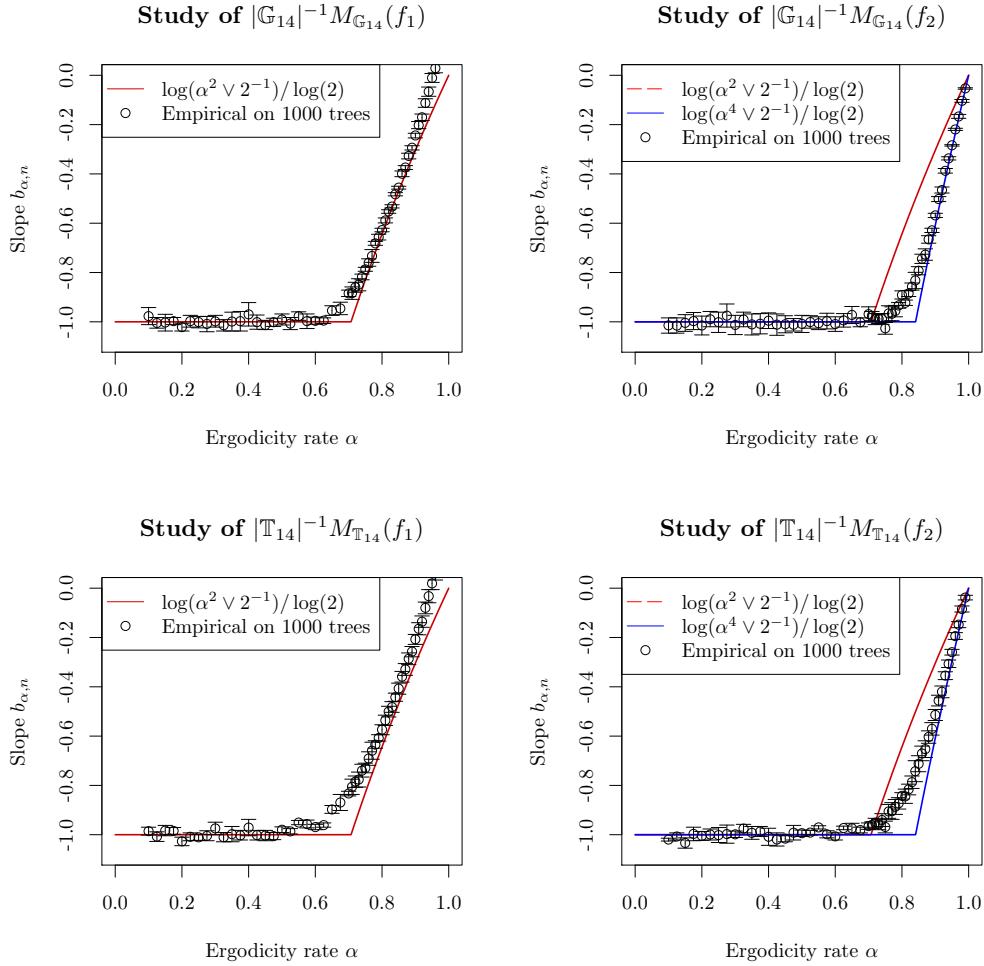


FIGURE 5.3 – Courbe (en noir) de la pente  $b_{\alpha,n}$  de la droite de régression de  $\log(\text{Var}(|\mathbb{A}_n|^{-1}M_{\mathbb{A}_n}(f)))$  en  $\log(|\mathbb{A}_n|)$  en fonction du taux d'ergodicité géométrique  $\alpha$ , pour  $n = 14$ ,  $\mathbb{A}_n \in \{\mathbb{G}_n, \mathbb{T}_n\}$  et  $f_1(x, y, z) = x^2y$ ,  $f_2(x, y, z) = xy$ . On a  $\mathcal{R}(\mathcal{P}f_1) \neq 0$  et  $\mathcal{R}(\mathcal{P}f_2) = 0$ . Pour les deux figures de gauche, on observe que la courbe empirique (en noir) est proche de la courbe (en rouge) de la fonction  $h_1(\alpha) = \log(\alpha^2 \vee 2^{-1})/\log(2)$  pour  $\alpha \in (0, 1)$ . Pour les deux figures de droite, on observe que la courbe empirique (en noir) s'écarte de la courbe de la fonction  $h_1(\alpha)$  pour  $\alpha > 1/\sqrt{2}$ , mais elle est proche de la courbe (en bleue) de la fonction  $h_2(\alpha) = \log(\alpha^4 \vee 2^{-1})/\log(2)$ .

**Remarque 5.9.** *À l'issue de tous ces calculs, une conclusion qui se dégage est que, à l'exception de quelques cas particuliers, les fonctionnelles de chaînes de Markov bifurcantes peuvent présenter deux types de comportement en fonction de la valeur du taux d'ergodicité géométrique (ou du taux de convergence en distance de Wasserstein) de la chaîne marquée avec un point de transition. Comme nous l'avons observer dans l'exemple précédent, le point de transition n'est en général pas universelle : il dépend des fonctions utilisées et de la nature du problème à traiter. Ces deux types de comportements sont l'illustration de ce que nous avons appelé “compétition entre lois et généalogie”. En effet, lorsque la loi du processus est prédominante, ce qui correspond aux “petites valeurs” du taux de convergence à l'équilibre de la chaîne marquée, les fonctionnelles de chaînes de Markov bifurcantes ont un comportement classique. Mais pour les “grandes valeurs” du taux de convergence à l'équilibre de la chaîne marquée, le comportement de ces fonctionnelles n'est plus classique ; on peut voir dans ce cas que la taille de la population a une grande influence sur leur comportement. Notons que cette nouvelle vision ouvre des perspectives nouvelles pour étendre les résultats du Chapitre 3 et pour étudier de nouvelles méthodes d'estimation non paramétrique pour les chaînes de Markov bifurcantes.*

## 5.4 Application à l'étude des estimateurs à noyau de la densité invariante et de la densité de la probabilité de transition

Afin de faciliter notre exposé, nous nous plaçons uniquement dans le cadre ponctuel ; les résultats et les techniques de preuves restent globalement les mêmes dans le cadre  $L^2$ . Dans la suite, nous posons  $S = \mathbb{R}^d$ ,  $d \geq 1$ , et nous munissons  $S$  de sa tribu borélienne  $\mathcal{B}_S$ . Nous posons  $\mathbb{A}_n \in \{\mathbb{G}_n, \mathbb{T}_n\}$ . Nous faisons les hypothèses supplémentaires suivantes.

**Hypothèse 5.8.** *La probabilité de transition  $\mathcal{P}$  admet une densité, que nous notons encore  $\mathcal{P}$ , par rapport à la mesure de Lebesgue.*

**Remarque 5.10.** *L'hypothèse 5.8 entraîne que les probabilités de transition  $\mathcal{P}_0$ ,  $\mathcal{P}_1$  et  $\mathcal{Q}$ , la mesure de probabilité invariante  $\mu$  et la mesure de probabilité  $\mu^\Delta$  (définie en (1.2)) admettent des densités, que nous notons encore  $\mathcal{P}_0$ ,  $\mathcal{P}_1$ ,  $\mathcal{Q}$ ,  $\mu$  et  $\mu^\Delta$  respectivement, par rapport à la mesure de Lebesgue (pour plus détails, nous renvoyons par exemple à [54], chap 6).*

**Hypothèse 5.9.** *La constante suivante est finie :*

$$C_0 = \sup_{x, x_0, x_1 \in S} (\mu(x) + \mathcal{Q}(x, x_0) + \mathcal{P}(x, x_0, x_1)).$$

**Remarque 5.11.** *Rappelons que l'intérêt de l'estimation des densités  $\mu$  et  $\mathcal{P}$  vient du fait que la probabilité invariante et la probabilité de transition d'une chaîne de Markov bifurcante sont inconnues en pratique, bien qu'elles jouent un rôle déterminant dans le*

comportement asymptotique et non asymptotique des fonctionnelles de chaînes de Markov bifurcantes.

#### 5.4.1 Définition des estimateurs de $\mu$ et $\mathcal{P}$

Nous observons  $\mathbb{X}^{\Delta n} = (X_u^{\Delta})_{u \in \mathbb{A}_n}$ , où  $\mathbb{A}_n \in \{\mathbb{G}_n, \mathbb{T}_n\}$  i.e. nous avons  $2^{n+2} - 1$  (ou  $3 \times 2^n$ ) variables aléatoires à valeurs dans  $S$ . Nous considérons une fonction intégrable  $K \in \mathcal{B}(S)$  telle que  $\int_S K(x) dx = 1$ . Notons que dans ce cas, on a  $K^{\otimes 3} \in \mathcal{B}(S^3)$  et  $\int_{S^3} K^{\otimes 3}(x, x_0, x_1) dx dx_0 dx_1 = 1$ . Soit  $h$  une fenêtre. Comme au Chapitre 2, nous définissons pour tout  $x, x_0, x_1 \in S$  :

$$\widehat{\mu}_{\mathbb{A}_n}(x) = \frac{1}{|\mathbb{A}_n| h^{d/2}} \sum_{u \in \mathbb{A}_n} K_h(x - X_u), \quad (5.22)$$

$$\begin{aligned} \widehat{\mu}_{\mathbb{A}_n}^{\Delta}(x, x_0, x_1) &= \frac{1}{|\mathbb{A}_n| h^{3d/2}} \sum_{u \in \mathbb{A}_n} K_h^{\otimes 3}(xx_0x_1 - X_u^{\Delta}), \\ \widehat{\mathcal{P}}_{\mathbb{A}_n}(x, x_0, x_1) &= \frac{\widehat{\mu}_{\mathbb{A}_n}^{\Delta}(xx_0x_1)}{\widehat{\mu}_{\mathbb{A}_n}(x)}, \end{aligned} \quad (5.23)$$

où  $K_h(x - y) = h^{-d/2} K(h^{-1}(x - y))$  et

$$K_h^{\otimes 3}(xx_0x_1 - yy_0y_1) = h^{-3d/2} K^{\otimes 3}(h^{-1}(x - y), h^{-1}(x_0 - y_0), h^{-1}(x_1 - y_1)),$$

où nous utilisons la convention que  $\widehat{\mathcal{P}}_{\mathbb{A}_n}(xx_0x_1) = 0$  si  $\widehat{\mu}_{\mathbb{A}_n}(x) = 0$ . Soulignons cependant si  $K$  est strictement positive, alors  $\widehat{\mu}_{\mathbb{A}_n}(x) > 0$  pour tout  $x \in S$ .

**Remarque 5.12.**

1. Nous avons omis  $n$  dans l'expression de la fenêtre “ $h$ ” afin de ne pas surcharger l'écriture. Aussi, les résultats que nous allons énoncer peuvent être aisément étendus aux cas des fenêtres anisotropes, i.e. en considérant un vecteur de fenêtres dont les composantes peuvent être différentes (voir par exemple (3.5)). Pour notre commodité, nous avons choisi de travailler avec les fenêtres isotropes.
2. La forme de l'expression de  $\widehat{\mu}$  donnée en (5.22) diffère légèrement de celles données en (2.6) et en (3.5). Nous avons adopté celle donnée en (5.22) afin d'être plus cohérent avec les expressions (5.9) et (5.20).

**Remarque 5.13.** Notons qu'en imitant (5.23), les densités  $\mathcal{P}_0$ ,  $\mathcal{P}_1$  et  $\mathcal{Q}$  peuvent respectivement être estimées par :

$$\widehat{P}_{\mathbb{A}_n, \eta}(x, y) = \frac{\widehat{\mu}_{\mathbb{A}_n}^{\Delta, \eta}(x, y)}{\widehat{\mu}_{\mathbb{A}_n}(x)} \quad \forall \eta \in \{0, 1\} \quad \text{et} \quad \widehat{\mathcal{Q}}_{\mathbb{A}_n}(x, y) = \frac{\widehat{\mu}_{\mathbb{A}_n}^{\mathcal{Q}}(x, y)}{\widehat{\mu}_{\mathbb{A}_n}(x)},$$

où  $\widehat{\mu}_{\mathbb{A}_n}$  est défini en (5.22) et

$$\begin{aligned}\widehat{\mu}_{\mathbb{A}_n}^{\Delta, \eta}(x, y) &= \frac{1}{|\mathbb{A}_n| h^d} \sum_{u \in \mathbb{A}_n} K_h(x - X_u) K_h(y - X_{u\eta}) \quad \text{et} \\ \widehat{\mu}_{\mathbb{A}_n}^Q(x, y) &= \frac{1}{|\mathbb{A}_{n+1}^*| h^d} \sum_{u \in \mathbb{A}_{n+1}^*} K_h(x - X_{u^-}) K_h(y - X_u),\end{aligned}$$

où  $u^-$  désigne la mère de  $u$  et  $\mathbb{A}_{n+1}^* = \mathbb{T}_{n+1} \setminus \mathbb{G}_0$  si  $\mathbb{A}_n = \mathbb{T}_n$  et  $\mathbb{A}_{n+1}^* = \mathbb{G}_{n+1}$  si  $\mathbb{A}_n = \mathbb{G}_n$ . L'étude de  $\widehat{P}_{\mathbb{A}_n, \eta}$  et  $\widehat{\mathbb{Q}}_{\mathbb{A}_n}$  se déduit aisément de celle de  $\widehat{\mathbb{P}}_{\mathbb{A}_n}$ .

#### 5.4.2 Principaux résultats

Les hypothèses suivantes sont valables pour l'étude de l'estimateur  $\widehat{\mu}$  de la densité invariante.

##### Hypothèse 5.10.

(i) Le noyau  $K$  vérifie

$$\|K\|_\infty < +\infty, \|K\|_1 < +\infty, \|K\|_2 < +\infty, \quad \text{et} \quad \lim_{|x| \rightarrow +\infty} |x|K(x) = 0.$$

(ii) Il existe  $\gamma \in (0, 1/d)$  tel que la fenêtre est définie par  $h_n = 2^{-n\gamma}$  et le taux d'ergodicité géométrique vérifie  $2\alpha^2 < 2^{d\gamma}$ .

**Hypothèse 5.11.** l'Hypothèse 5.10 est satisfaite et il existe  $s > 0$  tel que :

(i)  $\mu \in \mathcal{H}_{\mathbb{R}^d}^s$ , où  $s = (s, \dots, s) \in \mathbb{R}^d$ .

(ii) Le noyau  $K$  est d'ordre  $(\lfloor s \rfloor, \dots, \lfloor s \rfloor) \in \mathbb{N}^d$  : on a  $\int_{\mathbb{R}^d} |x|^s K(x) dx < \infty$  et  $\int_{\mathbb{R}} x_j^k K(x) dx_j = 0$  pour tout  $k \in \{1, \dots, \lfloor s \rfloor\}$  et  $j \in \{1, \dots, d\}$ .

(iii) **Contrôle de la fenêtre** : On a  $\gamma > 1/(2s + d)$ , c'est-à-dire  $\lim_{n \rightarrow \infty} |\mathbb{G}_n| h_n^{2s+d} = 0$ .

Pour l'étude de l'estimateur  $\widehat{\mathbb{P}}$  de la densité de la probabilité de transition, nous avons besoin des hypothèses supplémentaires suivantes, qui sont une modification des Hypothèses 5.10 et 5.11 dûe au fait qu'on passe d'un espace de dimension  $d$  à un espace de dimension  $3d$ .

##### Hypothèse 5.12.

(i) Le noyau  $K$  vérifie

$$\|K\|_4 < +\infty.$$

(ii) Il existe  $\gamma \in (0, 1/3d)$  tel que la fenêtre est définie par  $h_n = 2^{-n\gamma}$  et le taux d'ergodicité géométrique  $\alpha$  vérifie  $2\alpha^2 < 2^{3d\gamma}$ .

**Hypothèse 5.13.** L'hypothèse 5.12 est satisfaite et il existe  $s > 0$  tel que :

- (i)  $\mu^\Delta \in \mathcal{H}_{\mathbb{R}^{3d}}^{\mathbf{s}}$ , où  $\mathbf{s} = (s, \dots, s) \in \mathbb{R}^{3d}$ .
- (ii) **Contrôle de la fenêtre** : On a  $\gamma > 1/(2s + 3d)$ , c'est-à-dire  $\lim_{n \rightarrow \infty} |\mathbb{G}_n| h_n^{2s+3d} = 0$ .

Nous avons le résultat suivant.

**Théorème 5.6.** Soit  $X$  une chaîne de Markov bifurcante de probabilité de transition  $\mathcal{P}$  et de mesure initiale  $\nu$  telle que les Hypothèses 5.1, 5.3, 5.8 et 5.9 soient satisfaites.

- (i) Sous les Hypothèses additionnelles 5.10 et 5.11, on a, pour tout  $x \in \mathbb{R}^d$ ,

$$\widehat{\mu}_{\mathbb{A}_n}(x) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mu(x), \quad (5.24)$$

$$\sqrt{|\mathbb{A}_n| h_n^d} (\widehat{\mu}_{\mathbb{A}_n}(x) - \mu(x)) \xrightarrow[n \rightarrow \infty]{(d)} G, \quad (5.25)$$

où  $G$  est une variable aléatoire réelle gaussienne centrée de variance  $\|K\|_2^2 \mu(x)$ .

- (ii) Sous les Hypothèses additionnelles 5.10, 5.12 et 5.13, on a, pour tout  $x, x_0, x_1 \in \mathbb{R}^d$ ,

$$\begin{aligned} \widehat{\mathcal{P}}_{\mathbb{A}_n}(x, x_0, x_1) &\xrightarrow[n \rightarrow \infty]{\mathbb{P}} \mathcal{P}(x, x_0, x_1) \\ \sqrt{|\mathbb{A}_n| h_n^{3d}} (\widehat{\mathcal{P}}_{\mathbb{A}_n}(x, x_0, x_1) - \mathcal{P}(x, x_0, x_1)) &\xrightarrow[n \rightarrow \infty]{(d)} G, \end{aligned}$$

où  $G$  est une variable aléatoire réelle gaussienne centrée de variance  $\|K\|_2^6 \mathcal{P}(x, x_0, x_1)/\mu(x)$ .

Principales idées de la preuve (voir [21] Section 4.2 et [18] Section 4 pour plus de détails). Pour  $\mathbb{A}_n = \mathbb{T}_n$ , on a la décomposition suivante

$$\widehat{\mu}_{\mathbb{T}_n}(x) - \mu(x) = \frac{\sqrt{|\mathbb{G}_n|}}{|\mathbb{T}_n| h_n^{d/2}} N_{n,\emptyset}(\mathfrak{f}_n) + B_{h_n}(x),$$

où  $\mathfrak{f}_n = (f_{\ell,n}, \ell \in \mathbb{N})$  avec les fonctions  $f_{\ell,n}$  définies pour tout  $y \in \mathbb{R}^d$  par

$$f_{\ell,n}(y) = \begin{cases} K_h(x - y) = h^{-d/2} K\left(\frac{x-y}{h}\right) & \text{si } n \geq \ell \geq 0 \\ 0 & \text{sinon,} \end{cases}$$

la somme  $N_{n,\emptyset}$  est définie en (5.20) avec  $\mathfrak{f}_n$  à la place de  $\mathfrak{f}$  et le terme de biais  $B_{h_n}$  est défini par

$$B_{h_n}(x) = \frac{1}{|\mathbb{T}_n| h_n^{d/2}} \sum_{\ell=0}^n 2^{n-\ell} \langle \mu, f_{\ell,n} \rangle - \mu(x) = \langle \mu, h_n^{-d} K(h_n^{-1}(x - \cdot)) \rangle - \mu(x).$$

Alors (i) du Théorème 5.6 s'obtient en montrant que  $N_{n,\emptyset}(\mathfrak{f}_n)$  vérifie le résultat attendu et que

$$\lim_{n \rightarrow \infty} |\mathbb{T}_n|^{1/2} h_n^{d/2} B_{h_n}(x) = 0. \quad (5.26)$$

La preuve de (5.26) est classique. La preuve de la convergence en loi de  $N_{n,\emptyset}(\mathfrak{f}_n)$  suit les grandes lignes de celle obtenue au Théorème 5.1, avec  $\mathfrak{f}_n$  à la place de  $\mathfrak{f}$ .

Pour  $\mathbb{A}_n = \mathbb{G}_n$ , la démarche est la même avec les fonctions  $f_{\ell,n}$  définies par

$$f_{\ell,n}(y) = \begin{cases} K_h(x-y) = h^{-d/2}K\left(\frac{x-y}{h}\right) & \text{si } \ell = 0 \\ 0 & \text{sinon.} \end{cases}$$

La preuve de (ii) du Théorème 5.6 utilise de la décomposition suivante :

$$\begin{aligned} \sqrt{|\mathbb{A}_n| h_n^{3d}} (\widehat{\mathcal{P}}_{\mathbb{A}_n}(xx_0x_1) - \mathcal{P}(xx_0x_1)) &= (|\mathbb{A}_n| h_n^{3d})^{1/2} \left( \frac{\widehat{\mu}_{\mathbb{A}_n}^\Delta(xx_0x_1)}{\widehat{\mu}_{\mathbb{A}_n}(x)} - \frac{\mu^\Delta(xx_0x_1)}{\widehat{\mu}_{\mathbb{A}_n}(x)} \right) \\ &\quad - \frac{\mu^\Delta(xx_0x_1)}{\mu(x)\widehat{\mu}_{\mathbb{A}_n}(x)} (|\mathbb{A}_n| h_n^{3d})^{1/2} (\widehat{\mu}_{\mathbb{A}_n}(x) - \mu(x)). \end{aligned}$$

En utilisant des inégalités du type (3.6), la convergence (5.24) et le Lemme de Slutsky, on montre que le second terme du membre de droite de l'égalité précédente converge en probabilité vers 0. Le premier terme du membre de droite de l'égalité précédente vérifie le résultat attendu en établissant un résultat similaire à (5.25) pour  $\mu^\Delta$  à la place de  $\mu$  et en utilisant le Lemme de Slutsky.  $\square$

**Remarque 5.14.** *On voit ainsi apparaître deux régimes pour le choix de la fenêtre. En effet, comme on peut l'observer, les Hypothèses 5.10 (ii) et 5.12 (ii) sont satisfaites automatiquement si  $2\alpha^2 \leq 1$  et dans ce cas, le choix de la fenêtre est classique. Pour  $2\alpha^2 > 1$ , le choix de la fenêtre devient moins classique, puisque dans ce cas, la fenêtre est fonction du taux  $\alpha$  à travers les relations  $2\alpha^2 < 2^{d\gamma}$  et  $2\alpha^2 < 2^{3d\gamma}$ . On observe ainsi que le taux d'ergodicité géométrique de la chaîne marquée se comporte comme un paramètre de régularité, au même titre que la régularité  $\mathbf{s}$  des Hypothèses 5.11 et 5.13. Une nouvelle question émerge alors de cette observation : “construire une méthode adaptative prenant en compte le fait que  $\mathbf{s}$  et  $\alpha$  sont des paramètres inconnus”. L'exploration de cette question fait l'objet d'une étude que j'ai entrepris avec Marc Hoffmann et Angelina Roche. Soulignons que ces deux régimes dans le choix de la fenêtre sont en adéquation avec les conclusions des Sections 5.2 et 5.3, où nous avons observé que les chaînes de Markov bifurcantes ont un “comportement classique” pour les valeurs du taux  $\alpha$  plus petit qu'un certain seuil (en général  $1/\sqrt{2}$ ) et des comportements non habituels au delà de ce seuil. Cependant, le Théorème 5.6 semble nous dire qu'il y a une disparition des différents régimes pour les fluctuations des estimateurs  $\widehat{\mu}_{\mathbb{A}_n}$  et  $\widehat{\mathcal{P}}_{\mathbb{A}_n}$ . En fait, il n'en est rien puisque, pour  $\widehat{\mu}_{\mathbb{A}_n}$  par exemple, le théorème central limite a été établi pour  $2\alpha^2 < 2^{d\gamma}$ . Quand on pousse plus loin les calculs, on observe que pour  $2\alpha^2 \geq 2^{d\gamma}$ , l'erreur quadratique de  $\widehat{\mu}_{\mathbb{A}_n}$  est de l'ordre de  $\alpha^{2n}$  (différent de  $\sqrt{|\mathbb{A}_n| h_n^d}$ ). Il se trouve alors qu'on a toujours deux régimes. Mais ici, le point de la transition de phase, qui est une fonction de la régularité  $\mathbf{s}$  des fonctions inconnues, diffère des points évoqués aux Sections 5.2 et 5.3.*

### 5.4.3 Sélection de la fenêtre dans le cas particulier $S = \mathbb{R}$

Nous avons développé deux méthodes basiques pour sélectionner les fenêtres : une méthode de validation croisée et une méthode de type Silvermann. Pour les illustration numériques, nous renvoyons à [21], Sections 3.4 et 3.6 et [18], Section 5.3.

**Remarque 5.15.** *Notons que puisque l'estimateur de la densité de transition  $\mathcal{P}$  est un quotient, voir (5.23), on sélectionne indépendamment une fenêtre pour le numérateur et une fenêtre pour le dénominateur. Cette méthode, connue sous le nom de “l'approche à deux fenêtres”, nous a été inspirée par les travaux de Comte et Marie [38]. Nous renvoyons à plus tard une étude théorique approfondie de cette méthode pour les chaînes de Markov bifurcantes.*

#### Sélection des fenêtres par validation croisée

Nous choisissons les fenêtres qui minimisent les erreurs quadratiques moyennes intégrées

$$\mathbb{E} \left[ \int_{\mathbb{R}^3} (\widehat{\mu}_{\mathbb{G}_n}^\Delta - \mu^\Delta)^2 (xx_0x_1) dx dx_0 dx_1 \right] \quad \text{et} \quad \mathbb{E} \left[ \int_{\mathbb{R}} (\widehat{\mu}_{\mathbb{G}_n} - \mu)^2(x) dx \right],$$

où  $\widehat{\mu}_{\mathbb{G}_n}^\Delta$  et  $\widehat{\mu}_{\mathbb{G}_n}$  sont définies en (5.22) et (5.23). Ceci revient à minimiser les fonctions  $J^\Delta$  et  $J$  définies par

$$J^\Delta(h) = \mathbb{E} \left[ \int_{\mathbb{R}^3} (\widehat{\mu}_{\mathbb{G}_n}^\Delta)^2 (xx_0x_1) dx dx_0 dx_1 \right] - 2\mathbb{E} \left[ \int_{\mathbb{R}^3} (\widehat{\mu}_{\mathbb{G}_n}^\Delta \mu^\Delta)(x, x_0, x_1) dx dx_0 dx_1 \right]$$

et

$$J(h) = \mathbb{E} \left[ \int_{\mathbb{R}} (\widehat{\mu}_{\mathbb{G}_n})^2(x) dx \right] - 2\mathbb{E} \left[ \int_{\mathbb{R}} (\widehat{\mu}_{\mathbb{G}_n} \mu)(x) dx \right].$$

Nous renvoyons à [18], Section 5.1, pour plus de détails sur la méthode.

**Remarque 5.16.** *L'avantage de la méthode de validation croisée est qu'elle ne nécessite pas la connaissance du taux d'ergodicité géométrique de la chaîne marquée, même si on peut observer son influence dans les calculs. Dans l'avenir, une étude plus approfondie de cette méthode pour les chaînes de Markov bifurcantes permettra de voir comment agit le taux d'ergodicité géométrique de la chaîne marquée. Le principal inconvénient de cette méthode est le temps de calcul extrêmement couteux qu'il nécessite.*

#### Sélection de la fenêtre par une règle de type Silvermann

Nous avons également développer une règle de sélection “à la Silvermann”, [97], en nous appuyant sur le modèle BAR symétrique défini en (1.3) avec  $\rho = 0$ . Nous prenons avantage du fait que pour ce modèle, la densité  $\mathcal{P}$  de la probabilité de transition et les densités invariantes  $\mu$  et  $\mu^\Delta$  sont connues. Nous nous comportons comme si les densités  $\mu$  et  $\mu^\Delta$  sont inconnues, excepté que ces fonctions sont deux fois continûment dérивables. La méthode

repose alors sur le contrôle asymptotique du risque quadratique intégré des estimateurs  $\hat{\mu}$  et  $\hat{\mu}^\Delta$  des densités invariantes  $\mu$  et  $\mu^\Delta$ . Une estimation du paramètre d'ergodicité géométrique est également nécessaire. Pour cela, nous nous sommes inspirés des travaux de Gyori et Paulin [64] pour proposer l'estimateur suivant pour le paramètre d'ergodicité géométrique  $\alpha$  :

$$\hat{\alpha} = \left( \frac{\sum_{u \in \mathbb{G}_{n-m+1}} \sum_{v \in u\mathbb{G}_{m-1}} (X_u - \bar{X})(X_v - \bar{X})}{\sum_{u \in \mathbb{G}_n} (X_u - \bar{X})^2} \right)^{1/m} \quad \text{avec} \quad \bar{X} = \frac{1}{|\mathbb{G}_n|} \sum_{u \in \mathbb{G}_n} X_u,$$

où  $m \in \mathbb{N}$  est assez grand tel que  $m = \mathcal{O}(n)$ . En général, le choix  $m = \lfloor n/2 \rfloor + 1$  semble être pertinent. Ainsi, pour l'estimation de  $\mu$ , une approximation de la fenêtre optimale est

$$\hat{h} = |\mathbb{G}_n|^{-1/5} \hat{\sigma} \mathbf{1}_{\{2\hat{\alpha}^2 \leq 2^{1/5}\}} + (2\hat{\alpha}^2)^{-n} \hat{\sigma} \mathbf{1}_{\{2\hat{\alpha}^2 > 2^{1/5}\}}, \quad (5.27)$$

et pour l'estimation de  $\mu^\Delta$ , une approximation de la fenêtre optimale est

$$\hat{h} = |\mathbb{G}_n|^{-1/7} \hat{\sigma} \mathbf{1}_{\{2\hat{\alpha}^2 \leq 2^{3/7}\}} + (2\hat{\alpha}^2)^{-n/3} \hat{\sigma} \mathbf{1}_{\{2\hat{\alpha}^2 > 2^{3/7}\}}, \quad (5.28)$$

où  $\hat{\sigma}$  est l'estimateur standard de l'écart-type. Cette méthode est une adaptation de la règle du pouce développée dans [97]. Une des nouveautés ici est que nous prenons en compte l'estimation du taux d'ergodicité géométrique dans la procédure d'estimation. Nous renvoyons à [18], Section 5.2, pour plus de détails sur la méthode.

### Remarque 5.17.

1. *Contrairement au cas classique, les estimateurs définis en (5.27) et (5.28) ont deux parties, qui matérialisent la compétition entre la loi du processus et la généalogie. Ces estimateurs se généralisent aisément au cas où  $S = \mathbb{R}^d$ ,  $d \geq 1$ .*
2. *La principale difficulté dans notre règle “à la Silvermann” réside dans l'estimation du taux d'ergodicité géométrique où tout reste à faire. Plusieurs estimateurs du taux d'ergodicité géométrique ont été proposés dans la littérature des chaînes de Markov classiques. L'estimateur que nous proposons ici est obtenu en interprétant ce taux comme un coefficient de corrélation entre deux générations.*

# Chapitre 6

## Principes de déviations modérées pour les chaînes de Markov bifurcantes

Ce chapitre est issu des articles [25], écrit en collaboration avec Gorgui Gackou, et [16]. Nous complétons, d'une part les résultats de [23], où le principe de déviations modérées (PDM, en abrégé) avait été établi pour des fonctionnelles additives avec des fonctions centrées par rapport à la probabilité de transition  $\mathcal{P}$ ; et d'autre part, nous complétons les résultats du Chapitre 5.

### 6.1 Principe de déviation modérées pour les fonctionnelles additives de chaînes de Markov bifurcantes

Nous considérons une chaîne de Markov bifurcante  $(X_u, u \in \mathbb{T})$  à valeurs dans un espace mesurable  $(S, \mathcal{S})$ , de probabilité de transition  $\mathcal{P}$  et de mesure initiale  $\nu$ . Nous rappelons les définitions 2.2, 2.3 et 2.4. Nous faisons les hypothèses suivantes.

**Hypothèse 6.1.** *Il existe une mesure de probabilité  $\mu$  sur  $(S, \mathcal{S})$ , un nombre réel strictement positif  $M$  et  $\alpha \in (0, 1)$  tels que pour tout  $f \in \mathcal{B}_b(S)$  :*

$$|\mathcal{Q}^n f - \langle \mu, f \rangle| \leq M \alpha^n \|f\|_\infty \quad \text{pour tout } n \in \mathbb{N}.$$

**Hypothèse 6.2.** *Il existe une mesure de probabilité  $\mu$  sur  $(S, \mathcal{S})$ , un nombre réel strictement positif  $M$ ,  $\alpha \in (0, 1)$ , un ensemble fini d'indices non vides  $J$ , des valeurs propres complexes distinctes  $\{\alpha_j, j \in J\}$  de l'opérateur  $\mathcal{Q}$  avec  $|\alpha_j| = \alpha$ , des projections complexes non nulles  $\{\mathcal{R}_j, j \in J\}$  définies sur  $\mathbb{C}\mathcal{B}_b(S)$ , le  $\mathbb{C}$ -espace vectoriel engendré par  $\mathcal{B}_b(S)$ , tel que  $\mathcal{R}_j \circ \mathcal{R}_{j'} = \mathcal{R}_{j'} \circ \mathcal{R}_j = 0$  pour tout  $j \neq j'$  (de sorte que  $\sum_{j \in J} \mathcal{R}_j$  est aussi une projection définie sur  $\mathbb{C}\mathcal{B}_b(S)$ ) et une suite de nombres positifs  $(\beta_n, n \in \mathbb{N})$  convergeant vers 0, tels*

que pour tout  $f \in \mathcal{B}_b(S)$ , avec  $\theta_j = \alpha_j/\alpha$  :

$$\left| \mathcal{Q}^n(f) - \langle \mu, f \rangle - \alpha^n \sum_{j \in J} \theta_j^n \mathcal{R}_j(f) \right| \leq M \beta_n \alpha^n \|f\|_\infty \quad \text{pour tout } n \in \mathbb{N}.$$

**Remarque 6.1.** Notons que les Hypothèses 6.1 et 6.2 sont des versions fortes des Hypothèses 5.3, 5.4, 5.6 et 5.7. En particulier, si on suppose que  $S$  est un ensemble compact, alors les approches ponctuelle et  $L^2$  coïncident.

### 6.1.1 Principaux résultats

Soit  $\mathfrak{f} = (f_\ell, \ell \in \mathbb{N})$  une suite d'éléments de  $\mathcal{B}_b(S)$  telle que

$$\sup_{\ell \in \mathbb{N}} \{\|f_\ell\|_\infty\} = c_\infty < +\infty, \quad (6.1)$$

Nous rappelons la somme  $N_{n,\emptyset}(\mathfrak{f})$  définie en (5.9), les variances  $\Sigma^{\text{sub}}(\mathfrak{f})$  et  $\Sigma^{\text{crit}}(\mathfrak{f})$  définies en 5.12 et 5.17 respectivement. Nous considérons la fonction de taux  $I$  définie par

$$I(x) = \sup_{\lambda \in \mathbb{R}} \{\lambda x - \frac{1}{2} \lambda^2 \Sigma(\mathfrak{f})\} = \begin{cases} \frac{1}{2} \Sigma(\mathfrak{f})^{-1} x^2 & \text{si } \Sigma(\mathfrak{f}) \neq 0 \\ +\infty & \text{si } \Sigma(\mathfrak{f}) = 0, \end{cases} \quad (6.2)$$

où

$$\Sigma(\mathfrak{f}) = \begin{cases} \Sigma^{\text{sub}}(\mathfrak{f}) = \Sigma_1^{\text{sub}}(\mathfrak{f}) + 2\Sigma_2^{\text{sub}}(\mathfrak{f}) & \text{si } 2\alpha^2 < 1 \\ \Sigma^{\text{crit}}(\mathfrak{f}) = \Sigma_1^{\text{crit}}(\mathfrak{f}) + 2\Sigma_2^{\text{crit}}(\mathfrak{f}) & \text{si } 2\alpha^2 = 1. \end{cases}$$

Soit  $(b_n)_{n \in \mathbb{N}}$  une suite de nombres réels telle que :

$$\lim_{n \rightarrow \infty} b_n = \infty \quad \text{et} \quad \lim_{n \rightarrow \infty} \frac{b_n}{\sqrt{|\mathbb{G}_n|}} = 0,$$

et si  $2\alpha^2 = 1$ , on a en plus

$$\lim_{n \rightarrow \infty} b_n^2 n^2 2^{-n^a} = 0 \quad \text{pour un } a \in (0, 1).$$

Soit la suite  $(c_n)_{n \in \mathbb{N}}$  définie par

$$c_n = \begin{cases} b_n & \text{si } 2\alpha^2 < 1 \\ n^{1/2} b_n & \text{si } 2\alpha^2 = 1. \end{cases}$$

Alors nous avons le résultat suivant.

**Théorème 6.1.** Soit  $X$  une chaîne de Markov bifurcante de probabilité de transition  $\mathcal{P}$  et de mesure initiale  $\nu$  telle que

- pour  $2\alpha^2 < 1$ , l'Hypothèse 6.1 est satisfaite ;

— pour  $2\alpha^2 = 1$ , l'Hypothèse 6.2 est satisfaite.

Soit  $\mathfrak{f} = (f_\ell, \ell \in \mathbb{N})$  une suite d'éléments de  $\mathcal{B}_b(S)$  vérifiant (6.1). Alors  $c_n^{-1}N_{n,\emptyset}(\mathfrak{f})$  satisfait un principe de déviations modérées dans  $S$  de vitesse  $b_n^2$  et de fonction de taux  $I$  définie en (6.2). En particulier, on a

$$\lim_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}(|c_n^{-1}N_{n,\emptyset}(\mathfrak{f})| > \delta) = -I(\delta) \quad \forall \delta > 0.$$

Un aperçu rapide de la preuve (Voir [25], Section 5 et 6, pour plus de détails). Nous rappelons la décomposition 5.15. Alors, on termine la preuve en montrant ce qui suit :

$$\forall \delta > 0, \quad \lim_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}(|c_n^{-1}R_0(n)| > \delta) = -\infty; \quad (6.3)$$

$$\forall \delta > 0, \quad \lim_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}(|c_n^{-1}R_1(n)| > \delta) = -\infty; \quad (6.4)$$

$$c_n^{-1}\Delta_n(\mathfrak{f}) \quad \text{satisfait un PDM sur } S \text{ de vitesse } b_n^2 \text{ et de fonction de taux } I. \quad (6.5)$$

Les convergences (6.3) et (6.4) s'obtiennent à l'aide des inégalités de concentrations. La preuve de (6.5) utilise le principe de déviations modérées des martingales.  $\square$

### Remarque 6.2.

1. Notons que nous n'avons pas traiter le cas sur-critique  $2\alpha^2 > 1$ . Ce cas nécessite un centrage différent des fonctions. Nous renvoyons cette question à des travaux futurs.
2. Soit  $f \in \mathcal{B}_b(S)$ . Soit  $\mathfrak{f}_0$  et  $\mathfrak{f}$  définis en (5.10) et (5.11) respectivement. En prenant  $\mathfrak{f} = \mathfrak{f}_0$  ou  $\mathfrak{f} = \mathfrak{f}$ , on obtient que  $c_n^{-1}|\mathbb{G}_n|^{-1/2}M_{\mathbb{G}_n}(\tilde{f})$  et  $c_n^{-1}|\mathbb{T}_n|^{-1/2}M_{\mathbb{T}_n}(\tilde{f})$  satisfont un principe de déviations modérées dans  $S$  de vitesse  $b_n^2$  et de fonction de taux  $I$  définie en (6.2) où on a remplacé  $\Sigma(\mathfrak{f})$  par

$$\begin{cases} \Sigma_{\mathbb{G}}^{sub}(f)\mathbf{1}_{\{2\alpha^2 < 1\}} + \Sigma_{\mathbb{G}}^{crit}(f)\mathbf{1}_{\{2\alpha^2 = 1\}} & \text{pour } c_n^{-1}|\mathbb{G}_n|^{-1/2}M_{\mathbb{G}_n}(\tilde{f}) \\ \Sigma_{\mathbb{T}}^{sub}(f)\mathbf{1}_{\{2\alpha^2 < 1\}} + \Sigma_{\mathbb{T}}^{crit}(f)\mathbf{1}_{\{2\alpha^2 = 1\}} & \text{pour } c_n^{-1}|\mathbb{T}_n|^{-1/2}M_{\mathbb{T}_n}(\tilde{f}), \end{cases}$$

avec les variances asymptotiques  $\Sigma_{\mathbb{G}}^{sub}(f)$ ,  $\Sigma_{\mathbb{T}}^{sub}(f)$ ,  $\Sigma_{\mathbb{G}}^{crit}(f)$  et  $\Sigma_{\mathbb{T}}^{crit}(f)$  qui sont définies en (5.13) et (5.19).

### 6.1.2 Illustration numérique

Nous considérons la chaîne de Markov bifurcante  $(X_u, u \in \mathbb{T})$  à valeurs dans  $[0, 1]$ , de probabilité de transition  $\mathcal{P} = \mathcal{Q} \otimes \mathcal{Q}$  telle que la densité de  $\mathcal{Q}$ , que nous notons encore  $\mathcal{Q}$ ; est définie par

$$\mathcal{Q}(x, y) := (1-x) \frac{y(1-y)^2}{B(2, 3)} + x \frac{y^2(1-y)}{B(3, 2)}, \quad x, y \in [0, 1],$$

avec  $B(\alpha, \beta)$  la constante de normalisation de la loi Beta de paramètre  $\alpha$  et  $\beta$ . Pour notre commodité, nous choisissons  $X_\emptyset$  tel que  $\mathcal{L}(X_\emptyset) = Beta(2, 2)$ , où  $Beta(2, 2)$  est la loi Beta de paramètres  $(2, 2)$ . Le processus ainsi défini est stationnaire et sa loi invariante est une Beta de paramètres  $(2, 2)$ . Nous avons aussi  $\mathbb{E}[X_{u0}|X_u] = \mathbb{E}[X_{u1}|X_u] = X_u/5 + 2/5$ , (pour plus de détails, voir par exemple [87]). En utilisant le Theorem 2.1 dans [65], on montre que l’Hypothèse 6.1 est satisfaite avec  $\alpha = 1/5$ . Nous sommes donc dans le cas sous-critique. Nous allons premièrement illustrer le Théorème 6.1 avec la suite  $\mathbf{f} = (f, 0, 0, \dots)$  et la fonction  $f(x) = x$ . Dans ce cas, nous avons les résultats exacts suivants :

$$\langle \mu, f \rangle = \frac{1}{2}, \quad \mathcal{Q}^k \tilde{f}(x) = 5^{-k} (x - \frac{1}{2}) \quad \forall k \geq 0, \quad \Sigma_{\mathbb{G}}(f) = 6/115 \quad \text{and} \quad I(\delta) = \frac{115}{12} \delta^2.$$

Ensuite, nous illustrons le fait que la vitesse considérée dans le cas critique ne marche pas dans cet exemple (où nous sommes dans le cas sous-critique). Pour cela, nous simulons  $B = 50000$  échantillons ( $X^{(s)} = (X_u^{(s)}, u \in \mathbb{G}_{12}), s \in \{1, \dots, B\}$ ) d’une chaîne de Markov bifurcante à la  $n$ -ième génération, avec  $n = 12$ . Pour chaque échantillon  $X^{(s)}$ , nous calculons  $b_n^{-1} N_{n,\emptyset}^{(s)}(f) = b_n^{-1} |\mathbb{G}_n|^{-1/2} \sum_{u \in \mathbb{G}_n} (X_u^{(s)} - 1/2)$ . Enfin, pour différentes valeurs de  $\delta > 0$ , nous calculons  $b_n^{-2} \log(B^{-1} \sum_{s=1}^B \mathbf{1}_{\{|b_n^{-1} N_{n,\emptyset}^{(s)}(f)| > \delta\}})$ . Ce qui nous permet d’obtenir des valeurs empiriques pour la fonction de taux. Ensuite, sur le même repère, nous traçons la fonction de taux théorique et la fonction de taux empirique. Comme on peut l’observer sur la Figure 6.1, la fonction de taux empirique et la fonction de taux théorique coïncident, excepter sur la dernière figure où la fonction de taux empirique est proche de 0. En effet, pour cette figure, la vitesse considérée n’est pas valide en théorie dans le cas sous-critique ; elle est plutôt valide pour le cas critique. Notons que les différences observées entre les fonctions de taux empirique et théorique peuvent-être expliquées par le fait que la taille de l’échantillon n’est pas assez grande pour produire un nombre suffisant d’évènements de grandes déviations.

## 6.2 Application : Principe de déviations modérées pour l’estimateur à noyau de la densité invariante

Nous posons  $S = \mathbb{R}^d$ ,  $d \in \mathbb{N}^*$ . Les résultats et les techniques de la Section précédente nous permettent d’établir un principe de déviations modérées pour les estimateurs  $\hat{\mu}_{\mathbb{A}_n}$  définis en (5.22). Nous avons pour cela besoin, en plus des Hypothèses 5.10 et 5.11, de l’hypothèse supplémentaire suivante.

**Hypothèse 6.3.** *Avec les notations de l’Hypothèse 5.10 (ii), on suppose en plus que l’Hypothèse 6.1 est satisfaite avec*

$$\lim_{n \rightarrow +\infty} (2^{1-d\gamma} \alpha)^n = 0.$$

**Remarque 6.3.** *Contrairement à la Section 5.4.2 (voir en particulier Remarque 5.14) où le taux d’ergodicité géométrique  $\alpha$  de la chaîne marquée commence à influencer le choix de*

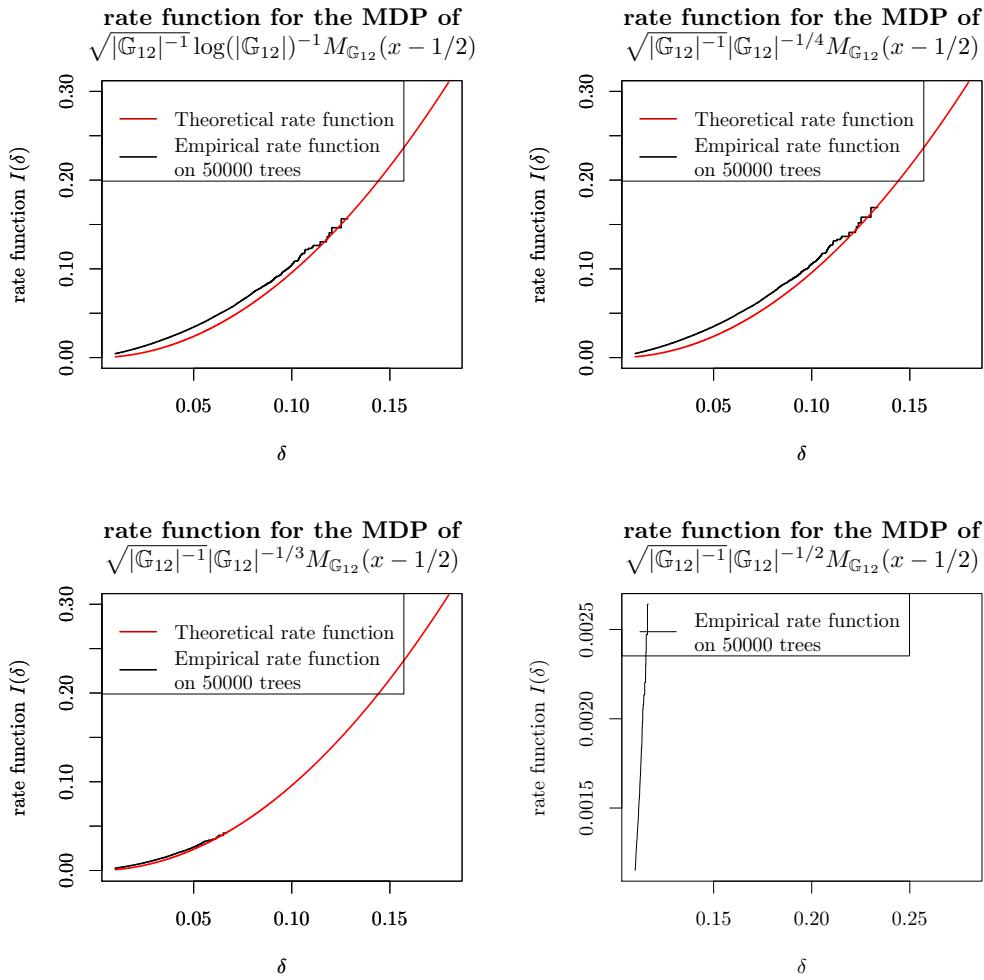


FIGURE 6.1 – Fonctions de taux théorique et exacte pour le principe de déviations modérées de  $|\mathbb{G}_n|^{-1/2} b_n^{-1} M_{\mathbb{G}_n}(x - 1/2)$ . Dans les trois premières figures, on peut observer que les courbes théoriques et empiriques coïncident, les différences pouvant être expliquées par le fait que la taille de l'échantillon n'est pas assez grande. Dans la dernière figure, on peut observer que la fonction de taux empirique est proche de 0, ce qui est normal puisque pour cette figure, nous avons considéré une vitesse valide pour le cas critique et pour notre processus nous sommes dans le cas sous-critique.

la fenêtre à partir de la valeur  $1/\sqrt{2}$ , nous observons, à travers l’Hypothèse 6.3, que pour le principe de déviations modérées, le taux d’ergodicité géométrique commence à impacter le choix de la fenêtre à partir de la valeur  $1/2$ . Ceci confirme en fait une observation que nous avions faite auparavant : “pour les études des estimateurs non paramétriques faisant intervenir des inégalités de concentrations, le point de transition se situe autour de  $\alpha = 1/2$ ”.

Soit  $(b_n, n \in \mathbb{N})$  une suite de nombres réels positifs telle que

$$\lim_{n \rightarrow +\infty} b_n = +\infty; \quad \lim_{n \rightarrow +\infty} \frac{n^{3/2} b_n}{\sqrt{|\mathbb{G}_n| h_n^d}} = 0; \quad \lim_{n \rightarrow +\infty} \frac{b_n}{\sqrt{|\mathbb{G}_n| h_n^{2s+d}}} = +\infty, \quad (6.6)$$

où  $s$  est le paramètre de régularité donné dans l’Hypothèse 5.11. Soit  $(\varpi_n, n \in \mathbb{N})$  une suite de nombres réels telle que  $\varpi_n \rightarrow 0$  quand  $n \rightarrow +\infty$ . Alors on a le résultat suivant.

**Théorème 6.2.** *Soit  $X$  une chaîne de Markov bifurcante de probabilité de transition  $\mathcal{P}$  et de mesure initiale  $\nu$  telle que les Hypothèses 5.8, 5.9, 5.10, 5.11 et 6.1 soient satisfaites. De plus, pour  $\alpha > 1/2$ , on suppose que l’Hypothèse 6.3 est satisfaite. Soit  $(b_n, n \in \mathbb{N})$  une suite de nombres réels positifs qui satisfait (6.6). Alors pour tout  $x \in \mathbb{R}^d$  et  $\mathbb{A}_n, \mathbb{A}_n^* \in \{\mathbb{T}_n, \mathbb{G}_n\}$ ,  $b_n^{-1} (\|K\|_2 \sqrt{\hat{\mu}_{\mathbb{A}_n^*}(x)} \vee \varpi_n)^{-1} \sqrt{|\mathbb{A}_n| h_n^d} (\hat{\mu}_{\mathbb{A}_n}(x) - \mu(x))$  satisfait un principe de déviations modérées dans  $\mathbb{R}^d$  de vitesse  $b_n^2$  et de fonction de taux  $I$  définie pour tout  $y \in \mathbb{R}$  par  $I(y) = y^2/2$ . En particulier, on a*

$$\lim_{n \rightarrow +\infty} \frac{1}{b_n^2} \log \mathbb{P} \left( b_n^{-1} \left( \|K\|_2 \sqrt{\hat{\mu}_{\mathbb{A}_n^*}(x)} \vee \varpi_n \right)^{-1} \sqrt{|\mathbb{A}_n| h_n^d} \left| (\hat{\mu}_{\mathbb{A}_n}(x) - \mu(x)) \right| > \delta \right) = -\frac{\delta^2}{2} \quad \forall \delta > 0.$$

*Idées de la preuve (voir [16], Section 4 et 5, pour plus de détails).* Nous utilisons les mêmes décompositions que celles de la preuve du Théorème 5.6. Nous utilisons ensuite les idées de la preuve de Théorème 6.1 pour obtenir le principe de déviations modérées.  $\square$

**Remarque 6.4.** *Le Théorème 6.2 nous permet d’obtenir un intervalle de confiance pour  $\mu(x)$  de longueur  $b_n / \sqrt{|\mathbb{A}_n| h_n^d}$  et de niveau asymptotiquement proche de  $1 - \exp(-(b_n^2 \delta^2)/2)$ .*

## Chapitre 7

# Comportement asymptotique des grandes valeurs du processus bifurcant autorégressif

Ce chapitre est issu de l'article [6] écrit en collaboration avec Vincent Bansaye. Nous nous intéressons au comportement à temps long du nombre d'individus ayant de grands traits dans le processus BAR stable (voir (1.4)). Nous voyons apparaître deux régimes pour le comportement trajectoriel de ce nombre. Ces deux régimes dépendent du fait que la valeur absolue de l'un des paramètres d'autorégression ( $a_0$  où  $a_1$ ) est plus grande que 1. Le thème central de notre étude porte donc sur les densités locales. En explorant ce thème, nous sommes amenés à étudier les grandes déviations de la chaîne marquée associée au processus BAR. En conséquence de notre étude sur les densités locales, nous obtenons un résultat asymptotique pour la valeur maximale du processus BAR sur une génération.

### 7.1 Cadre général

Nous considérons le processus BAR défini en (1.4) avec, pour simplifier,  $b_0 = b_1 = 0$  et  $X_0 = 0$ . Nous supposons que  $a_0, a_1 \in \mathbb{R}_+$ , avec  $a_0 a_1 < 1$ . La chaîne marquée  $Y$  associée à ce processus BAR est le processus autorégressif en environnement aléatoire défini par :

$$Y_0 = 0 \quad \text{et} \quad \forall n \geq 1, \quad Y_n = \theta_n Y_{n-1} + \varepsilon_n,$$

où l'environnement  $\Theta = (\theta_n, n \geq 1)$  est une suite de variables aléatoires i.i.d. telle que

$$\mathbb{P}(\theta_1 \in \{a_0, a_1\}) = 1, \quad \text{avec} \quad \mathbb{P}(\theta_1 = a_0) = \mathbb{P}(\theta_1 = a_1)$$

et  $(\varepsilon_n, n \geq 1)$  est une suite de variables aléatoires i.i.d. de loi normale standard, indépendante de  $\Theta$ . Alors on a, pour tout  $n \in \mathbb{N}^*$ ,

$$Y_n = \sum_{k=1}^n \left( \prod_{\ell=k+1}^n \theta_\ell \right) \varepsilon_k. \tag{7.1}$$

Ainsi,  $Y_n$  est une variable aléatoire gaussienne d'écart-type

$$A_n = A(\theta_1, \dots, \theta_n) = \sqrt{\sum_{k=1}^n \prod_{\ell=k+1}^n \theta_\ell^2}. \quad (7.2)$$

Notons que  $A_n$  a la même loi que  $A_n^* = A(\theta_n, \dots, \theta_1)$  et par monotonicité,  $A_n^*$  converge presque sûrement vers

$$A_\infty^* = \sqrt{\sum_{k=1}^{\infty} \prod_{\ell=1}^{k-1} \theta_\ell^2} \in (0, \infty].$$

Nous considérons la mesure ponctuelle aléatoire

$$Z_n = \sum_{i \in \mathbb{G}_n} \delta_{X_i}.$$

Le processus  $Z$  ainsi défini est un processus de branchement multitype où les types sont à valeurs dans  $\mathbb{R}$ . Pour deux réels  $a$  et  $b$  tels que  $a \leq b$ , on sait d'après le many-to-one-formula que

$$\mathbb{E}(Z_n([a, b])) = 2^n \mathbb{P}(Y_n \in [a, b]);$$

on a aussi que

$$\frac{1}{2^n} Z_n([a, b]) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \pi([a, b]),$$

où  $\pi$  est la loi de la variable aléatoire  $Y_\infty^*$  définie par

$$Y_\infty^* := \sum_{k=1}^{\infty} \left( \prod_{\ell=1}^{k-1} \theta_\ell \right) \varepsilon'_k,$$

avec  $(\varepsilon'_k, k \in \mathbb{N}^*)$  une suite de variables aléatoires i.i.d. gaussiennes, centrées et de variance  $\sigma^2$ , indépendante de l'environnement  $\Theta$ . Nous nous intéressons à l'étude de la densité locale

$$Z_n([\alpha_n, \infty)) = |\{i \in \mathbb{G}_n : X_i \geq \alpha_n\}|$$

lorsque  $\alpha_n \rightarrow \infty$ . Ceci nous conduit à étudier l'événement des grandes déviations  $\{Y_n \geq \alpha_n\}$  et la valeur maximale du processus sur une génération

$$M_n = \max\{X_i : i \in \mathbb{G}_n\}.$$

Nous voyons apparaître deux régimes dans notre étude : le régime  $a_0 = \max\{a_0, a_1\} < 1$  que nous appelons cas stable stricte et le régime  $a_1 < 1 < a_0$  que nous appelons cas stable faible.

## 7.2 Grandes déviations pour le processus $Y$

### 7.2.1 Grandes déviations et chemin des grandes déviations dans le régime stable strict

Cette section a deux objectifs. Le premier, évaluer les probabilités de grandes déviations du processus autorégressif à coefficients aléatoires  $(Y_n, n \in \mathbb{N})$ . Le deuxième, comprendre le comportement asymptotique des trajectoires de ce processus conditionnellement à l'évènement des grandes déviations. On a le résultat suivant.

**Théorème 7.1.** *Soit  $(\alpha_n)_{n \in \mathbb{N}}$  une suite croissante de nombres réels qui tend vers l'infini.*

*On a*

$$\lim_{n \rightarrow +\infty} \frac{1}{\alpha_n^2} \log \mathbb{P}(Y_n \geq \alpha_n) = -\frac{1 - a_0^2}{2}. \quad (7.3)$$

*De plus, pour toute suite d'entiers naturels  $(\ell_n)_n$  telle que  $n - \ell_n \rightarrow \infty$  et  $\ell_n = o(\log \alpha_n)$ , on a  $\forall \delta > 0$*

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left( \sup_{\ell \in \{0, \dots, \ell_n\}} \left| \frac{Y_{n-\ell}}{\alpha_n} - a_0^\ell \right| \leq \delta, \quad (\theta_{n-\ell_n}, \dots, \theta_n) = (a_0, \dots, a_0) \middle| Y_n \geq \alpha_n \right) = 1. \quad (7.4)$$

**Remarque 7.1.** *Ainsi, l'évènement des grandes déviations  $\{Y_n \geq \alpha_n\}$  est réalisé en sélectionnant l'environnement favorable  $a_0$  dans les dernières générations. La différence principale avec les marches aléatoires branchantes est que ici, les déviations ne sont pas linéaires, mais géométriques.*

*Idées principales de la preuve (voir [6], Section 2.1 pour plus de détails).* Premièrement, (7.3) est obtenu en utilisant (7.1) et les bornes supérieure et inférieure de la queue d'une gaussienne (voir par exemple [14]). Ensuite, (7.4) est obtenu en introduisant le temps

$$\tau_n = \sup\{i = 1, \dots, n : \theta_i = a_1\},$$

et en utilisant la structure autorégressive de  $Y_n$ . □

### 7.2.2 Grandes déviations dans le régime stable faible

Dans le régime stable faible, nous supposons qu'il existe  $\kappa > 0$  tel que

$$a_0^\kappa + a_1^\kappa = 2.$$

Soit  $\kappa_{0>0}$  et  $\gamma \in (0, 1)$  définis par

$$\log(a_0)a_0^{\kappa_0} + \log(a_1)a_1^{\kappa_0} = 0, \quad \gamma := (a_0^{\kappa_0} + a_1^{\kappa_0})/2 = \inf_{s \geq 0} \{(a_0^s + a_1^s)/2\}.$$

Nous introduisons la marche aléatoire  $(S_k)_{k \geq 1}$  définie par  $S_1 = 0$  et

$$S_k = \sum_{\ell=1}^{k-1} \log \theta_\ell \quad (k \geq 2).$$

Pour l'étude de l'événement  $\{Y_n \geq \alpha_n\}$ , nous supposons que le maximum  $\bar{S} = \sup_{k \geq 1} S_k$  est atteint avant le temps  $n$  lorsque  $\alpha_n$  est assez grand. Ceci correspond à dire les événements  $\{A_n^* \geq a_n\}$  et  $\{A_\infty^* \geq a_n\}$  sont comparables. Alors on a le résultat suivant.

**Théorème 7.2.** *Soit  $(\alpha_n, n \in \mathbb{N})$  une suite de nombres réels telle que  $\alpha_n^\kappa n^{-3/2} \gamma^n \rightarrow 0$  lorsque  $n \rightarrow \infty$ . Alors,*

$$\mathbb{P}(Y_n \geq \alpha_n) \asymp \mathbb{P}(A_n \geq \alpha_n) \asymp \alpha_n^{-\kappa}.$$

*Idées de la preuve (voir [6], Section 2.2 pour plus de détails).* Premièrement, le comportement asymptotique de  $\mathbb{P}(A_n \geq \alpha_n)$  est obtenu en utilisant les résultats de Kesten [72], Feller [57] (voir aussi [55]) et Borovkov [33]. La preuve de l'équivalence  $\mathbb{P}(Y_n \geq \alpha_n) \asymp \mathbb{P}(A_n \geq \alpha_n)$  utilise le fait que  $\mathcal{L}(Y_n | \mathcal{F}_{n-1}) = \mathcal{N}(0, a_n^2 W_n^2)$ , où  $W_n = A_n/a_n$ .  $\square$

**Remarque 7.2.** *Nous avons bon espoir que le régime transitoire  $a_1 \leq a_0 = 1$  et  $\alpha_n = \exp(n\rho)$  avec  $\rho > 1/\gamma$  peut-être étudié en utilisant une approche similaire de celle du Théorème 7.2. Notons que la principale difficulté avec ce régime est que, d'après nous,  $\{A_n^* \geq a_n\}$  et  $\{A_\infty^* \geq a_n\}$  ne sont pas comparables. Dans ces conditions, décrire le comportement de  $\{Y_n \geq \alpha_n\}$  nécessite plus de travail.*

### 7.3 Densités locales pour le processus BAR

Rappelons les constantes  $\kappa$  et  $\gamma$  définies à la Section 7.2.2. Les résultats de la Section 7.2 nous permettent d'obtenir les comportements asymptotiques suivants.

**Théorème 7.3.**

1. *Si  $a_1 \leq a_0 < 1$ , alors pour tout  $x \in [0, \sqrt{2 \log(2)/(1 - a_0^2)})$ ,*

$$\frac{1}{n} \log Z_n([x\sqrt{n}; +\infty)) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \log 2 - x^2(1 - a_0^2)/2.$$

2. *Si  $a_1 < 1 < a_0$ , alors pour  $c \in [0, +\infty)$  tel que  $c^\kappa < \min(2, 1/\gamma)$  et  $c > 1$ , on a*

$$\frac{1}{n} \log |\{i \in \mathbb{G}_n : X_i \geq c^n\}| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \log(2/c^\kappa).$$

**Remarque 7.3.** *Nous renvoyons à [6], Section 3, pour les détails de la preuve du Théorème 7.3. Quelques commentaires sont néanmoins nécessaires.*

1. Pour le régime stable strict, la preuve suit les idées de Biggins [13] pour les marches aléatoires branchantes. Nous considérons dans ce régime une sous population d'individus qui réalise la densité locale  $Z_n([x\sqrt{n}; +\infty))$ . Pour cela, grâce aux résultats de la Section 7.2.1, il suffit de suivre les individus qui subissent une déviation dans l'environnement  $a_0$  dans les dernières générations.
2. Pour le régime stable faible, la preuve nécessite également la sélection d'une sous population produisant les grandes valeurs  $[c^n, +\infty)$ . Mais cette sous population est très différente de celle considérée dans le cas stable strict. En effet, elle est caractérisée par  $A(\theta_1, \dots, \theta_n) \geq c^n$ , ce qui fait que seuls certaines lignées particulières de la population réalisent la densité locale  $Z_n([c^n, +\infty))$ . La preuve dans ce cas est plus complexe puisqu'elle nécessite l'étude du processus restreint à un sous arbre associé aux grandes déviations de  $Y_n$ . Ce phénomène, lié à la structure aléatoire de l'environnement hérité de l'arbre, peut-être comparé aux résultats obtenus dans le cas faiblement sous-critique du modèle de branchement de Kimmel [5]. Notons que dans [5], seul le comportement moyen est étudié. Notons aussi que la question de l'obtention d'une convergence sûre reste ouverte pour ce régime.
3. Ainsi, un examen minutieux des preuves permet de voir que, pour  $a_0 \neq a_1$ , les grandes valeurs du processus BAR peuvent être vues comme la combinaison de deux sources d'aléa : un terme additif gaussien qui apparaît le long de chaque lignée et le choix de la lignée aléatoire. Dans le régime stable strict, les densités locales proviennent des deux sources d'aléa dans les dernières générations. On observe ainsi un comportement proche de celui des marches aléatoires branchantes, à l'exception que les déviations se produisent dans les dernières générations. Dans le régime stable faible, seul le choix de la lignée intervient dans la formation des densités locales. Dans ce dernier cas, les déviations se produisent très tôt pour créer les grandes valeurs. Ce comportement est très différent de celui des marches aléatoires branchantes puisqu'une sous population "déterministe", fonction de  $A_n$  définie en (7.2), supporte les densités locales.
4. On voit ainsi apparaître une transition de phase dans le comportement des grandes valeurs du processus BAR. Ici, l'arbitre de la compétition est le paramètre d'autorégression  $a_0 = \max\{a_0, a_1\}$ . Le point de la transition est  $a_0 = 1$ . Le comportement des densités locales en ce point reste un problème ouvert.

Considérons la valeur maximale  $M_n$  du processus BAR à la génération  $n$  définie par  $M_n = \max\{X_i : i \in \mathbb{G}_n\}$ . Alors, comme conséquence du Théorème 7.3, nous avons le résultats suivant.

**Corollaire 7.1** (Valeur maximale pour le processus BAR). *On se place dans le contexte du Théorème 7.3.*

1. Si  $a_1 \leq a_0 < 1$ , alors on a

$$\frac{1}{\sqrt{n}} M_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \sqrt{\frac{2 \log(2)}{1 - a_0^2}}.$$

2. Si  $a_1 < 1 < a_0$ , alors on

$$\frac{1}{n} \log(M_n) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \frac{\log(2)}{\kappa}.$$

*Idée principale de la preuve.* La preuve repose principalement sur l'inégalité :

$$\mathbb{P}(M_n \geq \delta) \leq \mathbb{E}(Z_n[\delta, \infty)) = 2^n \mathbb{P}(Y_n \geq \delta) \quad \forall \delta > 0.$$

□

**Remarque 7.4.** La question de la valeur maximale a fait l'objet d'une grande attention et a été intensivement étudiée pour les marches aléatoires branchantes, voir par exemple [13, 34, 66, 74, 86]. Des estimations plus fines obtenues dernièrement ont permis de mieux comprendre les structures branchantes des processus [1, 2, 56, 68]. Le comportement de la valeur maximale est différente dans notre cas. Les techniques utilisées pour les marches aléatoires branchantes sont utilisées et adaptées ici, en particulier l'utilisation du many-to-one formula pour exploiter la description du chemin conduisant aux particules extrêmes dans le régime stable stricte [95] et la construction des processus de branchement auxiliaires pour le régime faiblement stable. Comme nous l'avons déjà mentionné, la principale nouveauté réside dans le régime faiblement stable  $a_1 < 1 < a_0$  où une loi des grands nombres d'un sous arbre bien choisi est nécessaire.

# Références

- [1] L. Addario-Berry and B. Reed. Minima in branching random walks. *Ann. Probab.*, 37(3) :1044–1079, 2009.
- [2] E. Aïdékon. Convergence in law of the minimum of a branching random walk. *Ann. Probab.*, 41(3A) :1362–1426, 2013.
- [3] H. Z. An and F. C. Huang. The geometrical ergodicity of nonlinear autoregressive models. *Statist. Sinica*, 6(4) :943–956, 1996.
- [4] S. Arlot and P. Massart. Data-driven calibration of penalties for least-squares regression. *Journal of Machine Learning Research*, 10(10) :245–279, 2009.
- [5] V. Bansaye. Proliferating parasites in dividing cells : Kimmel’s branching model revisited. *Ann. Appl. Probab.*, 18(3) :967–996, 2008.
- [6] V. Bansaye and S. V. Bitseki Penda. A phase transition for large values of bifurcating autoregressive models. *J Theor Probab*, 34 :2081–2116, 2021.
- [7] I. V. Basawa and J. Zhou. Non-Gaussian bifurcating models and quasi-likelihood estimation. *Adv. in Appl. Probab.*, 41(A) :55–64, 2004.
- [8] B. Bercu and V. Blandin. Limit theorems for bifurcating integer-valued autoregressive processes. *Statistical inference for stochastic processes*, 18 :33–67, 2015.
- [9] B. Bercu and V. Blandin. A rademacher–menchov approach for random coefficient bifurcating autoregressive processes. *Stochastic Processes and their Applications*, 125(4) :1218–1243, 2015.
- [10] B. Bercu, B. De Saporta, and A. Gégout-Petit. Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electronic Journal of Probability*, 14 :2492–2526, 2009.
- [11] R. Bhattacharya and C. Lee. On geometric ergodicity of nonlinear autoregressive models. *Statistics & Probability Letters*, 22(4) :311–315, 1995.
- [12] H. J. Bierens. Kernel estimators of regression functions. *Advances in econometrics : Fifth World Congress*, pages 99–144, 1987.
- [13] J. D. Biggins. The first- and last-birth problems for a multitype age-dependent branching process. *Advances in Appl. Probability*, 8(3) :446–459, 1976.

- [14] Z. W. Birnbaum. An inequality for Mill's ratio. *Ann. Math. Statistics*, 13 :245–246, 1942.
- [15] S. V. Bitseki Penda. Deviation inequalities for bifurcating Markov chains on Galton-Watson tree. *ESAIM : PS*, 19 :689–724, 2015.
- [16] S. V. Bitseki Penda. Moderate deviation principles for kernel estimator of invariant density in bifurcating markov chains. *Stochastic Processes and their Applications*, 158 :282–314, 2023.
- [17] S. V. Bitseki Penda. Central limit theorem for bifurcating markov chains : the mother-daughters triangles case. *Stochastics*, 96(4) :1453–1474, 2024.
- [18] S. V. Bitseki Penda. Kernel estimation of the transition density in bifurcating markov chains. *Journal of Statistical Planning and Inference*, 231 :106–138, 2024.
- [19] S. V. Bitseki Penda and J.-F. Delmas. Central limit theorem for bifurcating Markov chains under  $L^2$ -ergodic conditions. *Adv. in Appl. Probab.*, 54(4) :999–1031, 2022.
- [20] S. V. Bitseki Penda and J.-F. Delmas. Central limit theorem for bifurcating Markov chains under pointwise ergodic conditions. *The Annals of Applied Probability*, 32(5) :3817 – 3849, 2022.
- [21] S. V. Bitseki Penda and J.-F. Delmas. Central limit theorem for kernel estimator of invariant density in bifurcating Markov chains models. *J. Theoret. Probab.*, 36(3) :1591–1625, 2023.
- [22] S. V. Bitseki Penda and H. Djellout. Deviation inequalities and moderate deviations for estimators of parameters in bifurcating autoregressive models. *Ann. Inst. H. Poincaré Probab. Statist.*, 50(3) :806–844, 08 2014.
- [23] S. V. Bitseki Penda, H. Djellout, and A. Guillin. Deviation inequalities, moderate deviations and some limit theorems for bifurcating Markov chains with application. *Ann. Appl. Probab.*, 24(1) :235–291, 2014.
- [24] S. V. Bitseki Penda, M. Escobar-Bach, and A. Guillin. Transportation and concentration inequalities for bifurcating Markov chains. *Bernoulli*, 23(4B) :3213 – 3242, 2017.
- [25] S. V. Bitseki Penda and G. Gackou. Moderate deviation principles for bifurcating markov chains : case of functions dependent of one variable. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 19 :617–640, 2022.
- [26] S. V. Bitseki Penda, M. Hoffmann, and A. Olivier. Adaptive estimation for bifurcating Markov chains. *Bernoulli*, 23(4B) :3598–3637, 2017.
- [27] S. V. Bitseki Penda and A. Olivier. Autoregressive functions estimation in nonlinear bifurcating autoregressive models. *Stat. Inference Stoch. Process.*, 20(2) :179–210, 2017.
- [28] S. V. Bitseki Penda and A. Olivier. Moderate deviation principle in nonlinear bifurcating autoregressive models. *Statistics & Probability Letters*, 138 :20–26, 2018.

[29] S. V. Bitseki Penda and A. Roche. Local bandwidth selection for kernel density estimation in a bifurcating markov chain model. *Journal of Nonparametric Statistics*, 32(3) :535–562, 2020.

[30] V. Blandin. Asymptotic results for random coefficient bifurcating autoregressive processes. *Statistics*, 48(6) :1202–1232, 2014.

[31] S. G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1) :1–28, 1999.

[32] F. c. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probab. Theory Related Fields*, 137(3-4) :541–593, 2007.

[33] A. A. Borovkov. *Stochastic processes in queueing theory*, volume No. 4 of *Applications of Mathematics*. Springer-Verlag, New York-Berlin, 1976. Translated from the Russian by Kenneth Wickwire.

[34] M. D. Bramson. Maximal displacement of branching Brownian motion. *Comm. Pure Appl. Math.*, 31(5) :531–581, 1978.

[35] D. B. H. Cline. Stability of nonlinear stochastic recursions with application to nonlinear AR-GARCH models. *Adv. in Appl. Probab.*, 39(2) :462–491, 2007.

[36] A. Cohen. Wavelet methods in numerical analysis. In *Solution of Equation in  $\mathbb{R}^n$  (Part 3), Techniques of Scientific Computing (Part 3)*, volume 7 of *Handbook of Numerical Analysis*, pages 417–711. Elsevier, 2000.

[37] F. Comte and C. Lacour. Anisotropic adaptive kernel deconvolution. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(2) :569–609, 2013.

[38] F. Comte and N. Marie. On a Nadaraya-Watson estimator with two bandwidths. *Electron. J. Stat.*, 15(1) :2566–2607, 2021.

[39] R. Cowan and R. Staudte. The bifurcating autoregression model in cell lineage studies. *Biometrics*, 42(4) :769–783, December 1986.

[40] B. de Saporta, A. Gégout-Petit, and L. Marsalle. Parameters estimation for asymmetric bifurcating autoregressive processes with missing data. *Electronic Journal of Statistics*, 5(none) :1313 – 1353, 2011.

[41] B. de Saporta, A. Gégout-Petit, and L. Marsalle. Asymmetry tests for bifurcating auto-regressive processes with missing data. *Statist. Probab. Lett.*, 82(7) :1439–1444, 2012.

[42] B. de Saporta, A. Gégout-Petit, and L. Marsalle. Random coefficients bifurcating autoregressive processes. *ESAIM : PS*, 18 :365–399, 2014.

[43] B. de Saporta, A. Gégout-Petit, and L. Marsalle. Statistical study of asymmetry in cell lineage data. *Comput. Statist. Data Anal.*, 69 :15–39, 2014.

[44] J.-F. Delmas and L. Marsalle. Detection of cellular aging in a Galton-Watson process. *Stochastic Process. Appl.*, 120(12) :2495–2519, 2010.

[45] B. Delyon, B. de Saporta, N. Krell, and L. Robert. Investigation of asymmetry in *e. coli* growth rate, 2015.

[46] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[47] R. A. DeVore, S. V. Konyagin, and V. N. Temlyakov. Hyperbolic wavelet approximation. *Constr. Approx.*, 14(1) :1–26, 1998.

[48] H. Djellout, A. Guillin, and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.*, 32(3B) :2702–2732, 2004.

[49] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3) :425–455, 1994.

[50] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, 90(432) :1200–1224, 1995.

[51] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage : Asymptopia? *Journal of the Royal Statistical Society : Series B (Methodological)*, 57(2) :301–337, 1995.

[52] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Density estimation by wavelet thresholding. *Ann. Statist.*, 24(2) :508–539, 1996.

[53] M. Doumic, M. Hoffmann, N. Krell, and L. Robert. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli*, 21(3) :1760–1799, 2015.

[54] M. Duflo. *Random iterative models*, volume 34. Springer Science & Business Media, 2013.

[55] N. Enriquez, C. Sabot, and O. Zindy. A probabilistic representation of constants in Kesten’s renewal theorem. *Probab. Theory Related Fields*, 144(3-4) :581–613, 2009.

[56] G. Faraud, Y. Hu, and Z. Shi. Almost sure convergence for stochastically biased random walks on trees. *Probab. Theory Related Fields*, 154(3-4) :621–660, 2012.

[57] W. Feller. *An introduction to probability theory and its applications, vol 2*. John Wiley & Sons, 2008.

[58] N. Fournier and A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields*, 162(3-4) :707–738, 2015.

[59] A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation : Oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39(3) :1608 – 1632, 2011.

[60] A. V. Goldenshluger and O. V. Lepski. General selection rule from a family of linear estimators. *Theory of Probability & Its Applications*, 57(2) :209–226, 2013.

- [61] N. Gozlan. Integral criteria for transportation-cost inequalities. *Electron. Comm. Probab.*, 11 :64–77, 2006.
- [62] J. Guyon. Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.*, 17(5-6) :1538–1569, 2007.
- [63] J. Guyon, A. Bize, G. Paul, E. Stewart, J.-F. Delmas, and F. Taddéi. Statistical study of cellular aging. In *ESAIM : Proceedings*, volume 14, pages 100–114. EDP Sciences, 2005.
- [64] B. M. Gyori and D. Paulin. Hypothesis testing for Markov chain Monte Carlo. *Stat. Comput.*, 26(6) :1281–1292, 2016.
- [65] M. Hairer and J. C. Mattingly. Yet another look at harris’ ergodic theorem for markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, pages 109–117. Springer, 2011.
- [66] J. M. Hammersley. Postulates for subadditive processes. *Ann. Probability*, 2 :652–680, 1974.
- [67] W. Härdle and P. Vieu. Kernel regression smoothing of time series. *Journal of Time Series Analysis*, 13(3) :209–232, 1992.
- [68] Y. Hu and Z. Shi. Minimal position and critical martingale convergence in branching random walks, and directed polymers on disordered trees. *Ann. Probab.*, 37(2) :742–789, 2009.
- [69] F. A. Hughes, A. R. Barr, and P. Thomas. Patterns of interdivision time correlations reveal hidden cell cycle factors. *Elife*, 11 :e80927, 2022.
- [70] P. Katajisto, J. Döhla, C. L. Chaffer, N. Penttinkoski, N. Marjanovic, S. Iqbal, R. Zoncu, W. Chen, R. A. Weinberg, and D. M. Sabatini. Asymmetric apportioning of aged mitochondria between daughter cells is required for stemness. *Science*, 348(6232) :340–343, 2015.
- [71] G. Kerkyacharian and D. Picard. Thresholding algorithms, maxisets and well-concentrated bases. *Test*, 9(2) :283–344, 2000. With comments, and a rejoinder by the authors.
- [72] H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Math.*, 131 :207–248, 1973.
- [73] M. Kimmel. Quasistationarity in a branching model of division-within-division. In *Classical and modern branching processes (Minneapolis, MN, 1994)*, volume 84 of *IMA Vol. Math. Appl.*, pages 157–164. Springer, New York, 1997.
- [74] J. F. C. Kingman. The first birth problem for an age-dependent branching process. *Ann. Probability*, 3(5) :790–801, 1975.
- [75] C. Lacour. Adaptive estimation of the transition density of a particular hidden markov chain. *Journal of Multivariate Analysis*, 99(5) :787–814, 2008.
- [76] C. Lacour and P. Massart. Minimal penalty for Goldenshluger-Lepski method. *Stochastic Process. Appl.*, 126(12) :3774–3789, 2016.

[77] K. Marton. A simple proof of the blowing-up lemma. *IEEE Trans. Inform. Theory*, 32(3) :445–446, 1986.

[78] K. Marton. Bounding  $\bar{d}$ -distance by informational divergence : a method to prove measure concentration. *The Annals of Probability*, 24(2) :857 – 866, 1996.

[79] K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.*, 6(3) :556–571, 1996.

[80] K. Marton. Measure concentration for a class of random processes. *Probab. Theory Related Fields*, 110(3) :427–439, 1998.

[81] K. Marton. Measure concentration and strong mixing. *Studia Sci. Math. Hungar.*, 40(1-2) :95–113, 2003.

[82] Y. Meyer. *Ondelettes et opérateurs. I*. Actualités Mathématiques. [Current Mathematical Topics]. Hermann, Paris, 1990. Ondelettes. [Wavelets].

[83] A. Mokkadem and M. Pelletier. Confidence bands for densities, logarithmic point of view. *ALEA Lat. Am. J. Probab. Math. Stat.*, 2 :231–266, 2006.

[84] M. Mura, C. Feillet, R. Bertolusso, F. Delaunay, and M. Kimmel. Mathematical modelling reveals unexpected inheritance and variability patterns of cell cycle parameters in mammalian cells. *PLoS computational biology*, 15(6) :e1007054, 2019.

[85] E. A. Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1) :141–142, 1964.

[86] M. Nakashima. Minimal position of branching random walks in random environment. *J. Theoret. Probab.*, 26(4) :1181–1217, 2013.

[87] M. K. Pitt, C. Chatfield, and S. G. Walker. Constructing first order stationary autoregressive models via latent processes. *Scand. J. Statist.*, 29(4) :657–663, 2002.

[88] A. Puhalskii. Large deviations of semimartingales via convergence of the predictable characteristics. *Stochastics Stochastics Rep.*, 49(1-2) :27–85, 1994.

[89] G. Rebelles. Pointwise adaptive estimation of a multivariate density under independence hypothesis. *Bernoulli*, 21(4) :1984 – 2023, 2015.

[90] Y.-X. Ren, R. Song, and R. Zhang. Central limit theorems for supercritical branching Markov processes. *J. Funct. Anal.*, 266(3) :1716–1756, 2014.

[91] Y.-X. Ren, R. Song, and R. Zhang. Central limit theorems for supercritical superprocesses. *Stochastic Process. Appl.*, 125(2) :428–457, 2015.

[92] Y.-X. Ren, R. Song, and R. Zhang. Limit theorems for some critical superprocesses. *Illinois J. Math.*, 59(1) :235–276, 2015.

[93] Y.-X. Ren, R. Song, and R. Zhang. Central limit theorems for supercritical branching nonsymmetric Markov processes. *Ann. Probab.*, 45(1) :564–623, 2017.

[94] Y.-X. Ren, R. Song, and R. Zhang. Functional central limit theorems for supercritical superprocesses. *Acta Appl. Math.*, 147 :137–175, 2017.

- [95] M. I. Roberts. A simple path to asymptotics for the frontier of a branching Brownian motion. *Ann. Probab.*, 41(5) :3518–3541, 2013.
- [96] P. M. Robinson. Nonparametric estimators for time series. *Journal of Time Series Analysis*, 4(3) :185–207, 1983.
- [97] B. W. Silverman. *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1986.
- [98] D. A. Sinclair and L. Guarente. Extrachromosomal rDNA circles— a cause of aging in yeast. *Cell*, 91(7) :1033–1042, 1997.
- [99] 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(2) :null, 02 2005.
- [100] J. J. Trujillo, M. Rivero, and B. Bonilla. On a Riemann-Liouville generalized Taylor’s formula. *J. Math. Anal. Appl.*, 231(1) :255–265, 1999.
- [101] A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.
- [102] D. Usero. Fractional taylor series for caputo fractional derivatives. construction of numerical schemes. *Preprint* [http://www.fdi.ucm.es/profesor/lvazquez/calcfrac/docs/paper\\_Usero.pdf](http://www.fdi.ucm.es/profesor/lvazquez/calcfrac/docs/paper_Usero.pdf), 2008.
- [103] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.
- [104] G. S. Watson. Smooth regression analysis. *Sankhyā : The Indian Journal of Statistics, Series A*, pages 359–372, 1964.
- [105] J. Worms. *Principes de déviations modérées pour des martingales et applications statistiques*. PhD thesis, Université de Marne-la-Vallée (1991-2019), 2000.
- [106] J. Worms. Moderate deviations of some dependent variables. I. Martingales. *Math. Methods Statist.*, 10(1) :38–72, 2001.
- [107] J. Worms. Moderate deviations of some dependent variables. II. Some kernel estimators. *Math. Methods Statist.*, 10(2) :161–193, 2001.