Inégalités de déviations, principe de déviations modérées et théorèmes limites pour des processus indexés par un arbre binaire et pour des modèles markoviens

Siméon Valère Bitseki Penda

N° d'Ordre : D.U. 2291

UNIVERSITÉ BLAISE PASCAL
U.F.R. Sciences et Technologies

ÉCOLE DOCTORALE DES SCIENCES FONDAMENTALES
N° 730

THÈSE

Présentée pour obtenir le grade de

DOCTEUR D'UNIVERSITÉ

Spécialité :
MATHÉMATIQUES APPLIQUÉES

Par **BITSEKI PENDA Siméon Valère**
D.E.A Mathématiques appliquées Université de Yaoundé I

Inégalités de déviations, Principes de déviations modérées et théorèmes limites pour des processus indexés par un arbre binaire et pour des modèles markoviens

Soutenue publiquement le 20 Novembre 2012, devant la commission d'examen.

Président :	WU Liming	Professeur	
Examinateurs :	BERCU Bernard	Professeur	
	CHAFAÏ Djalil	Professeur	
	DELMAS Jean-François	Professeur	(Rapporteur)
	GAMBOA Fabrice	Professeur	(Rapporteur)
	GUILLIN Arnaud	Professeur	(Directeur)
	SERLET Laurent	Professeur	

Remerciements

L'heure est enfin venu pour moi de rédiger le dernier acte de cette thèse. Je dois avouer que cet acte n'est pas si facile. Car, comment exprimer par des “simples” mots toute sa gratitude envers quelqu'un ? Quelques lignes dans une thèse d'une centaine de pages suffisent-elles à dire combien on est reconnaissant ?

Débuter la thèse n'a pas été facile pour moi. Je vais même dire que cela a été énormément compliqué. C'est la raison pour laquelle je remercie Arnaud GUILLIN (mon Directeur de thèse). Je ne le fais pas parce qu'un Doctorant doit (normalement) remercier son encadreur. Je le fais parce qu'il y a des choses qu'on ne peut taire. Par ses formidables qualités humaines, Arnaud a su rendre ma thèse viable. Son optimisme à toute épreuve, ainsi sa vision assez simple des choses m'ont permis, finalement, d'aller jusqu'au bout de ma thèse. Je te remercie aussi Arnaud pour ta patience. Je sais que ça pas toujours été facile de supporter tous les raisonnements erronés que j'ai très souvent produit.

Grâce à Etienne PARDOUX j'ai fait mes premiers pas dans le “monde stochastique”. Je te remercie Etienne pour toute la confiance que tu as mis en moi, ainsi que pour toute la volonté que tu as mis afin que cette thèse débute. J'en profite également pour remercier toute l'équipe de probabilité de l'Université Gaston Berger de Saint-Louis du Sénégal.

A la racine de cette thèse il y a David BEKOLLE. Je te remercie David pour tout le soutien que tu m'as accordé durant toutes ses années. Je te suis énormément reconnaissant car sans ton aide je n'aurais sûrement jamais commencer cette thèse.

Je remercie Jean-François DELMAS et Fabrice GAMBOA qui ont accepté de rapporter ce travail. Leurs critiques et leurs remarques ont permis une nette amélioration de la qualité du document. Je remercie également Bernard BERCU, Djalil CHAFAÏ, Laurent SERLET et Liming WU qui me font l'honneur de faire partie de mon jury de thèse.

J'ai souvent passé de longues heures à travailler avec Hacène DJELLOUT. Je te remercie Hacène pour tout le temps que tu m'as accordé. Je te remercie pour toutes les tasses de café que tu m'as offert. Je te remercie également pour toute l'aide que tu m'as souvent apporté sur le plan administratif.

Un merci à Frédéric PROÏA avec qui j'ai eu beaucoup de plaisir à travailler. J'ai notamment appris auprès de lui quelques astuces sur latex.

L'occasion m'est donné ici de remercier les membres du Laboratoires de Mathématiques. A Karine, Valérie, Marie-Paule, Damien, Cédric, Annick, pour leur disponibilité. Merci d'avoir souvent prêté une attention particulière à mes questions parfois stupides.

Un merci aux doctorants et ATER du Laboratoire de Mathématiques. A Romuald, Xinyu, Colin, Damien, Stéphane, Christelle, Manon, je vous témoigne ma profonde sympathie.

Merci à ma famille pour son soutien sans faille. Merci de me porter si haut dans

vos cœurs. Merci à toi Alain pour tout ce que tu nous a apporté.

A ma mère, pour tous les sacrifices qu'elle a consenti pour nous élever.

A la mémoire de mon père.

Résumé

Le contrôle explicite de la convergence des sommes convenablement normalisées de variables aléatoires, ainsi que l'étude du principe de déviations modérées associé à ces sommes constituent les thèmes centraux de cette thèse. Nous étudions principalement deux types de processus.

Premièrement, nous nous intéressons aux processus indexés par un arbre binaire, aléatoire ou non. Ces processus ont été introduits dans la littérature afin d'étudier le mécanisme de la division cellulaire. Au chapitre 2, nous étudions les chaînes de Markov bifurcantes. Ces chaînes peuvent être vues comme une adaptation des chaînes de Markov "usuelles" dans le cas où l'ensemble des indices à une structure binaire. Sous des hypothèses d'ergodicité géométrique uniforme et non-uniforme d'une chaîne de Markov induite, nous fournissons des inégalités de déviations et un principe de déviations modérées pour les chaînes de Markov bifurcantes. Au chapitre 3, nous nous intéressons aux processus bifurcants autorégressifs d'ordre p ($p \geq 1$). Ces processus sont une adaptation des processus autorégressifs linéaires d'ordre p dans le cas où l'ensemble des indices à une structure binaire. Nous donnons des inégalités de déviations, ainsi qu'un principe de déviations modérées pour les estimateurs des moindres carrés des paramètres "d'autorégression" de ce modèle. Au chapitre 4, nous traitons des inégalités de déviations pour des chaînes de Markov bifurcantes sur un arbre de Galton-Watson. Ces chaînes sont une généralisation de la notion de chaînes de Markov bifurcantes au cas où l'ensemble des indices est un arbre de Galton-Watson binaire. Elles permettent dans le cas de la division cellulaire de prendre en compte la mort des cellules. Les hypothèses principales que nous faisons dans ce chapitre sont : l'ergodicité géométrique uniforme d'une chaîne de Markov induite et la non-extinction du processus de Galton-Watson associé.

Au chapitre 5, nous nous intéressons aux modèles autorégressifs linéaires d'ordre 1 ayant des résidus corrélés. Plus particulièrement, nous nous concentrons sur la statistique de Durbin-Watson. La statistique de Durbin-Watson est à la base des tests de Durbin-Watson, qui permettent de détecter l'autocorrélation résiduelle dans des modèles autorégressifs d'ordre 1. Nous fournissons un principe de déviations modérées pour cette statistique.

Les preuves du principe de déviations modérées des chapitres 2, 3 et 5 reposent essentiellement sur le principe de déviations modérées des martingales. Les inégalités de déviations sont établies principalement grâce à l'inégalité d'Azuma-Bennet-Hoeffding et l'utilisation de la structure binaire des processus.

Le chapitre 6 est né de l'importance qu'a l'ergodicité explicite des chaînes de Markov au chapitre 2. L'ergodicité géométrique explicite des processus de Markov à temps discret et continu ayant été très bien étudiée dans la littérature, nous nous sommes penchés sur l'ergodicité sous-exponentielle des processus de Markov à temps continu. Nous fournissons alors des taux explicites pour la convergence sous exponentielle d'un processus de Markov à temps continu vers sa mesure de probabilité d'équilibre. Les hypothèses principales que nous utilisons sont : l'existence d'une fonction de Lyapunov et d'une condition de minoration. Les preuves reposent en grande partie sur la construction du couplage et le contrôle explicite de la queue du temps de couplage.

Mots clés : Chaînes de Markov bifurcantes, processus bifurcant autorégressif, processus de Markov, théorèmes limites, ergodicité, inégalités de déviations, principe de déviations modérées, martingale, vieillissement cellulaire, estimateurs des moindres carrés, statistique de Durbin-Watson, processus autorégressif d'ordre 1, autocorrélation résiduelle, conditions de Lyapunov, condition de minoration,

Table des matières

Chapitre 1

Présentation générale

1.1 Introduction

Cette thèse s'intéresse au comportement asymptotique et non-asymptotique des processus indexés par les données d'un arbre binaire, au comportement asymptotique des modèles d'autorégression linéaire dont les résidus sont autocorrélés, ainsi qu'à l'ergodicité sous exponentielle des processus de Markov. Les thèmes suivants sont abordés.

(i) Inégalités de déviations. Soit $(X_n, n \in \mathbb{N})$ une suite de variables aléatoires définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. Posons $S_n = \sum_{i=1}^n X_i$ et supposons que $(S_n/n, n \in \mathbb{N})$ converge en probabilité vers une quantité s, aléatoire ou non. Il s'agit, pour $\delta > 0$, de déterminer des bornes de la forme

$$\mathbb{P}\left(\left|\frac{S_n}{n} - s\right| > \delta\right) \leq h(\delta, n, c),$$

où c est une constante dépendant de certains paramètres de la suite de variables aléatoires (X_n), et h est une fonction réelle positive telle que pour δ et c fixés, $h(\delta, n, c)$ converge vers 0 quand n tend vers l'infini. Nous parlerons d'inégalité de déviations exponentielles si $h(\delta, n, c) = \exp\left(-C(\delta, n, c)\right)$, C étant une fonction réelle positive.

(ii) Principe de déviations modérées (PDM en abrégé). On suppose à présent que la suite $(S_n/n, n \in \mathbb{N})$ définie ci-dessus converge en probabilité vers 0. Soit $(b_n, n \in \mathbb{N})$ une suite de vitesse telle que $\sqrt{n} = o(b_n)$ et $b_n = o(n)$. Il s'agit grosso modo, pour $\delta > 0$, de déterminer l'équivalence asymptotique

$$\frac{n}{b_n^2} \log \mathbb{P}\left(\frac{S_n}{b_n} \geq \delta\right) \sim -I(\delta),$$

où $I(\cdot)$ est une fonction semi-continue inférieurement.

(iii) Bornes explicites pour l'ergodicité sous exponentielles des processus de Markov à temps continu. Soit $(X_t, t \geq 0)$ un processus de Markov de fonction de transition $(P^t, t \geq 0)$. On suppose que $(X_t, t \geq 0)$ admet une probabilité invariante π. Il s'agit alors de déterminer des fonctions r et g telles que

$$r(t)\|P^t(x,\cdot) - \pi(\cdot)\| \leq g(x),$$

où $\|\cdot\|$ est une norme appropriée définie sur l'ensemble des probabilités.

Les inégalités de déviations sont d'un grand intérêt en théorie comme en pratique. Elles sont un moyen pour démontrer certains théorèmes limites, notamment la loi forte des grands nombres via le lemme de Borel Cantelli. Nous utilisons d'ailleurs ce fait au chapitre 2 pour compléter la loi forte des grands nombres établie par Guyon [66]. Elles permettent également d'établir des convergences "super-exponentielles" en probabilités, outils essentiels pour l'obtention du principe de déviations modérées pour les martingales (nous reviendrons plus loin sur la convergence "super-exponentielle" en probabilité et son lien avec le PDM des martingales). Ce fait constitua d'ailleurs pour nous la motivation première pour l'étude des inégalités de déviations faite ici. Nous établissons et utilisons intensivement ces inégalités aux chapitres 2 et 3, afin d'obtenir la convergence "super-exponentielle" en probabilité de certaines sommes proprement normalisées et des estimateurs des modèles que nous étudions. Mais, c'est sans doute dans le domaine de la statistique "non-asymptotique" que ces inégalités trouvent leur plus grande utilité dans la pratique. En effet, en théorie de l'estimation par exemple, fournir des intervalles de confiance non asymptotique, c'est-à-dire n'utilisant pas un théorème limite, peut s'avérer très utile quand l'échantillon dont on dispose est de petite taille, ou de taille insuffisante pour appliquer le théorème limite en question. Dans de pareilles situations, les inégalités de déviations sont très bien adaptées dans la mesure où les bornes qu'elles fournissent dépendent explicitement de l'ordre de grandeur de l'échantillon étudié. Les inégalités de déviations ont été largement étudiées dans la littérature. L'inégalité de déviations la plus célèbre est certainement l'inégalité de Tchebychev. Bennett et Hoeffding, [16], [70], ont établi des inégalités de déviations exponentielles, qui portent leurs noms, pour des sommes de variables aléatoires bornées i.i.d. Plus tard, Azuma [6] a établi des inégalités de déviations exponentielles pour des martingales à différences bornées. La technique que nous utilisons dans cette thèse pour établir nos inégalités de déviations exponentielles (chapitres 2 et 3) est très proche de celle utilisée par ces trois auteurs. Pour établir les inégalités de déviations non exponentielles du chapitre 2, nous nous servons de l'approche classique utilisée pour démontrer l'inégalité de Tchebychev. Nous n'abordons pas dans cette thèse la question d'inégalités de concentration plus générale (à la : McDiarmid [79], Ledoux [74], Ledoux et Talagrand [75] par exemple), nous laissons cet aspect pour des travaux futurs.

Nous avons signaler tantôt en quel sens les inégalités de déviations peuvent aider à obtenir des PDMs. Mais qu'est que c'est exactement un PDM ? Notons tout d'abord que techniquement parlant, un PDM est juste un principe de grandes déviations (PGD en abrégé). Pour se fixer les idées, commençons donc par définir ce qu'on entend par principe de grande déviations. Les définitions sont issues du livre de Dembo et Zeitouni [35]. Soit S un espace métrique complet et séparable (espace polonais) muni de sa σ-algèbre borélienne $\mathcal{S}$. Tout au long de cette thèse, nous entendrons par vitesse une suite réelle positive croissant vers l'infini. Toutes les suites de variables aléatoires que nous considérons dans cette Section sont définies sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.

Définition 1.1.1. *Une fonction de taux est une application $I : S \to [0, +\infty]$ semi-continue inférieurement, c'est-à-dire dont les ensembles de niveau $\{x : I(x) \leq \alpha\}$ pour $\alpha \in [0, \infty)$ sont des parties fermées de S. Lorsque les ensembles de niveau sont compacts, on dit que I est une bonne fonction de taux.*

Définition 1.1.2. *Une suite (Z_n) de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeurs dans S satisfait à un PGD de vitesse (v_n) et de fonction de taux $I(\cdot)$ si l'on a, pour tout $A \in \mathcal{S}$,*

$$-\inf_{x \in A^\circ} I(x) \leq \liminf_{n \to \infty} \frac{1}{v_n} \log \mathbb{P}\left(Z_n \in A\right) \leq \limsup_{n \to \infty} \frac{1}{v_n} \log \mathbb{P}\left(Z_n \in A\right) \leq -\inf_{x \in \bar{A}} I(x),$$

où A° et $\bar{A}$ désignent respectivement l'intérieur et l'adhérence de A.

Un PDM est un résultat intermédiaire entre un théorème central limite (TCL en abrégé) et un PGD. Supposons que $Z_n = S_n/v_n$ converge en probabilité vers 0 et que $\sqrt{v_n} Z_n = S_n/\sqrt{v_n}$ converge en loi vers une loi normale centrée.

Définition 1.1.3. *(Z_n) satisfait à un principes de déviations modérées dans S de vitesse (a_n) et de fonction de taux $I(\cdot)$ si*

- *$a_n = o(v_n)$ et la suite $(\sqrt{v_n/a_n} Z_n) = (S_n/\sqrt{a_n v_n})$ satisfait à un PGD de vitesse (a_n) et de fonction de taux $I(\cdot)$.*

Nous utiliserons très souvent l'abus de langage "$(\sqrt{v_n/a_n} Z_n) = (S_n/\sqrt{a_n v_n})$ satisfait un PDM de vitesse (a_n) et de fonction de taux $I(\cdot)$" pour dire que "(Z_n) satisfait un PDM de vitesse (a_n) et de fonction de taux $I(\cdot)$". Afin d'être complet, définissons ce qu'on entend par convergence "super-exponentielle" que nous avons précédemment évoquée.

Définition 1.1.4. *Si d est une distance rendant S complet, on dit que (Z_n) converge (v_n)-super-exponentiellement vite en probabilité vers une variable aléatoire Z si*

$$\limsup_{n \to +\infty} \frac{1}{v_n} \log \mathbb{P}\left(d(Z_n, Z) > \delta\right) = -\infty \quad \forall \delta > 0.$$

Cette convergence sera souvent notée $Z_n \overset{\text{superexp}}{\underset{v_n}{\Longrightarrow}} Z$ tout au long de cette thèse.

Nous reviendrons à la fin de ce chapitre sur l'utilité de cette notion pour le PDM des martingales.

Ces deux dernières décennies, l'étude du principe de déviations modérées à pris une place de choix dans l'étude du comportement asymptotique des systèmes dynamiques et des modèles de prévisions [116], [63], [62], [64], [45], [39], [112], cette liste n'étant pas exhaustive. Plusieurs raisons expliquent cet intérêt. Tout d'abord, un principe de déviations modérées peut quelques fois être prouvé la où l'obtention d'un PGD est difficile, voire impossible. C'est le cas par exemple des modèles étudiés dans cette thèse. De plus, l'obtention d'un PGD nécessite parfois des hypothèses drastiques sur les paramètres du modèle, contrairement à l'obtention d'un PDM. Un autre intérêt de l'étude des déviations modérés est la forme très souvent accessible de sa fonction de taux, ce qui n'est pas souvent le cas pour certaines fonctions de taux de PGD rencontrées dans la littérature. En effet, la proximité des déviations modérées avec le TCL confère généralement au taux du PDM une forme quadratique. Nous renvoyons à [113] pour d'autres motivations des PDM.

C'est l'étude du PDM pour les modèles bifurcants (chapitres 2 et 3) qui nous amena à nous intéresser au PDM des martingales. Signalons que pour les martingales, les PDMs ont largement été étudiées. En 1996, Dembo [34], donna un critère pour l'obtention du PDM pour les martingales à sauts bornés. Son résultat fut prolongée plus tard par Worms [113], [114], [115] et par Djellout [38], qui utilisèrent intensivement la troncature et les résultats de Puhalskii afin de se libérer de l'hypothèse de "bornitude" imposée par Dembo. Nous utilisons dans cette thèse des idées similaires à celles utilisées par Djellout et Worms pour démontrer nos résultats sur le PDM (chapitres 3 et 5). Ceci nous permet au chapitre 5 d'étendre les résultats obtenus par Worms sur les suites autorégressives à bruit sous-gaussiens au cas où le bruit vérifie une condition dite de "Chen-Ledoux". Notons que même si elle est similaire sur la forme à celle utilisée par Djellout et Worms, la méthodologie que nous utilisons aux chapitres 3 et 5 pour établir le PDM est assez différente de la leur sur le fond. Bien sur, nous n'affirmons pas avoir découvert une nouvelle façon d'établir des PDMs puisque sous la forme, l'idée reste toujours la même : faire une troncature, appliquer les résultats de Puhalskii et démontrer la négligeabilité au sens des grandes déviations d'un terme de reste.

Les inégalités de déviations et les PDMs ci-dessus, particulièrement celle du chapitre 2, sont très liés à l'ergodicité d'une certaine chaîne de Markov associée au modèle (nous l'appelons pour l'instant chaîne de Markov induite), plus précisément à la façon dont cette chaîne de Markov converge vers sa probabilité d'équilibre. C'est ce fait qui nous poussa à nous intéresser aux bornes explicites pour l'ergodicité des chaînes de Markov (chapitre 6). En effet, comme nous le verrons au chapitre 2, les inégalités de déviations que nous obtenons sont d'autant plus précises que la borne pour l'ergodicité de la chaîne de Markov induite est explicite. Même si les résultats sur les inégalités de déviations établis dans cette thèse le sont sous une hypothèse d'ergodicité géométrique (uniforme et non uniforme) de la chaîne de Markov induite, nous pensons que sous une hypothèse d'ergodicité sous exponentielle de la chaîne de Markov induite, il serait également possible d'obtenir des inégalités de déviations (sûrement moins bonne).

Nous allons à présent faire une présentation des différents concepts qui sont étudiés dans cette thèse. Nous donnerons les points de motivation ayant conduits à leur introduction dans la littérature, ainsi qu'un résumé des principaux résultats qu'apportent cette thèse par rapport aux résultats existants.

1.2 Chaînes de Markov bifurcantes

En 2006 dans ses travaux de thèse, Guyon introduisait le concept de "chaînes de Markov bifurcantes" (CMBs). Le but était d'étudier l'effet du vieillissement sur la reproduction de Escherichia Coli (E. Coli en abrégé). E. Coli est une bactérie en forme de tige qui se reproduit en se divisant en deux. Chaque cellule produit ainsi deux filles après chaque division. Une qui possède l'extrémité la plus récente de la mère, et qu'on appelle fille de type nouveau pôle. Et l'autre qui possède l'extrémité la plus ancienne de la mère, et qu'on appelle fille de type ancien pôle (voir Figure 1.1). Dans la suite, on conviendra de nommer les individus de type nouveau pôle

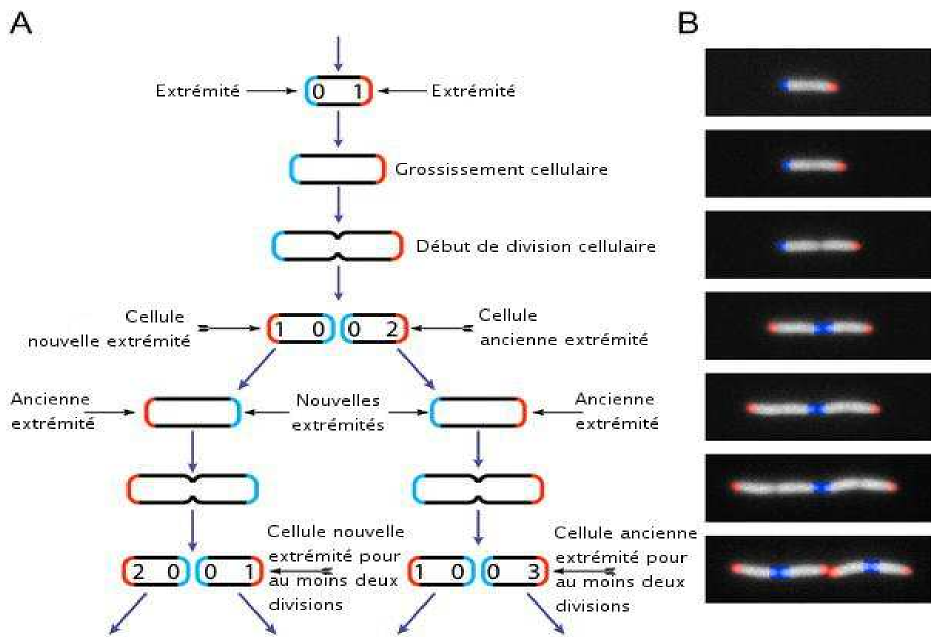

FIG. 1.1 – Cycle cellulaire d'E. Coli (source http ://commons.wikimedia.org)

par "individu de type 0", et les individus de type ancien pôle par "individu de type 1". En fait, chaque cellule fille possède deux extrémités. Une issue de la division de la cellule mère (nouveau pôle), et une autre qui existait déjà (vieux pôle). L'âge d'une cellule est donnée par l'âge de son vieux pôle, au sens du nombre de divisions depuis lesquelles ce vieux pôle existe. L'objectif fixé par Guyon était de montrer que les cellules de type 0 croisent plus vite que les cellules de type 1. A cette fin, Guyon & Al [67] ont proposé le modèle gaussien suivant. La cellule initiale est désignée par 1. Les filles de la cellule d'indice n sont désignées par $2n$- la cellule fille de type 0 et $2n+1$- la cellule fille de type 1. Soit X_n le taux de croissance de l'individu n. Alors les X_n vérifient la relation

$$\mathcal{L}(X_1) = \nu, \qquad \text{et} \qquad \forall n \geq 1, \quad \begin{cases} X_{2n} = \alpha_0 X_n + \beta_0 + \varepsilon_{2n} \\ \\ X_{2n+1} = \alpha_1 X_n + \beta_1 + \varepsilon_{2n+1}, \end{cases} \tag{1.2.1}$$

où ν est une mesure de probabilité sur $\mathbb{R}$; $\alpha_0, \alpha_1 \in (-1,1)$; $\beta_0, \beta_1 \in \mathbb{R}$ et $\big((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1\big)$ est une suite de vecteurs gaussiens bidimensionnels indépendants, centrés et de matrice de covariance

$$\Gamma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \sigma^2 > 0, \quad \rho \in (-1,1).$$

Notons que le processus défini par (1.2.1) est un exemple type de CMB, appelé processus bifurcant autorégressif d'ordre 1 (BAR(1) pour "first order bifurcating autoregressive process"). Afin de savoir si la dynamique des taux de croissance des cellules de type 0 est différente de celle des cellules de type 1, il convient :

- d'estimer les paramètres $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$, σ^2 et ρ,
- de tester l'hypothèse nulle $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$ contre son alternative $H_1 = \{(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)\}$.

Ainsi, rejeter H_0 reviendrait à admettre que la dynamique de croissance des cellules de type 0 est différente de celle des cellules de type 1. Ceci reviendrait également à admettre la notion de vieillissement cellulaire. L'étude asymptotique rigoureuse des estimateurs de θ, ρ et σ, ainsi que des tests d'hypothèse permettant de détecter le vieillissement cellulaire, a été depuis lors le principal point de motivation pour l'étude des CMBs.

Afin d'introduire plus formellement les CMBs, donnons au préalable quelques notations.

Soit $\mathbb{T}$ un arbre binaire régulier (voir Figure 1.2). Nous verrons $\mathbb{T}$ comme étant l'ensemble de toutes les générations d'une population donnée. Chaque colonne de $\mathbb{T}$ représente alors une génération dans la population. Chaque individu de la population est vu comme un entier naturel non nul. L'individu initial est 1, et $n \geq 1$ est la mère de $2n$ et $2n+1$. On note $\mathbb{G}_q = \{2^q, 2^q+1, \cdots, 2^{q+1}-1\}$ l'ensemble des individus de la q-ième génération et $\mathbb{T}_r = \cup_{q=0}^r \mathbb{G}_q$ le sous-arbre constitué des r+1 premières générations. Leurs cardinaux respectifs sont $|\mathbb{G}_q| = 2^q$ et $|\mathbb{T}_r| = 2^{r+1}-1$. La génération d'un individu donné n est $\mathbb{G}_{r_n}$ où $r_n = \lfloor \log_2 n \rfloor$, $\lfloor x \rfloor$ désignant la partie entière du réel x. On peut constater que la généalogie de E.Coli

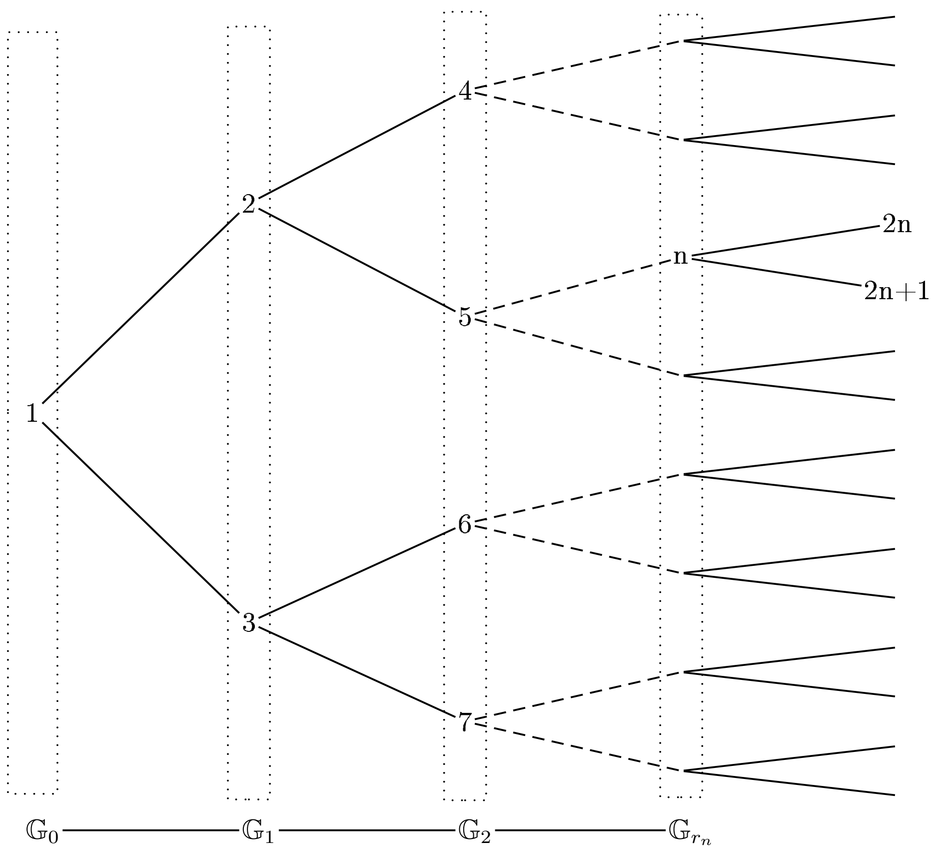

FIG. 1.2 – L'arbre binaire $\mathbb{T}$ associé à l'étude des CMBs

est entièrement décrite par $\mathbb{T}$. Pour un individu $n \in \mathbb{T}$, on s'intéresse à une quantité aléatoire X_n (taux de croissance, taille, poids...), à valeurs dans un espace métrique S muni de sa σ-algèbre borélienne $\mathcal{S}$. Dans le modèle (1.2.1) par exemple, $S = \mathbb{R}$. On suppose que cette quantité ne dépend de celle de ses ancêtres qu'à travers celle de sa mère. Afin de décrire cette structure de dépendance, introduisons la probabilité de transition suivante.

Pour tout entier $p \geq 2$, on munit S^p de la σ-algèbre produit $\mathcal{S}^p$. On appelle $\mathbb{T}$-probabilité de transition toute application $P : S \times \mathcal{S}^2 \to [0, 1]$ telle que

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

Soient $x, y, z \in S$. $P(x, dy, dz)$ donne la probabilité que le couple des quantités liées aux filles soit dans un voisinage de y et de z, sachant que la quantité liée à la mère est égale à x.

Pour une $\mathbb{T}$-probabilité de transition P, soient P_0, P_1 et Q les probabilités de transitions définies sur $S \times \mathcal{S}$ par :

$$\forall B \in \mathcal{S}, \;\; P_0(\cdot, B) = P(\cdot, B \times S), \;\; P_1(\cdot, B) = P(\cdot, S \times B) \;\text{ et }\; Q = (P_0 + P_1)/2.$$

Dans le modèle de croissance de E.Coli décrit précédemment, P_0 peut être vu comme étant la transition liée aux cellules de type 0, tandis que P_1 est la transition liée aux cellules de type 1.

Pour tout $p \geq 1$, on désigne par $\mathcal{B}(S^p)$ (resp. $\mathcal{B}_b(S^p)$, $\mathcal{C}(S^p)$, $\mathcal{C}_b(S^p)$) l'ensemble des fonctions numériques définies sur S^p qui sont, $\mathcal{S}^p$-mesurable (resp. $\mathcal{S}^p$-mesurables et bornées, continues, continues et bornées). Pour $f \in \mathcal{B}(S^3)$, lorsqu'elle existe, on désigne par Pf la fonction définie par

$$x \mapsto Pf(x) = \int_{S^2} f(x, y, z) P(x, dy, dz).$$

Nous sommes à présent en mesure de fournir une définition rigoureuse de ce qu'on entend par CMB.

Définition 1.2.1. *Soit $(X_n, n \in \mathbb{T})$ une suite de variables aléatoires définies sur un espace probabilisé filtré $(\Omega, \mathcal{F}, (\mathcal{F}_r, r \in \mathbb{N}), \mathbb{P})$ et à valeurs dans $(S, \mathcal{S})$. Soit ν une probabilité sur $(S, \mathcal{S})$. Soit P une $\mathbb{T}$-probabilité de transition sur $(S, \mathcal{S}^2)$. On dit que $(X_n, n \in \mathbb{T})$ est une $(\mathcal{F}_r)$-chaîne de Markov bifurcante (CMB) ou P-CMB de distribution initiale ν et de $\mathbb{T}$-probabilité de transition P si*

- *X_n est $\mathcal{F}_{r_n}$-mesurable pour tout $n \in \mathbb{T}$,*
- *$\mathcal{L}(X_1) = \nu$,*
- *pour tout $r \in \mathbb{N}$ et pour toute famille $(f_n, n \in \mathbb{G}_r) \subseteq \mathcal{B}_b(S^3)$*

$$\mathbb{E}\left[\prod_{n \in \mathbb{G}_r} f_n(X_n, X_{2n}, X_{2n+1}) \Big| \mathcal{F}_r\right] = \prod_{n \in \mathbb{G}_r} Pf_n(X_n).$$

Ceci signifie que, étant données les quantités $(X_n, n \in \mathbb{T}_r)$ liées aux individus des générations 0 à r, les quantités liées aux individus de la génération $\mathbb{G}_{r+1}$ sont construites en tirant 2^r couples indépendants (X_{2n}, X_{2n+1}) suivant $P(X_n, \cdot)$, $n \in \mathbb{G}_r$. Sauf mention contraire, la filtration implicitement utilisée sera $\mathcal{F}_r = \sigma(X_i, i \in \mathbb{T}_r)$. Notons que pour $f \in \mathcal{B}_b(S^3)$, $\mathbb{E}\left[f(X_n, X_{2n}, X_{2n+1}) | \mathcal{F}_{r_n}\right]$ se réduit à un conditionnement par rapport à la variable X_n, de telle sorte que $\mathbb{E}\left[f(X_n, X_{2n}, X_{2n+1}) | \mathcal{F}_{r_n}\right] = \mathbb{E}\left[f(X_n, X_{2n}, X_{2n+1}) | X_n\right]$. Ceci signifie que chaque X_n dépend du passé seulement à travers la quantité liée à la mère. Une conséquence de cette définition est que conditionnellement à $\mathcal{F}_{r_n}$, les triplets $\{(X_i, X_{2i}, X_{2i+1}), i \in \mathbb{G}_{r_n}\}$ sont indépendants ; ce fait est intensivement utilisé au chapitre 2 pour démontrer les inégalités de déviations exponentielles.

Dans le but d'étudier les estimateurs des paramètres résultants du modèle (1.2.1), il convient de définir des moyennes empiriques liées à l'observation des états d'une CMB $(X_n, n \in \mathbb{T})$. Soit donc $f \in \mathcal{B}(S)$ et I un sous-ensemble fini de $\mathbb{T}$. Posons $M_I(f) = \sum_{i\in I} f(X_i)$ et $\overline{M}_I(f) = |I|^{-1}M_I(f)$, où $|I|$ désigne le cardinal de I. Dans le cas où $f \in \mathcal{B}(S^3)$, désignons par $\Delta_i = (X_i, X_{2i}, X_{2i+1})$ la quantité associée au triangle mère-filles $(i, 2i, 2i+1)$. Et posons $M_I(f) = \sum_{i\in I} f(\Delta_i)$ et $\overline{M}_I(f) = |I|^{-1}M_I(f)$. Plusieurs moyennes empiriques peuvent ainsi être considérées.

- On peut moyenniser sur la r-ième génération, c'est-à-dire, on calcule $\overline{M}_{\mathbb{G}_r}(f)$.
- On peut moyenniser sur les $r+1$ première génération, c'est-à- dire on calcule $\overline{M}_{\mathbb{T}_r}(f)$.

On peut aussi vouloir moyenniser sur les n "premiers" individus, c'est-à-dire calculer $n^{-1}\sum_{i=1}^{n} f(X_i)$. Mais cependant, il n'existe pas a priori un ordre naturel dans la population. En effet, pour une population d'E.Coli par exemple, on ne peut déterminer quelle cellule se divise en premier, les divisions pouvant être simultanées. Pour résoudre ce problème, il convient de créer un ordre aléatoire dans la population qui respecte l'ordre généalogique. Et en plus cet ordre doit être construit de telle sorte que, le comportement asymptotique des moyennes empiriques issues de cette construction coincide avec le comportement asymptotique de $\overline{M}_{\mathbb{G}_r}(\cdot)$ et $\overline{M}_{\mathbb{T}_r}(\cdot)$. La solution est donnée par la construction suivante.

Désignons par $\mathfrak{G}$ l'ensemble de toutes les permutations de $\mathbb{N}^*$ qui laissent invariant chaque $\mathbb{G}_r$. Indépendamment de $X = (X_n, n \in \mathbb{T})$, tirons uniformément une permutation Π dans $\mathfrak{G}$. Ceci revient à tirer uniformément la restriction de Π sur $\mathbb{G}_r$ parmi les $2^r!$ permutations de $\mathbb{G}_r$. En particulier, $(\Pi(2^r), \Pi(2^r+1), \cdots, \Pi(2^{r+1}-1))$, peut être vu comme un tirage sans remise des éléments de $\mathbb{G}_r$. Π permet ainsi de définir un ordre aléatoire sur $\mathbb{T}$ qui preserve l'ordre généalogique. Désignons alors par $(\Pi(i), 1 \le i \le n)$ l'ensemble des n "premiers" individus de $\mathbb{T}$. Moyenniser sur les n "premiers" individus revient alors à calculer $\overline{M}_n^{\Pi}(f) = n^{-1}M_n^{\Pi}(f)$, où

$$M_n^{\Pi}(f) = \sum_{i=1}^{n} f(\tilde{\Delta}_{\Pi(i)}),$$

avec $f(\tilde{\Delta}_{\Pi(i)}) = f(\Delta_{\Pi(i)})$ si $f \in \mathcal{B}(S^3)$ et $f(\tilde{\Delta}_{\Pi(i)}) = f(X_{\Pi(i)})$ si $f \in \mathcal{B}(S)$.

Nous ferons souvent usage de la filtration $(\mathcal{H}_n)_{n\ge 0}$ définie par $\mathcal{H}_0 = \sigma(X_1)$ et $\mathcal{H}_n = \sigma\left(\Delta_{\Pi(i)}, \Pi(i+1), 1 \le i \le n\right)$.

Revenons un instant au modèle BAR(1) décrit ci-dessus (1.2.1). Supposons qu'on observe les $r+2$ premières générations. Alors, avec les notations ci-dessus, l'estimateur des moindres carrés- qui dans ce cadre coincide avec l'estimateur du maximum de vraisemblance- $\widehat{\theta}^r = \left(\widehat{\alpha}_0^r, \widehat{\beta}_0^r, \widehat{\alpha}_1^r, \widehat{\beta}_1^r\right)$ du paramètre $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$

est donné par, pour $\eta \in \{0,1\}$

$$\begin{cases} \widehat{\alpha}^r_\eta = \dfrac{|\mathbb{T}_r|^{-1} \sum\limits_{i\in\mathbb{T}_r} X_i X_{2i+\eta} - \left(|\mathbb{T}_r|^{-1} \sum\limits_{i\in\mathbb{T}_r} X_i\right)\left(|\mathbb{T}_r|^{-1} \sum\limits_{i\in\mathbb{T}_r} X_{2i+\eta}\right)}{|\mathbb{T}_r|^{-1} \sum\limits_{i\in\mathbb{T}_r} X_i^2 - \left(|\mathbb{T}_r|^{-1} \sum\limits_{i\in\mathbb{T}_r} X_i\right)^2} \\ \widehat{\beta}^r_\eta = |\mathbb{T}_r|^{-1} \sum\limits_{i\in\mathbb{T}_r} X_{2i+\eta} - \widehat{\alpha}^r_\eta |\mathbb{T}_r|^{-1} \sum\limits_{i\in\mathbb{T}_r} X_i. \end{cases} \tag{1.2.2}$$

Les estimateurs de la variance conditionnelle σ^2 et de la corrélation conditionnelle entre les sœurs ρ sont donnés par

$$\begin{cases} \widehat{\sigma}^2_r = \frac{1}{2\mathbb{T}_r} \sum\limits_{i\in\mathbb{T}_r} (\widehat{\varepsilon}^2_{2i} + \widehat{\varepsilon}^2_{2i+1}) \\ \widehat{\rho}_r = \frac{1}{\widehat{\sigma}^2_r} \sum\limits_{i\in\mathbb{T}_r} \widehat{\varepsilon}_{2i}\widehat{\varepsilon}_{2i+1}, \end{cases} \tag{1.2.3}$$

où les résidus sont définis par $\widehat{\varepsilon}_{2i+\eta} = X_{2i+\eta} - \widehat{\alpha}^r_\eta X_i - \widehat{\beta}^r_\eta$, avec $\eta \in \{0,1\}$.

Des théorèmes limites pour les moyennes empiriques $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_{\mathbb{T}_r}(f)$ et $\overline{M}^{\Pi}_n(f)$ ont été établis par Guyon [66] sous divers hypothèses de régularité sur f, ainsi que d'ergodicité de la chaîne de Markov de transition Q. Il s'agit notamment de la loi des grands nombres, ainsi que du théorème centrale limite. Afin de rappeler ces résultats, introduisons au préalable quelques notations. Désignons par $(Y_r, r \in \mathbb{N})$ la chaîne de Markov sur S de probabilité de transition Q, avec $Y_0 = X_1$. La chaîne $(Y_r, r \in \mathbb{N})$ corresponds aux valeurs d'une lignée aléatoire prise dans la population. Désignons par :

- $f \otimes g$ l'application $(x,y) \mapsto f(x)g(y)$,
- Q^p la puissance p-ième de Q définie récursivement par les formules $Q^0(x,\cdot) = \delta_x$ et $Q^{p+1}(x,B) = \int_S Q(s,dy)Q^p(y,B)$ pour tout $B \in \mathcal{S}$; Q^p est une probabilité de transition sur $(S,\mathcal{S})$.
- νQ est la probabilité sur $(S,\mathcal{S})$ définie par $\nu Q(B) = \int_S \nu(dx)Q(x,B)$; νQ^p est la loi de Y_p,
- $(Qf)(x) = \int_S f(y)Q(x,dy)$ quand elle est définie ;
- (νf) ou (ν, f) l'intégrale $\int_S f d\nu$ quand elle est définie.

Soit F un sous-espace vectoriel de $\mathcal{B}(S)$ tel que

(h.i) F contient les fonctions constantes,

(h.ii) $F^2 \subset F$,

(h.iii) $F \otimes F \subset L^1(P(x,.))$ pour tout $x \in S$, et $P(F \otimes F) \subset F$,

(h.iv) il existe une probabilité μ sur $(S,\mathcal{S})$ telle que $F \subset L^1(\mu)$ et $\lim\limits_{r\to\infty} \mathbb{E}_x\Big[f(Y_r)\Big] = (\mu, f)$ pour tous $x \in S$ et $f \in F$,

(h.v) pour tout $f \in F$, il existe $g \in F$ tel que pour tout $r \in \mathbb{N}$, $|Q^r f| \le g$,

(h.vi) $F \subset L^1(\nu)$,

où on a utilisé les notations $F^2 = \{f^2/f \in F\}$, $F \otimes F = \{f \otimes g/f, g \in F\}$ et $PE = \{Pf/f \in E\}$ si P agit sur E. Notons que dans le cas où la chaîne de Markov $(Y_r, r \in \mathbb{N})$ est ergodique (c'est-à-dire si (h.iv) est vérifiée pour toute fonction continue bornée f), alors, les hypothèses (h.i)-(h.vi) sont satisfaites pour $F = \mathcal{C}_b(S)$.

Dans le cadre du modèle (1.2.1), soit $\mathcal{C}_{pol}(\mathbb{R})$ l'ensemble des fonctions continues et à croissance polynomiale, c'est-à-dire l'ensemble des fonctions continues $f : \mathbb{R} \to \mathbb{R}$ telles qu'ils existent $c \geq 0$ et $m \in \mathbb{N}$ tels que, pour tout $x \in \mathbb{R}$,

$$|f(x)| \leq c(1 + |x|^m).$$

Alors $\mathcal{C}_{pol}(\mathbb{R})$ vérifie les hypothèses (h.i)-(h.v). De plus, si ν admet des moments finis de tout ordre, alors $\mathcal{C}_{pol}(\mathbb{R})$ vérifie aussi l'hypothèse (h.vi).

Le théorème qui suit synthétise l'ensemble des principaux résultats établis par Guyon.

Théorème 1.2.2. *Soit F vérifiant les hypothèses (h.i)-(h.vi).*

(a) Pour tout $f \in F$, les trois moyennes empiriques $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_{\mathbb{T}_r}(f)$ et $\overline{M}_n^{\Pi}(f)$ convergent vers (μ, f) en moyenne quadratique.

(b) Soit $f \in \mathcal{B}(S^3)$ tel que Pf et Pf^2 existent et appartiennent à F. Alors les trois moyennes empiriques $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_{\mathbb{T}_r}(f)$ et $\overline{M}_n^{\Pi}(f)$ converge vers (μ, f) en moyenne quadratique.

(c) Soit $f \in F$. Supposons qu'il existe $h \in F$ tel que

$$P\left(\sum_{r\in\mathbb{N}} |Q^r(f - (\mu, f)) \otimes Q^r(f - (\mu, f))|\right) \leq h. \tag{1.2.4}$$

Alors les moyennes empiriques $\overline{M}_{\mathbb{G}_r}(f)$ et $\overline{M}_{\mathbb{T}_r}(f)$ convergent presque sûrement vers (μ, f).

(d) Soit $f \in \mathcal{B}(S^3)$ tel que Pf et Pf^2existent et appartiennent à F. Supposons qu'il existe $h \in F$ tel que

$$P\left(\sum_{r\in\mathbb{N}} |Q^r(Pf - (\mu, Pf)) \otimes Q^r(Pf - (\mu, Pf))|\right) \leq h. \tag{1.2.5}$$

Alors les moyennes empiriques $\overline{M}_{\mathbb{G}_r}(f)$ et $\overline{M}_{\mathbb{T}_r}(f)$ convergent presque sûrement vers (μ, Pf).

(e) Soit $f \in \mathcal{B}(S^3)$ tel que Pf, Pf^2 et Pf^4 existent et appartiennent à F. Alors, $n^{-1/2}M_n^{\Pi}(f - Pf)$ et $|\mathbb{T}_r|^{-1/2}M_{\mathbb{T}_r}(f - Pf)$ convergent en loi vers la loi normale $\mathcal{N}(0, s^2)$, où $s^2 = (\mu, Pf^2 - (Pf)^2)$.

Une condition suffisante pour avoir (1.2.4) et (1.2.5) est l'existence d'une fonction $c \in F$ et une suite de nombres positifs $(\kappa_r, r \in \mathbb{N})$ telles que

$$\sum_{r \in \mathbb{N}} \kappa_r < \infty \quad \text{et} \quad \forall x \in S, \forall r \in \mathbb{N}, |Q^r f(x)| \leq c(x)\kappa_r. \tag{1.2.6}$$

Dans le cadre du modèle (1.2.1), la condition (1.2.6) est vérifiée avec $\kappa_r = \alpha^r$, où $\alpha = \max\{|\alpha_0|, |\alpha_1|\}$, pour les éléments de $\mathcal{C}^1_{pol}(\mathbb{R})$, l'ensemble des fonctions $f : \mathbb{R} \to \mathbb{R}$ de classe $\mathcal{C}^1$ telles que $|f| + |f'|$ est borné par un polynôme.

En appliquant ensuite ces résultats sur le modèle (1.2.1), Guyon a pu montrer que les estimateurs $\widehat{\theta}^r$, $\widehat{\sigma}^2_r$ et $\widehat{\rho}_r$ définis en (1.2.2) et (1.2.3) sont fortement convergents quand r tend vers l'infini. Il a aussi montré la normalité asymptotique de $\widehat{\theta}^r - \theta$.

Nous complétons et étendons ces résultats au chapitre 2 sous d'hypothèses d'ergodicité géométrique (uniforme et non-uniforme) de la chaîne de Markov $(Y_r, r \in \mathbb{N})$. Plus précisément :

• Dans la Section 2.2, en plus des hypothèses (h.i)-(h.vi), nous faisons l'hypothèse suivante :

(H1) Supposons que pour toute fonction $f \in F$ telle que $(\mu, f) = 0$, il existe $g \in F$ telle que pour tout $r \in \mathbb{N}$ et pour tout $x \in S$, $|Q^r f(x)| \leq \alpha^r g(x)$ pour un $\alpha \in (0, 1)$.

Sous les hypothèses (h.i)-(h.vi) et **(H1)**, nous commençons par établir une inégalité des moments pour la moyenne empirique $\overline{M}_{\mathbb{G}_r}(f)$, $f \in \mathcal{B}(S)$. Nous montrons alors le théorème suivant.

Théorème 1.2.3. *Sous les hypothèses (h.i)-(h.vi) et* ***(H1)****, soit $f \in F$ tel que $(\mu, f) = 0$. Alors, pour tout $r \in \mathbb{N}$*

$$\mathbb{E}\left[\left(\overline{M}_{\mathbb{G}_r}(f)\right)^4\right] \leq \begin{cases} c\left(\frac{1}{4}\right)^r & si \quad \alpha^2 < \frac{1}{2} \\ cr^2\left(\frac{1}{4}\right)^r & si \quad \alpha^2 = \frac{1}{2} \\ c\alpha^{4r} & si \quad \alpha^2 > \frac{1}{2} \end{cases} \tag{1.2.7}$$

où la constante positive c dépend de α et f.

La preuve du Théorème 1.2.3 repose essentiellement sur deux points : un calcul explicite du moment d'ordre 4 de $\overline{M}_{\mathbb{G}_r}(f)$ via des conditionnements successifs et un contrôle des différents termes apparaissant dans ce calcul, via l'utilisation des hypothèses (h.i)-(h.vi) et **(H1)**. Pour le calcul explicite du moment d'ordre 4 de $\overline{M}_{\mathbb{G}_r}(f)$, nous utilisons intensément (avec une légère modification toutefois) le calcul explicite du moment d'ordre deux de $\overline{M}_{\mathbb{G}_r}(f)$ fait par Guyon. Notons que la méthode que nous utilisons pour ce calcul généralise très bien celle utilisée par Guyon. De plus, il est possible par notre procédé de calculer les moments d'ordre supérieur (au prix d'énorme calculs certes).

Comme conséquence du Théorème 1.2.3, nous avons le corollaire suivant.

Corollaire 1.2.4. *Supposons les hypothèses (h.i)-(h.vi) et **(H1)** vérifiées.*

(i) Pour tout $f \in F$, pour tout $\delta > 0$ et pour tout $r \in \mathbb{N}$, on a

$$\mathbb{P}\Big(\big|\overline{M}_{\mathbb{T}_r}(f) - (\mu, f)\big| > \delta\Big) \leq \begin{cases} \frac{c}{\delta^4}\left(\frac{1}{4}\right)^{r+1} & si \quad \alpha^2 < \frac{1}{2} \\ \\ \frac{c}{\delta^4} r^2 \left(\frac{1}{4}\right)^{r+1} & si \quad \alpha^2 = \frac{1}{2} \\ \\ \frac{c}{\delta^4}\alpha^{4(r+1)} & si \quad \alpha^2 > \frac{1}{2}, \end{cases}$$

où la constante positive c dépend de α et f.

(ii) Soit $f \in \mathcal{B}(S^3)$ tel que Pf, Pf^2 et Pf^4 existent et appartiennent à F. Alors pour tout $\delta > 0$ et pour tout $r \in \mathbb{N}$, on a

$$\mathbb{P}\Big(\big|\overline{M}_{\mathbb{T}_r}(f - Pf)\big| > \delta\Big) \leq \frac{c}{\delta^4}\left(\frac{1}{4}\right)^{r+1}$$

où la constante positive c dépend de α et f.

Le Théorème 1.2.3 nous permet par la suite d'établir une loi forte des grands nombres pour la moyenne empirique $\overline{M}_n^{\Pi}(f)$ lorsque $f \in \mathcal{B}(S)$. Nous complétons ainsi les résultats de Guyon où la loi forte des grands nombres n'avait été donnée que pour $\overline{M}_{\mathbb{G}_r}(f)$ et $\overline{M}_{\mathbb{T}_r}(f)$. Nous établissons aussi dans la foulée, à l'aide de cette loi des grands nombres, une loi du logarithme itéré et un théorème de la limite centrale presque sûre fonctionnelle pour $M_n^{\Pi}(f - Pf)$ lorsque $f \in \mathcal{B}(S^3)$. Ces résultats sont contenus dans le théorème suivant.

Théorème 1.2.5. *Supposons que F vérifie les hypothèses (h.i)-(h.vi) et que l'hypothèse **(H1)** est satisfaite avec $\alpha \in (0, \sqrt[4]{8}/2)$. Alors :*

(i) Pour tout $f \in F$, $\overline{M}_n^{\Pi}(f)$ converge presque sûrement vers (μ, f) quand n tend vers l'infini.

(ii) Pour tout $f \in \mathcal{B}(S^3)$ tel que Pf, Pf^2 et Pf^4 existent et appartiennent à F, on a

$$\limsup_{n\to\infty} \frac{M_n^{\Pi}(f - Pf)}{\sqrt{2n \log\log n}} = \sqrt{(\mu, Pf^2 - (Pf)^2)} \quad a.s.$$

(iii) Pour tout $f \in \mathcal{B}(S^3)$ tel que Pf, Pf^2 et Pf^4 existent et appartiennent à F, $M_n^{\Pi}(f - Pf)$ vérifie un théorème de la limite centrale presque sûre fonctionnelle quand n tend vers l'infini. En particulier, posons

$$s^2 = (\mu, Pf^2 - (Pf)^2).$$

Alors

$$\lim_{n\to+\infty} \frac{1}{\log(s^2 n)} \sum_{k=1}^{n} \frac{1}{k+1} \delta_{\left\{\frac{M_k^{\Pi}(f-Pf)}{s\sqrt{k}}\right\}} = \mathcal{N}(0,1) \quad p.s., \tag{1.2.8}$$

où $\mathcal{N}(0,1)$ *est la loi normale standard sur* $\mathbb{R}$*, et la convergence est la convergence étroite des mesures sur* $\mathbb{R}$*.*

Remarque 1.2.6. *Soit* $(M_n, n \geq 0)$ *une martingale réelle centrée. On note* $\mathbb{F}$ *sa filtration naturelle.* $(M_n, n \geq 0)$ *vérifie un théorème de la limite centrale presque sûre fonctionnelle si : pour une suite croissante* $(V_n)_{n\geq 0}$ *de variables aléatoires* $\mathbb{F}$*-prévisible et pour presque tout* ω*, les mesures aléatoires pondérées :*

$$W_N(\omega, \bullet) = (\log V_N^2)^{-1} \sum_{n=1}^{N} \left(1 - \frac{V_n^2}{V_{n+1}^2}\right) \delta_{\{\Psi_n(\omega)\in\bullet\}}$$

associées aux processus continus $\Psi_n(\omega) = \{\Psi_n(\omega, t), 0 \leq t \leq 1\}$ *définis par :*

$$\Psi_n(\omega, t) = V_n^{-1}\{M_k + (V_{k+1}^2 - V_k^2)^{-1}(tV_n^2 - V_k^2)(M_{k+1} - M_k)\},$$

lorsque $V_k^2 \leq tV_n^2 < V_{k+1}^2$, $0 \leq k \leq n-1$ *convergent étroitement vers la mesure de Wiener sur* $\mathcal{C}([0,1], \mathbb{R})$*.*

Remarque 1.2.7.

(1) Rappelons q'une suite de mesures finies $(\nu_n)_{n\in\mathbb{N}}$ *sur* $(S, \mathcal{S})$ *converge étroitement vers une mesure finie* ν *sur* $(S, \mathcal{S})$ *si*

$$\int_S f d\nu_n \underset{n\to+\infty}{\longrightarrow} \int_S f d\nu$$

pour toute fonction réelle continue bornée f *sur* S*.*

(2) La limite (1.2.8) peut encore être réécrite sous la forme

$$\lim_{n\to+\infty} \frac{1}{\log(s^2 n)} \sum_{k=1}^{n} \frac{1}{k+1} \mathbf{1}_{\{M_k^{\Pi}(f-Pf)\leq sx\sqrt{k}\}} = \Phi(x) \quad p.s.,$$

où $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{t^2}{2}) dt$ *est la fonction cumulative de la loi normale standard.*

À la fin de la Section 2.2, nous établissons des inégalités de déviations (non exponentielles) pour les moyennes empiriques $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_{\mathbb{T}_r}(f)$ et $\overline{M}_n^{\Pi}(f)$. Ces inégalités nous sont utiles par la suite pour montrer un PDM pour $\overline{M}_n^{\Pi}(f - Pf)$ lorsque $f \in \mathcal{B}(S^3)$. Nous utilisons une inégalité de Markov classique pour établir les inégalités de déviations. Nous nous sommes restreints à un contrôle des moments d'ordre 2. La raison principale pour cela fut que nous nous rendîmes compte que l'utilisation des moments d'ordres supérieurs n'améliorait pas la vitesse dans le PDM. Mais toutefois, le procédé que nous avons développer pour calculer les moments de $\overline{M}_{\mathbb{G}_r}(f)$ peut être utilisé pour améliorer la qualité des inégalités de déviations lorsque $f \in \mathcal{B}(S)$. Nous synthétisons dans le théorème qui suit quelque un de ces résultats.

Théorème 1.2.8. *Sous les hypothèses (h.i)-(h.vi) et **(H1)** on a :*
(i) Pour tout $f \in F$, pour tout $\delta > 0$ et pour tout $n \in \mathbb{N}$,

$$\mathbb{P}\Big(\big|\overline{M}_n^{\Pi}(f) - (\mu, f)\big| > \delta\Big) \leq \begin{cases} \frac{c}{\delta^2}\left(\frac{1}{2}\right)^{r_n+1} & \text{if} \quad \alpha^2 < \frac{1}{2} \\ \frac{c}{\delta^2} r_n \left(\frac{1}{2}\right)^{r_n+1} & \text{if} \quad \alpha^2 = \frac{1}{2} \\ \frac{c}{\delta^2}\alpha^{2(r_n+1)} & \text{if} \quad \alpha^2 > \frac{1}{2}, \end{cases}$$

où la constante positive c dépend de α et f.
(ii) Soit $f \in \mathcal{B}(S^3)$ tel que Pf, Pf^2 et Pf^4 existent et appartiennent à F. Soit (b_n) une suite croissante de nombre réels positifs telle que $\sqrt{n} = o(b_n)$ et $b_n = o(\sqrt{n \log n})$. Si

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log\left(n \operatorname*{ess\,sup}_{1\leq k \leq c^{-1}(b_{n+1})} \mathbb{P}\Big(\big|(f - Pf)\left(\Delta_{\Pi(k)}\right)\big| > b_n \Big| \mathcal{H}_{k-1}\Big)\right) = -\infty,$$

où $c^{-1}(b_{n+1}) := \inf\left\{k \in \mathbb{N} : \frac{k}{b_k} \geq b_{n+1}\right\}$, alors $\left(M_n^{\Pi}(f - Pf)/b_n\right)$ satisfait un PDM dans $\mathbb{R}$ de vitesse b_n^2/n et de bonne fonction de taux $I(x) = \dfrac{x^2}{2(\mu, Pf^2 - (Pf)^2)}$.

Remarque 1.2.9. *Soulignons que la dichotomie autour de la valeur $\alpha^2 = 1/2$, (Théorème 1.2.3 et Théorème 1.2.8), apparaît naturellement dans les calculs. Elle est liée à la structure binaire du processus.*

• Nous effectuons dans la Section 2.3 le même travail qu'à la fin de la Section 2.2, c'est-à-dire nous établissons des inégalités de déviations ainsi qu'un PDM. Mais cette fois, nous travaillons avec des fonctions bornées (appartenant à $\mathcal{B}_b(S)$ ou à $\mathcal{B}_b(S^3)$). Nous supposons en plus que la chaîne de Markov $(Y_r, r \in \mathbb{N})$ est uniformément géométriquement ergodique. Plus précisément, nous faisons l'hypothèse suivante.

(H2) Il existe une mesure de probabilité μ sur $(S, \mathcal{S})$ telle que pour tout $f \in \mathcal{B}_b(S)$, il existe une constante positive c telle que

$$|Q^r(f)(x) - (\mu, f)| \leq c\alpha^r \quad \text{pour un } \alpha \in (0,1) \text{ et pour tout } x \in S.$$

Nous rappelons que dans ces conditions, $\mathcal{B}_b(S)$ vérifie les hypothèses (h.i)-(h.vi). Sous l'hypothèse **(H2)**, nous obtenons alors des inégalités de déviations exponentielles pour les moyennes empiriques $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_{\mathbb{T}_r}(f)$ et $\overline{M}_n^{\Pi}(f)$, ainsi qu'un PDM pour $\overline{M}_n^{\Pi}(f - Pf)$ lorsque $f \in \mathcal{B}_b(S^3)$. On a plus précisément le résultat partiel suivant.

Théorème 1.2.10. *On suppose l'hypothèse* ***(H2)*** *satisfaite.*

(i) Pour tout $f \in \mathcal{B}_b(S)$ *et pour tout* $\delta > 0$ *on a*

$$\mathbb{P}\Big(\overline{M}_n^{\Pi}(f) - (\mu, f) > \delta\Big) \leq \begin{cases} \exp\left(c''\delta\right)\exp\left(-c'\delta^2 n\right), & \forall n \in \mathbb{N}, \quad si\ \alpha < \frac{1}{2}, \\ \exp\left(2c'\delta(r_n+1)\right)\exp\left(-c'\delta^2 n\right), & \forall n \in \mathbb{N}, \quad si\ \alpha = \frac{1}{2}, \\ \exp\left(-c'\delta^2 n\right), \forall n \in \mathbb{N}\ tel\ que\ r_n > r_0, & si\ \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}, \\ \exp\left(-c'\delta^2 \frac{n}{r_n+1}\right), \forall n \in \mathbb{N}\ tel\ que\ r_n > r_0, & si\ \alpha = \frac{\sqrt{2}}{2}, \\ \exp\left(-c'\delta^2 \frac{1}{\alpha^{2(r_n+1)}}\right), \forall n \in \mathbb{N}^*\ tel\ que\, r_n > r_0 - 2, & si\ \alpha > \frac{\sqrt{2}}{2}, \end{cases}$$

où $r_0 := \log\left(\frac{\delta}{c_0}\right) / \log(\alpha)$*, et* c_0*,* c' *et* c'' *sont des constantes positives dépendants de* α*,* $\|f\|_\infty$ *et* c*.*

(ii) Soit (b_n) *une suite croissante de nombres positifs réels telle que*

- $\frac{b_n}{\sqrt{n}} \longrightarrow +\infty$,
- *si* $\alpha^2 < \frac{1}{2}$*, la suite* (b_n) *est telle que* $b_n/n \longrightarrow 0$*,*
- *si* $\alpha^2 = \frac{1}{2}$*, la suite* (b_n) *est telle que* $(b_n \log n)/n \longrightarrow 0$*,*
- *si* $\alpha^2 > \frac{1}{2}$*, la suite* (b_n) *est telle que* $(b_n \alpha^{r_n+1})/\sqrt{n} \longrightarrow 0$*.*

Alors, pour tout $f \in \mathcal{B}_b(S^3)$*,* $\Big(M_n^{\Pi}(f - Pf)/b_n\Big)$ *satisfait à un PDM dans* S *de vitesse* b_n^2/n *et de bonne fonction de taux* $I(x) = \dfrac{x^2}{2(\mu, Pf^2 - (Pf)^2)}$*.*

Notons que les inégalités obtenus pour $\overline{M}_{\mathbb{G}_r}(f)$ et $\overline{M}_{\mathbb{T}_r}(f)$ sont du même style que celles obtenues dans le Théorème 1.2.10. Les résultats obtenus dans cette Section sont nettement meilleurs que ceux de la Section 2.2. Mais le prix à payer pour cette amélioration est l'introduction d'hypothèses fortes (ergodicité géométrique uniforme, fonctionnelles bornées). La preuve de la partie (i) du Théorème 1.2.10 utilise intensivement l'inégalité d'Azuma-Bennet-Hoeffding, le fait que conditionnellement à $\mathcal{F}_{r_n}$ les triplets $\{(X_i, X_{2i}, X_{2i+1}), i \in \mathbb{G}_{r_n}\}$ sont indépendants (ceci est une conséquence de la propriété de Markov) et des conditionnements successifs pour contrôler les moments exponentiels des sommes $M_n^{\Pi}(f)$. Pour la preuve de la partie (ii), nous utilisons le fait que $(M_n^{\Pi}(f - Pf))$ est une martingale. Nous appliquons alors les résultats connus sur le PDM des martingales [34], [38], [113]. Le point essentiel ici est de montrer la convergence super-exponentielle du crochet de la martingale. Cette convergence est obtenue à l'aide des inégalités de déviations de la partie (i).

Remarque 1.2.11. *Une fois de plus,soulignons que la dichotomie autour des valeurs $\alpha = 1/2$ et $\alpha^2 = 1/2$, (Théorème 1.2.10), apparaît naturellement dans les calculs. Elle est liée à la structure binaire du processus.*

• Dans la Section 2.4, nous appliquons tous ces résultats au cas particulier des processus BAR(1). Nous travaillons dans deux cadres. Dans le premier, nous supposons que le processus BAR(1) est défini comme (1.2.1), c'est-à-dire avec un bruit gaussien. Dans le deuxième, nous supposons que la suite de bruits, ainsi que la valeur initiale X_1 sont à valeurs dans un ensemble compact ; cette hypothèse entraîne clairement que le processus $(X_n, n \in \mathbb{T})$ est borné. Dans les deux cadres, nous montrons que les estimateurs $\widehat{\theta}^r$ et $(\widehat{\sigma}_r^2, \widehat{\rho}_r)$ définis par les équations (1.2.2) et (1.2.3) convergent super-exponentiellement vite en probabilité respectivement vers θ et (σ^2, ρ), les vitesses étant définies comme au Théorème 1.2.8 pour le cadre gaussien et comme au Théorème 1.2.10 pour le cadre borné. Dans le cadre gaussien, nous établissons des inégalités de déviations non exponentielles pour $\widehat{\theta}^r - \theta$ en utilisant le contrôle des moments d'ordre 4 fait à la Section 2.2 (voir aussi le Corollaire 1.2.4). Dans le cadre borné, les inégalités de déviations que nous obtenons pour $\widehat{\theta}^r - \theta$ sont exponentielles. Nous prouvons ensuite dans les deux cadres un PDM pour $\widehat{\theta}^r - \theta$, ainsi que pour le test statistique

$$\chi_r^{(1)} = \frac{|\mathbb{T}_r|}{2\widehat{\sigma}_r^2} \left\{ (\widehat{\alpha}_0^r - \widehat{\alpha}_1^r)^2 (\widehat{\mu}_{2,r}^2 - \widehat{\mu}_{1,r}^2) + \left((\widehat{\alpha}_0^r - \widehat{\alpha}_1^r)\widehat{\mu}_{1,r} + \widehat{\beta}_0^r - \widehat{\beta}_1^r \right)^2 \right\},$$

($\widehat{\mu}_{2,r}$ $\widehat{\mu}_{1,r}$ étant deux fonctions continues de $\widehat{\theta}^r$ et $\widehat{\sigma}_r^2$) permettant de discriminer entre l'hypothèse nulle $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$ et son alternative $H_1 = \{(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)\}$. Comme pour la convergence super-exponentielle, la vitesse du PDM dans le cadre gaussien est définie comme au Théorème 1.2.8, et la vitesse dans le cadre borné est définie comme au Théorème 1.2.10.

1.2.1 Problèmes ouverts et perspectives

Comme nous l'avons signaler précédemment, les inégalités de déviations exponentielles obtenues dans ce chapitre le sont sous d'hypothèses fortes (fonctions tests bornées, ergodicité géométrique uniforme). Il serait donc assez intéressant d'affaiblir ces hypothèses.

Afin de prendre en compte des fonctions tests lipschitziennes et se libérer par la même occasion de l'hypothèse d'ergodicité géométrique uniforme, nous envisageons de faire appel à la théorie des inégalités de transports. L'idée est de suivre une approche similaire à celle utilisée par Djellout-Guillin-Wu [42]. La principale difficulté qui se profile à l'horizon est celle de la fonction coût à utiliser. En effet, comme très souvent avec les CMBs, à cause de la structure de branchement qui leur est associée, une application grossière des résultats usuels nous mène à une "impasse". Il faudra donc sans doute définir une distance qui prennent en compte cette structure ; mais comment le faire ? Telle est la question que nous nous proposons de répondre dans des travaux futurs.

Les PDMs que nous avons énoncé aux Théorèmes 1.2.8 et 1.2.10 le sont pour des fonctions f de trois variables avec un centrage par rapport à Pf, P étant la $\mathbb{T}$-probabilité de transition. Pour des fonctions f dépendants d'une variable et pour un centrage par rapport (μ, f), les PDMs ne sont pas abordés dans cette thèse. Soulignons que pour ce cas, la plupart des méthodes classiques connues pour établir des PDMs échouent. La faute est due à la structure de branchement liée aux CMBs qui, dans le cas de la décomposition de Gordin par exemple, fait exploser le terme de reste. L'obtention d'un PDM pour les CMBs lorsque la fonctions f dépend d'une variable est encore un problème ouvert.

1.3 Processus bifurcants autorégressifs d'ordre p

En 2009, Bercu & Al [18] ont étudié une extension du modèle (1.2.1) en utilisant une approche martingale. Au lieu d'une simple dépendance par rapport au taux de croissance de la mère, ils ont supposé que le taux de croissance d'un individu peut dépendre de celui de ses ancêtres jusqu'à p générations dans le passé. Ils ont donc considéré le modèle bifurcant autorégressif d'ordre p (BAR(p) pour "p-th order bifurcating autoregressive process") suivant. Soit $p \in \mathbb{N}^*$. Pour tout $n \geq 2^{p-1}$,

$$\begin{cases} X_{2n} = a_0 + \sum_{k=1}^{p} a_k X_{\lfloor \frac{n}{2^{k-1}} \rfloor} + \varepsilon_{2n} \\ X_{2n+1} = b_0 + \sum_{k=1}^{p} b_k X_{\lfloor \frac{n}{2^{k-1}} \rfloor} + \varepsilon_{2n+1}, \end{cases} \tag{1.3.1}$$

où $\lfloor x \rfloor$ désigne la partie entière de x. L'état initial du processus est donné par $\{X_k, 1 \leq k \leq 2^{p-1}-1\}$. La suite $\{(\varepsilon_{2n}, \varepsilon_{2n+1}), n \in \mathbb{N}^*\}$ représente la suite des bruits du processus. Les paramètres $(a_0, a_1, \cdots, a_p)$ et $(b_0, b_1, \cdots, b_p)$ sont inconnus. Ce processus BAR(p) peut s'écrire sous la forme plus condensée suivante. Pour tout $n \geq 2^{p-1}$,

$$\begin{cases} \mathbb{X}_{2n} = A\mathbb{X}_n + \eta_{2n} \\ \mathbb{X}_{2n+1} = B\mathbb{X}_n + \eta_{2n+1}, \end{cases} \tag{1.3.2}$$

où $\mathbb{X}_n = \left(X_n, X_{\lfloor \frac{n}{2} \rfloor}, \cdots, X_{\lfloor \frac{n}{2^{p-1}} \rfloor}\right)^t$, $\eta_{2n} = (a_0 + \varepsilon_{2n})e_1$, $\eta_{2n+1} = (b_0 + \varepsilon_{2n+1})e_1$, avec $e_1 = (1, 0, \cdots, 0)^t \in \mathbb{R}^p$. A et B désignent les matrices compagnons

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_p \\ 1 & 0 & \cdots & 0 \\ 0 & . & \cdots & . \\ 0 & \vdots & 1 & \vdots \end{pmatrix} \qquad \text{et} \quad B = \begin{pmatrix} b_1 & b_2 & \cdots & b_p \\ 1 & 0 & \cdots & 0 \\ 0 & . & \cdots & . \\ 0 & \vdots & 1 & \vdots \end{pmatrix}.$$

Nous supposons que les matrices A et B satisfont la propriété de contraction

$$\beta = \max(||A||, ||B||) < 1, \tag{1.3.3}$$

où pour toute matrice M, M^t, $\|M\|$ et $\text{Tr}(M)$ représentent respectivement la transposée, la norme euclidienne et la trace de M. Nous utilisons les même notations que dans la Section précédente pour l'arbre binaire $\mathbb{T}$. On désigne par $\mathbb{T}_{n,p} = \{k \in \mathbb{T}_n, k \geq 2^p\}$ le sous-arbre constitué de tous les individus de la génération p à la génération n. On peut observer que, pour tout $n \geq 1$, $\mathbb{T}_{n,0} = \mathbb{T}_n$ et pour tout $p \geq 1$, $\mathbb{T}_{p,p} = \mathbb{G}_p$. Soit

$$\theta = \begin{pmatrix} a_0 & b_0 \\ a_1 & b_1 \\ \vdots & \vdots \\ a_p & b_p \end{pmatrix}.$$

Le processus BAR(p) peut encore être écrit, pour tout $n \geq 2^{p-1}$, sous la forme matricielle

$$Z_n = \theta^t Y_n + V_n, \tag{1.3.4}$$

où

$$Z_n = \begin{pmatrix} X_{2n} \\ X_{2n+1} \end{pmatrix}, \quad Y_n = \begin{pmatrix} 1 \\ \mathbb{X}_n \end{pmatrix}, \quad V_n = \begin{pmatrix} \varepsilon_{2n} \\ \varepsilon_{2n+1} \end{pmatrix}.$$

En supposant qu'on observe tous les individus jusqu'à la génération n, c'est-à-dire qu'on observe entièrement le sous arbre $\mathbb{T}_n$, l'estimateur des moindres carrés de θ est donné par

$$\widehat{\theta}_n = S_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1,p-1}} Y_k Z_k^t, \tag{1.3.5}$$

où S_n est la matrice carré d'ordre $p+1$ définie par

$$S_n = \sum_{k \in \mathbb{T}_{n,p-1}} Y_k Y_k^t = \sum_{k \in \mathbb{T}_{n,p-1}} \begin{pmatrix} 1 & \mathbb{X}_k^t \\ \mathbb{X}_k & \mathbb{X}_k \mathbb{X}_k^t \end{pmatrix}. \tag{1.3.6}$$

Sans nuire à la généralité, nous supposons que pour tout $n \geq p-1$, S_n est inversible. Dans le cas contraire, il suffit de travailler avec $I_{p+1} + S_n$, où I_{p+1} désigne la matrice identité d'ordre $p+1$. En assimilant respectivement θ et $\widehat{\theta}_n$ aux vecteurs

$$\text{vec}(\theta) = \begin{pmatrix} a_0 \\ \vdots \\ a_p \\ b_0 \\ \vdots \\ b_p \end{pmatrix} \quad \text{et} \quad \text{vec}(\widehat{\theta}_n) = \begin{pmatrix} \widehat{a}_{0,n} \\ \vdots \\ \widehat{a}_{p,n} \\ \widehat{b}_{0,n} \\ \vdots \\ \widehat{b}_{p,n} \end{pmatrix},$$

on obtient en utilisant (1.3.5)

$$\widehat{\theta}_n = \Sigma_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1,p-1}} \text{vec}(Y_k Z_k^t) = \Sigma_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1,p-1}} \begin{pmatrix} X_{2k} \\ X_k \mathbb{X}_{2k} \\ X_{2k+1} \\ X_k \mathbb{X}_{2k+1} \end{pmatrix}, \tag{1.3.7}$$

où $\Sigma_n = I_2 \otimes S_n$, $\otimes$ représentant le produit de Kronecker des matrices. Il s'ensuit, en utilisant (1.3.4), que

$$\widehat{\theta}_n - \theta = \Sigma_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1,p-1}} \begin{pmatrix} \varepsilon_{2k} \\ \varepsilon_{2k} \mathbb{X}_k \\ \varepsilon_{2k+1} \\ \varepsilon_{2k+1} \mathbb{X}_k \end{pmatrix}. \tag{1.3.8}$$

Soit $\mathbb{F} = (\mathcal{F}_n)$ la filtration naturelle associée au processus BAR(p), c'est-à-dire, $\mathcal{F}_n = \sigma\{X_k, k \in \mathbb{T}_n\}$. Supposons que la suite de bruit $\{(\varepsilon_{2n}, \varepsilon_{2n+1}), n \in \mathbb{N}^*\}$ vérifie les hypothèses suivantes.

(h.I) Il existe $\sigma^2 > 0$ tel que, pour tout $n \geq p-1$ et pour tout $k \in \mathbb{G}_{n+1}$, ε_k appartient à L^2 avec

$$\mathbb{E}\left[\varepsilon_k | \mathcal{F}_n\right] = 0 \quad \text{et} \quad \mathbb{E}\left[\varepsilon_k^2 | \mathcal{F}_n\right] = \sigma^2 \quad \text{p.s.}$$

(h.II) Il existe $|\rho| < \sigma^2$ tel que, pour tout $n \geq p-1$ et pour tout $k \neq l \in \mathbb{G}_{n+1}$ avec $\lfloor k/2 \rfloor = \lfloor l/2 \rfloor$,

$$\mathbb{E}\left[\varepsilon_k \varepsilon_l | \mathcal{F}_n\right] = \rho \quad \text{p.s.}$$

(h.III) Pour tout $n \geq p-1$ et pour tout $k \in \mathbb{G}_{n+1}$, $\varepsilon_k \in L^4$ et

$$\sup_{n \geq p-1} \sup_{k \in \mathbb{G}_{n+1}} \mathbb{E}\left[\varepsilon_k^4 | \mathcal{F}_n\right] < \infty \quad \text{p.s.}$$

(h.IV) Il existe $\tau^4 > 0$ et $\nu^2 < \tau^4$ tel que, pour tout $n \geq p-1$ et pour tout $k \neq l \in \mathbb{G}_{n+1}$ avec $\lfloor k/2 \rfloor = \lfloor l/2 \rfloor$,

$$\mathbb{E}\left[\varepsilon_k^4 | \mathcal{F}_n\right] = \tau^4 \quad \text{et} \quad \mathbb{E}\left[\varepsilon_k^2 \varepsilon_l^2 | \mathcal{F}_n\right] = \nu^2 \quad \text{p.s.}$$

(h.V) Pour tout $n \geq p-1$ et pour tout $k \in \mathbb{G}_{n+1}$, ε_k appartient à L^8 avec

$$\sup_{n\geq p-1} \sup_{k\in\mathbb{G}_{n+1}} \mathbb{E}\left[\varepsilon_k^8|\mathcal{F}_n\right] < \infty \quad \text{p.s.}$$

En supposant encore qu'on observe entièrement tous les individus de $\mathbb{T}_n$, les estimateurs $\widehat{\sigma}_n^2$ et $\widehat{\rho}_n$ de σ^2 et ρ sont donnés par

$$\widehat{\sigma}_n^2 = \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k\in\mathbb{T}_{n-1,p-1}} ||\widehat{V}_k||^2 = \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k\in\mathbb{T}_{n-1,p-1}} (\widehat{\varepsilon}_{2k}^2 + \widehat{\varepsilon}_{2k+1}^2), \tag{1.3.9}$$

où pour tout $n \geq p-1$ et pour tout $k \in \mathbb{G}_n$, $\widehat{V}_k^t = (\widehat{\varepsilon}_{2k}, \widehat{\varepsilon}_{2k+1})^t$ avec

$$\begin{cases} \widehat{\varepsilon}_{2k} = X_{2k} - \widehat{a}_{0,n} - \sum_{i=1}^p \widehat{a}_{i,n} X_{\lfloor \frac{k}{2^{i-1}} \rfloor} \\ \widehat{\varepsilon}_{2k+1} = X_{2k+1} - \widehat{b}_{0,n} - \sum_{i=1}^p \widehat{b}_{i,n} X_{\lfloor \frac{k}{2^{i-1}} \rfloor}, \end{cases}$$

et

$$\widehat{\rho}_n = \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k\in\mathbb{T}_{n-1,p-1}} \widehat{\varepsilon}_{2k}\widehat{\varepsilon}_{2k+1}. \tag{1.3.10}$$

En utilisant la théorie des martingales et sous les hypothèses (h.I)-(h.V), Bercu & Al [18] ont montré que les estimateurs $\widehat{\theta}_n$, $\widehat{\sigma}_n^2$ et $\widehat{\rho}_n$ sont fortements consistants. Ils ont également établi la normalité asymptotique de $\widehat{\theta}_n - \theta$, $\widehat{\sigma}_n^2 - \sigma$ et $\widehat{\rho}_n - \rho$. Plus précisément, pour une certaine matrice symétrique définie positive Λ, ils ont établi les résultats suivants.

Théorème 1.3.1. *Supposons que (ε_n) satisfait aux hypothèses (h.I)-(h.III). Alors $\widehat{\theta}_n$, $\widehat{\sigma}_n^2$ et $\widehat{\rho}_n$ convergent respectivement presque sûrement vers θ, σ^2 et ρ. De plus, si (ε_n) satisfait aux hypothèses supplémentaires (h.IV)-(h.V), alors on a*

$$|\mathbb{T}_{n-1}|^{1/2}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Lambda),$$

$$|\mathbb{T}_{n-1}|^{1/2}(\widehat{\sigma}_n^2 - \sigma) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \frac{\tau^4 - 2\sigma^4 + \nu^2}{2}),$$

et

$$|\mathbb{T}_{n-1}|^{1/2}(\widehat{\rho}_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu^2 - \rho^2).$$

Nous prolongeons ces résultats au Chapitre 3 de cette thèse en étudiant les inégalités de déviations et un PDM pour $\widehat{\theta}_n - \theta$, ainsi que des vitesses exponentielles de convergence en probabilité pour les estimateurs $\widehat{\sigma}_n^2$ et $\widehat{\rho}_n$. Plus précisément, sous des hypothèses appropriées sur le bruit $\{(\varepsilon_{2n}, \varepsilon_{2n+1}), n \in \mathbb{N}^*\}$, on a le théorème suivant (pour une vitesse (b_n^2) dépendant des hypothèses et pour une certaine matrice inversible Λ).

Théorème 1.3.2.

(i) La suite $\left(\sqrt{|\mathbb{T}_{n-1}|}(\widehat{\theta}_n - \theta)/b_{|\mathbb{T}_{n-1}|}\right)_{n\geq 1}$ *satisfait un PDM dans* $\mathbb{R}^{2(p+1)}$ *de vitesse* $b^2_{|\mathbb{T}_{n-1}|}$ *et de bonne fonction de taux* $I_\theta(x) = \frac{1}{2}x^t\Lambda^{-1}x.$

(ii)

$$\widehat{\sigma}_n^2 \underset{b^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} \sigma^2 \quad et \quad \widehat{\rho}_n \underset{b^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} \rho.$$

La preuve du PDM utilise une approche martingale. Comme nous l'avons déjà mentionner, la technique consiste à faire une troncature et à utiliser les résultats de Puhalskii sur la partie bornée ; on montre ensuite la négligeabilité de la partie non bornée au sens des grandes déviations. Le plus gros du problème revient toujours à montrer la convergence super-exponentielle du crochet de martingales. Pour le faire, nous avons établi des inégalités de déviations pour des sommes convenablement normalisées des variables aléatoires $(\varepsilon_n^a, n \in \mathbb{T})$ et $(\beta^{r_n}\varepsilon_n^a, n \in \mathbb{T})$, avec $a \in [1, 4+\epsilon)$, pour un $\epsilon > 0$. Ces inégalités de déviations peuvent être vu comme une adaptation des inégalités de Hoeffding dans le cas où la suite de variables aléatoires est indexée par les données d'un arbre binaire. En plus de la convergence super-exponentielle du crochet de la martingale, ces calculs nous permettent aussi d'obtenir des inégalités de déviations pour $\widehat{\theta}_n - \theta$. La convergence super-exponentielle de la partie (ii) du Théorème 1.3.2 s'obtient grâce à la formule de Taylor-Lagrange (la δ-méthode) et à la convergence super-exponentielle du crochet de la martingale. On peut donc constater qu'implicitement, les inégalités de déviations jouent encore un rôle déterminant pour obtenir ce résultat.

1.3.1 Problèmes ouverts et perspectives

En considérant le modèle BAR(1), la question qui se pose est celle de savoir quelle est le comportement du point de vu des déviations modérées des estimateurs de a_1 et b_1 dans le cas "explosif" (c'est-à-dire quand $|a_1| > 1$ ou $|b_1| > 1$) ? En s'inspirant des travaux de Worms [113], nous avons grand espoir d'apporter une réponse à cette question dans des travaux futurs.

Une autre question d'intérêt est d'étudier le comportement asymptotique, ainsi qu'un PDM pour les estimateurs des paramètres du modèle BAR(1) lorsqu'on se trouve à la limite du cas stable. Cela revient à considérer par exemple que $a_1 = a_1(n) = 1 - a/k_n$ et $b_1 = b_1(n) = 1 - b/k_n$, où a et b sont des réels positifs et $(k_n)_{n\in\mathbb{N}}$ est une suite croissante convergeant vers l'infini plus lentement que n. Soulignons que dans le cas des modèles AR(1), des problèmes similaires ont déjà été étudiés (voir [88] pour plus de détails).

Dernièrement, De Saporta & Al [31] ont proposé un modèle BAR(1) avec des coefficients aléatoires et prenant en compte les données manquantes. Ils ont pour cela modéliser les données disponibles à l'aide d'un arbre de Galton-Watson. Dans le modèle biologique décrit pour E. Coli par exemple, prendre en compte les données manquantes revient juste à considérer la possibilité pour une cellule de mourir.

Dans leur travail, De Saporta & Al ont fait une étude asymptotique des estimateurs des moindres carrés modifiés des paramètres de leur modèle (convergence presque sûre, loi forte quadratique et normalité asymptotique). Dans nos travaux futurs, nous envisageons d'établir un PDM pour ces estimateurs. La principale difficulté qui se dégage d'ores et déjà est le fait que les ensembles d'indices qu'on manipule sont aléatoires. De plus, les crochets des martingales qui y apparaissent convergent vers des valeurs aléatoires. Ceci fait que les résultats de Dembo, Djellout et Worms ne sont plus valables.

1.4 Chaînes de Markov bifurcantes sur un arbre de Galton-Watson

Dans le modèle de reproduction d'E. Coli présenté précédemment, la mort des cellules n'est pas prise en compte. Il résulte pourtant des expériences biologiques que les cellules mortes représentent parfois une proportion non négligeable de la population. Pour prendre en compte ce fait, Guyon a supposé dans ces expériences numériques que, dans l'échantillon étudié, seuls les individus de la dernière génération sont susceptibles de mourir. En première approximation, cette hypothèse à bien un sens. En effet, dans un échantillon donné, le nombre d'individu de la dernière génération représente la moitié de la population totale. Soulignons au passage que l'application $\Pi(\cdot)$ introduite précédemment permet de tenir compte de cette approximation.

Afin de prendre en compte de façon plus rigoureuse la mort des cellules, Delmas et Marsalle [33] ont utilisé un arbre de Galton-Watson binaire pour modéliser le taux de croissance des cellules. Le modèle qu'ils ont proposé est le suivant. Soit $\mathbb{T}$ l'arbre binaire régulier défini à la Section 1.2. Désignons par X_n le taux de croissance de la cellule "n".

- Avec probabilité $p_{1,0}$, n donne naissance à deux cellules $2n$ et $2n+1$ qui auront chacune une descendance. Le taux de croissance des filles X_{2n} et X_{2n+1} est alors lié à celui de leur mère via les equations d'autorégression (1.2.1).
- Avec probabilité p_0, n donne naissance à deux filles et seule la fille de type nouveau pôle, $2n$, aura une descendance. Son taux de croissance X_{2n} est alors lié à celui de sa mère via l'équation

$$X_{2n} = \alpha'_0 X_n + \beta'_0 + \varepsilon'_{2n}, \tag{1.4.1}$$

  où $\alpha'_0 \in (-1, 1)$, $\beta'_0 \in \mathbb{R}$ et $(\varepsilon'_{2n}, n \in \mathbb{T})$ est une suite de variables aléatoires i.i.d, de loi gaussienne centrée et de variance $\sigma_0^2 > 0$.
- Avec probabilité p_1, n donne naissance à deux filles et seule la fille de type ancien pôle, $2n+1$, aura une descendance. Son taux de croissance X_{2n+1} est alors lié à celui de sa mère via l'équation

$$X_{2n+1} = \alpha'_1 X_n + \beta'_1 + \varepsilon'_{2n+1}, \tag{1.4.2}$$

où $\alpha'_1 \in (-1, 1)$, $\beta'_1 \in \mathbb{R}$ et $(\varepsilon'_{2n+1}, n \in \mathbb{T})$ est une suite de variables aléatoires i.i.d, de loi gaussienne centrée et de variance $\sigma_1^2 > 0$.

– Avec probabilité $1 - p_{1,0} - p_1 - p_0$, n donne naissance à deux filles qui n'auront pas de descendance.
– Les suites $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \in \mathbb{T})$, $(\varepsilon'_{2n}, n \in \mathbb{T})$ et $(\varepsilon'_{2n+1}, n \in \mathbb{T})$ sont indépendantes.

Les cellules qui ont une descendance (c'est-à-dire celles qui se divisent) sont dites vivantes ; et celle qui n'ont pas de descendance ou qui n'existent pas sont dites mortes. Par exemple sur la figure 1.3, la cellule initiale donne naissance à deux

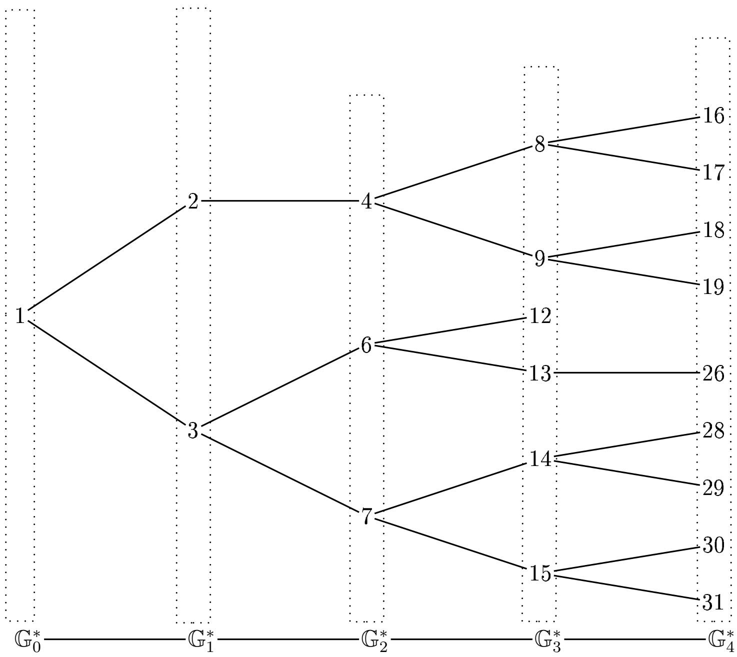

FIG. 1.3 – Arbre de Galton-Watson binaire associé à la division cellulaire de E. Coli.

cellules qui à leur tour se reproduisent ; cette situation se produit avec probabilité $p_{1,0}$. La cellule "1" est dite vivante. La cellule "2" donne naissance à deux cellules ("4" et "5") et seule la cellule "4" se reproduit à sont tour ; cette situation survient avec probabilité p_0. La cellule "4" est dite vivante, tandis que "5" et toute la descendance qu'elle pouvait avoir (c'est-à-dire "$5 \times 2^k + l$", pour tout $k \in \mathbb{N}^*$ et $l \in \{0, 1, \cdots, 2^k - 1\}$) sont dites mortes. La cellule "12" donne naissance à deux cellules ("24" et "25") qui ne se reproduisent pas ; cette situation se produit avec probabilité $1 - p_{1,0} - p_0 - p_1$. La cellule "12" est dite vivante, tandis que les cellules "24" et "25" et toute leur descendance sont dites mortes. Soulignons que sur la figure 1.3, on ne représente que les cellules vivantes.

En suivant la dénomination de De Saporta & al [32], le processus (X_n) ainsi défini sera appelé processus bifurcant du premier ordre (BAR(1)) avec données manquantes. Les données manquantes sont celles liées au cellules mortes. Sur la figure 1.3 par exemple, X_5 est une donnée manquante. Le processus ainsi défini est un exemple type de chaîne de Markov bifurcante sur un arbre de Galton-Watson.

Afin de donner une définition plus rigoureuse de ce qu'on entend par CMB sur un arbre de Galton-Watson (CMB sur un arbre de GW en abrégé), commençons par donner quelques notations. Soit S un espace métrique muni de sa σ-algèbre borélienne $\mathcal{S}$. Les notations et définitions que nous avons faites à la Section 1.2 restent valables. Soit $\overline{S} = S \cup \{\partial\}$. Dans le modèle biologique présenté précédemment, ∂ représente la valeur arbitraire qu'on attribue aux données manquantes et S est l'espace d'état des quantités liés aux cellules vivantes. Soit $\overline{\mathcal{S}}$ la σ-algèbre engendré par $\mathcal{S}$ et ∂. Soit P^* une $\mathbb{T}$-probabilité de transition définie sur $\overline{S} \times \overline{\mathcal{S}}$ telle que

$$P^*(\partial, \{(\partial, \partial)\}) = 1. \tag{1.4.3}$$

Dans le modèle biologique précédent, 1.4.3 signifie que seul un être vivant peut transmettre la vie. Nous avons alors la définition suivante.

Définition 1.4.1 (CMB sur un arbre de GW, voir [33]). *Soit $X = (X_n, n \in \mathbb{T})$ une P^*-CMB sur $(\overline{S}, \overline{\mathcal{S}})$, avec P^* vérifiant (1.4.3). Alors, le processus $(X_n, n \in \mathbb{T}^*)$, avec $\mathbb{T}^* = \{n \in \mathbb{T} : X_n \neq \partial\}$, est appelé une CMB sur un arbre de GW. La P^*-CMB est dite spatialement homogène (ou simplement homogène) si $p_{1,0} = P^*(x, S \times S)$, $p_0 = P^*(x, S \times \{\partial\})$, et $p_1 = P^*(x, \{\partial\} \times S)$ ne dépendent pas de $x \in S$. Une P^*-CMB homogène est dite sur-critique si $m > 1$, où $m = 2p_{1,0} + p_1 + p_0$.*

Le nom de CMB sur arbre de GW vient du fait que la condition 1.4.3 et l'homogénéité spatiale entraîne que $\mathbb{T}^*$ est un arbre de GW.

Soient P_0^* et P_1^* les restrictions de la première et de la seconde marginale de P^* sur S, c'est-à-dire

$$P_0^* = P^*\left(\cdot, \left(\cdot \bigcap S\right) \times \bar{S}\right) \quad \text{et} \quad P_1^* = P^*\left(\cdot, \bar{S} \times \left(\cdot \bigcap S\right)\right).$$

Désignons par $(Y_n, n \in \mathbb{N})$ la chaîne de Markov sur S de valeur initiale $Y_0 = X_1$ et de probabilité de transition $Q = \frac{1}{m}(P_0^* + P_1^*)$.

Pour tout sous ensemble $J \subset \mathbb{T}$, soit

$$J^* = J \cap \mathbb{T}^* = \{j \in J : X_j \neq \partial\}$$

le sous ensemble des cellules vivantes dans J et $|J|$ le cardinal de J. Afin d'étudier les estimateurs des paramètres liés au modèle BAR(1) avec données manquantes, définissons les quantités suivantes liées aux CMBs sur un arbre de GW. Pour $i \in \mathbb{T}$, soit $\Delta_i = (X_i, X_{2i}, X_{2i+1})$. Pour un sous ensemble fini $J \subset \mathbb{T}$, posons

$$M_J(f) = \begin{cases} \sum_{i \in J} f(X_i) & \text{pour } f \in \mathcal{B}(\bar{S}), \\ \sum_{i \in J} f(\Delta_i) & \text{pour } f \in \mathcal{B}(\bar{S}^3), \end{cases} \tag{1.4.4}$$

avec la convention qu'une somme sur un ensemble vide est nulle. Définissons aussi les moyennes de f sur J suivantes :

$$\overline{M}_J(f) = \frac{1}{|J|} M_J(f) \text{ si } |J| > 0 \text{ et } \widetilde{M}_J(f) = \frac{1}{\mathbb{E}\left[|J|\right]} M_J(f) \text{ si } \mathbb{E}\left[|J|\right] > 0. \quad (1.4.5)$$

Revenons un instant sur le modèle BAR(1) avec données manquantes défini au début de cette Section. Soit

$$\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1, \alpha'_0, \beta'_0, \alpha'_1, \beta'_1). \quad (1.4.6)$$

Avec les notations ci-dessus, supposons qu'on observe les $r+2$ premières générations. Soient $\mathbb{T}_r^{0,1}$ la sous population de $\mathbb{T}_r^*$ constituée des cellules qui ont deux filles vivantes, $\mathbb{T}_r^0$ (resp.$\mathbb{T}_r^1$) la sous population de $\mathbb{T}_r^*$ constituée des cellules dont seule la fille de type nouveau (resp. ancien) pôle est vivante :

$$\mathbb{T}_r^{1,0} = \left\{i \in \mathbb{T}_r^* : \Delta_i \in S^3\right\}, \quad \mathbb{T}_r^0 = \{i \in \mathbb{T}_r^* : \Delta_i \in S^2 \times \{\partial\}\} \quad \text{et}$$
$$\mathbb{T}_r^1 = \{i \in \mathbb{T}_r^* : \Delta_i \in S \times \{\partial\} \times S\}.$$

Alors l'estimateur des moindres carrés, (qui dans ce cas coincide avec l'estimateur du maximum de vraisemblance),

$$\widehat{\theta}_r = (\widehat{\alpha}_0^r, \widehat{\beta}_0^r, \widehat{\alpha}_1^r, \widehat{\beta}_1^r, \widehat{\alpha}_0'^r, \widehat{\beta}_0'^r, \widehat{\alpha}_1'^r, \widehat{\beta}_1'^r)$$

de θ (1.4.6), est donné par, pour $\eta \in \{0,1\}$,

$$\widehat{\alpha}_\eta^r = \frac{|\mathbb{T}_r^{1,0}|^{-1} \sum_{i \in \mathbb{T}_r^{1,0}} X_i X_{2i+\eta} - \left(|\mathbb{T}_r^{1,0}|^{-1} \sum_{i \in \mathbb{T}_r^{1,0}} X_i\right)\left(|\mathbb{T}_r^{1,0}|^{-1} \sum_{i \in \mathbb{T}_r^{1,0}} X_{2i+\eta}\right)}{|\mathbb{T}_r^{1,0}|^{-1} \sum_{i \in \mathbb{T}_r^{1,0}} X_i^2 - \left(|\mathbb{T}_r^{1,0}|^{-1} \sum_{i \in \mathbb{T}_r^{1,0}} X_i\right)^2},$$

$$\widehat{\beta}_\eta^r = |\mathbb{T}_r^{1,0}|^{-1} \sum_{i \in \mathbb{T}_r^{1,0}} X_{2i+\eta} - \widehat{\alpha}_\eta^r |\mathbb{T}_r^{1,0}|^{-1} \sum_{i \in \mathbb{T}_r^{1,0}} X_i,$$

$$\widehat{\alpha}_\eta'^r = \frac{|\mathbb{T}_r^\eta|^{-1} \sum_{i \in \mathbb{T}_r^\eta} X_i X_{2i+\eta} - \left(|\mathbb{T}_r^\eta|^{-1} \sum_{i \in \mathbb{T}_r^\eta} X_i\right)\left(|\mathbb{T}_r^\eta|^{-1} \sum_{i \in \mathbb{T}_r^\eta} X_{2i+\eta}\right)}{|\mathbb{T}_r^\eta|^{-1} \sum_{i \in \mathbb{T}_r^\eta} X_i^2 - \left(|\mathbb{T}_r^\eta|^{-1} \sum_{i \in \mathbb{T}_r^\eta} X_i\right)^2},$$

$$\widehat{\beta}_\eta'^r = |\mathbb{T}_r^\eta|^{-1} \sum_{i \in \mathbb{T}_r^\eta} X_{2i+\eta} - \widehat{\alpha}_\eta'^n |\mathbb{T}_r^\eta|^{-1} \sum_{i \in \mathbb{T}_r^\eta} X_i.$$

Il est facile de voir que $(|\mathbb{G}_k^*|, k \in \mathbb{N})$ est un processus de GW dont le nombre moyen d'enfant d'un individu est m. Il est bien connu, voir [5], que dans le cas sur-critique, (c'est-à-dire $m > 1$), il existe une variable aléatoire positive W telle que

$$m^{-q}|\mathbb{G}_q^*| \to W \quad \text{p.s. quand } q \to \infty. \tag{1.4.7}$$

Des lois de grands nombres pour les moyennes 1.4.5, avec $J = \mathbb{G}_r^*$ et $J = \mathbb{T}_r^*$, ont été établies par Delmas et Marsalle [33] . Plus particulièrement, le résultat simplifié suivant à été démontré.

Théorème 1.4.2. *Soit $(X_i, i \in \mathbb{T}^*)$ une P^*-CMB spatialement homogène et sur-critique sur un arbre de GW. Soit $f \in \mathcal{B}_b(S)$ (resp. $f \in \mathcal{B}(S^3)$). Supposons qu'ils existent une constante réelle strictement positive c, une mesure de probabilité μ sur $(S, \mathcal{S})$, et une suite de nombres réels positifs $(a_r, r \in \mathbb{N})$ telle que $\sum_{r\in\mathbb{N}} a_r^2 < \infty$, et pour tout $x \in S$ et $r \in \mathbb{N}$, $|Q^r f(x) - \langle \mu, f\rangle| \le c a_r$ (resp. $|Q^r(P^*f)(x) - \langle \mu, P^*f\rangle| \le c a_r$). Alors les suites $(\widetilde{M}_{\mathbb{G}_q^*}(f), q \in \mathbb{N})$ et $(\widetilde{M}_{\mathbb{T}_r^*}(f), r \in \mathbb{N})$ convergent p.s. vers $\langle \mu, f\rangle W$ (resp. $\langle \mu, P^*f\rangle W$), où W est défini par (1.4.7) ; et les suites $(\overline{M}_{\mathbb{G}_q^*}\mathbf{1}_{\{|\mathbb{G}_q^*|>0\}}, q \in \mathbb{N})$ et $(\overline{M}_{\mathbb{T}_r^*}\mathbf{1}_{\{|\mathbb{G}_r^*|>0\}}, r \in \mathbb{N})$ convergent p.s. vers $\langle \mu, f\rangle \mathbf{1}_{\{W\neq 0\}}$ (resp. $\langle \mu, P^*f\rangle \mathbf{1}_{\{W\neq 0\}}$).*

Notons que toute fonction f définies sur S est prolongée sur $\overline{S}$ en posant $f(\partial) = 0$. Sous d'hypothèses convenables, nous précisons au Chapitre 4, l'ordre de convergence dans le Théorème 1.4.2 par l'obtention des inégalités de déviations pour les moyennes $\widetilde{M}_{\mathbb{G}_r^*}(f)$, $\widetilde{M}_{\mathbb{T}_r^*}(f)$, $\overline{M}_{\mathbb{G}_r^*}(f)$ et $\overline{M}_{\mathbb{T}_r^*}(f)$. Plus précisément, nous faisons les hypothèses suivantes.

(H3) : Il existe une mesure de probabilité μ sur $(S, \mathcal{S})$ telle que pour tout $f \in \mathcal{B}_b(S)$ vérifiant $\langle \mu, f\rangle = 0$, il existe $c > 0$ tel que pour tout $k \in \mathbb{N}$ et pour tout $x \in S$, $|Q^k f(x)| \le c\alpha^k$, avec $\alpha \in (0, 1)$.

(H4) : $m > \sqrt{2}$.

(H5) : $p_{1,0} + p_0 + p_1 = 1$.

désignons par $\mathbb{H}_r$ l'un des ensembles $\mathbb{G}_r$ ou $\mathbb{T}_r$. Posons $h_r = (m^2/2)^r$ si $\mathbb{H}_r = \mathbb{G}_r$ et $h_r = (m^2/2)^{r+1}$ si $\mathbb{H}_r = \mathbb{T}_r$. Posons également $t_r := \mathbb{E}[|\mathbb{T}_r^*|]$. Il est facile de voir que

$$t_r = \frac{m^{r+1} - 1}{m - 1}.$$

Alors, sous les hypothèses (**H3**)-(**H5**) on a ce qui suit.

Theorem 1.4.3. *Pour tout $f \in \mathcal{B}_b(S)$, pour tout $\delta > 0$, $a > 0$ et pour tout $b > 0$*

tels que $b < a/(\delta+1)$*, on a*

$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f\rangle W > \delta\Big) \leq$$

$$\begin{cases} \exp\left(c''\delta\right)\exp\left(-c'\delta^2 h_r\right) + A_r, \qquad \forall r \in \mathbb{N}, \quad si\ m\alpha < 1, \\ \\ \exp\left(c''\delta\right)\exp\left(-c'\delta^2 (m^2/2)^r\right) + A_r, \forall r \in \mathbb{N}, si\ \alpha m = 1\ \ et\ \mathbb{H}_r = \mathbb{G}_r, \\ \\ \exp\left(c''\delta(r+1)\right)\exp\left(-c'\delta^2 (m^2/2)^{r+1}\right) + A_r, \forall r \in \mathbb{N}, si\ \alpha m = 1\ \ et\ \mathbb{H}_r = \mathbb{T}_r \\ \\ \exp\left(-c'\delta^2 h_r\right) + A_r, \forall r \in \mathbb{N}\ tel\ que\ r > r_0,\ si\ 1 < m\alpha < \sqrt{2}, \\ \\ \exp\left(-\frac{c'\delta^2 h_r}{r}\right) + A_r, \forall r \in \mathbb{N}\ tel\ que\ r > r_0,\ si\ m\alpha = \sqrt{2}, \\ \\ \exp\left(-\frac{c'\delta^2}{\alpha^{2r}}\right) + A_r, \forall r \in \mathbb{N}^*\ tel\ que\ r > r_0,\ si\ m\alpha > \sqrt{2}, \end{cases}$$

où,

- *pour tout* $r \in \mathbb{N}$,

$$A_r = \begin{cases} c'\exp\left(-c''\delta^{2/3}(m^{1/3})^r\right) & si\ \mathbb{H}_r = \mathbb{G}_r \\ \exp\left(c'\delta^{2/3}\right)\exp\left(-c''\delta^{2/3}\left(t_r/(r+1)^2\right)^{1/3}\right) & si\ \mathbb{H}_r = \mathbb{T}_r, \end{cases}$$

- $r_0 := \log\left(\frac{\delta}{c_0}\right)/\log(\alpha) - k_0$*, avec* $k_0 \in \{0,1\}$*,*
- c_0*,* c' *and* c'' *sont des constantes positives qui dépendent de* α*,* m*, et* c*.*

et

$$\mathbb{P}\Big(\overline{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f\rangle > \delta \Big| W \geq a\Big)$$

$$\leq \begin{cases} \exp\left(c''\delta b\right)\exp\left(-c'(\delta b)^2 h_r\right) + B_r, \qquad \forall r \in \mathbb{N}, \quad si\ m\alpha < 1, \\ \\ \exp\left(c''\delta b\right)\exp\left(-c'(\delta b)^2 (m^2/2)^r\right) + B_r, \forall r \in \mathbb{N}, si\ \alpha m = 1\ \ and\ \mathbb{H}_r = \mathbb{G}_r, \\ \\ \exp\left(2c'\delta b(r+1)\right)\exp\left(-c'(\delta b)^2 (m^2/2)^{r+1}\right) + B_r, \forall r \in \mathbb{N}, si\ \alpha m = 1 \\ \qquad\qquad\qquad et\ \mathbb{H}_r = \mathbb{T}_r \\ \exp\left(-c'(\delta b)^2 h_r\right) + B_r, \forall r \in \mathbb{N}\ tel\ que\ r > r_0,\ si\ 1 < m\alpha < \sqrt{2}, \\ \\ \exp\left(-\frac{c'(\delta b)^2 h_r}{r}\right) + B_r, \forall r \in \mathbb{N}\ tel\ que\ r > r_0,\ si\ m\alpha = \sqrt{2}, \\ \\ \exp\left(-\frac{c'(\delta b)^2}{\alpha^{2r}}\right) + B_r, \forall r \in \mathbb{N}^*\ tel\ que\ r > r_0,\ si\ m\alpha > \sqrt{2}, \end{cases}$$

où,

– *pour tout $r \in \mathbb{N}$,*

$$B_r = \begin{cases} c' \exp\left(-c''(\delta b)^{2/3}(m^{1/3})^r\right) & si\ \mathbb{H}_r = \mathbb{G}_r \\ \exp\left(c'(\delta b)^{2/3}\right) \exp\left(-c''(\delta b)^{2/3}\left(t_r/(r+1)^2\right)^{1/3}\right) & si\ \mathbb{H}_r = \mathbb{T}_r, \end{cases}$$

– $r_0 := \log\left(\frac{\delta b}{c_0}\right) / \log(\alpha) - k_0$, *avec* $k_0 \in \{0, 1\}$,

– c_0, c' *and* c'' *sont des constantes positives qui dépendent de* α, m, a, *et* c.

La preuve du Théorème 1.4.3 utilise la même méthodologie que la démonstration des inégalités de déviations exponentielles du Chapitre 2. Mais toutefois, elle fait en plus appel aux inégalités de déviations pour les chaînes de Galton-Watson. Dans la Section 4.3, ces résultats sont utilisés pour obtenir des inégalités de déviations pour $\widehat{\theta}^r - \theta$, où θ est donné par (1.4.6), et $\widehat{\theta}^r$ est l'estimateur des moindres carrés de θ. Nous travaillons principalement sous une hypothèse entraînant que le processus BAR(1) avec données manquantes est à valeurs dans un ensemble compact.

Remarque 1.4.4. *Une fois de plus, La dichotomie autour des valeurs* $m\alpha = \sqrt{2}$ *et* $m\alpha = 1$ *apparaît naturellement dans les calculs. Elle est liée à la structure binaire du processus.*

1.4.1 Problèmes ouverts et perspectives

Comme pour le cas des CMBs, les inégalités de déviations sont obtenues ici à l'aide des hypothèses fortes. Nous envisageons dans des travaux futurs d'assouplir ces hypothèses. Notamment, nous essayerons d'obtenir des inégalités de déviations avec des fonctions Lipschitziennes.

Nous avons également supposé que $m > \sqrt{2}$. Le cas $1 < m \leq \sqrt{2}$ n'est pas traité ici.

Un autre problème d'intérêt est l'obtention d'un principe de déviations modérées pour $\overline{M}_{\mathbb{T}_r^*}(f)$. Comme nous l'avons déjà signaler, l'utilisation des méthodes classiques, de même que des méthodes développées dans cette thèse échouent. La cause principale étant que les quantités d'intérêt convergent vers des quantités aléatoires.

1.5 La statistique de Durbin-Watson

Dans un modèle d'autorégression classique on a

$$X_n = \theta X_{n-1} + \varepsilon_n,$$

où (X_n) désigne une suite de quantités aléatoires observables et (ε_n) une suite de variables aléatoires i.i.d non observables. Mais très souvent, il arrive que le bruit

(ε_n) soit aussi autocorrélé, c'est-à-dire qu'il vérifie aussi une équation d'autorégression (d'ordre 1 pour simplifier). La question qui se pose est alors : comment détecter que le bruit générateur d'un modèle d'autorégression est autocorrélé ? Une réponse à cette question peut être obtenue en effectuant des tests de Durbin-Watson. Les tests de Durbin-Watson reposent sur la statistique de Durbin-Watson, introduite au milieu du siècle dernier par Durbin et Watson [50], [51], [52], pour détecter la présence d'autocorrélation significative dans les résidus issus d'une régression linéaire classique.

Afin d'introduire cette statistique, intéressons nous à l'autocorrélation résiduelle d'un modèle autorégressif d'ordre 1 (AR(1)). Le modèle est donné, pour $n \geq 1$, par

$$\begin{cases} X_n &= \theta X_{n-1} + \varepsilon_n \\ \varepsilon_n &= \rho \varepsilon_{n-1} + V_n, \end{cases} \tag{1.5.1}$$

où X_0 et ε_0 sont arbitrairement choisis et de carré intégrable. Nous supposons que $|\theta| < 1$ et $|\rho| < 1$, ce qui assure la stabilité du modèle. La suite $\{V_n, n \in \mathbb{N}^*\}$ est une suite de variable aléatoire i.i.d, centrées et de variance strictement positive. En supposant que les quantités $\{X_k, 0 \leq k \leq n\}$ sont observées, l'estimateur des moindres carrés $\widehat{\theta}_n$ du paramètre θ est donné par

$$\widehat{\theta}_n = \frac{\sum_{k=1}^n X_k X_{k-1}}{\sum_{k=1}^n X_{k-1}^2}. \tag{1.5.2}$$

L'estimateur $\widehat{\rho}_n$ du paramètre d'autocorrélation ρ est obtenu en construisant le résidu empirique

$$\widehat{\varepsilon}_k = X_k - \widehat{\theta}_n X_{k-1} \quad \text{pour } 1 \leq k \leq n.$$

On a alors

$$\widehat{\rho}_n = \frac{\sum_{k=1}^n \widehat{\varepsilon}_k \widehat{\varepsilon}_{k-1}}{\sum_{k=1}^n \widehat{\varepsilon}_{k-1}^{\,2}}. \tag{1.5.3}$$

Enfin, la statistique de Durbin-Watson est définie, pour tout $n \geq 1$, par

$$\widehat{D}_n = \frac{\sum_{k=1}^n (\widehat{\varepsilon}_k - \widehat{\varepsilon}_{k-1})^2}{\sum_{k=0}^n \widehat{\varepsilon}_k^{\,2}}. \tag{1.5.4}$$

La valeur de $\widehat{D}_n$ varie entre 0 et 4. Une valeur très proche de 2 révèle une absence d'autocorrélation. Une valeur proche de 0 signifie la présence d'une autocorrélation avec un paramètre d'autocorrélation positif. Une valeur proche de 4 indique une autocorrélation avec un paramètre d'autocorrélation négatif. Une littérature abondante est disponible sur la puissance du test de Durbin-Watson par rapport aux autres procédures de détection d'autocorrélation, on peut citer entre autre [105], [98], [24]. Récemment, Bercu et Proïa [19] ont pu montrer par des simulations numériques que les tests de Durbin-Watson sont supérieurs aux tests utilisés régulièrement par les économètres (les tests de Ljung-Box et de Box-Pierce). Ils ont également montré, en utilisant la théorie des martingales, certaines propriétés de convergence forte et de normalité asymptotique pour les estimateurs (1.5.2), (1.5.3) et (1.5.4). Afin de rappeler leurs résultats, faisons les notations suivantes. On pose :

$$\theta^* = \frac{\theta + \rho}{1 + \theta\rho}, \tag{1.5.5}$$

$$\sigma_\theta^2 = \frac{(1-\theta^2)(1-\theta\rho)(1-\rho^2)}{(1+\theta\rho)^3}, \tag{1.5.6}$$

$$\rho^* = \theta\rho\theta^*, \tag{1.5.7}$$

$$\sigma_\rho^2 = \frac{(1-\theta\rho)}{(1+\theta\rho)^3}\big((\theta+\rho)^2(1+\theta\rho)^2 + (\theta\rho)^2(1-\theta^2)(1-\rho^2)\big), \tag{1.5.8}$$

$$\Gamma = \begin{pmatrix} \sigma_\theta^2 & \theta\rho\sigma_\theta^2 \\ \theta\rho\sigma_\theta^2 & \sigma_\rho^2 \end{pmatrix}. \tag{1.5.9}$$

On a alors le théorème suivant.

Théorème 1.5.1. *On a la convergence presque sûre des estimateurs et de la statistique de Durbin-Watson :*

$$\lim_{n\to\infty} \widehat{\theta}_n = \theta^*, \quad \lim_{n\to\infty} \widehat{\rho}_n = \rho^*, \quad \lim_{n\to\infty} \widehat{D}_n = D^* \quad \text{a.s.}$$

De plus, si $\mathbb{E}[V_1^4] < \infty$*, on a la normalité asymptotique*

$$\sqrt{n}\left(\widehat{\theta}_n - \theta^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_\theta^2), \quad \sqrt{n}\left(\widehat{\rho}_n - \rho^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_\rho^2),$$

$$\sqrt{n}\begin{pmatrix} \widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^* \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma), \quad \sqrt{n}\left(\widehat{D}_n - D^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_D^2),$$

avec $\sigma_D^2 = 4\sigma_\rho^2$.

Au Chapitre 5, nous donnons une version plus précise. Nous établissons un principe de déviations modérées pour les estimateurs définis en (1.5.2) et (1.5.3), ainsi que pour la statistique de Durbin-Watson sous divers hypothèses sur le bruit (V_n) et sur les conditions initiales X_0 et ε_0. Plus précisément, soit (b_n) une suite de nombres réels positifs vérifiant $1 = o(b_n^2)$ et $b_n^2 = o(n)$.

- Dans la Section 5.2, nous supposons que le bruit (V_n) est gaussien. Nous supposons en plus qu'il existe $t > 0$ tel que $\mathbb{E}\Big[\exp(t\varepsilon_0^2)\Big] < \infty$, et $\mathbb{E}\Big[\exp(tX_0^2)\Big] < \infty$.
- Dans la Section 5.3, nous supposons des hypothèses de type Chen-Ledoux. Plus précisément, pour $a = 2$ et $a = 4$, nous supposons que

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log n\mathbb{P}\Big(|V_1|^a > b_n\sqrt{n}\Big) = -\infty, \ \frac{|\varepsilon_0|^a}{b_n\sqrt{n}} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0 \text{ et } \frac{|X_0|^a}{b_n\sqrt{n}} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0.$$

Nous montrons alors les résultats suivants.

Théorème 1.5.2. *Sous les hypothèses ci-dessus, les suites*

$$\left(\frac{\sqrt{n}}{b_n}\left(\widehat{\theta}_n - \theta^*\right)\right)_{n\geq 1}, \quad \left(\frac{\sqrt{n}}{b_n}\left(\widehat{\rho}_n - \rho^*\right)\right)_{n\geq 1} \quad \textit{et} \quad \left(\frac{\sqrt{n}}{b_n}\left(\widehat{D}_n - D^*\right)\right)_{n\geq 1}$$

satisfont à un PDM dans $\mathbb{R}$ de vitesse b_n^2 et de bonne fonction de taux respective

$$I_\theta(x) = \frac{x^2}{2\sigma_\theta^2}, \quad I_\rho(x) = \frac{x^2}{2\sigma_\rho^2} \quad \textit{et} \quad I_D(x) = \frac{x^2}{2\sigma_D^2},$$

où σ_θ^2, σ_ρ^2 sont donnés par (1.5.6) et (1.5.8) et $\sigma_D^2 = 4\sigma_\rho^2$.

Comme aux Chapitres 2 et 3, nous utilisons encore une approche martingale pour démontrer ces résultats. Dans le cadre gaussien de la Section 5.2, la convergence super-exponentielle du crochet des différentes martingales est obtenue grâce au Théorème de Cramer et aux inégalités de probabilités exponentielles établies par Bercu et Touati [20]. Nous utilisons ensuite le PDM des séries régressives établi par Worms [112] pour obtenir notre PDM. Dans le cadre où le bruit vérifie des hypothèses de type "Chen-Ledoux", nous utilisons la même approche que celle du Chapitre 3, mais cette fois, la convergence super-exponentielle du crochet des différentes martingales est obtenue à l'aide des inégalités de Bercu et Touati [20]. Les résultats que nous obtenons dans ce chapitre permettent au passage d'étendre ceux obtenus par worms [112], au cas où le bruit du modèle autorégressif d'ordre 1 en dimension 1 vérifie une condition de type "chen-Ledoux". Signalons que dans [112], le cas traité est celui où le bruit est sous-gaussien.

1.5.1 Problèmes ouverts et perspectives

Dans le cadre gaussien, nous pensons qu'il est certainement possible de montrer un PGD pour la statistique de Durbin-Watson. En s'inspirant des travaux de Worms [113] et ceux de Bercu [17], nous nous y attellerons dans nos travaux futurs.

1.6 Borne explicite pour l'ergodicité des processus de Markov

Trouver des bornes explicites pour l'ergodicité des processus de Markov à temps discret et continu est devenu, au cours de ces deux dernières décennies un enjeux majeur en théorie des probabilités. Ceci est dû, pour les cas discret à l'utilisation des chaînes de Markov dans les méthodes numériques (algorithmes MCCM...) [58], [99]. Dans le cas continu, ceci est en partie dû à l'utilisation des diffusions de Langevin pour les simulations de Monte Carlo [92]. L'obtention des inégalités de déviations pour les chaînes de Markov bifurcantes nous a conduit à nous intéresser aux bornes explicites pour l'ergodicité des processus de Markov. En effet, comme nous l'avons déjà signaler, ces inégalités de déviations sont d'autant plus précises que les bornes pour l'ergodicité de la chaîne de Markov induite sont explicites. Plusieurs auteurs ont travaillé sur l'ergodicité, explicite ou non des processus de Markov, on peut citer entre autres [47], [95], [44], [48], [55], [43], [92]. Nous allons nous attarder ici sur les travaux de Roberts-Rosenthal [92] et de Douc-Fort-Guillin [43], qui constituent la base de notre travail. Tout d'abord, commençons par donner quelques définitions et notations.

Soit $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, (X_t)_{t\geq 0}, (\mathbb{P}_x)_{x\in X})$ une famille de Markov définie sur un espace métrique séparable et localement compact X, muni de sa σ-algèbre borélienne $\mathcal{B}(X)$: $(\Omega, \mathcal{F})$ est un espace mesurable, $(X_t)_{t\geq 0}$ est un processus de Markov sur X par rapport à la filtration $(\mathcal{F}_t)_{t\geq 0}$ et $\mathbb{P}_x$ (resp. $\mathbb{E}_x$) désigne la probabilité canonique (resp. l'espérance) associée au processus de Markov de loi initial δ_x, la masse de Dirac en x. Nous supposons que $(X_t)_{t\geq 0}$ est un processus de Markov fort homogène et à trajectoires càdlàg. On désigne par $(P^t)_{t\geq 0}$ la fonction de transition associée sur $(X, \mathcal{B}(X))$. Pour un $t \geq 0$ P^t agit sur les fonctions mesurables bornées f et les mesures σ-finies μ sur X via les formules

$$P^t f(x) = \int_X P^t(x, dy) f(y), \quad \mu P^t(A) = \int_X \mu(dx) P^t(x, A).$$

Une mesure de probabilité π sur $\mathcal{B}(X)$ est dite invariante si

$$\pi = \pi P^t \quad \forall t \geq 0.$$

Le processus de Markov $(X_t)_{t\geq 0}$ est ϕ-irréductible pour une mesure σ-finie ϕ si,

$$\phi(A) > 0 \Rightarrow \mathbb{E}_x \left[\int_0^\infty \mathbf{1}_A(X_s) ds \right] > 0 \forall x \in X.$$

Un processus ϕ-irréductible possède une mesure d'irréductibilité maximale ψ, au sens où toute autre mesure irréductible φ est absolument continue par rapport à ψ.

Le processus $(X_t)_{t\geq 0}$ est Harris-récurrent si, pour une mesure σ-finie μ,

$$\mu(A) > 0 \Rightarrow \int_0^\infty \mathbf{1}_A(X_s) ds = \infty \quad \mathbb{P}_x\text{-presque partout} \quad \forall x \in X.$$

Un processus Harris-récurrent est trivialement ϕ-irréductible ; de plus [59], s'il est càdlàg, il admet une mesure invariante π ; si π est une probabilité invariante, alors le processus est dit Harris-récurrent positif. Différents critères pour un processus d'être Harris récurrent positif ont été donnés [82], [83].

Le processus $(X_t)_{t\geq 0}$ est ergodique s'il existe une mesure de probabilité invariante π et

$$\lim_{t\to\infty} \|P^t(x,\cdot) - \pi(\cdot)\|_{TV} = 0, \quad \forall x \in X,$$

où la norme en variation totale d'une mesure signée μ est donnée par

$$\|\mu\|_{TV} = \sup_{A\in\mathcal{B}(X)} \mu(A) - \inf_{A\in\mathcal{B}(X)} \mu(A).$$

Notons qu'un processus Harris-récurrent positif n'est pas forcément ergodique. En effet, des conditions supplémentaires sur le processus sont nécessaires pour obtenir l'ergodicité. De plus, pour les simulations numériques, connaître à quelle vitesse P^t converge vers π est d'une importance capitale. Plus précisément, en supposant que le processus est ergodique, on souhaiterait déterminer deux fonctions $g : X \to [0,\infty)$ et $r : [0,+\infty) \to (0,+\infty)$, calculables explicitement, telles que

$$\|P^t(x,\cdot) - \pi(\cdot)\|_{TV} \leq \frac{g(x)}{r(t)} \quad \forall x \in X. \tag{1.6.1}$$

Lorsque $r(t) = \rho^t$ avec $\rho > 1$, on parle d'ergodicité exponentielle ou géométrique. Si $r(t)$ est un polynôme, on parle d'ergodicité polynomiale. Cette dernière fait partie d'une classe plus grande qu'on appelle ergodicité sous-géométrique. Nous allons à présent énoncer les résultats obtenus par Roberts et Rosenthal [92] sur les bornes explicites pour l'ergodicité exponentielle des processus de Markov à temps continu. Commençons par les définitions suivantes.

Définition 1.6.1 (Ensemble small, Ensemble petite).
On dit qu'un sous-ensemble non vide $C \subseteq X$ est (t^*,ε)-small (ou simplement small), pour un temps strictement positif t^* et $\varepsilon > 0$, s'il existe une mesure de probabilité $Q(\cdot)$ sur X vérifiant la condition de minoration

$$P^{t^*}(x,\cdot) \geq \varepsilon Q(\cdot) \;\; \forall x \in C. \tag{1.6.2}$$

Le sous-ensemble C est ν_a-petite (ou simplement petite) s'ils existent une mesure de probabilité a sur la σ-algèbre borélienne de $[0,+\infty)$ et une mesure σ-finie non triviale ν_a sur $\mathcal{B}(X)$ telles que

$$\int_0^\infty P^t(x,\cdot)a(dt) \geq \nu_a(\cdot) \;\; \forall x \in C.$$

On vérifie aisément que tout ensemble small est petite.

Un outil essentiel pour l'obtention des bornes explicites est le couplage, c'est-à-dire la construction conjointe de deux processus de Markov dont les marginales devront se comporter comme la loi du processus initial.

1.6.1 Construction du couplage

Supposons que $C \subseteq X$ est (t^*, ϵ)-small. Soit (X_t) et (X'_t) deux processus que nous allons construire par couplage. Étant donné une suite de variables aléatoires i.i.d $Z_1, Z_2, ..., Z_i \sim B(1, \epsilon)$, on construit (X_t, X'_t) et le temps d'arrêt aléatoire

$$\widetilde{T} = \inf \left\{ \widetilde{\tau}_i^{t^*} + t^*, \left(X_{\widetilde{\tau}_i^{t^*}}, X'_{\widetilde{\tau}_i^{t^*}} \right) \in C \times C, Z_i = 1 \right\}.$$

Soit

$$\widetilde{\tau}_{C\times C}^{t^*} = \inf \left\{ t \geq t^*, (X_t, X'_t) \in C \times C \right\}, \quad \widetilde{\tau}_1^{t^*} = \widetilde{\tau}_{C\times C}^{0},$$

et

$$\widetilde{\tau}_i^{t^*} = \inf \left\{ t \geq \widetilde{\tau}_{i-1}^{t^*} + t^*, (X_t, X'_t) \in C \times C \right\}, \quad i \geq 2.$$

Pour chaque temps $\widetilde{\tau}_i^{t^*}$, si X_t et X'_t n'ont pas encore été couplés, alors on procède comme suit :

1. si $Z_i = 1$, on pose
$$X_{\widetilde{\tau}_i^{t^*}+t^*} = X'_{\widetilde{\tau}_i^{t^*}+t^*} \sim Q(\cdot),$$
et on déclare que les processus ont été couplés, et par la définition du temps de couplage, on a $X_{\widetilde{\tau}_i^{t^*}+t^*} = X'_{\widetilde{\tau}_i^{t^*}+t^*}$,
2. si $Z_i = 0$, on pose
$$X_{\widetilde{\tau}_i^{t^*}+t^*} \sim \frac{1}{1-\epsilon} \left(P^{t^*}(X_{\widetilde{\tau}_i^{t^*}}, \cdot) - \epsilon Q(\cdot) \right)$$
et
$$X'_{\widetilde{\tau}_i^{t^*}+t^*} \sim \frac{1}{1-\epsilon} \left(P^{t^*}(X'_{\widetilde{\tau}_i^{t^*}}, \cdot) - \epsilon Q(\cdot) \right)$$
conditionnellement indépendant.

Dans les deux cas, on remplit les valeurs X_t et X'_t pour $\widetilde{\tau}_i^{t^*} < t < \widetilde{\tau}_i^{t^*} + t^*$ conditionnellement indépendantes, en utilisant les lois conditionnelles adaptées des données $X_{\widetilde{\tau}_i^{t^*}}$, $X'_{\widetilde{\tau}_i^{t^*}}$, $X_{\widetilde{\tau}_i^{t^*}+t^*}$, $X'_{\widetilde{\tau}_i^{t^*}+t*}$. $\widetilde{T}$ est le temps de couplage. Il est facile de voir que (X_t) et (X'_t) suivent marginalement les probabilités de transitions $P^t(\cdot, \cdot)$ (voir Roberts-Rosenthal [92, 94]).

On a alors le résultat suivant qui donne une borne exponentielle explicite de la distance en variation totale entre $\mathcal{L}(X_t)$ et la probabilité invariante $\pi(\cdot)$.

Théorème 1.6.2. *Supposons que le processus de Markov $(X_t)_{t\geq 0}$ admet une probabilité invariante $\pi(\cdot)$. Supposons que $C \subset X$ est (t^*, ε)-small, pour un temps t^* strictement positif et $\varepsilon > 0$. Supposons en plus qu'il existe $\delta > 0$ et une fonction $h : X \times X \to [1, \infty)$ tels que*

$$\mathbb{E}_{x,y} \left[\exp \left(\delta \tau_{C\times C} \right) \right] \leq h(x, y), \quad (x, y) \notin C \times C, \tag{1.6.3}$$

où $\tau_{C\times C} = \inf\{t \geq 0, (X_t, X'_t) \in C \times C\}$, (X_t) *et* (X'_t) *sont définis conjointement comme à la construction du couplage 1.6.1. Posons*

$$A = \sup_{(x,y)\in C\times C} \mathbb{E}_{x,y}\left[h(X_{t^*}, X'_{t^*})\right]$$

et supposons que $A < \infty$. *Alors pour* $t > 0$ *et pour tout* $0 < r < 1/t^*$,

$$\|\mathcal{L}(X_t) - \pi(\cdot)\|_{TV} \leq (1-\varepsilon)^{\lfloor rt \rfloor} + \exp\left(-\delta(t-t^*)\right) A^{\lfloor rt \rfloor - 1}\mathbb{E}\left[h\left(X_0, X'_0\right)\right].$$

Une des principales difficultés dans l'application de ce théorème est la vérification de la condition (1.6.3). Nous reviendrons plus tard sur une condition suffisante permettant de l'obtenir. La preuve de ce théorème repose en grande partie sur le contrôle du temps de couplage $\widetilde{T}$ définie dans la construction précédente.

Dans le cas où la convergence de $P^t(x,\cdot)$ vers $\pi(\cdot)$ est sous-exponentielle, Douc-Fort-Guillin [43] ont donné des bornes non explicites pour $\|P^t(x,\cdot) - \pi(\cdot)\|_{TV}$. Considérons la condition de dérive suivante pour un ensemble petite C.

$\mathbf{D}(\mathbf{C}, \mathbf{V}, \phi, b)$: Il existe un ensemble petite C, une fonction càdlàg $V : X \longrightarrow [1,\infty)$, une fonction $\phi : [1,\infty) \longrightarrow (0,\infty)$ croissante, positive, concave, dérivable et une constante $b < \infty$ tels que pour tous $s \geq 0$, $x \in X$,

$$\mathbb{E}_x\Big[V(X_s)\Big] + \mathbb{E}_x\left[\int_0^s \phi \circ V(X_u)du\right] \leq V(x) + b\mathbb{E}_x\left[\int_0^s 1_C(X_u)du\right]. \quad (1.6.4)$$

Notons que (1.6.4) est équivalent à la condition que la fonctionnelle

$$s \longmapsto V(X_s) - V(X_0) + \int_0^s \phi \circ V(X_u)du - b\int_0^s 1_C(X_u)du$$

est une $\mathbb{P}_x$-surmartingale par rapport à la filtration $(\mathcal{F}_t)_{t\geq 0}$ pour tout $x \in X$.

La stabilité et la convergence sous-exponentielle ont été étudiées dans [43] sous l'hypothèse principale $\mathbf{D}(\mathbf{C}, \mathbf{V}, \phi, b)$. Plus précisément, Douc & al ont montré ce qui suit.

Théorème 1.6.3 ([43], Proposition 3.9 et Théorème 3.10).
Supposons $\mathbf{D}(\mathbf{C}, \mathbf{V}, \phi, b)$ *et* $\sup_C V < \infty$. *Alors le processus est Harris récurrent positif avec la mesure de probabilité invariante* π *telle que* $\pi(\phi \circ V) < \infty$. *Si de plus une chaîne squelette est irréductible et* $\lim_{+\infty} \phi' = 0$, *alors, il existe une constante c finie telle que pour tout* $t > 0$ *et tout* $x \in X$,

$$\phi \circ H_\phi^{-1}(t)\|P^t(x,\cdot) - \pi(\cdot)\|_{TV} \leq cV(x) \quad (1.6.5)$$

où la fonction H_ϕ *est définie par*

$$H_\phi(u) = \int_1^u \frac{ds}{\phi(s)} \quad \forall u \geq 1.$$

Remarque 1.6.4. *Rappelons que pour $m > 0$, la m-squelette du processus de Markov $(X_t, t \geq 0)$ est la chaîne de Markov à temps discret $(X_{km}, k \in \mathbb{N})$, de probabilité de transition P^m. En gros, la m-squelette de $(X_t, t \geq 0)$ corresponds au processus $(X_t, t \geq 0)$ échantillonné aux instants m, $2m$, $3m$, $\cdots$.*

Comme pour (1.6.3), la condition de dérive $\mathbf{D}(\mathbf{C}, \mathbf{V}, \phi, b)$ n'est pas facile à obtenir. Il est alors important de fournir des conditions suffisantes et aisément maniables qui permettent de vérifier ces deux conditions. On a besoin pour cela de quelques notations. Soit $\mathcal{D}(\mathcal{A})$ l'ensemble des fonctions mesurables $f : X \to \mathbb{R}$ ayant la propriété suivante : il existe une fonction mesurable $h : X \to \mathbb{R}$ telle que la fonction $t \mapsto h(X_t)$ est intégrable $\mathbb{P}_x$-presque partout pour tout $x \in X$ et le processus

$$t \mapsto f(X_t) - f(X_0) - \int_0^t h(X_s) ds$$

est une $\mathbb{P}_x$-martingale locale pour tout $x \in X$. On écrit alors $h = \mathcal{A}f$, et on dit que f est dans le domaine du générateur étendue $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$.

Définition 1.6.5 (Fonctions de Lyapunov).

(i) On dit q'une fonction $V : X \to [1, +\infty)$ appartenant à $\mathcal{D}(\mathcal{A})$ est une fonction de Lyapunov s'ils existent un ensemble petite C, des constantes $\lambda > 0$ et $b < \infty$ telles que

$$\mathcal{A}V(x) \leq -\lambda V(x) + b 1_C(x), \quad \forall x \in X. \tag{1.6.6}$$

(ii) Soit $\Phi : [1, \infty) \to (0, \infty)$ une fonction croissante, concave, dérivable et telle que $\lim_{t \to +\infty} \phi(t) = +\infty$, $\lim_{t \to +\infty} \phi'(t) = 0$, $\log \phi(t)/t \downarrow 0$ quand $t \to +\infty$. On dit qu'une fonction $V : X \to [1, +\infty)$ appartenant à $\mathcal{D}(\mathcal{A})$ est une Φ-fonction de Lyapunov s'ils existent un ensemble petite C, une constante $b < \infty$ tels que

$$\mathcal{A}V(x) \leq -\Phi \circ V(x) + b 1_C(x), \quad \forall x \in X. \tag{1.6.7}$$

Les conditions de Lyapunov (1.6.6) et (1.6.7) sont les plus utilisées dans la littérature pour vérifier les conditions (1.6.3) et $\mathbf{D}(\mathbf{C}, \mathbf{V}, \phi, b)$. On a plus précisément.

Théorème 1.6.6 ([92] Corollaire 4 et [43] Théorème 3.11).

(i) Soit C un ensemble (t^*, ε)-small. Supposons la condition (1.6.6) satisfaite avec V, λ et b. Alors, en posant $B = \inf_{x \notin C} V(x)$, on a que la condition (1.6.3) est vérifiée avec $h(x, y) = 1/2(V(x) + V(y))$ et $\delta = \lambda - b/B$.

(ii) Soit C un ensemble petite fermé. Supposons la condition (1.6.7) satisfaite avec Φ, V, λ et b. Alors la condition $\mathbf{D}(\mathbf{C}, \mathbf{V}, \Phi, b)$ est vérifiée.

Nous serons capable au cours de cette thèse, de construire des fonctions de Lyapunov pour des processus couplés, ce qui sera la clé pour obtenir nos résultats.

Notons qu'il existe un lien, (que nous n'abordons pas ici), entre les conditions de Lyapounov et les inégalités fonctionnelles (Poincaré, Poincaré faible). Nous renvoyons à [8] pour plus de précisions sur ce sujet.

Notons que la constante c qui apparaît dans l'equation (1.6.5) n'est pas déterminée. En suivant les idées de Roberts-Rosenthal [92] et de Douc-fort-Guillin [43], nous nous proposons, au Chapitre 6 de cette thèse, de fournir un contrôle explicite de $\|\mathcal{L}(X_t)-\pi(\cdot)\|_{TV}$ sous des hypothèses entraînant l'ergodicité sous exponentielles du processus de Markov $(X_t)_{t\geq 0}$ vers sa mesure de probabilité d'équilibre $\pi(\cdot)$. Plus précisément, soient $\Phi : [1,\infty) \to (0,\infty)$ une fonction croissante, concave, dérivable et telle que $\lim_{t\to+\infty}\phi(t) = +\infty$, $\lim_{t\to+\infty}\phi'(t) = 0$, $\log\phi(t)/t \downarrow 0$ quand $t \to +\infty$ et $H_\Phi(\cdot)$ la fonction définie par $H_\Phi(u) = \int_1^u \frac{ds}{\Phi(s)}$ $\forall u \geq 1$. Soit $C \in X$ un ensemble (t^*, ε)-small pour un temps strictement positif t^* et $\varepsilon > 0$. Alors nous avons

Théorème 1.6.7. *Étant donné un processus de Markov $(X_t)_{t\geq 0}$ de probabilités de transition $(P^t(x,\cdot))$, supposons la condition de Lyapunov 1.6.7 vérifiée avec Φ, C et V tel que $\sup_C V < \infty$. Soit π la probabilité invariante. Si $\pi(V) < \infty$, alors*

$$||\mathcal{L}(X_s) - \pi||_{TV} \leq \sum_{0\leq n, s\geq nt^*} (1-\epsilon)^n \widetilde{A}_1 \frac{H_\Phi^{-1}(\lambda \widetilde{A}_0 n)}{H_\Phi^{-1}(\lambda(s-t^*))} \leq \frac{\widetilde{A}_1 \widetilde{A}_{\epsilon,\Phi}}{H_\Phi^{-1}(\lambda(s-t^*))}$$

où

$$\widetilde{A}_0 =: \frac{2}{\lambda\Phi(1)} \sup_{x\in C}\left\{ V(x) - 1 + \frac{b}{\phi(1)}\int_0^{t^*} \phi\circ H_\phi^{-1}(\lambda s) ds)\right\} < \infty,$$

$$\widetilde{A}_1 = \max\left\{ \frac{\mathbb{E}\Big[V(X_0)\Big] + \pi(V) - 2}{H_\Phi^{-1}(\lambda \widetilde{A}_0)}, 1\right\},$$

$$\widetilde{A}_{\epsilon,\Phi} = \frac{\sup_{n\geq 1} \frac{H_\Phi^{-1}(\lambda\widetilde{A}_0 n)}{(\frac{1-a_\epsilon}{1-\epsilon})^n}}{a_\epsilon} \quad \forall 0 < a_\varepsilon < \varepsilon$$

et

$$d_0 = \inf_{x\notin C} V(x), \quad 0 < \lambda \leq 1 - \frac{b}{\Phi(d_0)}.$$

La preuve du Théorème 1.6.7 repose en grande partie sur le contrôle de la queue du temps de couplage défini au paragraphe 1.6.1.

La suite de cette thèse est constituée de cinq chapitres, faisant chacun l'objet d'un article soumis (ou en passe de l'être) dans une revue internationale. Ceci justifie notamment le fait que les différents chapitres sont rédigés en langue Anglaise. Au cours du déroulement de notre thèse certains de ces chapitres ont été motivés par d'autres. Ainsi, le Chapitre 2 a motivé l'étude du Chapitre 3 et du Chapitre 6. Le Chapitre 3 a motivé l'étude du Chapitre 5. Néanmoins, la lecture des différents chapitres peut être fait de façon totalement indépendante.

Avant de clore ce chapitre introductif, nous allons dans la Section qui suit présenter quelques résultats fondamentaux sur le principe de déviations modérées pour des Martingales.

1.7 Quelques résultats fondamentaux sur le Principe de déviations modérées pour des martingales

Nous allons à présent énoncer des résultats que nous utiliserons régulièrement dans cette thèse. Le premier nous donne une condition nécessaire et suffisante pour qu'une somme normalisée de variables aléatoires indépendantes satisfasse un PDM. Dans toute la suite de cette Section, (b_n) est une suite croissante de nombres réels telle que

$$\frac{b_n}{\sqrt{n}} \uparrow \infty \quad \text{et} \quad \frac{b_n}{n} \downarrow 0. \tag{1.7.1}$$

Théorème 1.7.1 (Eichelsbacher et Löwe [53]). *Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires i.i.d. Les propositions suivantes sont équivalentes :*

- *Les variables aléatoires (X_i) satisfont $\mathbb{E}[X_1] = 0$ et*

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log\left(n\mathbb{P}\left(|X_1| > b_n\right)\right) = -\infty. \tag{1.7.2}$$

- *La suite $(\frac{1}{b_n}\sum_{i=1}^n X_i)$ satisfait un PDM de vitesse (b_n^2/n) et de bonne fonction de taux*

$$I(x) = \frac{x^2}{2\mathbb{E}[X_1^2]}.$$

Remarque 1.7.2.

- *La condition (1.7.2) est appelée "condition de type Chen-Ledoux".*
- *La condition de type Chen-Ledoux (1.7.2) entraîne que $\mathbb{E}[X_1^2] < \infty$. Le Théorème 1.7.1 a donc bien un sens.*

Nous allons à présent introduire deux théorèmes fondamentaux utiles à l'obtention du PDM pour les martingales. Le premier est une version simplifiée du théorème de Puhalskii [91], et le second est une version simplifiée du théorème de Djellout [38]. Nous ne donnons que les versions simplifiées de ces résultats car c'est sous ces formes que nous les utiliserons dans cette thèse.

Théorème 1.7.3 (Puhalskii [91]). *Soit $(m_j^n)_{1\leq j\leq n}$ un tableau triangulaire de différences martingales à valeurs dans $\mathbb{R}^d$, par rapport à la filtration $(\mathcal{F}_n)_{n\geq 1}$. Supposons que :*

(P1) *Il existe une matrice symétrique et semi-définie positive Q, telle que*

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\Big[m_k^n (m_k^n)' \big| \mathcal{F}_{k-1}\Big] \overset{\text{superexp}}{\underset{b_n^2/n}{\Longrightarrow}} Q,$$

(P2) *Il existe une constante $c > 0$ telle que, pour chaque $1 \leq k \leq n$, $|m_k^n| \leq c\frac{n}{b_n}$* p.s.,

(P3) *Pour tout $a > 0$, on a la condition de Lindeberg exponentielle*

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\Big[|m_k^n|^2 \mathrm{I}_{\{|m_k^n|\geq a\frac{n}{b_n}\}}\Big|\mathcal{F}_{k-1}\Big] \overset{\text{superexp}}{\underset{b_n^2/n}{\Longrightarrow}} 0.$$

Alors, $(\sum_{k=1}^{n} m_k^n/b_n)_{n\geq 1}$ satisfait un PDM sur $\mathbb{R}^d$ de vitesse b_n^2/n et de fonction de taux

$$\Lambda^*(v) = \sup_{\lambda\in\mathbb{R}^d}\left(\lambda' v - \frac{1}{2}\lambda' Q\lambda\right).$$

En particulier, si Q est inversible, $\Lambda^*(v) = \frac{1}{2}v'Q^{-1}v$.

Théorème 1.7.4 (Djellout [38]). *Soit $(M_n)_{n\in\mathbb{N}}$ une martingale de carré intégrable, à valeurs dans $\mathbb{R}$, adaptée à une filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ et nulle en 0. Soit $c(n) := \frac{\sqrt{n}}{b_n}$. On définit la fonction réciproque $c^{-1}(t)$ par*

$$c^{-1}(t) := \inf\{n \in \mathbb{N} : c(n) \geq t\}.$$

Sous les hypothèses suivantes

(D1) *il existe $Q \in \mathbb{R}_+^*$ tel que $\dfrac{\langle M\rangle_n}{n} \overset{\text{superexp}}{\underset{b_n^2/n}{\Longrightarrow}} Q$, où $\langle M\rangle_n$ désigne le crochet de la martingale M_n ;*

(D2) $\displaystyle\limsup_{n\to+\infty}\frac{n}{b_n^2}\log\left(n \operatorname*{ess\,sup}_{1\leq k\leq c^{-1}(b_{n+1})} \mathbb{P}(|M_k - M_{k-1}| > b_n|\mathcal{F}_{k-1})\right) = -\infty;$

(D3) *pour tout $a > 0$,* $\displaystyle\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left(|M_k - M_{k-1}|^2\mathbf{1}_{\{|M_k-M_{k-1}|\geq a\frac{n}{b_n}\}}\Big|\mathcal{F}_{k-1}\right) \overset{\text{superexp}}{\underset{b_n^2/n}{\Longrightarrow}} 0;$

$(M_n/b_n)_{n\in\mathbb{N}}$ satisfait un PDM dans $\mathbb{R}$ de vitesse b_n^2/n et de fonction de taux $I(x) = x^2/2Q$.

Remarque 1.7.5. *Les conditions **D3** et **P3** sont appelées condition de Linderberg exponentielle. La vérification de l'hypothèse **D1** (convergence super-exponentielle du crochet) est fondamentale pour établir les PDMs des martingales. Pour les martingales à différences bornées par exemple, les hypothèses **D2** et **D3** sont automatiquement vérifiées, de telle sorte que pour montrer un PDM, il suffit de montrer la convergence super-exponentielle du crochet, c'est-à-dire **D1**. Un outil pour démontrer cette convergence super-exponentielle du crochet est par exemple d'établir des inégalités de déviation, c'est ainsi que nous procédons par exemple au Chapitre 2 et au Chapitre 3.*

La stratégie pour prouver nos différents résultats sur le PDM sera la suivante.

- Si la martingale est à différences bornées, alors il suffit juste de montrer la convergence super-exponentielle du crochet et d'appliquer par exemple le théorème de Djellout.

– Si la martingale n'est pas à différence bornée, nous faisons une troncature. Pour la partie bornée, nous montrons qu'un PDM est satisfait en utilisant le théorème de Puhalskii. Ensuite nous montrons que la partie non bornée est négligeable au sens des déviations modérées

Rappelons à présent une suite de résultats fondamentaux de la théorie des grandes déviations et que nous utilisons intensivement aux chapitres 2, 3 et 5.

En théorie de la convergence faible, le théorème de continuité ("continuous mapping theorem" en anglais) permet de transporter une propriété de convergence d'un espace vers un autre via une fonction continue. L'analogue de ce théorème dans la théorie des grandes déviations est le principe de contraction, [35] Théorème 4.2.1. Plus précisément on a

Théorème 1.7.6. *Soient S et S' deux espaces polonais. Soit $f : S \to S'$ une fonction continue. Si une suite de variables aléatoires (Z_n) satisfait à un PGD de bonne fonction de taux I et si nous définissons*

$$I'(y) = \inf \left\{ I(x) : x \in S, y = f(x) \right\},$$

alors $(f(Z_n))$ satisfait un PGD sur S' de bonne fonction de taux I'.

Originellement établi par Varadhan [107], le Théorème 1.7.6 a connu plusieurs extensions [35], [89] [90], [54], [3], [36]. Nous renvoyons à [57] pour une discussion autour de toutes ces extensions.

Le résultat suivant, dû à Worms [113] se démontre en partie à l'aide du principe de contraction ci-dessus. Il nous sera très utile pour démontrer le PDM pour les modèles bifurcants autorégressifs (Chapitre 2 et 3), ainsi que pour la statistique de Durbin-Watson (Chapitre 5).

Théorème 1.7.7 ([113] Lemme 4.1). *Soient (M_n) une suite de matrice aléatoire de taille $p \times d$, C une matrice de covariance $p \times p$ déterministe inversible, (v_n) et (a_n) des vitesses telles que $a_n = o(v_n)$, et (C_n) une suite de matrices aléatoires symétriques et inversibles $p \times p$ telle que*

$$\frac{C_n}{v_n} \underset{a_n}{\overset{\text{superexp}}{\Longrightarrow}} C.$$

– *Alors on a aussi*

$$v_n C_n^{-1} \underset{a_n}{\overset{\text{superexp}}{\Longrightarrow}} C^{-1}.$$

Ainsi, si $M_n/v_n \underset{a_n}{\overset{\text{superexp}}{\Longrightarrow}} 0$, *alors* $C_n^{-1} M_n \underset{a_n}{\overset{\text{superexp}}{\Longrightarrow}} 0$.

– *Soit $I(\cdot)$ une fonction de taux sur $\mathbb{R}^{pd}$. Si la suite $\left(vecM_n/\sqrt{a_n v_n}\right)_{n\geq 1}$ satisfait à un PGD de vitesse (a_n) et de fonction de taux $I(\cdot)$, alors la suite $\left(\sqrt{v_n/a_n} vec(C_n^{-1} M_n)\right)_{n\geq 1}$ satisfait à un PGD de même vitesse et de fonction de taux*

$$I'(x) = I(vec(Cmatx)).$$

Remarque 1.7.8. *Rappelons les notations suivantes. On note $\mathcal{M}_{p,d}$ l'ensemble des matrices de taille $p \times d$, à coefficients réels. Un élément $M \in \mathcal{M}_{p,d}$ peut être assimilé de manière naturelle à un vecteur de $\mathbb{R}^{pd}$, que l'on note*

$$vecA = (A_{11}, \cdots, A_{p1}, A_{12}, \cdots, A_{pd})^t.$$

Inversement, à un vecteur $\bar{A} = (\bar{A}_k)_{1 \le k \le pd} \in \mathbb{R}^{pd}$, on associe la matrice $mat\bar{A}$ définie par $(mat\bar{A})_{ij} = \bar{A}_{i+(j-1)p}$.

Les résultats qui suivent, essentiellement le Théorème 4.2.13 et le Théorème 4.2.16 de Dembo et Zeitouni [35], traitent de l'équivalence et de l'approximation au sens des grandes déviations. Afin de les énoncer, donnons des définitions (essentiellement la Définition 4.2.10 et la Définition 4.2.14 de Dembo et Zeitouni [35]).

Définition 1.7.9. *Soit d une distance rendant S complet. Soit (v_n) une vitesse.*
- *Deux suites de variables aléatoires (Z_n) et (Z'_n) à valeurs dans S sont dites exponentiellement équivalentes pour la vitesse (v_n) si*

$$d(Z_n, Z'_n) \overset{\text{superexp}}{\underset{v_n}{\Longrightarrow}} 0.$$

  Nous utiliserons l'abréviation suivante pour cette notion :

$$Z_n \overset{\text{superexp}}{\underset{v_n}{\sim}} Z'_n.$$

- *Une suite de variables aléatoires $(Z_{n,R})_{n \in \mathbb{N}, R \in \mathbb{R}_+}$ à valeurs dans S est une approximation exponentiellement bonne de (Z_n) pour la vitesse (v_n) si*

$$\lim_{R \to \infty} \limsup_{n \to \infty} \frac{1}{v_n} \log \mathbb{P}\left(d(Z_n, Z_{n,R}) > \delta\right) = -\infty.$$

Théorème 1.7.10. *Soit d une distance rendant S complet. Les suite de variables aléatoires considérées sont à valeurs dans S.*
- *Si (Z_n) et (Z'_n) sont exponentiellements équivalents pour la vitesse (v_n) et si (Z_n) satisfait un PGD de bonne fonction de taux $I(\cdot)$, alors (Z'_n) satisfait le même PGD.*
- *Supposons que $(Z_{n,R})_{n \in \mathbb{N}, R \in \mathbb{R}_+}$ est une approximation exponentiellement bonne de (Z_n) et que pour tout $R > 0$, $(Z_{n,R})_{n \in \mathbb{N}}$ satisfait un PGD de fonction de taux $I_R(\cdot)$. Soit $I(\cdot)$ la fonction de taux définie par*

$$I(x) = \sup_{\delta > 0} \liminf_{R \to \infty} \inf_{y \in B_{x,\delta}} I_R(y),$$

  où $B_{x,\delta}$ désigne la boule $\{y : d(x,y) < \delta\}$. Supposons que $I(\cdot)$ est une bonne fonction de taux et que pour tout ensemble fermé A,

$$\inf_{y \in A} I(y) \le \limsup_{R \to \infty} \inf_{y \in A} I_R(y).$$

  Alors (Z_n) satisfait un PGD de bonne fonction de taux $I(\cdot)$.

La dernière partie du théorème précèdent, bien que plus compliquée à établir s'applique très facilement dans certaines situations. Elle est intensivement utilisée dans les Chapitres 3 et 5.

Chapitre 2

Deviation inequalities, Moderate deviations and some limit theorems for bifurcating Markov chains with application

2.1 Introduction

Bifurcating Markov chains (BMC) are an adaptation of (usual) Markov chains to the data of a regular binary tree (see below for a more precise definition). In other terms, it is a Markov chain for which the index set is a regular binary tree. They are appropriate for example in the modeling of cell lineage data when each cell in one generation gives birth to two offspring in the next one. Recently, they have received a great deal of attention because of the experiments of biologists on aging of *Escherichia Coli* (see [100], [67]). E. Coli is a rod-shaped bacterium which reproduces by dividing in the middle, thus producing two cells, one which already existed and that we call old pole progeny, and the other which is new and that we call new pole progeny. The aim of their experiments was to look for evidence of aging in E. Coli. In this section, we will introduce the model that allowed the authors of [67] to study the aging of E. Coli and we refer to their works for further motivations and insights on the data leading to the model studied here. This model is a typical example of bifurcating Markovian dynamics and it has been the motivation for the rigorous mathematical study of BMC in [66]. This also motivates Section 2.2 and Section 2.3 in the sequel, where we give a rigorous asymptotic (and non asymptotic) study of BMC under geometric ergodicity and uniform geometric ergodicity assumptions.

2.1.1 The model

Let $\mathbb{T}$ be a binary regular tree in which each vertex is seen as a positive integer different from 0, see Figure 2.1. For $r \in \mathbb{N}$, let

$$\mathbb{G}_r = \left\{2^r, 2^r + 1, \cdots, 2^{r+1} - 1\right\}, \quad \mathbb{T}_r = \bigcup_{q=0}^{r} \mathbb{G}_q,$$

which denote respectively the r-th column and the first $(r+1)$ columns of the tree. Then, the cardinality $|\mathbb{G}_r|$ of $\mathbb{G}_r$ is 2^r and that of $\mathbb{T}_r$ is $|\mathbb{T}_r| = 2^{r+1} - 1$. A column

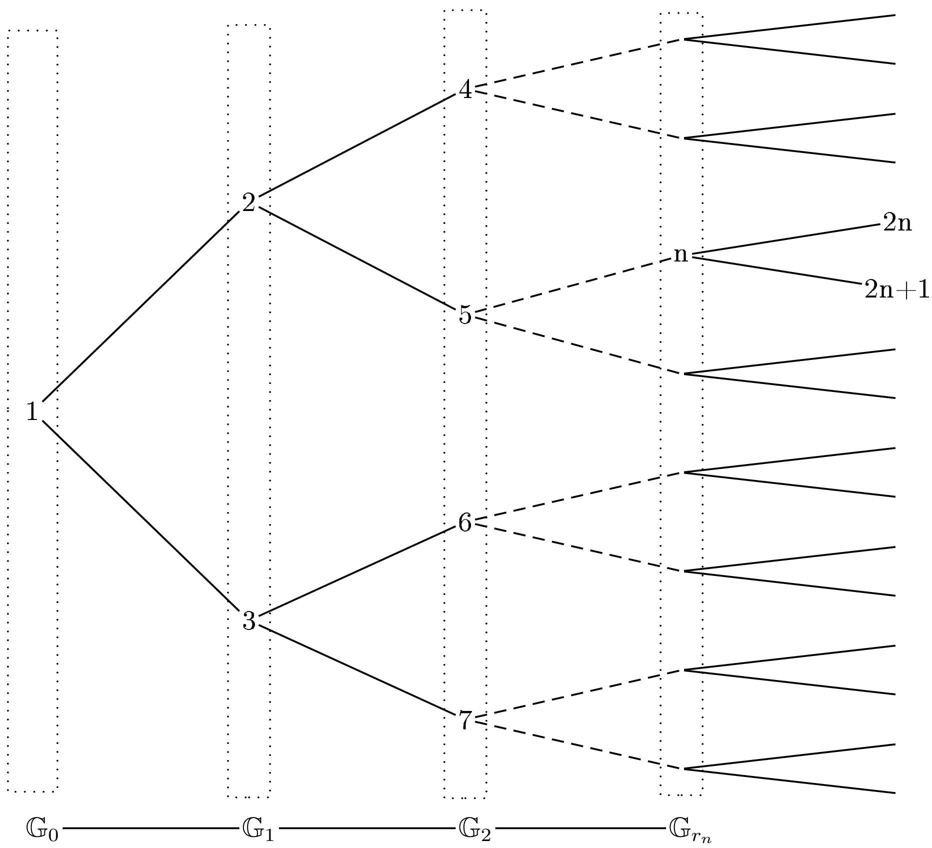

Figure 2.1: The binary tree $\mathbb{T}$

of a given integer n is $\mathbb{G}_{r_n}$ with $r_n = \lfloor \log_2 n \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of the real number x.

The genealogy of the cells is described by this tree. In the sequel we will thus see $\mathbb{T}$ as a given population. Then the vertex n, the column $\mathbb{G}_r$ and the first $(r+1)$ columns $\mathbb{T}_r$ designate respectively individual n, the r-th generation and the first $(r+1)$ generations. The initial individual is denoted 1.

Guyon & Al. ([67], [66]) proposed the following linear Gaussian model to describe the evolution of the growth rate of the population of cells derived from an initial individual

$$\mathcal{L}(X_1) = \nu, \qquad \text{and} \qquad \forall n \geq 1, \quad \begin{cases} X_{2n} = \alpha_0 X_n + \beta_0 + \varepsilon_{2n} \\ \\ X_{2n+1} = \alpha_1 X_n + \beta_1 + \varepsilon_{2n+1}, \end{cases} \tag{2.1.1}$$

where X_n is the growth rate of individual n, n is the mother of $2n$ (the new pole progeny cell) and $2n+1$ (the old pole progeny cell), ν is a distribution probability on $\mathbb{R}$, $\alpha_0, \alpha_1 \in (-1, 1)$; $\beta_0, \beta_1 \in \mathbb{R}$ and $\big((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1\big)$ forms a sequence of i.i.d bivariate random variables with law $\mathcal{N}_2(0, \Gamma)$, where

$$\Gamma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \sigma^2 > 0, \quad \rho \in (-1, 1).$$

The processes (X_n) defined by (2.1.1) are typical examples of BMC which are called the first order bifurcating autoregressive processes (BAR(1)). The BAR(1) processes are an adaptation of autoregressive processes, when the data have a binary tree structure. They were first introduced by Cowan and Staudte [28] for cell lineage data where each individual in one generation gives rise to two offsprings in the next generation. We will not discuss here extensions to m-ary tree, which follows more or less from the same method, or Markov chains on Galton-Watson trees that we shall study in chapter 4.

In [66], Guyon, after establishing the first results on the theory of BMC, proves laws of large numbers and central limit theorem for the least-square estimators $\widehat{\theta}^r = (\widehat{\alpha}_0^r, \widehat{\beta}_0^r, \widehat{\alpha}_1^r, \widehat{\beta}_1^r)$ of the 4-dimensional parameter $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$, (see Section 2.4 for a more precise definition). He also gives some statistical tests which allow to check if the model is symmetric or not (roughly $\alpha_0 = \alpha_1$ or not), and if the new pole and the old pole populations are even distinct in mean, which allows him to conclude a statistical evidence in aging in E. Coli. Let us also mention [18], where Bercu & Al. using the martingale approach give asymptotic analysis of the least squares estimators of the unknown parameters of a general asymmetric pth-order BAR processes.

In this chapter, we will give moderate deviation principle (MDP) for this estimator and the statistical tests done by Guyon. We will also give deviation inequalities for $\widehat{\theta}^r - \theta$, which are important for a rigorous (non asymptotic) statistical study. This will be done in two cases: the Gaussian case as described above and the case where the noise and the initial state X_1 are assumed to take values in a compact set. Note that the latter case implies that the BAR(1) process defined by (2.1.1) valued in compact set.

We are now going to give a rigorous definition of BMC. We refer to [66] for more details.

2.1.2 Definitions

For an individual $n \in \mathbb{T}$, we are interested in the quantity X_n (it may be the weight, the growth rate,$\cdots$) with values in the metric space S endowed with its Borel σ-field $\mathcal{S}$.

Definition 2.1.1 ($\mathbb{T}$-transition probability, see ([66])). *We call $\mathbb{T}$-transition probability any mapping $P : S \times \mathcal{S}^2 \to [0,1]$ such that*

- *$P(.,A)$ is measurable for all $A \in \mathcal{S}^2$,*
- *$P(x,.)$ is a probability measure on $(S^2, \mathcal{S}^2)$ for all $x \in S$.*

For a $\mathbb{T}$-transition probability P on $S \times \mathcal{S}^2$, we denote by P_0, P_1 and Q respectively the first and the second marginal of P, and the mean of P_0 and P_1, that is $P_0(x,B) = P(x, B \times S)$, $P_1(x,B) = P(x, S \times B)$ for all $x \in S$ and $B \in \mathcal{S}$ and $Q = \dfrac{P_0 + P_1}{2}$.

For $p \geq 1$, we denote by $\mathcal{B}(S^p)$ (resp. $\mathcal{B}_b(S^p)$), the set of all $\mathcal{S}^p$-measurable (resp. $\mathcal{S}^p$-measurable and bounded) mappings $f : S^p \to \mathbb{R}$. For $f \in \mathcal{B}(S^3)$, we denote by $Pf \in \mathcal{B}(S)$ the function

$$x \mapsto Pf(x) = \int_{S^2} f(x,y,z) P(x, dy, dz), \text{ when it is defined.}$$

Definition 2.1.2 (Bifurcating Markov Chains, see ([66])). *Let $(X_n, n \in \mathbb{T})$ be a family of S-valued random variables defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_r, r \in \mathbb{N}), \mathbb{P})$. Let ν be a probability on $(S, \mathcal{S})$ and P be a $\mathbb{T}$-transition probability. We say that $(X_n, n \in \mathbb{T})$ is a $(\mathcal{F}_r)$-bifurcating Markov chain with initial distribution ν and $\mathbb{T}$-transition probability P if*

- *X_n is $\mathcal{F}_{r_n}$-measurable for all $n \in \mathbb{T}$,*
- *$\mathcal{L}(X_1) = \nu$,*
- *for all $r \in \mathbb{N}$ and for all family $(f_n, n \in \mathbb{G}_r) \subseteq \mathcal{B}_b(S^3)$*

$$\mathbb{E}\left[\prod_{n \in \mathbb{G}_r} f_n(X_n, X_{2n}, X_{2n+1}) \Big| \mathcal{F}_r \right] = \prod_{n \in \mathbb{G}_r} Pf_n(X_n).$$

In the following, when unprecised, the filtration implicitly used will be $\mathcal{F}_r = \sigma(X_i, i \in \mathbb{T}_r)$. We denote by $(Y_r, r \in \mathbb{N})$ the Markov chain on S with $Y_0 = X_1$ and transition probability Q. The chain $(Y_r, r \in \mathbb{N})$ corresponds to a random lineage taken in the population.

We denote by $\mathfrak{G}$ the set of all permutations of $\mathbb{N}^*$ that leaves each $\mathbb{G}_r$ invariant. We draw a permutation Π uniformly on $\mathfrak{G}$, independently of $X = (X_n, n \in \mathbb{T})$.

Drawing Π "uniformly" on $\mathfrak{G}$ means drawing the restriction of Π on $\mathbb{G}_r$ uniformly among the $(2^r)!$ permutations of $\mathbb{G}_r$. In particular,

$$\big(\Pi(2^r), \Pi(2^r+1), \cdots, \Pi(2^{r+1}-1)\big)$$

can be viewed as a random drawing of all the elements of $\mathbb{G}_r$ without replacement. Notice that Π allows to define a random order on $\mathbb{T}$ which preserves the genealogical order. For example, $(\Pi(i), 1 \leq i \leq n)$ denotes the set of the "first" n individuals of $\mathbb{T}$. Π was introduced by Guyon in order to sample over the "first" n individuals. As mentioned in [66], this choice of Π allows to preserve the same asymptotic behavior for the empirical means resulting from the sampling over (say) the rth generation, the first $(r+1)$ generations or the "first" n individuals. In general, the choice of another permutation does not preserve the asymptotic behavior of these empirical means. We refer to [66] Section 2.2, for more details.

In all the chapter, we will denote by:

- $f \otimes g$ the mapping $(x, y) \mapsto f(x)g(y)$,
- Q^p the pth iterated of Q recursively defined by the formulas $Q^0(x, \cdot) = \delta_x$ and $Q^{p+1}(x, B) = \int_S Q(s, dy)Q^p(y, B)$ for all $B \in \mathcal{S}$; Q^p is a transition probability in $(S, \mathcal{S})$.
- νQ the distribution on $(S, \mathcal{S})$ defined by $\nu Q(B) = \int_S \nu(dx)Q(x, B)$; νQ^p is the law of Y_p,
- $(Qf)(x) = \int_S f(y)Q(x, dy)$ when it is defined;
- (νf) or (ν, f) the integral $\int_S f d\nu$ when it is defined.

For all $i \in \mathbb{T}$, we set $\Delta_i = (X_i, X_{2i}, X_{2i+1})$. We introduce the following empirical quantities:

$$\begin{cases} \overline{M}_{\mathbb{G}_r}(f) = \frac{1}{|\mathbb{G}_r|} \sum\limits_{i \in \mathbb{G}_r} f(\widetilde{\Delta}_i), \\ \overline{M}_{\mathbb{T}_r}(f) = \frac{1}{|\mathbb{T}_r|} \sum\limits_{i \in \mathbb{T}_r} f(\widetilde{\Delta}_i), \\ \overline{M}_n^{\Pi}(f) = \frac{1}{n} \sum\limits_{i=1}^{n} f(\widetilde{\Delta}_{\Pi(i)}), \end{cases} \tag{2.1.2}$$

where $f(\widetilde{\Delta}_i) = f(\Delta_i) = f(X_i, X_{2i}, X_{2i+1})$ if $f \in \mathcal{B}(S^3)$ and $f(\widetilde{\Delta}_i) = f(X_i)$ if $f \in \mathcal{B}(S)$.

Guyon in [66] studied limit theorems of the empirical means (2.1.2), namely the law of large numbers (L^2 and almost sure versions) and the central limit theorems for (2.1.2) when $f \in \mathcal{B}(S^3)$ but centered by the conditional expectation rather than by the limit mean. An extension of the BMC has been proposed in [33], in which the authors studied a model of BMC with missing data. To take into account the possibility for a cell to die, the authors in [33] use Galton-Watson tree instead of a regular tree. And they give a weak law of large number, an invariance principle and the central limit result for the average over one generation or up to one generation. As previously mentioned, this setting will be considered in chapter 4. One can also mention the work of De Saporta & Al. [32] dealing with bifurcating autoregressive processes with missing data in the estimation procedure of the parameters of the asymmetric BAR process. They use a two type Galton-Watson process to model the genealogy and give convergence and asymptotic normality of their estimators. It is important to remark that the non-asymptotic study of deviation inequalities has not been considered at all in these works, despite their practical interest.

2.1.3 Objectives

Our objectives in this chapter are:

- to give some limit theorems for BMC that complete those done in [66] (LLN, LIL,...);
- to give probability inequalities and deviation inequalities for the empirical means (2.1.2), i.e. for $f \in \mathcal{B}(S)$ and all $x > 0$

$$\mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(f) - (\mu, f) \geq x\right) \leq e^{-C(x,r)}$$

  where $C(x,r)$ will crucially depends on our set of assumptions on f and on the ergodic property of Q but valid for (nearly) all r;
- to study moderate deviation principle (MDP) for BMC, i.e. for some range of speed $\sqrt{r} \ll b_r \ll r$ (depending on assumptions) and for $f \in \mathcal{C}_b(S^3)$ with $Pf = 0$

$$\frac{b^2_{|\mathbb{T}_r|}}{|\mathbb{T}_r|} \log \mathbb{P}\left(\frac{1}{b_{|\mathbb{T}_r|}} M_{|\mathbb{T}_r|}(f) \geq x\right) \sim -\frac{x^2}{2\sigma^2};$$

- to obtain the MDP and deviation inequalities for the estimator of bifurcating autoregressive process, which are important for a rigorous statistical study.

All these results will be obtained under hypothesis of geometric ergodicity or uniform geometric ergodicity, meaning that Q^r converges (uniformly) exponentially fast to a limiting measure.

The limit theorems, proved in this chapter, include strong law of large numbers for the empirical average $\overline{M}^{\Pi}_n(f)$ with $f \in \mathcal{B}(S)$ (this case is not studied in [66]), the law of the iterated logarithm and the almost sure functional central limit theorem. Strong law of large numbers will be obtained via control of 4th order moments. We thus generalize the computation of 2nd order moments made by Guyon in [66]. It will be noted that the technique we will use can be applied to compute the other higher order moments but at the price of huge and tedious computations.

Deviation inequalities will be obtained in the setting of unbounded functions, by using the classical Markov inequality and under geometric ergodicity assumption. The results are however at this point quite restrictive.

Exponential deviation inequalities will be shown for bounded functions and under a uniform geometric ergodicity assumption. Their proof intensively uses Azuma-Bennet-Hoeffding inequality [6], [16], [70], which requires bounded random variables. Extension to unbounded functions and weaker ergodicity assumptions will be done in a further work, using transportation inequalities in the spirit of [41].

The MDP will be mainly deduced from these inequalities and general results on moderate deviations of martingales (see [38], recalled in the appendix 2.B). Its speed will depend on whether uniform geometric ergodicity or only geometric ergodicity is satisfied.

Before presenting the plan of the chapter, let us recall the definition of a moderate deviation principle (MDP): let $(b_n)_{n\geq 0}$ be a positive sequence such that

$$\frac{b_n}{n} \underset{n\to\infty}{\longrightarrow} 0, \qquad \text{and} \qquad \frac{b_n^2}{n} \underset{n\to\infty}{\longrightarrow} \infty.$$

We say that a sequence of centered random variables $(M_n)_n$ with topological state space $(S, \mathcal{S})$ satisfies a MDP with speed b_n^2/n and rate function $I : S \to \mathbb{R}_+^*$ if for each $A \in \mathcal{S}$,

$$\begin{aligned} -\inf_{x\in A^o} I(x) \leq \liminf_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{n}{b_n} M_n \in A\right) \\ \leq \limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{n}{b_n} M_n \in A\right) \leq -\inf_{x\in\overline{A}} I(x), \end{aligned}$$

here A^o and $\overline{A}$ denote the interior and closure of A respectively.

The MDP can thus be seen as an intermediate behavior between the central limit theorem ($b_n = b\sqrt{n}$) and Large deviation ($b_n = bn$). Usually, the MDP exhibit a simpler rate function inherited from the approximated Gaussian process, and holds for a larger class of dependent random variables than the large deviation principle.

This chapter is organized as follows. Section 2.2 states the moments control inequalities and their consequences. We shall state in this section a first result on the MDP for BMC in a general framework (that is for unbounded functions), but with a very restricted range of speed. Section 2.3 deals with the exponential inequalities and their consequences. In this section, we shall generalize the

MDP done in section 2.2 allowing for a larger range of speed, but under more stringent assumptions. In section 2.4, we will focus particularly on the first order bifurcating autoregressive processes. The proofs of some inequalities are technical so postponed in Appendix 2.A. Appendix 2.B is devoted to definitions and limit theorems for martingales used intensively in this chapter, included here for completeness.

2.2 Moments control and consequences

Let F be a vector subspace of $\mathcal{B}(S)$ such that

(i) F contains the constants,

(ii) $F^2 \subset F$,

(iii) $F \otimes F \subset L^1(P(x,.))$ for all $x \in S$, and $P(F \otimes F) \subset F$,

(iv) there exists a probability μ on $(S, \mathcal{S})$ such that $F \subset L^1(\mu)$ and $\lim_{r\to\infty} \mathbb{E}_x\Big[f(Y_r)\Big] = (\mu, f)$ for all $x \in S$ and $f \in F$,

(v) for all $f \in F$, there exists $g \in F$ such that for all $r \in \mathbb{N}$, $|Q^r f| \le g$,

(vi) $F \subset L^1(\nu)$,

where we have used the notation $F^2 = \{f^2/f \in F\}$, $F \otimes F = \{f \otimes g/f, g \in F\}$ and $PE = \{Pf/f \in E\}$ whenever an operator P acts on a set E.

The following hypothesis is about the geometric ergodicity of Q:

(H1) Assume that for all $f \in F$ such that $(\mu, f) = 0$, there exists $g \in F$ such that for all $r \in \mathbb{N}$ and for all $x \in S$, $|Q^r f(x)| \le \alpha^r g(x)$ for some $\alpha \in (0,1)$, that is the Markov chain $(Y_r, r \in \mathbb{N})$ is geometrically ergodic.

Recall that under this hypothesis, Guyon [66] has shown the weak law of large numbers for the three empirical average $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_{\mathbb{T}_r}(f)$ and $\overline{M}_n^{\Pi}(f)$ (in [66]: see Theorem 11 when $f \in F$ and Theorem 12 when $f \in \mathcal{B}(S^3)$) and the strong law of large numbers only for $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_{\mathbb{T}_r}(f)$ (in [66]: see Theorem 14 and Corollary 15 when $f \in F$ and Theorem 18 when $f \in \mathcal{B}(S^3)$).

When $f \in \mathcal{B}(S^3)$ and under the additional hypothesis Pf^2 and Pf^4 exist and belong to F, he proved the central limit theorem for $\overline{M}_{\mathbb{T}_r}(f)$ and $\overline{M}_n^{\Pi}(f)$ (in [66]: see Theorem 19 and Corollary 21). Recall that the central limit theorem for the three empirical means (2.1.2) when $f \in \mathcal{B}(S)$ is still an open question, see [33] for more precision.

In this section, we complete these results by showing the strong law of large numbers for $\overline{M}_n^{\Pi}(f)$, when $f \in F$. We prove also the law of the iterated logarithm (LIL) and almost sure functional central limit theorem (ASFCLT) for $\overline{M}_n^{\Pi}(f)$ when $f \in \mathcal{B}(S^3)$.

2.2.1 Control of the 4-th order moments

In order to establish limit theorems below, let us state the following

Theorem 2.2.1. *Let F satisfy (i)-(vi). Let $f \in F$ such that $(\mu, f) = 0$. We assume hypothesis (**H1**). Then for all $r \in \mathbb{N}$*

$$\mathbb{E}\left[\left(\overline{M}_{\mathbb{G}_r}(f)\right)^4\right] \le \begin{cases} c\left(\frac{1}{4}\right)^r & \text{if} \quad \alpha^2 < \frac{1}{2} \\ cr^2\left(\frac{1}{4}\right)^r & \text{if} \quad \alpha^2 = \frac{1}{2} \\ c\alpha^{4r} & \text{if} \quad \alpha^2 > \frac{1}{2} \end{cases} \tag{2.2.1}$$

where the positive constant c depends on α and f (and may differ line by line).

Proof. First note that $f(X_i) \in L^4$ for all $i \in \mathbb{G}_r$. Indeed, let $(z_1, \cdots, z_r) \in \{0,1\}^r$ the unique path in the binary tree from the root 1 to i. Then,

$$\mathbb{E}\Big[f^4(X_i)\Big] = \nu P_{z_1} \cdots P_{z_r} f^4,$$

and from hypothesis (ii), (iii) and (vi) we conclude that $\nu P_{z_1} \cdots P_{z_r} f^4 < \infty$.

Now, the the proof divides into two parts.

Part 1. Computation of $\mathbb{E}\left[\left(\overline{M}_{\mathbb{G}_r}(f)\right)^4\right]$.

Independently on X, let us draw four independent indices I_r, J_r, K_r and L_r uniformly from $\mathbb{G}_r$. Then

$$\mathbb{E}\Big[\left(\overline{M}_{\mathbb{G}_r}(f)\right)^4\Big] = \mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\Big].$$

For all $p \in \{0, \cdots, r\}$, let us define the following events:

- E_0^p : "The ancestors of I_r, J_r, K_r and L_r are different in $\mathbb{G}_p$".
- E_1^p : "Exactly two of I_r, J_r, K_r and L_r have the same ancestor in $\mathbb{G}_p$".
- E_2^p : "I_r, J_r, K_r and L_r have the same ancestor two by two in $\mathbb{G}_p$".
- E_3^p : "Exactly three of I_r, J_r, K_r and L_r have the same ancestor in $\mathbb{G}_p$".
- E_4^p : "I_r, J_r, K_r and L_r have the same ancestor in $\mathbb{G}_p$".

We also consider the following events whose for each fixed $p \le r$, probability depend only on p.

- $E_0^{'p}$: "Draw uniformly four independent indices from $\mathbb{G}_p$ which are different".
- $E_1^{'p}$: "Draw uniformly four independent indices from $\mathbb{G}_p$ such that two are the same and the others are different ".

- $E_2^{'p}$: "Draw uniformly four independent indices from $\mathbb{G}_p$ which are the same two by two".
- $E_3^{'p}$: "Draw uniformly four independent indices from $\mathbb{G}_p$ such that exactly three are the same.".
- $E_4^{'p}$: "Draw uniformly four independent indices from $\mathbb{G}_p$ which are all the same".

In the sequel we do the convention that E_0^{r+1} is a certain event. Then after successive conditioning by events E_i^p for $p \in \{0, \cdots, r\}$ and $i \in \{0, \cdots, 4\}$, we have

$$\begin{aligned}
\mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\Big] &= \mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|E_0^2\Big] \times \mathbb{P}\big(E_0^2\big) \\
&+ \sum_{p=2}^{r} \mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|E_0^{p+1}, E_1^p\Big] \times \mathbb{P}\big(E_1^p \cap E_0^{p+1}\big) \\
&+ \sum_{p=2}^{r} \mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|E_0^{p+1}, E_2^p\Big] \times \mathbb{P}\big(E_2^p \cap E_0^{p+1}\big) \\
&+ \mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|E_3^r\Big] \times \mathbb{P}\big(E_3^r\big) \\
&+ \mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|E_4^r\Big] \times \mathbb{P}\big(E_4^r\big).
\end{aligned} \tag{2.2.2}$$

Let us notice that

- for all $i \in \{1, 2, 3, 4\}$, E_i^r and $E_i^{'r}$ have the same probability;
- the realization of "$E_1^p \cap E_0^{p+1}$" can be seen as "draw uniformly four independent indices from $\mathbb{G}_p$ such that two are the same and others are different and the two indices which are the same take different paths at $\mathbb{G}_{p+1}$". Thus, "$E_1^p \cap E_0^{p+1}$" has the same probability that "$E_1^{'p} \cap A_{p,p+1}$", where "$A_{p,p+1}$" is the event "the indices which are the same in $\mathbb{G}_p$ take different paths at $\mathbb{G}_{p+1}$";
- similarly, the realization of "$E_2^p \cap E_0^{p+1}$" may be interpreted as "draw uniformly four independent indices from $\mathbb{G}_p$ which are the same two by two and all the indices take different path at $\mathbb{G}_{p+1}$". Thus, "$E_2^p \cap E_0^{p+1}$" has the same probability that "$E_2^{'p} \cap A_{p,p+1}$", where "$A_{p,p+1}$" is the event "the indices which are the same in $\mathbb{G}_p$ take different paths at $\mathbb{G}_{p+1}$",
- For all $p \in \{0, \cdots, r\}$, we have

$$\mathbb{P}(E_1^{'p}) = \frac{6(2^p - 1)(2^p - 2)}{2^{3p}}, \quad \mathbb{P}(E_2^{'p}) = \frac{3(2^p - 1)}{2^{3p}},$$
$$\mathbb{P}(E_3^{'p}) = \frac{4(2^p - 1)}{2^{3p}}, \quad \mathbb{P}(E_4^{'p}) = \frac{1}{2^{3p}}.$$

We may then deduce that

$$\mathbb{P}\big(E_0^2\big) = \frac{3}{32}; \quad \mathbb{P}\big(E_3^r\big) = \frac{4(2^r-1)}{2^{3r}}, \quad \mathbb{P}\big(E_4^r\big) = \frac{1}{2^{3r}},$$

and for $p \in \big\{2, \cdots, r-1\big\}$,

$$\mathbb{P}\big(E_1^p \cap E_0^{p+1}\big) = \mathbb{P}\big(E_1^{'p}\big)\mathbb{P}\big(A_{p,p+1}\big|E_1^{'p}\big) = \frac{3(2^p-1)(2^p-2)}{2^{3p}},$$

and

$$\mathbb{P}\big(E_2^p \cap E_0^{p+1}\big) = \mathbb{P}\big(E_2^{'p}\big)\mathbb{P}\big(A_{p,p+1}\big|E_2^{'p}\big) = \frac{3}{4}\frac{2^p-1}{2^{3p}}.$$

We are now going to compute each term which appears in (2.2.2). We do the following convention $P(Q^{-1}f \otimes Q^{-1}f) = f^2$. In the sequel, we will use intensively, with a slight modification, the calculations made by Guyon [66] in order to compute conditional expectations related to the event "draw uniformly two independent indices from $\mathbb{G}_p$", for $p \in \{0, \cdots, r\}$.

(a) We have that

$$\mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})|E_4^r\Big] = \nu Q^r f^4.$$

(b) Conditionally on E_3^r, we may assume that the indices I_r, K_r and L_r are the same. We then have using the calculations made by Guyon [66]

$$\mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|E_3^r\Big] = \mathbb{E}\Big[f^3(X_{I_r})f(X_{J_r})\big|E_3^r\Big]$$
$$= \tfrac{2^r}{2^r-1}\left\{\sum_{p=0}^{r-1} 2^{-p-2}\nu Q^p P\left(Q^{r-p-1}f^3 \otimes Q^{r-p-1}f + Q^{r-p-1}f \otimes Q^{r-p-1}f^3\right)\right\}.$$

(c) Let $p \in \big\{2, \cdots, r\big\}$. Conditionally on E_2^p and E_0^{p+1} we may assume that I_r and J_r have the same ancestor at $\mathbb{G}_p$ and K_r and L_r have the same ancestor at $\mathbb{G}_p$. For simplification, we will use the following notation

$$Q_\otimes^k f := Q^k f \otimes Q^k f, \tag{2.2.3}$$

we thus have

$$\mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|E_0^{p+1}, E_2^p\Big]$$
$$= \mathbb{E}\Bigg[\mathbb{E}\Big[\mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|\mathcal{F}_{p+1}\Big]\Big|\mathcal{F}_p\Big]\Bigg|E_0^{p+1}, E_2^p\Bigg]$$
$$= \mathbb{E}\Big[P\left(Q_\otimes^{r-p-1}f\right)(X_{I_r\wedge_p J_r})P\left(Q_\otimes^{r-p-1}f\right)(X_{K_r\wedge_p L_r})\big|E_0^{p+1}, E_2^p\Big]$$
$$= \tfrac{2^p}{2^p-1}\sum_{l=0}^{p-1} 2^{-l-1}\nu Q^l P\left(\left(Q^{p-l-1}P\left(Q_\otimes^{r-p-1}f\right)\right) \otimes \left(Q^{p-l-1}P\left(Q_\otimes^{r-p-1}f\right)\right)\right),$$

where $I_r \wedge_p J_r$(resp. $K_r \wedge_p L_r$) denotes the common ancestor of I_r and J_r which is in $\mathbb{G}_p$ (resp. the common ancestor of K_r and L_r which is in $\mathbb{G}_p$).

(d) Let $p \in \{2, \cdots, r\}$. Now conditionally on E_1^p and E_0^{p+1} we may assume that it is K_r and L_r which have the same ancestor in $\mathbb{G}_p$. We denote by $p(I_r)$ and $p(J_r)$ respectively the ancestor of I_r and J_r which are in $\mathbb{G}_p$. As before, the common ancestor of K_r and L_r which are in $\mathbb{G}_p$ is denoted by $K_r \wedge_p L_r$. At this step, we may repeat the successive conditioning that we have done in the beginning but this time for indices $p(I_r)$, $p(J_r)$ and $K_r \wedge_p L_r$. This leads us to

$$\begin{aligned}
&\mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\Big|E_0^{p+1}, E_1^p\Big] \\
&= \mathbb{E}\left[Q^{r-p}f(X_{p(I_r)})Q^{r-p}f(X_{p(J_r)})P\left(Q_{\otimes}^{r-p-1}f\right)(X_{K_r\wedge_p L_r})\Big|E_0^{p+1}, E_1^p\right] \\
&= \tfrac{2^{2p}}{(2^p-1)(2^p-2)}\sum_{l=2}^{p-1}\tfrac{1}{2^{l+1}}\tfrac{1}{2}\sum_{m=0}^{l-1}2^{-m-1} \\
&\Big\{\nu Q^m P\left(\left(Q^{l-m-1}P\left(Q_{\otimes}^{r-l-1}f\right)\right) \otimes Q^{p-m-1}P\left(Q_{\otimes}^{r-p-1}f\right)\right) \\
&+ \nu Q^m P\left(\left(Q^{p-m-1}P\left(Q_{\otimes}^{r-p-1}f\right)\right) \otimes \left(Q^{l-m-1}P\left(Q_{\otimes}^{r-l-1}f\right)\right)\right) \\
&+ \nu Q^m P\left(\left(Q^{l-m-1}P\left(Q^{r-l-1}f \otimes Q^{p-l-1}P\left(Q_{\otimes}^{r-p-1}f\right)\right)\right) \otimes \left(Q^{r-m-1}f\right)\right) \\
&+ \nu Q^m P\left(Q^{r-m-1}f \otimes \left(Q^{l-m-1}P\left(Q^{r-l-1}f \otimes Q^{p-l-1}P\left(Q_{\otimes}^{r-p-1}f\right)\right)\right)\right) \\
&+ \nu Q^m P\left(\left(Q^{l-m-1}P\left(Q^{p-l-1}P\left(Q_{\otimes}^{r-p-1}f\right) \otimes Q^{r-l-1}f\right)\right) \otimes \left(Q^{r-m-1}f\right)\right) \\
&+ \nu Q^m P\left(\left(Q^{r-m-1}f\right) \otimes \left(Q^{l-m-1}P\left(Q^{p-l-1}P\left(Q_{\otimes}^{r-p-1}f\right) \otimes Q^{r-l-1}f\right)\right)\right)\Big\} \\
&+ \tfrac{3}{48}\tfrac{2^{2p}}{(2^p-1)(2^p-2)}\Big\{\nu P\left(P\left(Q_{\otimes}^{r-2}f\right) \otimes Q^{p-1}P\left(Q_{\otimes}^{r-p-1}f\right)\right) \\
&\quad + \nu P\left(Q^{p-1}P\left(Q_{\otimes}^{r-p-1}f\right) \otimes P\left(Q_{\otimes}^{r-2}f\right)\right) \\
&\quad + \nu P\left(P\left(Q^{r-2}f \otimes Q^{p-2}P\left(Q_{\otimes}^{r-p-1}f\right)\right) \otimes Q^{r-1}f\right) \\
&\quad + \nu P\left(P\left(Q^{p-2}P\left(Q_{\otimes}^{r-p-1}f\right) \otimes Q^{r-2}f\right) \otimes Q^{r-1}f\right) \\
&\quad + \nu P\left(Q^{r-1}f \otimes P\left(Q^{r-2}f \otimes P\left(Q_{\otimes}^{r-p-1}f\right)\right)\right) \\
&\quad + \nu P\left(Q^{r-1}f \otimes P\left(Q^{p-2}P\left(Q_{\otimes}^{r-p-1}f\right) \otimes Q^{r-2}f\right)\right)\Big\}.
\end{aligned}$$

(e) Finally,

$$\begin{aligned}
&\mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\Big|E_0^2\Big] \\
&\qquad = \mathbb{E}\Big[\mathbb{E}\Big[\mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\Big|\mathcal{F}_2\Big]\Big|\mathcal{F}_1\Big]\Big|E_0^2\Big] \\
&\qquad = \mathbb{E}\Big[P\left(Q_{\otimes}^{r-2}f\right)(X_2)P\left(Q_{\otimes}^{r-2}f\right)(X_3)\Big|E_0^2\Big] \\
&\qquad = \nu P\left(P\left(Q_{\otimes}^{r-2}f\right) \otimes P\left(Q_{\otimes}^{r-2}f\right)\right).
\end{aligned}$$

Gathering together all these terms each multiplied by their respective probability, we obtain an explicit expression for $\mathbb{E}\Big[\left(\overline{M}_{\mathbb{G}_r}(f)\right)^4\Big]$.

Part 2. Rate.

We are now going to give some rate for the different terms that appear in the expression of $\mathbb{E}\Big[\left(\overline{M}_{\mathbb{G}_r}(f)\right)^4\Big]$.

Throughout this part, we will use intensively the following to bound quantities which appear in the expression of $\mathbb{E}\Big[\left(\overline{M}_{\mathbb{G}_r}(f)\right)^4\Big]$:

- Let $f \in F$ such that $(\mu, f) = 0$. Then from (i)-(vi) and hypothesis **(H1)**, there exists a positive constant c such that $\forall\, l, m, n \in \mathbb{N}$,

$$\nu Q^l P\left(Q^m f \otimes Q^n f\right) \leq \alpha^{m+n} \nu Q^l P\left(g \otimes g\right) \leq c\alpha^{m+n},$$

  where g is given in hypothesis **(H1)**.

In the sequel c denotes a positive constant which depends on f and c_1 denotes a positive constant which depend on α. The constants c and c_1 may vary from one line to another and from one expression to another.

(a) For the first term appearing in (2.2.2), we have

$$\mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|E_0^2\Big] \times \mathbb{P}\big(E_0^2\big) \leq c_1 c\alpha^{4r}.$$

(b) For the fifth term appearing in (2.2.2), we have

$$\mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|E_4^r\Big] \times \mathbb{P}\big(E_4^r\big) \leq c\left(\frac{1}{2}\right)^{3r},$$

where, from (ii), (v) and (vi), c is such that $\nu Q^r f^4 < c$,

(c) For the fourth term appearing in (2.2.2), we have

$$\mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|E_3^r\Big] \times \mathbb{P}\big(E_3^r\big) \leq cc_1\alpha^r\left(\frac{1}{4}\right)^r\sum_{p=0}^{r-1}\left(\frac{1}{2\alpha}\right)^p,$$

where, from (ii), (iii), (v) and (vi), c is such that for all $p, q \in \mathbb{N}$

$$\max\left(\nu Q^p P\left(Q^q f^3 \otimes g\right), \nu Q^p P\left(g \otimes Q^q f^3\right)\right) < c,$$

and from hypothesis **(H1)**, g is such that for all $p \in \big\{1, \cdots, r-1\big\}$

$$Q^{r-p-1} f \leq \alpha^{r-p-1} g. \tag{2.2.4}$$

Now depending on the value of α, we obtain that

$$\mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|E_3^r\Big]\times\mathbb{P}\big(E_3^r\big) \leq \begin{cases} c_1 c\left(\left(\frac{\alpha}{4}\right)^r + \left(\frac{1}{2^3}\right)^r\right) & \text{if} \quad \alpha \neq \frac{1}{2} \\ c_1 c r\left(\frac{1}{2^3}\right)^r & \text{if} \quad \alpha = \frac{1}{2}. \end{cases}$$

(d) Let us denote the third term appearing in (2.2.2) by

$$A_r := \sum_{p=2}^{r} \mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|E_0^{p+1}, E_2^p\Big] \times \mathbb{P}\big(E_2^p \cap E_0^{p+1}\big).$$

So we have

$$A_r \leq c_1 c \left(\left(\frac{1}{4}\right)^r + \alpha^{4r} \sum_{p=2}^{r-1} \left(\frac{1}{4\alpha^4}\right)^p \right),$$

where, from (ii), (iii), (v) and (vi), c is such that
for all $p \in \{2, \cdots, r-1\}$, $q \in \{0, \cdots, r-1\}$, $l \in \{0, \cdots, p-1\}$

$$\max\Big(\nu Q^q P\left(Q_\otimes^{r-q-1} f^2\right), \nu Q^l P\left(Q_\otimes^{p-l-1} P(g \otimes g)\right)\Big) < c,$$

and g is defined as before (2.2.4) and the notation $Q_\otimes$ is given in (2.2.3).

Now depending on the value of α, we obtain that

- if $\alpha^2 \neq \frac{1}{2}$ then $A_r \leq c_1 c \left(\left(\frac{1}{4}\right)^r + \alpha^{4r} \right)$;
- if $\alpha^2 = \frac{1}{2}$ then $A_r \leq c_1 c\,(r-1) \left(\frac{1}{4}\right)^r$.

(e) For the second term appearing in (2.2.2), we have when $p = r$

- if $\alpha = \frac{1}{2}$ then

$$\mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|E_1^r\Big] \times \mathbb{P}\big(E_1^r\big) \leq c_1 c \left(\frac{1}{4}\right)^r;$$

- if $\alpha \neq \frac{1}{2}$
  - if $\alpha^2 = \frac{1}{2}$ then

$$\mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|E_1^r\Big] \times \mathbb{P}\big(E_1^r\big) \leq c_1 (r-1) \left(\frac{1}{4}\right)^r;$$

  - if $\alpha^2 \neq \frac{1}{2}$ then

$$\mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|E_1^r\Big] \times \mathbb{P}\big(E_1^r\big) \leq c_1 c \left(\left(\frac{\alpha^2}{2}\right)^r + \left(\frac{1}{4}\right)^r \right),$$

where, from (ii), (iii), (v) and (vi), c is such that for all
$l \in \{2, \cdots, r-1\}$, $q \in \{0, \cdots, l-1\}$,

$$\max\big(\nu Q^q P\left(Q^{l-q-1} P\left(g \otimes g\right) \otimes Q^{r-q-1} f^2\right), \nu Q^q P\left(Q^{l-q-1} P\left(g \otimes Q^{r-l-1} f^2\right) \otimes g\right)\big) < c,$$

and g is defined as before (2.2.4).

(f) For the second terms appearing in (2.2.2), and for the remaining term in the sum ($p \neq r$), let us denote by :

$$B_r := \sum_{p=2}^{r-1} \mathbb{E}\Big[f(X_{I_r})f(X_{J_r})f(X_{K_r})f(X_{L_r})\big|E_0^{p+1}, E_1^p\Big] \times \mathbb{P}\big(E_1^p \cap E_0^{p+1}\big).$$

So we have

- if $\alpha = \frac{1}{2}$ then $B_r \leq c_1 c \left(\frac{1}{4}\right)^r$;
- if $\alpha \neq \frac{1}{2}$
  - if $\alpha^2 = \frac{1}{2}$ then $B_r \leq c_1 c r^2 \left(\frac{1}{4}\right)^r$;
  - if $\alpha^2 \neq \frac{1}{2}$ then $B_r \leq c_1 c \left(\alpha^{4r} + \left(\frac{\alpha^2}{2}\right)^r + \left(\frac{1}{4}\right)^r\right)$,

where c is defined in the same way as before.

Now the results of the Theorem 2.2.1 follow from **(a)**-**(f)** of **Part 2**. $\square$

It leads us to an extension of Theorem 2.2.1 to the two empirical averages $\overline{M}_{\mathbb{T}_r}(f)$ and $\overline{M}_n^{\Pi}(f)$.

Corollary 2.2.2. *Let F satisfy (i)-(vi). Let $f \in F$ such that $(\mu, f) = 0$. We assume that hypothesis (**H1**) is fulfilled. Then for all $r \in \mathbb{N}$ and $n \in \mathbb{N}$*

$$\mathbb{E}\Big[\left(\overline{M}_{\mathbb{T}_r}(f)\right)^4\Big] \leq \begin{cases} c\left(\frac{1}{4}\right)^{r+1} & \text{if} \quad \alpha^2 < \frac{1}{2}, \\ cr^2\left(\frac{1}{4}\right)^{r+1} & \text{if} \quad \alpha^2 = \frac{1}{2}, \\ c\alpha^{4(r+1)} & \text{if} \quad \alpha^2 > \frac{1}{2}, \end{cases} \tag{2.2.5}$$

and

$$\mathbb{E}\Big[\left(\overline{M}_n^{\Pi}(f)\right)^4\Big] \leq \begin{cases} c\left(\frac{1}{4}\right)^{r_n+1} & \text{if} \quad \alpha^2 < \frac{1}{2}, \\ cr_n^2\left(\frac{1}{4}\right)^{r_n+1} & \text{if} \quad \alpha^2 = \frac{1}{2}, \\ c\alpha^{4(r_n+1)} & \text{if} \quad \alpha^2 > \frac{1}{2}, \end{cases} \tag{2.2.6}$$

where the positive constant c depends on α and f and may differ line by line.

Proof. The proof follows the same steps as in the proof of **Part 2** and **Part 3** of Theorem 2.2.11, and uses the results of the proof of Theorem 2.2.5 to get the control of the 4th order moment in incomplete generation. See section 2.2.2 and 2.A.1 for more details. □

Remark 2.2.3. *If $f \in \mathcal{B}(S^3)$ is such that Pf^2 and Pf^4 exist and belong to F, with $Pf = 0$, then we have for all $r \in \mathbb{N}$ and for some positive constant c*

$$\mathbb{E}\Big[\left(\overline{M}_{\mathbb{G}_r}(f)\right)^4\Big] \leq \frac{c}{|\mathbb{G}_r|^2}. \tag{2.2.7}$$

Indeed, let $M_{\mathbb{G}_r}(f) = \sum_{i\in\mathbb{G}_r} f(\Delta_i)$. We have

$$\begin{aligned}\mathbb{E}\Big[\left(M_{\mathbb{G}_r}(f)\right)^4\Big] &= \mathbb{E}\Big[M_{\mathbb{G}_r}(f^4)\Big] + 6\mathbb{E}\left[\sum_{i\neq j\in\mathbb{G}_r} f^2(\Delta_i)f^2(\Delta_j)\right] \\ &\quad + 4\mathbb{E}\left[\sum_{i\neq j\in\mathbb{G}_r} f^3(\Delta_i)f(\Delta_j)\right] + 12\mathbb{E}\left[\sum_{i\neq j\neq k\in\mathbb{G}_r} f^2(\Delta_i)f(\Delta_j)f(\Delta_k)\right] \\ &\quad +24\mathbb{E}\left[\sum_{i\neq j\neq k\neq l\in\mathbb{G}_r} f(\Delta_i)f(\Delta_j)f(\Delta_k)f(\Delta_l)\right] \\ &= \mathbb{E}\left[\sum_{i\in\mathbb{G}_r} Pf^4(X_i)\right] + 6\mathbb{E}\left[\sum_{i\neq j\in\mathbb{G}_r} Pf^2(X_i)Pf^2(X_j)\right],\end{aligned}$$

where the last equality was obtained after conditioning by $\mathcal{F}_r$ and using the fact that $Pf = 0$. Now, dividing by $|\mathbb{G}_r|^4$ leads us to

$$\begin{aligned}\mathbb{E}\Big[\left(\overline{M}_{\mathbb{G}_r}(f)\right)^4\Big] &= \tfrac{6}{|\mathbb{G}_r|^2}\mathbb{E}\left[\tfrac{1}{|\mathbb{G}_r|^2}\sum_{i\neq j\in\mathbb{G}_r} Pf^2(X_i)Pf^2(X_j)\right] \\ &\qquad + \tfrac{1}{|\mathbb{G}_r|^3}\mathbb{E}\left[\tfrac{1}{|\mathbb{G}_r|}\sum_{i\in\mathbb{G}_r} Pf^4(X_i)\right] \\ &\leq \tfrac{6}{|\mathbb{G}_r|^2}\mathbb{E}\left[\left(\overline{M}_{\mathbb{G}_r}(Pf^2)\right)^2\right] + \tfrac{1}{|\mathbb{G}_r|^3}\mathbb{E}\Big[\overline{M}_{\mathbb{G}_r}(Pf^4)\Big]\end{aligned}$$

and (2.2.7) then follows from the control of $\left(\mathbb{E}\left[\left(\overline{M}_{\mathbb{G}_r}(Pf^2)\right)^2\right]\right)_r$ and $\left(\mathbb{E}\Big[\overline{M}_{\mathbb{G}_r}(Pf^4)\Big]\right)_r$ (see [66]).

Remark 2.2.4. *From remark 2.2.3, we deduce that if $f \in \mathcal{B}(S^3)$ is such that Pf^2 and Pf^4 exist and belong to F, with $Pf = 0$, then we have for all $r \in \mathbb{N}$ and for some positive constant c*

$$\mathbb{E}\Big[\left(\overline{M}_{\mathbb{T}_r}(f)\right)^4\Big] \leq c\left(\frac{1}{4}\right)^{r+1}. \tag{2.2.8}$$

Indeed, from equality

$$\overline{M}_{\mathbb{T}_r}(f) = \sum_{q=0}^{r} \frac{|\mathbb{G}_q|}{|\mathbb{T}_r|} \overline{M}_{\mathbb{G}_q}(f),$$

we deduce that

$$\mathbb{E}\left[\left(\overline{M}_{\mathbb{T}_r}(f)\right)^4\right] \leq \left(\sum_{q=0}^{r} \frac{|\mathbb{G}_q|}{|\mathbb{T}_r|} \left\|\overline{M}_{\mathbb{G}_q}(f)\right\|_4\right)^4,$$

where $\|\cdot\|_4$ *stands for the* L^4*-norm. We then infer from (2.2.7) that*

$$\mathbb{E}\left[\left(\overline{M}_{\mathbb{T}_r}(f)\right)^4\right] \leq c \left(\sum_{q=0}^{r} \frac{(\sqrt{2})^q}{2^{r+1}}\right)^4$$

for some positive constant c. (2.2.8) then follows from the last inequality.

2.2.2 Strong law of large numbers on incomplete subtree

We now turn to prove the strong law of large numbers for $\overline{M}_n^{\Pi}(f)$, completing the work of Guyon [66], where the LLN was proved only for the two averages $\overline{M}_{\mathbb{T}_r}(f)$ and $\overline{M}_{\mathbb{G}_r}(f)$.

Theorem 2.2.5. *Let* F *satisfy (i)-(vi). Let* $f \in F$ *such that* $(\mu, f) = 0$*. We assume that hypothesis (**H1**) is fulfilled with* $\alpha \in \left(0, \frac{\sqrt[4]{8}}{2}\right)$*. Then* $\overline{M}_n^{\Pi}(f)$ *almost surely converges to 0 as* n *goes to* ∞*.*

Proof. From the decomposition

$$\overline{M}_n^{\Pi}(f) = \sum_{q=0}^{r_n-1} \frac{2^q}{n} \overline{M}_{\mathbb{G}_q}(f) + \frac{1}{n} \sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)}),$$

it is enough to check that

$$\sum_{n=1}^{\infty} \mathbb{E}\left[\left(\frac{1}{n} \sum_{i=2^{r_n}}^{n} f\left(X_{\Pi(i)}\right)\right)^4\right] < \infty.$$

Indeed, since $\overline{M}_{\mathbb{G}_q}(f)$ almost surely converges to 0 (Corollary 15 in [66]), we deduce that the first term of the right hand side of the previous decomposition almost surely converges to 0 (Lemma 13 in [66]). We have

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=2^{r_n}}^{n} f\left(X_{\Pi(i)}\right)\right)^4\right] = \frac{1}{n^4}\mathbb{E}\left[\sum_{i=2^{r_n}}^{n} f^4\left(X_{\Pi(i)}\right)\right] \\
+\frac{6}{n^4}\mathbb{E}\left[\sum_{i,j=2^{r_n};i\neq j}^{n} f^2\left(X_{\Pi(i)}\right) f^2\left(X_{\Pi(j)}\right)\right] \\
+\frac{4}{n^4}\mathbb{E}\left[\sum_{i,j=2^{r_n};i\neq j}^{n} f^3\left(X_{\Pi(i)}\right) f\left(X_{\Pi(j)}\right)\right] \\
+\frac{12}{n^4}\mathbb{E}\left[\sum_{i,j,k=2^{r_n};i\neq j\neq k}^{n} f^2(X_{\Pi(i)})f(X_{\Pi(j)})f(X_{\Pi(k)})\right] \\
+\frac{24}{n^4}\mathbb{E}\left[\sum_{i,j,k,l=2^{r_n};i\neq j\neq k\neq l}^{n} f\left(X_{\Pi(i)}\right) f\left(X_{\Pi(j)}\right) f\left(X_{\Pi(k)}\right) f\left(X_{\Pi(l)}\right)\right].
\end{aligned}
\tag{2.2.9}
$$

We will control each term appearing in the decomposition (2.2.9). For the first term in the right hand side of (2.2.9), using (ii), (v) and (vi) we have for some positive constant c

$$
\mathbb{E}\left[\sum_{i=2^{r_n}}^{n} f^4\left(X_{\Pi(i)}\right)\right] = (n-2^{r_n}+1)\nu Q^{r_n} f^4 \leq c(n-2^{r_n}+1),
$$

which implies that

$$
\frac{1}{n^4}\mathbb{E}\left[\sum_{i=2^{r_n}}^{n} f^4\left(X_{\Pi(i)}\right)\right] = O\Big(\frac{1}{n^3}\Big). \tag{2.2.10}
$$

Recall the following: for i, j, k and $l \in \{2^{r_n}, \cdots, n\}$

- if $i \neq j$, then $r_n \geq 1$. Independently on (X, Π), draw two independent indices I_{r_n} and J_{r_n} uniformly from $\mathbb{G}_{r_n}$. Then the law of $(\Pi(i), \Pi(j))$ is the conditional law of (I_{r_n}, J_{r_n}) given $\{I_{r_n} \neq J_{r_n}\}$;
- if $i \neq j \neq k$, then $r_n \geq 2$. Independently on (X, Π), draw three independent indices I_{r_n}, J_{r_n} and K_{r_n} uniformly from $\mathbb{G}_{r_n}$. Then the law of $(\Pi(i), \Pi(j), \Pi(k))$ is the conditional law of $(I_{r_n}, J_{r_n}, K_{r_n})$ given $\{I_{r_n} \neq J_{r_n} \neq K_{r_n}\}$;
- if $i \neq j \neq k \neq l$, then $r_n \geq 2$. Independently on (X, Π), draw four independent indices $I_{r_n}, J_{r_n}, K_{r_n}$ and L_{r_n} uniformly from $\mathbb{G}_{r_n}$. Then the law of $(\Pi(i), \Pi(j), \Pi(k)), \Pi(l))$ is the conditional law of $(I_{r_n}, J_{r_n}, K_{r_n}, L_{r_n})$ given $\{I_{r_n} \neq J_{r_n} \neq K_{r_n} \neq J_{r_n}\}$.

Now we have to control the second and third term of (2.2.9). We have to check that

$$\frac{1}{n^4}\mathbb{E}\left[\sum_{i,j=2^{r_n};i\neq j}^{n} f^2\left(X_{\Pi(i)}\right)f^2\left(X_{\Pi(j)}\right)\right]=O\Big(\frac{1}{n^2}\Big), \tag{2.2.11}$$

and

$$\frac{1}{n^4}\mathbb{E}\left[\sum_{i,j=2^{r_n};i\neq j}^{n} f^3\left(X_{\Pi(i)}\right)f\left(X_{\Pi(j)}\right)\right]=o\Big(\frac{1}{n^2}\Big). \tag{2.2.12}$$

Indeed from the previous reminder and (i)-(vi), we have for some positive constant c

$$\begin{aligned}\mathbb{E}&\left[\sum_{i,j=2^{r_n};i\neq j}^{n} f^2\left(X_{\Pi(i)}\right)f^2\left(X_{\Pi(j)}\right)\right]\\ &\qquad=\frac{(n-2^{r_n})(n-2^{r_n}+1)}{(1-2^{-r_n})}\sum_{p=0}^{r_n-1}2^{-p-1}\nu Q^pP\left(Q_{\otimes}^{r_n-p-1}f^2\right)\\ &\qquad\leq c(n-2^{r_n})(n-2^{r_n}+1),\end{aligned}$$

which implies (2.2.11). In the same way and using in addition hypothesis **(H1)** we obtain that

$$\begin{aligned}&\mathbb{E}\left[\sum_{i,j=2^{r_n};i\neq j}^{n} f^3\left(X_{\Pi(i)}\right)f\left(X_{\Pi(j)}\right)\right]\\ &=\frac{(n-2^{r_n})(n-2^{r_n}+1)}{(1-2^{-r_n})}\sum_{p=0}^{r_n-1}2^{-p-2}\nu Q^pP\left(Q^{r_n-p-1}f^3\otimes Q^{r_n-p-1}f\right.\\ &\qquad\qquad\left.+Q^{r_n-p-1}f\otimes Q^{r_n-p-1}f^3\right)\\ &\leq\begin{cases}c\,2^{-r_n}(n-2^{r_n})(n-2^{r_n}+1) & \text{if}\quad \alpha<\frac{1}{2}\\ c\,r_n2^{-r_n}(n-2^{r_n})(n-2^{r_n}+1) & \text{if}\quad \alpha=\frac{1}{2}\\ c\,\alpha^{r_n}(n-2^{r_n})(n-2^{r_n}+1) & \text{if}\quad \alpha>\frac{1}{2},\end{cases}\end{aligned}$$

which implies (2.2.12). Let us deal with the remaining term of (2.2.9)

$$\begin{aligned}\frac{1}{n^4}\mathbb{E}&\left[\sum_{i,j,k=2^{r_n};i\neq j\neq k}^{n} f^2(X_{\Pi(i)})f(X_{\Pi(j)})f(X_{\Pi(k)})\right]\\ &=\frac{(n-2^{r_n}-1)(n-2^{r_n})(n-2^{r_n}+1)}{\mathbb{P}(I_{r_n}\neq J_{r_n}\neq K_{r_n})\times n^4}\\ &\qquad\times\mathbb{E}\Big[f^2\left(X_{I_{r_n}}\right)f\left(X_{J_{r_n}}\right)f\left(X_{K_{r_n}}\right)\mathbf{1}_{\{I_{r_n}\neq J_{r_n}\neq K_{r_n}\}}\Big].\end{aligned}$$

Then, we get an explicit expression for the last expectation similar to that obtained in part **(d)** of the calculus of $\mathbb{E}\Big[(\overline{M}_{\mathbb{G}_r}(f))^4\Big]$ with a slight modification of the functions. Calculating the rate of this expression, we obtain

$$\sum_{n=4}^{\infty}\frac{1}{n^4}\mathbb{E}\left[\sum_{i,j,k=2^{r_n};i\neq j\neq k}^{n} f^2(X_{\Pi(i)})f(X_{\Pi(j)})f(X_{\Pi(k)})\right]$$
$$\leq c\sum_{n=1}^{\infty}\frac{1}{n}\alpha^{2r_n}+c\sum_{n=1}^{\infty}\sum_{p=2}^{r_n-1}\sum_{l=0}^{p-1}\frac{1}{n}\frac{1}{2^p}\frac{1}{2^{l+1}}\alpha^{2r_n-2p}$$
$$+c\sum_{n=1}^{\infty}\sum_{p=2}^{r_n-1}\sum_{l=0}^{p-1}\frac{1}{n}\frac{1}{2^p}\frac{1}{2^{l+1}}\alpha^{2r_n-p-l},$$

for some positive c. Now it is not hard to see that the right hand side is finite.

Finally, to check that the series of general term

$$\frac{1}{n^4}\mathbb{E}\left[\sum_{i,j,k,l=2^{r_n};i\neq j\neq k\neq l}^{n} f\left(X_{\Pi(i)}\right)f\left(X_{\Pi(j)}\right)f\left(X_{\Pi(k)}\right)f\left(X_{\Pi(l)}\right)\right]$$

is finite, it is enough, according to the calculation of rates we have done in **Part 2** of the proof of Theorem 2.2.1, to check that $\sum_{n=1}^{\infty}\alpha^{4r_n}<\infty$, which is the case if $\alpha\in\left(0,\frac{\sqrt[4]{8}}{2}\right)$ and this ends the proof of Theorem 2.2.5. □

Remark 2.2.6. *Note that this theorem can be improved, but the price to pay is enormous computations related to the calculation of higher moments. If f is bounded, this result is true for every $\alpha\in(0,1)$ as we will see in section 2.3.*

2.2.3 Law of the iterated logarithm (LIL)

Using the LIL for martingales (see Theorem 2.B.3 of Stout in the Appendix 2.B), we are going to prove a LIL for the BMC. This will be done when f depends on the mother-daughters triangle (Δ_i). We use the notations $M_n^{\Pi}(f)=\sum_{i=1}^{n}f(\Delta_{\Pi(i)})$ and $M_{\mathbb{T}_r}(f)=\sum_{i\in\mathbb{T}_r}f(\Delta_i)$.

Theorem 2.2.7. *Let F satisfy (i)-(vi). Let $f\in\mathcal{B}(\mathcal{S}^3)$ such that $Pf=0$, Pf^2 and Pf^4 exist and belong to F. We assume that hypothesis **(H1)** is fulfilled. Then*

$$\limsup_{n\to\infty}\frac{M_n^{\Pi}(f)}{\sqrt{2\langle M^{\Pi}(f)\rangle_n\log\log\langle M^{\Pi}(f)\rangle_n}}=1\quad a.s.$$

And in particular

$$\limsup_{r\to\infty}\frac{M_{\mathbb{T}_r}(f)}{\sqrt{2|\mathbb{T}_r|\log\log|\mathbb{T}_r|}}=\sqrt{(\mu,Pf^2)}\quad a.s.$$

Proof. We will check the hypothesis of Stout Theorem's 2.B.3. Let $f \in \mathcal{B}\big(\mathcal{S}^3\big)$. We introduce the following filtration $(\mathcal{H}_n)_{n\geq 0}$ defined by $\mathcal{H}_0 = \sigma(X_1)$ and $\mathcal{H}_n = \sigma\big(\Delta_{\Pi(i)}, \Pi(i{+}1), 1 \leq i \leq n\big)$. Let $\big(M_n^{\Pi}(f)\big)_{n\geq 0}$ defined by $M_0^{\Pi}(f) = 0$ and $M_n^{\Pi}(f) = \sum_{i=1}^{n} f(\Delta_{\Pi(i)})$. Then since $Pf = 0$, $\big(M_n^{\Pi}(f)\big)$ is a $\mathcal{H}_n$-martingale with $\mathbb{E}\Big[M_1^{\Pi}(f)\Big] = 0$. The bracket of the above martingale is given by

$$\langle M^{\Pi}(f)\rangle_n = \sum_{i=0}^{n} Pf^2(X_{\Pi(i)}) = M_n^{\Pi}(Pf^2).$$

We have the following decomposition

$$\frac{\langle M^{\Pi}(f)\rangle_n}{n} = \overline{M}_n^{\Pi}(Pf^2) = \sum_{q=0}^{r_n-1} \frac{2^q}{n}\overline{M}_{\mathbb{G}_q}(Pf^2) + \frac{1}{n}\sum_{i=2^{r_n}}^{n} Pf^2(X_{\Pi(i)}).$$

Since

$$\forall q \leq r_n - 1,\ \frac{2^q}{2^{r_n+1}} \leq \frac{2^q}{n} \leq \frac{2^q}{2^{r_n}} \qquad \text{and} \qquad \frac{1}{n}\sum_{i=2^{r_n}}^{n} Pf^2(X_{\Pi(i)}) \leq \overline{M}_{\mathbb{G}_{r_n}}(Pf^2),$$

we deduce that

$$\sum_{q=0}^{r_n-1} \frac{2^q}{2^{r_n+1}}\overline{M}_{\mathbb{G}_q}(Pf^2) \leq \overline{M}_n^{\Pi}(Pf^2) \leq \sum_{q=0}^{r_n} \frac{2^q}{2^{r_n}}\overline{M}_{\mathbb{G}_q}(Pf^2).$$

From the strong law of large numbers of $\overline{M}_{\mathbb{G}_q}(Pf^2)$ (see [66], Corollary 15) and from Lemma 5.2 of [32], we infer that

$$\sum_{q=0}^{r_n-1} \frac{2^q}{2^{r_n+1}}\overline{M}_{\mathbb{G}_q}(Pf^2) \xrightarrow{a.s} \frac{(\mu, Pf^2)}{2} \qquad \text{and} \qquad \sum_{q=0}^{r_n} \frac{2^q}{2^{r_n}}\overline{M}_{\mathbb{G}_q}(Pf^2) \xrightarrow{a.s} 2(\mu, Pf^2).$$

Using these results, we thus deduce that $\langle M^{\Pi}(f)\rangle_n = O(n)$ and $n = O\Big(\langle M^{\Pi}(f)\rangle_n\Big)$ a.s.. This implies in particular that $\langle M^{\Pi}(f)\rangle_n \underset{n\to\infty}{\longrightarrow} \infty$ a.s.

Now let $K_n = \frac{\sqrt{2}}{\sqrt{\log\log(n)}}$ in Theorem 2.B.3, we have

$$\begin{aligned} R := \sum_{n=1}^{\infty} \frac{2\log\log\langle M^{\Pi}(f)\rangle_n}{K_n^2\langle M^{\Pi}(f)\rangle_n}\mathbb{E}\left[f^2(\Delta_{\Pi(n)})\mathbf{1}_{\left\{f^2(\Delta_{\Pi(n)}) > \frac{K_n^2\langle M^{\Pi}(f)\rangle_n}{2\log\log\langle M^{\Pi}(f)\rangle_n}\right\}}\Big|\mathcal{H}_{n-1}\right] \\ \leq \sum_{n=1}^{\infty} \frac{4(\log\log\langle M^{\Pi}(f)\rangle_n)^2}{K_n^4(\langle M^{\Pi}(f)\rangle_n)^2} Pf^4(X_{\Pi(n)}) \quad a.s., \end{aligned}$$

since $\langle M^{\Pi}(f)\rangle_n = O(n)$ a.s. so that for $R < \infty$ a.s., it is enough to check that

$$\sum_{n=1}^{\infty} \frac{Pf^4(X_{\Pi(n)})}{n^{\delta}} < \infty \quad \text{a.s.} \qquad \text{with any } 1 < \delta < 2. \tag{2.2.13}$$

Now, according to (v) and (vi), there exists a positive constant c such that for all $n \geq 1$, $\mathbb{E}\Big[Pf^4\left(X_{\Pi(n)}\right)\Big] = \nu Q^{r_n} Pf^4 \leq c$, and (2.2.13) follows. Applying Theorem 2.B.3, we have

$$\limsup_{n\to\infty} \frac{M_n^{\Pi}(f)}{\sqrt{2\langle M^{\Pi}(f)\rangle_n \log\log\langle M^{\Pi}(f)\rangle_n}} = 1 \quad \text{a.s.}$$

Now, for $n = |\mathbb{T}_r|$ we have the following

$$\frac{M_{\mathbb{T}_r}(f)}{\sqrt{2\langle M^{\Pi}(f)\rangle_{|\mathbb{T}_r|}\log\log\langle M^{\Pi}(f)\rangle_{|\mathbb{T}_r|}}} = \sqrt{\frac{|\mathbb{T}_r|\frac{\langle M^{\Pi}(f)\rangle_{|\mathbb{T}_r|}}{|\mathbb{T}_r|}}{2\log\log\langle M^{\Pi}(f)\rangle_{|\mathbb{T}_r|}}} \times \frac{M_{\mathbb{T}_r}(f)}{|\mathbb{T}_r|\frac{\langle M^{\Pi}(f)\rangle_{|\mathbb{T}_r|}}{|\mathbb{T}_r|}}$$

and since $\frac{\langle M^{\Pi}(f)\rangle_{|\mathbb{T}_r|}}{|\mathbb{T}_r|} = \overline{M}_{\mathbb{T}_r}(Pf^2) \underset{r\to\infty}{\longrightarrow} (\mu, Pf^2)$ a.s. (see Theorem 18 in [66]) , we get

$$\limsup_{r\to\infty} \frac{M_{\mathbb{T}_r}(f)}{\sqrt{2|\mathbb{T}_r|\log\log|\mathbb{T}_r|}} = \sqrt{(\mu, Pf^2)} \quad \text{a.s},$$

which ends the proof. □

Remark 2.2.8. *Let us note that using Theorem 2.2.5, we can prove that if hypothesis **(H1)** is fulfilled with* $\alpha \in \left(0, \frac{\sqrt[4]{8}}{2}\right)$ *then,*

$$\limsup_{n\to\infty} \frac{M_n^{\Pi}(f)}{\sqrt{2n\log\log n}} = \sqrt{(\mu, Pf^2)} \quad a.s.,$$

and via the computation of 2k-th order moments of $\overline{M}_{\mathbb{G}_r}(g)$*, with* $k > 2$ *and* $g \in \mathcal{B}(S)$*, it is possible to prove the latter for all* $\alpha \in (0,1)$*. But, as already emphasized, this comes at the price of enormous computations.*

2.2.4 Almost-sure functional central limit theorem (ASFCLT)

We are now going to prove an ASFCLT theorem for the BMC $(X_n, n \in \mathbb{T})$. Here again, this will be done when f depends on the mother-daughters triangle by using the ASFCLT for discrete time martingale. We refer to Theorem 2.B.4 of Chaabane in the Appendix 2.B for the definition of an ASFCLT.

Theorem 2.2.9. *Let F satisfy (i)-(vi). Let* $f \in \mathcal{B}(\mathcal{S}^3)$ *such that* $Pf = 0$*,* Pf^2 *and* Pf^4 *exist and belong to F. We assume that hypothesis **(H1)** is fulfilled with* $\alpha \in \left(0, \frac{\sqrt[4]{8}}{2}\right)$*. Then* $M_n^{\Pi}(f)$ *verify an ASFCLT, when n goes to* ∞*.*

Proof. We use Theorem 2.B.4. Let $(\mathcal{H}_n)_{n\in\mathbb{N}}$ be the filtration defined as in subsection 2.2.3. Then $(M_n^{\Pi}(f))$ is a $\mathcal{H}_n$ martingale. We have to check the hypotheses of Theorem 2.B.4. For all $n \geq 1$, let $V_n = s\sqrt{n}$ where $s^2 = (\mu, Pf^2)$. Then according to Theorem 2.2.5

$$\frac{\langle M^{\Pi}(f)\rangle_n}{V_n^2} = V_n^{-2} M_n^{\Pi}(Pf^2) \underset{n\to\infty}{\longrightarrow} 1 \quad \text{a.s.}$$

Let $\varepsilon > 0$. We have

$$\sum_{n\geq 1} \frac{1}{V_n^2}\mathbb{E}\left[f^2(\Delta_{\Pi(n)})\mathbf{1}_{\left\{\left|f\left(\Delta_{\Pi(n)}\right)\right|>\varepsilon V_n\right\}}\middle|\mathcal{H}_{n-1}\right] \leq \frac{1}{\varepsilon^2 s^4}\sum_{n\geq 1}\frac{Pf^4\left(X_{\Pi(n)}\right)}{n^2} \quad \text{a.s.}$$

According to (v) and (vi), there exists a positive constant c such that for all $n \geq 1$, $\mathbb{E}\Big[Pf^4\left(X_{\Pi(n)}\right)\Big] = \nu Q^{r_n} Pf^4 \leq c$, and therefore, $\forall \varepsilon > 0$

$$\sum_{n\geq 1} \frac{1}{V_n^2}\mathbb{E}\left[f^2(\Delta_{\Pi(n)})\mathbf{1}_{\left\{\left|f\left(\Delta_{\Pi(n)}\right)\right|>\varepsilon V_n\right\}}\middle|\mathcal{H}_{n-1}\right] < \infty \quad \text{a.s.}$$

Finally, we have

$$\sum_{n\geq 1} \frac{1}{V_n^4}\mathbb{E}\left[f^4\left(\Delta_{\Pi(n)}\right)\mathbf{1}_{\left\{\left|f\left(\Delta_{\Pi(n)}\right)\right|\leq V_n\right\}}\middle|\mathcal{H}_{n-1}\right] \leq \frac{1}{s^4}\sum_{n\geq 1}\frac{Pf^4\left(X_{\Pi(n)}\right)}{n^2}, \quad \text{a.s.}$$

which as before is a.s. finite and the proof is then complete. ☐

Remark 2.2.10. *As before, let us note that this result can be extended to the general case $\alpha \in (0,1)$, but at the price of enormous computation related to the computation of $2k$-order moments, $k > 2$, for $\overline{M}_{\mathbb{G}_r}(g)$, $g \in \mathcal{B}(S)$.*

2.2.5 Deviation inequalities for BMC

We are now going to give some deviation inequalities under $(i)-(vi)$ and **(H1)** for the empirical means (2.1.2) when $f \in \mathcal{B}(S)$ with $(\mu, f) = 0$ and when $f \in \mathcal{B}(S^3)$ with $(\mu, Pf) = 0$. This will help us in the sequel to obtain a MDP result in a general framework, that is for functional of BMC with unbounded test functions. Let us recall that the main disadvantage of this "weak" set of assumptions is that the range of speed for the MDP is very restricted. However, we still work under geometric ergodicity assumption and general test function, which will not be the case when we would want to extend the MDP (see Section 2.3). Note that we postpone to the Appendix A nearly all the proofs of this section, these proofs being quite long and technical.

Theorem 2.2.11. *Let F satisfy conditions (i)-(vi). We assume that **(H1)** is fulfilled. Let $f \in F$ such that $(\mu, f) = 0$. Then we have for all $\delta > 0$ and all $r \in \mathbb{N}$ and all $n \in \mathbb{N}$*

$$\mathbb{P}\Big(\big|\overline{M}_{\mathbb{G}_r}(f)\big| > \delta\Big) \leq \begin{cases} \frac{c}{\delta^2}\left(\frac{1}{2}\right)^r & \text{if} \quad \alpha^2 < \frac{1}{2}; \\ \frac{c}{\delta^2} r \left(\frac{1}{2}\right)^r & \text{if} \quad \alpha^2 = \frac{1}{2}; \\ \frac{c}{\delta^2}\alpha^{2r} & \text{if} \quad \alpha^2 > \frac{1}{2}; \end{cases} \tag{2.2.14}$$

$$\mathbb{P}\Big(\big|\overline{M}_n^{\Pi}(f)\big| > \delta\Big) \leq \begin{cases} \frac{c}{\delta^2}\left(\frac{1}{2}\right)^{r_n+1} & \text{if} \quad \alpha^2 < \frac{1}{2}; \\ \frac{c}{\delta^2} r_n \left(\frac{1}{2}\right)^{r_n+1} & \text{if} \quad \alpha^2 = \frac{1}{2}; \\ \frac{c}{\delta^2}\alpha^{2(r_n+1)} & \text{if} \quad \alpha^2 > \frac{1}{2}; \end{cases} \tag{2.2.15}$$

and

$$\mathbb{P}\Big(\big|\overline{M}_{\mathbb{T}_r}(f)\big| > \delta\Big) \leq \begin{cases} \frac{c}{\delta^2}\left(\frac{1}{2}\right)^{r+1} & \text{if} \quad \alpha^2 < \frac{1}{2}; \\ \frac{c}{\delta^2} r \left(\frac{1}{2}\right)^{r+1} & \text{if} \quad \alpha^2 = \frac{1}{2}; \\ \frac{c}{\delta^2}\alpha^{2(r+1)} & \text{if} \quad \alpha^2 > \frac{1}{2}; \end{cases} \tag{2.2.16}$$

where the positive constant c depends on f and α and may differ term by term.

Proof. See section 2.A.1 in the Appendix 2.A. □

We shall also need an extension of Theorem 2.2.11 to the case when f does not only depend on an individual X_i, but on the mother-daughters triangle (Δ_i).

Theorem 2.2.12. *Let F satisfy conditions (i)-(vi). We assume that **(H1)** is fulfilled. Let $f \in \mathcal{B}(S^3)$ such that Pf and Pf^2 exists and belong to F and $(\mu, Pf) = 0$. Then we have the same conclusion as the Theorem 2.2.11 for the three empirical averages given in (2.1.2): $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_{\mathbb{T}_r}(f)$ and $\overline{M}_n^{\Pi}(f)$.*

Proof. See section 2.A.2 in the Appendix 2.A. □

We thus have the following first result on the super-exponential convergence in probability, whose definition we present now

Definition 2.2.13. *Let (E, d) a metric space. Let (Z_n) be a sequence of random variables valued in E, Z be a random variable valued in E and (v_n) be a rate. We say that Z_n converges v_n-superexponentially fast in probability to Z if for all $\delta > 0$*

$$\limsup_{n\to\infty} \frac{1}{v_n} \log \mathbb{P}(d(Z_n, Z) > \delta) = -\infty.$$

This "exponential convergence" with speed v_n will be shortened as

$$Z_n \overset{\text{superexp}}{\underset{v_n}{\Longrightarrow}} Z.$$

We may now set

Proposition 2.2.14. *Let F satisfy conditions (i)-(vi).Let $f \in \mathcal{B}\big(S^3\big)$ such that Pf and Pf^2 exists and belong to F and $(\mu, Pf) = 0$. We assume that **(H1)** is fulfilled. Let (b_n) be a sequence of increasing positive real numbers such that*

$$\frac{b_n}{\sqrt{n}} \longrightarrow +\infty, \qquad \frac{b_n}{\sqrt{n \log n}} \longrightarrow 0, \qquad \frac{n}{b_n} \text{ is non-decreasing.} \tag{2.2.17}$$

Then

$$\overline{M}_n^{\Pi}(f) \overset{\text{superexp}}{\underset{\frac{b_n^2}{n}}{\Longrightarrow}} 0.$$

Proof. The proof is a direct consequence of Theorem 2.2.12. □

2.2.6 Moderate deviations for BMC

Now, using the MDP for martingale (see e.g [38], [113]), we are going to prove a MDP for BMC. We will use Proposition 2.B.5, in the Appendix 2.B.

Theorem 2.2.15. *Let F satisfy conditions (i)-(vi). We assume that **H1** is satisfied. Let $f \in \mathcal{B}\big(S^3\big)$ such that Pf^2 and Pf^4 exist and belong to F. Assume that $Pf = 0$. Let (b_n) be a sequence of increasing positive real numbers satisfying (2.2.17). If*

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log\left(n \operatorname*{ess\,sup}_{1 \le k \le c^{-1}(b_{n+1})} \mathbb{P}\Big(\big| f\left(\Delta_{\Pi(k)}\right) \big| > b_n \big| \mathcal{H}_{k-1} \Big) \right) = -\infty, \tag{2.2.18}$$

where $c^{-1}(b_{n+1}) := \inf\left\{k \in \mathbb{N} : \frac{k}{b_k} \ge b_{n+1}\right\}$; then $\Big(M_n^{\Pi}(f)/b_n\Big)$ satisfies a MDP in $\mathbb{R}$ with the speed b_n^2/n and the rate function $I(x) = \dfrac{x^2}{2(\mu, Pf^2)}$.

Proof. Firstly, note that under the hypothesis $M_n^{\Pi}(f)$ is a $\mathcal{H}_n$-martingale, with $\mathcal{H}_0 = \sigma(X_1)$ and $\mathcal{H}_n = \sigma\big(\Delta_{\Pi(i)}, \Pi(i+1), 1 \le i \le n\big)$. From Proposition 2.B.5 in the Appendix 2.B, we only have to check conditions (**C1**) and (**C3**).

On the one hand, (2.2.15) applied to $Pf^4 - (\mu, Pf^4)$ implies that for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^{n} Pf^4(X_{\Pi(i)}) > (\mu, Pf^4) + \delta \right) = -\infty$$

and this implies the exponential Lindeberg condition (see for e.g [113]), that is condition (**C3**).

On the other hand, we have $\langle M^{\Pi}(f)\rangle_n = M_n^{\Pi}(Pf^2)$ and (2.2.15) applied to $Pf^2 - (\mu, Pf^2)$ implies that

$$\overline{M}_n^{\Pi}\left(Pf^2 - (\mu, Pf^2)\right) \overset{\text{superexp}}{\underset{b_n^2/n}{\Longrightarrow}} 0,$$

that is condition (**C1**). □

Remark 2.2.16. *One of the main difficulties in the application of this Theorem lies in the verification of (2.2.18). Note however that in the range of speed considered it is sufficient to have some uniform control in X_i of some moment of $f(X_i, X_{2i}, X_{2i+1})$ conditionally on X_i, which leads to condition of the type $P|f|^k$ bounded for some $k \geq 2$. It is of course the case if f is bounded.*

Remark 2.2.17. *In the special case of model (2.1.1), we have (see Section 2.4), for f such that $Pf = 0$ and for all k*

$$\mathbb{E}\left[\exp\left(\lambda\frac{b_n}{n}f\left(\Delta_{\Pi(k)}\right)\right)\bigg|\mathcal{H}_{k-1}\right] = \exp\left(\frac{b_n^2}{n}\left(\frac{\lambda^2 Pf^2}{2n}\right)\left(X_{\Pi(k)}\right)\right).$$

This condition implies that a MDP is satisfied for $(M_n^{\Pi}(f)/b_n)$. Indeed, if this relation is satisfied, we then have that for $\lambda \in \mathbb{R}$ the quantity

$$G_n(\lambda) = \frac{\lambda^2}{2n}\sum_{k=1}^{n} Pf^2(X_{\Pi(k)}) = \frac{\lambda^2}{2}\overline{M}_n^{\Pi}(Pf^2)$$

is an upper and lower cumulant (see e.g [113]), and we may apply Gärtner-Ellis type methodology. In addition, due to (2.2.15) applied to $Pf^2 - (\mu, Pf^2)$, we have for $\lambda \in \mathbb{R}$

$$G_n(\lambda) \overset{\text{superexp}}{\underset{b_n^2/n}{\Longrightarrow}} \frac{\lambda^2(\mu, Pf^2)}{2},$$

which implies that $\left(M_n^{\Pi}(f)/b_n\right)$ satisfies a MDP in $\mathbb{R}$ with the speed b_n^2/n and the rate function $I(x) = \dfrac{x^2}{2(\mu, Pf^2)}$.

2.3 Exponential deviation inequalities for BMC and consequences

We give here stronger deviation inequalities than the one obtained in the previous section, namely exponential deviation inequalities. Of course, it requires more stringent assumptions.

2.3.1 Exponential deviation inequalities

Let us consider the following hypothesis.

(H2) There exists a probability μ on $(S, \mathcal{S})$ such that, for all $f \in \mathcal{B}_b(S)$ with $(\mu, f) = 0$, there exists a positive constant c such that

$$|Q^r f(x)| \leq c\alpha^r \quad \text{for some } \alpha \in (0,1) \text{ and for all } x \in S.$$

One can easily check that, under hypothesis **(H2)**, $\mathcal{B}_b(S)$ fulfills hypothesis (i)-(vi) of the previous section.

Under this assumption, we will prove exponential deviation inequalities for $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_{\mathbb{T}_r}(f)$ and $\overline{M}_n^{\Pi}(f)$ when $f \in \mathcal{B}_b(S)$ with $(\mu, f) = 0$ (resp. $f \in \mathcal{B}_b\big(S^3\big)$ with $(\mu, Pf) = 0$).

Theorem 2.3.1. *Let $f \in \mathcal{B}_b(S)$ such that $(\mu, f) = 0$. Assume that* **(H2)** *is satisfied. Then we have for all $\delta > 0$*

$$\mathbb{P}\Big(\overline{M}_{\mathbb{G}_r}(f) > \delta\Big) \leq \begin{cases} \exp\left(c''\delta\right)\exp\left(-c'\delta^2|\mathbb{G}_r|\right), \quad \forall r \in \mathbb{N}, & \textit{if } \alpha \leq \frac{1}{2}, \\ \exp\left(-c'\delta^2|\mathbb{G}_r|\right), \forall r \in \mathbb{N} \textit{ such that } r > r_0, & \textit{if } \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}, \\ \exp\left(-c'\delta^2\frac{|\mathbb{G}_r|}{r}\right), \forall r \in \mathbb{N} \textit{ such that } r > r_0, & \textit{if } \alpha^2 = \frac{1}{2}, \\ \exp\left(-c'\delta^2\frac{1}{\alpha^{2r}}\right), \forall r \in \mathbb{N} \textit{ such that } r > r_0, & \textit{if } \alpha^2 > \frac{1}{2}, \end{cases} \tag{2.3.1}$$

$$\mathbb{P}\Big(\overline{M}_{\mathbb{T}_r}(f) > \delta\Big) \leq \begin{cases} \exp\left(c''\delta\right)\exp\left(-c'\delta^2|\mathbb{T}_r|\right), \quad \forall r \in \mathbb{N}, & \textit{if } \alpha < \frac{1}{2}, \\ \exp\left(2c'\delta(r+1)\right)\exp\left(-c'\delta^2|\mathbb{T}_r|\right), \quad \forall r \in \mathbb{N}, & \textit{if } \alpha = \frac{1}{2}, \\ \exp\left(-c'\delta^2|\mathbb{T}_r|\right), \forall r \in \mathbb{N} \textit{ such that } r > r_0 - 1, & \textit{if } \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}, \\ \exp\left(-c'\delta^2\frac{|\mathbb{T}_r|}{r+1}\right), \forall r \in \mathbb{N} \textit{ such that } r > r_0 - 1, & \textit{if } \alpha = \frac{\sqrt{2}}{2}, \\ \exp\left(-c'\delta^2\frac{1}{\alpha^{2(r+1)}}\right), \forall r \in \mathbb{N}^* \textit{ such that } r > r_0 - 3, & \textit{if } \alpha > \frac{\sqrt{2}}{2}, \end{cases} \tag{2.3.2}$$

and

$$\mathbb{P}\Big(\overline{M}^{\Pi}_n(f) > \delta\Big) \leq$$

$$\begin{cases} \exp\left(c''\delta\right)\exp\left(-c'\delta^2 n\right), & \forall n \in \mathbb{N}, \quad \text{if } \alpha < \frac{1}{2}, \\ \exp\left(2c'\delta(r_n+1)\right)\exp\left(-c'\delta^2 n\right), & \forall n \in \mathbb{N}, \quad \text{if } \alpha = \frac{1}{2}, \\ \exp\left(-c'\delta^2 n\right), \forall n \in \mathbb{N} \text{ such that } r_n > r_0, \ \text{if } \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}, & \\ \exp\left(-c'\delta^2 \frac{n}{r_n+1}\right), \forall n \in \mathbb{N} \text{ such that } r_n > r_0, \ \text{if } \alpha = \frac{\sqrt{2}}{2}, & \\ \exp\left(-c'\delta^2 \frac{1}{\alpha^{2(r_n+1)}}\right), \forall n \in \mathbb{N}^* \text{ such that } r_n > r_0 - 2, \text{if } \alpha > \frac{\sqrt{2}}{2}, & \end{cases} \tag{2.3.3}$$

where $r_0 := \log\left(\frac{\delta}{c_0}\right) / \log(\alpha)$, *and* c_0, c' *and* c'' *are positive constants which depend on* α *and* f *and differ line by line (see the proofs for the dependence).*

Proof. The details of the proof are in Section 2.A.3 in the Appendix 2.A. It relies mainly on successive conditioning, using carefully the uniform geometric ergodicity assumption to get rid of the conditioning. □

The condition about α less than 1/2 or greater is of course linked to the binary structure of the tree. The extension to $m-$ary tree will follow from the same ideas.

Theorem 2.3.2. *Let* $f \in \mathcal{B}_b\big(S^3\big)$ *such that* $(\mu, Pf) = 0$. *Assume that* **(H2)** *is satisfied. Then we have the same conclusions, for the three empirical averages* $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}^{\Pi}_n(f)$ *and* $\overline{M}_{\mathbb{T}_r}(f)$, *as in the Theorem 2.3.1.*

Proof. See section 2.A.4 in the Appendix 2.A. □

Now, using Borel Cantelli Theorem and (2.3.3), we state easily the following

Corollary 2.3.3. *Let* $f \in \mathcal{B}_b(S)$ *such that* $(\mu, f) = 0$ *(resp.* $f \in \mathcal{B}_b(S^3)$ *and* $(\mu, Pf) = 0$*). Assume that* **(H2)** *is satisfied. Then* $\overline{M}^{\Pi}_n(f)$ *almost surely converges to* 0 *as* n *goes to* ∞.

Remark 2.3.4. *Of course uniform ergodicity, and bounded test functions are surely a very strong set of assumptions but it is not so difficult to verify if the Markov chains daughters lie in a compact set. We are convinced that it is possible to consider the geometric ergodic case and bounded test functions but to the price of tedious calculations that we will pursue in an other work. We will also investigate the use of transportation inequalities, leading to deviation inequality for Lipschitz test functions under some Wasserstein contraction property for the kernel* P, *in the* *spirit of the Theorems 2.5 or 2.11 in [41] (that is $W_1^d(P(x,\cdot), P(\tilde{x},\cdot)) \leq rd(x,\tilde{x})$ for every x, $\tilde{x}$ in S, for some $r < 1$ and d a metric on S). Given two probability measures ν_1 and ν_2 on $(S,\mathcal{S})$, we recall that*

$$W_1^d(\nu_1,\nu_2) = \sup_{f:\|f\|_{Lip}\leq 1} \left| \int f d\nu_1 - \int f d\nu_2 \right| \quad \textit{and} \quad \|f\|_{Lip} = \sup_{x\neq y} \frac{|f(x)-f(y)|}{d(x,y)}.$$

2.3.2 Moderate deviation principle for BMC

We introduce the following assumption on the speed of the MDP.

Assumption 1. *Let (b_n) be an increasing sequence of positive real numbers such that*

$$\frac{b_n}{\sqrt{n}} \longrightarrow +\infty,$$

and

- *if $\alpha^2 < \frac{1}{2}$, the sequence (b_n) is such that $b_n/n \longrightarrow 0$,*
- *if $\alpha^2 = \frac{1}{2}$, the sequence (b_n) is such that $(b_n \log n)/n \longrightarrow 0$,*
- *if $\alpha^2 > \frac{1}{2}$, the sequence (b_n) is such that $(b_n \alpha^{r_n+1})/\sqrt{n} \longrightarrow 0$.*

Using the MDP for martingale with bounded jumps (see e.g [34], [38]), we can now state the following

Theorem 2.3.5. *Let $f \in \mathcal{B}_b(S^3)$ such that $Pf = 0$. Assume that* **(H2)** *is satisfied. Let (b_n) be a sequence of real numbers satisfying the Assumption 1, then $\left(M_n^{\Pi}(f)/b_n\right)$ satisfies a MDP in S with the speed b_n^2/n and rate function $I(x) = \dfrac{x^2}{2(\mu, Pf^2)}$.*

Proof. The proof easily follows from the previous exponential probability inequalities and the MDP for martingale with bounded jumps (see e.g [34], [38], [113]). □

Remark 2.3.6. *Taking particularly $n = |\mathbb{T}_r|$, and (b_n) be a sequence of real numbers satisfying the Assumption 1, we get that for all $f \in \mathcal{B}_b(S^3)$, $\left(M_{\mathbb{T}_r}(f)/b_{|\mathbb{T}_r|}\right)$ satisfies a MDP in $\mathbb{R}$ with the speed $b_{|\mathbb{T}_r|}^2/|\mathbb{T}_r|$ and the rate function $I(x) = \frac{x^2}{2(\mu, Pf^2)}$.*

2.4 Application: First order Bifurcating autoregressive processes

In this section, we seek to apply the results of the previous sections to the following bifurcating autoregressive process with memory 1 defined by

$$\mathcal{L}(X_1) = \nu, \qquad \text{and} \qquad \forall n \geq 1, \quad \begin{cases} X_{2n} = \alpha_0 X_n + \beta_0 + \varepsilon_{2n} \\ \\ X_{2n+1} = \alpha_1 X_n + \beta_1 + \varepsilon_{2n+1}, \end{cases} \tag{2.4.1}$$

where $\alpha_0, \alpha_1 \in (-1, 1)$; $\beta_0, \beta_1 \in \mathbb{R}$, $\Big((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1\Big)$ forms a sequence of i.i.d. bivariate random variables and ν a probability measure on $\mathbb{R}$.

Several extensions of the model have been proposed and various estimators are studied in the literature for the unknown parameters, see for instance [13],[10], [11], [12], [14], [15]. See [18] for relevant references.

In all this section, we assume that the distribution ν has finite moments of all orders.

In the sequel, we will study (2.4.1) in two settings:

- the Gaussian setting which corresponds to the case where

  $$\Big((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1\Big)$$

  forms a sequence of i.i.d bivariate random variables with law $\mathcal{N}_2(0, \Gamma)$ with

  $$\Gamma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \sigma^2 > 0, \quad \rho \in (-1, 1); \tag{2.4.2}$$

- the bounded setting which corresponds to the case where X_1 and

  $$\Big((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1\Big),$$

  which forms a sequence of centered i.i.d. bivariate random variables, valued in a compact set. Let us note that in this case, $(X_n, n \in \mathbb{T})$ takes its values in a compact set.

Our main goal is to give deviation inequalities and MDP for the estimator of the 4-dimensional unknown parameter $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$ and for the statistical test defined in [66].

To estimate the 4-parameter $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$, as well as σ^2 and ρ, assume we observe a complete subtree $\mathbb{T}_{r+1}$. The least squares estimator $\widehat{\theta}^r = \big(\widehat{\alpha}_0^r, \widehat{\beta}_0^r, \widehat{\alpha}_1^r, \widehat{\beta}_1^r\big)$

of θ is given by (see [66]), for $\eta \in \{0,1\}$

$$\begin{cases} \widehat{\alpha}^r_\eta = \dfrac{|\mathbb{T}_r|^{-1} \sum_{i\in\mathbb{T}_r} X_i X_{2i+\eta} - \left(|\mathbb{T}_r|^{-1} \sum_{i\in\mathbb{T}_r} X_i\right)\left(|\mathbb{T}_r|^{-1} \sum_{i\in\mathbb{T}_r} X_{2i+\eta}\right)}{|\mathbb{T}_r|^{-1} \sum_{i\in\mathbb{T}_r} X_i^2 - \left(|\mathbb{T}_r|^{-1} \sum_{i\in\mathbb{T}_r} X_i\right)^2} \\ \widehat{\beta}^r_\eta = |\mathbb{T}_r|^{-1} \sum_{i\in\mathbb{T}_r} X_{2i+\eta} - \widehat{\alpha}^r_\eta |\mathbb{T}_r|^{-1} \sum_{i\in\mathbb{T}_r} X_i. \end{cases} \tag{2.4.3}$$

Notice that in the Gaussian case, this least squares estimator corresponds to the maximum likelihood estimator.

We also need to introduce the estimators of the conditional variance σ^2 and the conditional sister-sister correlation ρ. These estimators are naturally given by

$$\begin{cases} \widehat{\sigma}^2_r = \frac{1}{2\mathbb{T}_r} \sum_{i\in\mathbb{T}_r} (\widehat{\varepsilon}^2_{2i} + \widehat{\varepsilon}^2_{2i+1}) \\ \widehat{\rho}_r = \frac{1}{\widehat{\sigma}^2_r} \sum_{i\in\mathbb{T}_r} \widehat{\varepsilon}_{2i}\widehat{\varepsilon}_{2i+1} \end{cases} \tag{2.4.4}$$

where the residues are defined by $\widehat{\varepsilon}_{2i+\eta} = X_{2i+\eta} - \widehat{\alpha}^r_\eta X_i - \widehat{\beta}^r_\eta$, with $\eta \in \{0,1\}$.

Let us denote by $\mathcal{C}_{pol}(\mathbb{R})$ (resp. $\mathcal{C}_{pol}(\mathbb{R}^3)$) the set of all continuous functions $f : \mathbb{R} \to \mathbb{R}$ (resp. $f : \mathbb{R}^3 \to \mathbb{R}$) such that $|f|$ is bounded above by a polynomial. From [66], we know that $\mathcal{C}_{pol}(\mathbb{R})$ fulfills hypothesis (i)-(vi).

We will take $F = \mathcal{C}^1_{pol}(\mathbb{R})$ the set of all $\mathcal{C}^1$ functions $f : \mathbb{R} \to \mathbb{R}$ such that $|f|+|f'|$ is bounded above by a polynomial. Then, one can check that F fulfills hypothesis (i)-(vi). Moreover, for all $f \in F$, hypothesis **(H1)** holds with $\alpha = \max(|\alpha_0|, |\alpha_1|)$. Let μ be the unique stationary distribution of the induced Markov chain $(Y_r, r \in \mathbb{N})$, see [66] for more details.

Let us denote by $\mathcal{C}^1_{pol}(\mathbb{R}^3)$ the set of all $\mathcal{C}^1$ functions $f : \mathbb{R}^3 \to \mathbb{R}$ such that $|f| + |f'|$ is bounded above by a polynomial. We shall denote by $\mathbf{x}$ (resp. $\mathbf{x}^2$, $\mathbf{xy}$, $\mathbf{y} \cdots$) the element of $\mathcal{C}^1_{pol}(\mathbb{R}^3)$ defined by $(x,y,z) \mapsto x$ (resp. x^2, xy, y, $\cdots$).

We define two continuous functions $\mu_1 : \Theta \to \mathbb{R}$ and $\mu_2 : \Theta \times \mathbb{R}^*_+ \to \mathbb{R}$ by writing

$$(\mu, \mathbf{x}) = \mu_1(\theta) \quad \text{and} \quad (\mu, \mathbf{x}^2) = \mu_2(\theta, \sigma^2), \tag{2.4.5}$$

where $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1) \in \Theta = (-1,1) \times \mathbb{R} \times (-1,1) \times \mathbb{R}$.

To segregate between $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$ and its alternative $H_1 = \{(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)\}$, we shall use the test statistic

$$\chi^{(1)}_r = \frac{|\mathbb{T}_r|}{2\widehat{\sigma}^2_r(1-\widehat{\rho}_r)} \left\{ (\widehat{\alpha}^r_0 - \widehat{\alpha}^r_1)^2 (\widehat{\mu}^2_{2,r} - \widehat{\mu}^2_{1,r}) + \left((\widehat{\alpha}^r_0 - \widehat{\alpha}^r_1)\widehat{\mu}_{1,r} + \widehat{\beta}^r_0 - \widehat{\beta}^r_1 \right)^2 \right\},$$

where we write $\widehat{\mu}_{1,r} = \mu_1(\widehat{\theta}^r)$ and $\widehat{\mu}_{2,r} = \mu_2(\widehat{\theta}_r, \widehat{\sigma}_r)$.

As usual the Gaussian setting has specific properties that allow easier calculations and more general assumptions.

2.4.1 The Gaussian setting

We introduce the following assumption on the speed of the MDP. Let (b_n) be an increasing sequence of positive real numbers such that

$$\frac{b_n}{\sqrt{n}} \longrightarrow +\infty \quad \text{and} \qquad \frac{b_n}{\sqrt{n \log n}} \to 0. \tag{2.4.6}$$

Proposition 2.4.1. *Let (b_n) be a sequence of real numbers satisfying (2.4.6). Then*

$$\widehat{\theta}^r \underset{\frac{b^2_{|\mathbb{T}_r|}}{|\mathbb{T}_r|}}{\overset{\text{superexp}}{\Longrightarrow}} \theta.$$

Proof. We will treat the case of $\widehat{\alpha}_0^r$ given in (2.4.3). The others $\widehat{\beta}_0^r, \widehat{\alpha}_1^r$ and $\widehat{\beta}_1^r$ given in (2.4.3) may be treated in a similar way. Note that $\widehat{\alpha}_0^r = \dfrac{C_r}{B_r}$, where

$$C_r = \overline{M}_{\mathbb{T}_r}(\mathbf{xy}) - \overline{M}_{\mathbb{T}_r}(\mathbf{x})\overline{M}_{\mathbb{T}_r}(\mathbf{y}) \quad \text{and} \quad B_r = \overline{M}_{\mathbb{T}_r}(\mathbf{x}^2) - \overline{M}_{\mathbb{T}_r}(\mathbf{x})^2.$$

Now, using Lemma 2.B.2 and Proposition 2.2.14, it follows that

$$\widehat{\alpha}_0^r \underset{\frac{b^2_{|\mathbb{T}_r|}}{|\mathbb{T}_r|}}{\overset{\text{superexp}}{\Longrightarrow}} \alpha_0.$$

□

We recall that in the BAR model (2.4.1), we use $\alpha = \max\{|\alpha_0|, |\alpha_1|\}$, and $b := \mu_2(\theta, \sigma^2) - \mu_1(\theta)^2$, where μ_1 and μ_2 are given in (2.4.5), so we have the following deviation inequality

Proposition 2.4.2. *For all $\delta > 0$, for all $r \in \mathbb{N}$ and for all $\gamma < \min\left(\frac{c_1 b}{1+\delta}, \frac{c_1 b}{1+\sqrt{\delta}}, \frac{c_1 b}{1+\sqrt[4]{\delta}}\right)$, where c_1 is a positive constant which depends on μ_1, we have*

$$\mathbb{P}\left(\left\|\widehat{\theta}^r - \theta\right\| > \delta\right) \leq \begin{cases} \frac{c}{\gamma^{4q}\delta^{4-p}} \left(\frac{1}{4}\right)^{r+1} & \text{if} \quad \alpha^2 < \frac{1}{2}, \\ \frac{c}{\gamma^{4q}\delta^{4-p}} r^2 \left(\frac{1}{4}\right)^{r+1} & \text{if} \quad \alpha^2 = \frac{1}{2}, \\ \frac{c}{\gamma^{4q}\delta^{4-p}} \alpha^{4(r+1)} & \text{if} \quad \alpha^2 > \frac{1}{2}, \end{cases} \tag{2.4.7}$$

where the constant c depends on α, μ_1, μ_2 and differs line by line, $p = p(\delta) \in \{0, 2, 4\}$ and $q = q(\delta) \in \{0, 1\}$.

Remark 2.4.3. *The values of p and q in Proposition 2.4.2 depend on the order of δ. For example, if δ is small enough, we have $p = 0$ and $q = 0$.*

Proof. See section 2.A.5 in the Appendix 2.A. □

Remark 2.4.4. *The Proposition 2.4.2 can be improved by calculating the $2k$-th order moments, with $k > 2$, as in the proof of Theorem 2.2.1. But, as we have said, this comes at the price of enormous computation.*

Proposition 2.4.5. *Let (b_n) be a sequence of real numbers satisfying (2.4.6). Then*

$$(\widehat{\sigma}_r^2, \widehat{\rho}_r) \underset{\frac{b^2_{|\mathbb{T}_r|}}{|\mathbb{T}_r|}}{\overset{\text{superexp}}{\Longrightarrow}} (\sigma^2, \rho).$$

Proof. Let us first deal with σ_r^2 given in (2.4.4). The proof is based on the δ-method. We have (see e.g [66])

$$\widehat{\sigma}_r^2 = \frac{1}{2}\overline{M}_{\mathbb{T}_r}(f(.,\theta)) + D_r$$

where $f(x,y,z,\theta) = (y - \alpha_0 x - \beta_0)^2 + (z - \alpha_1 x - \beta_1)^2$ and

$$D_r = \frac{1}{2|\mathbb{T}_r|}\sum_{i\in\mathbb{T}_r}(f(\Delta_i, \widehat{\theta}^r) - f(\Delta_i, \theta)).$$

By Taylor-Lagrange formula, we can find $g \in \mathcal{C}_{pol}(\mathbb{R}^3)$ such that (see [66])

$$|D_r| \leq \frac{1}{2}\|\widehat{\theta}^r - \theta\| \left(1 + \|\theta\| + \|\widehat{\theta}^r - \theta\|\right) \overline{M}_{\mathbb{T}_r}(g).$$

Now, Proposition 2.2.14 and Proposition 2.4.1 lead us to

$$\widehat{\sigma}_r^2 \underset{\frac{b^2_{|\mathbb{T}_r|}}{|\mathbb{T}_r|}}{\overset{\text{superexp}}{\Longrightarrow}} \sigma^2.$$

The proof for $\widehat{\rho}_r$ given in (2.4.4) is similar. □

Proposition 2.4.6. *Let (b_n) be a sequence of real numbers satisfying (2.4.6). Then the sequence $\left(|\mathbb{T}_r|(\widehat{\theta}^r - \theta)/b_{|\mathbb{T}_r|}\right)$ satisfies the MDP on $\mathbb{R}^4$ with the speed $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ and the rate function I given by*

$$I(x) = \frac{1}{2}x^t(\Sigma')^{-1}x,$$

where

$$\Sigma' = \sigma^2 \begin{pmatrix} K & \rho K \\ \rho K & K \end{pmatrix} \text{ with } K = \frac{1}{\mu_2(\theta,\sigma^2) - \mu_1(\theta)^2}\begin{pmatrix} 1 & -\mu_1(\theta) \\ -\mu_1(\theta) & \mu_2(\theta,\sigma^2) \end{pmatrix}.$$

Proof. We first observe that

$$\frac{|\mathbb{T}_r|}{b_{|\mathbb{T}_r|}}\left(\widehat{\theta}^r - \theta\right) = M(A_r, B_r).\frac{U^r(f)}{b_{|\mathbb{T}_r|}}$$

where $f=(f_1,f_2,f_3,f_4)^t=(\mathbf{xy},\mathbf{y},\mathbf{xz},\mathbf{z})^t$, $U^r(f)=M_{\mathbb{T}_r}(f-Pf)$, $A_r=\overline{M}_{\mathbb{T}_r}(\mathbf{x})$, $B_r=\overline{M}_{\mathbb{T}_r}(\mathbf{x}^2)-\overline{M}_{\mathbb{T}_r}(\mathbf{x})^2$ and

$$M(A_r,B_r)=\begin{pmatrix}\frac{1}{B_r} & \frac{-A_r}{B_r} & 0 & 0\\ \frac{-A_r}{B_r} & \frac{B_r+A_r^2}{B_r} & 0 & 0\\ 0 & 0 & \frac{1}{B_r} & \frac{-A_r}{B_r}\\ 0 & 0 & \frac{-A_r}{B_r} & \frac{B_r+A_r^2}{B_r}\end{pmatrix}.$$

For the sake of simplicity we wrote $Pf=(Pf_1,Pf_2,Pf_3,Pf_4)^t$, where P denotes the $\mathbb{T}-$transition probability associated to BAR(1) process in the Gaussian case, which is given by

$$P(x,dy,dz)=\frac{1}{2\pi\sigma^2(1-\rho^2)}\exp\left(-\frac{1}{2}\begin{pmatrix}y-\alpha_0x-\beta_0\\ z-\alpha_1x-\beta_1\end{pmatrix}^t\Gamma^{-1}\begin{pmatrix}y-\alpha_0x-\beta_0\\ z-\alpha_1x-\beta_1\end{pmatrix}\right)dydz$$

where Γ is the covariance matrix defined in (2.4.2).

On the one hand, from Proposition 2.2.14

$$A_r\underset{\frac{b^2_{|\mathbb{T}_r|}}{|\mathbb{T}_r|}}{\overset{\text{superexp}}{\Longrightarrow}}a:=\mu_1(\theta)\qquad\text{and}\quad B_r\underset{\frac{b^2_{|\mathbb{T}_r|}}{|\mathbb{T}_r|}}{\overset{\text{superexp}}{\Longrightarrow}}b:=\mu_2(\theta,\sigma^2)-\mu_1(\theta)^2,$$

so that by Lemma 2.B.2, we obtain

$$M(A_r,B_r)\underset{\frac{b^2_{|\mathbb{T}_r|}}{|\mathbb{T}_r|}}{\overset{\text{superexp}}{\Longrightarrow}}M(a,b):=\begin{pmatrix}K & 0\\ 0 & K\end{pmatrix}.$$

On the other hand, let $\lambda=(\lambda_1,\lambda_2,\lambda_3,\lambda_4)^t\in\mathbb{R}^4$. For all $x\in\mathbb{R}$, we have that

$$\begin{aligned}P\exp\big(\lambda^t(f-Pf)\big)(x)&=\int_{\mathbb{R}^2}\exp\left(\sum_{i=1}^4\lambda_i(f_i-Pf_i)\right)(x,y,z)P(x,dy,dz)\\ =&\int_{\mathbb{R}^2}\exp\left(\lambda^t\begin{pmatrix}xy-x(\alpha_0x+\beta_0)\\ y-\alpha_0x-\beta_0\\ xz-x(\alpha_1x+\beta_1)\\ z-\alpha_1x-\beta_1\end{pmatrix}\right)P(x,dy,dz)\\ =&\exp\left(-\begin{pmatrix}\alpha_0x+\beta_0\\ \alpha_1x+\beta_1\end{pmatrix}^t\begin{pmatrix}\lambda_1x+\lambda_2\\ \lambda_3x+\lambda_4\end{pmatrix}\right)\\ &\times\int_{\mathbb{R}^2}\exp\left(\begin{pmatrix}\lambda_1x+\lambda_2\\ \lambda_3x+\lambda_4\end{pmatrix}^t\begin{pmatrix}y\\ z\end{pmatrix}\right)P(x,dy,dz).\end{aligned}$$

We know that

$$\int_{\mathbb{R}^2} \exp\left(\begin{pmatrix}\lambda_1 x+\lambda_2\\ \lambda_3 x+\lambda_4\end{pmatrix}^t \begin{pmatrix} y\\ z\end{pmatrix}\right) P(x,dy,dz) = \exp\left(\begin{pmatrix}\alpha_0 x+\beta_0\\ \alpha_1 x+\beta_1\end{pmatrix}^t \begin{pmatrix}\lambda_1 x+\lambda_2\\ \lambda_3 x+\lambda_4\end{pmatrix}\right)$$
$$\times \exp\left(\frac{1}{2}\begin{pmatrix}\lambda_1 x+\lambda_2\\ \lambda_3 x+\lambda_4\end{pmatrix}^t \Gamma \begin{pmatrix}\lambda_1 x+\lambda_2\\ \lambda_3 x+\lambda_4\end{pmatrix}\right).$$

Let $\Xi(x)$ denotes the square matrix with entries $(Pf_if_j - Pf_iPf_j)(x)$, for $1 \leq i,j \leq 4$. So we obtain that

$$\begin{aligned} P\exp\big(\lambda^t(f-Pf)\big)(x) &= \exp\left(\frac{1}{2}\begin{pmatrix}\lambda_1 x+\lambda_2\\ \lambda_3 x+\lambda_4\end{pmatrix}^t \Gamma \begin{pmatrix}\lambda_1 x+\lambda_2\\ \lambda_3 x+\lambda_4\end{pmatrix}\right) \\ &= \exp\left(\frac{1}{2}\sum_{i,j=1}^{4}\lambda_i\lambda_j(Pf_if_j - Pf_iPf_j)(x)\right) \\ &= \exp\left(\frac{1}{2}\lambda^t\Xi(x)\lambda\right). \end{aligned}$$

Recall that the filtration $(\mathcal{H}_n)_{n\geq 0}$ is defined by

$$\mathcal{H}_0 = \sigma(X_1) \quad \text{and} \quad \mathcal{H}_n = \sigma\big(\Delta_{\Pi(i)}, \Pi(i+1), 1 \leq i \leq n\big).$$

Therefore, from the previous calculations, we deduce that for all $k \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E}\left[\exp\big(\lambda^t(f-Pf)(\Delta_{\Pi(k)})\big)\Big|\mathcal{H}_{k-1}\right] &= P\big(\exp\big(\lambda^t(f-Pf)\big)\big)(X_{\Pi(k)}) \\ &= \exp\left(\frac{1}{2}\lambda^t\Xi(X_{\Pi(k)})\lambda\right). \end{aligned}$$

Now, recall that $(M_n^{\Pi}(f-Pf))_{n\in\mathbb{N}}$ is a $(\mathcal{H}_n)-$martingale and by straightforward calculations, its increasing process is given by $\langle M^{\Pi}(f-Pf)\rangle_n = \sum_{k=1}^{n}\Xi(X_{\Pi(k)})$. From the foregoing, we infer that

$$\left(\exp\left(\lambda^t M_n^{\Pi}(f-Pf) - \frac{\lambda^t\langle M^{\Pi}(f-Pf)\rangle_n\lambda}{2}\right)\right)_{n\in\mathbb{N}}$$

is a $(\mathcal{H}_n)$-martingale. It then follows that for all $\lambda \in \mathbb{R}^4$,

$$G_n(\lambda) = \frac{1}{2n}\lambda^t\langle M^{\Pi}(f-Pf)\rangle_n\lambda$$

is an upper and lower cumulant. Moreover, from Proposition 2.2.14 and Lemma 2.B.2

$$G_n(\lambda) \underset{\frac{b^2_{|\mathbb{T}_r|}}{|\mathbb{T}_r|}}{\overset{\text{superexp}}{\Longrightarrow}} \frac{1}{2}\lambda^t\Sigma\lambda \qquad \text{where} \quad \Sigma = \sigma^2\begin{pmatrix}K^{-1} & \rho K^{-1}\\ \rho K^{-1} & K^{-1}\end{pmatrix}.$$

We thus deduce that (see e.g [113]) $\left(M_n^{\Pi}(f)/b_n\right)$ satisfies a MDP on $\mathbb{R}^4$ with speed b_n^2/n and the rate function

$$J(x) = \frac{1}{2} x^t \Sigma^{-1} x. \tag{2.4.8}$$

Taking $n = |\mathbb{T}_r|$, it follows that $\left(U^r(f)/b_{|\mathbb{T}_r|}\right)$ satisfies a MDP with speed $b_{|\mathbb{T}_r|}^2/|\mathbb{T}_r|$ and the rate function J given in (2.4.8). Finally, using the contraction principle(see e.g [35]) as in ([112]), we get the result. □

Let us now consider the test statistic.

Proposition 2.4.7. *Let (b_n) a sequence of real numbers satisfying (2.4.6). Then under the null hypothesis $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$, $\frac{|\mathbb{T}_r|^{1/2}}{b_{|\mathbb{T}_r|}} (\chi_r^{(1)})^{1/2}$ satisfies a MDP on $\mathbb{R}$ with speed $b_{|\mathbb{T}_r|}^2/|\mathbb{T}_r|$ and the rate function*

$$I'(y) = \begin{cases} \frac{y^2}{2} & \text{if } y \in \mathbb{R}_+ \\ +\infty & \text{otherwise.} \end{cases}$$

Under the alternative hypothesis H_1 of H_0, we have for all $A > 0$

$$\limsup_{r \to \infty} \frac{|\mathbb{T}_r|}{b_{\mathbb{T}_r}^2} \log \mathbb{P}\left(\chi_r^{(1)} < A\right) = -\infty.$$

Proof. We have

$$H_0 = \{g(\theta) = 0\} \quad \text{where} \quad g(\theta) = (\alpha_0 - \alpha_1, \beta_0 - \beta_1)^t.$$

From Proposition 2.4.6,

$$\left(|\mathbb{T}_r|(\widehat{\theta}^r - \theta)/b_{|\mathbb{T}_r|}\right)$$

satisfies a MDP on $\mathbb{R}^4$ with speed $b_{|\mathbb{T}_r|}^2/|\mathbb{T}_r|$ and the rate function

$$I(x) = \frac{1}{2} x^t (\Sigma')^{-1} x,$$

so that, using the delta method for the MDP (see e.g [56], Theorem 3.1) we conclude that

$$\left(|\mathbb{T}_r|(g(\widehat{\theta}^r) - g(\theta))/b_{|\mathbb{T}_r|}\right)$$

satisfies a MDP on $\mathbb{R}^2$ with speed $b_{|\mathbb{T}_r|}^2/|\mathbb{T}_r|$ and the rate function

$$J(y) = \inf\left\{I(x); y = g'(\theta)x\right\}.$$

Identification of this rate function by usual optimization argument leads us to

$$J(x) = \frac{1}{2} x^t (\Sigma'')^{-1} x, \quad \text{where} \quad \Sigma'' = 2\sigma^2(1-\rho)K. \tag{2.4.9}$$

Under the null hypothesis H_0, we have $g(\theta) = 0$, so that

$$\left(|\mathbb{T}_r| g(\widehat{\theta}^r)/b_{|\mathbb{T}_r|}\right)$$

satisfies a MDP on $\mathbb{R}^2$ with speed $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ and rate function J given in (2.4.9).

Now, since $K = K(\theta, \sigma)$ is a continuous function of (θ, σ) (see [66]), so that, letting $\widehat{K}_r = K(\widehat{\theta}^r, \widehat{\sigma}_r)$, Lemma 2.B.2, Proposition 2.4.6 and Proposition 2.4.5 entail that

$$\widehat{\Sigma}''_r = 2\widehat{\sigma}^2_r(1 - \widehat{\rho}_r)\widehat{K}_r \underset{\frac{b^2_{|\mathbb{T}_r|}}{|\mathbb{T}_r|}}{\overset{\text{superexp}}{\Longrightarrow}} \Sigma''.$$

It follows using the contraction principle (see e.g [112]) that

$$\left(|\mathbb{T}_r| \widehat{\Sigma''}_r^{-1/2} g(\widehat{\theta}^r)/b_{|\mathbb{T}_r|}\right)$$

satisfies a MDP on $\mathbb{R}^2$ with speed $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ and the rate function

$$J'(y) = \frac{\|y\|^2}{2}.$$

In particular,

$$\left\| \frac{|\mathbb{T}_r|}{b_{|\mathbb{T}_r|}} \widehat{\Sigma''}_r^{-1/2} g(\widehat{\theta}^r) \right\| = \frac{|\mathbb{T}_r|^{1/2}}{b_{|\mathbb{T}_r|}} \sqrt{\chi_r^{(1)}}$$

satisfies a MDP with speed $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ and the rate function I' given in the Proposition 2.4.7.

Now, under the alternative hypothesis H_1,

$$\frac{\chi_r^{(1)}}{|\mathbb{T}_r|} = g(\widehat{\theta}^r)^t \widehat{\Sigma''}_r^{-1} g(\widehat{\theta}^r) \underset{\frac{b^2_{|\mathbb{T}_r|}}{|\mathbb{T}_r|}}{\overset{\text{superexp}}{\Longrightarrow}} g(\theta)^t (\Sigma'')^{-1} g(\theta) > 0,$$

so that $\chi_r^{(1)}$ converges $\frac{b^2_{|\mathbb{T}_r|}}{|\mathbb{T}_r|}$-superexponentially fast to $+\infty$. This concludes the proof of the Proposition 2.4.7. $\square$

2.4.2 Compact case: the uniformly ergodic setting.

We recall that the model under study in this section is the model (2.4.1) where we assume that the noise and initial state X_1 take their values in a compact set. The results will be given without proofs, since the proofs are similar to those done in the previous section. The novelty here is that the range of speed is improved in comparison to the previous section. However, we suppose that the process takes its values in a compact set, which is not the case in the previous section.

We take $F = \mathcal{C}_b^1(\mathbb{R})$ the set of all $\mathcal{C}^1$ functions bounded on $\mathbb{R}$. Therefore, one can easily check (as in [66], proof of Proposition 28) that hypothesis (**H2**) is satisfied with $\alpha = \max(|\alpha_0|, |\alpha_1|)$. We use the same notations as in the previous section.

Let us begin by the fact that the estimator of θ converges super exponentially fast to the true parameter.

Proposition 2.4.8. *Let (b_n) a sequence of real numbers satisfying the Assumption 1. Then we have*

$$\widehat{\theta}^r \underset{\frac{b^2_{|\mathbb{T}_r|}}{|\mathbb{T}_r|}}{\overset{\text{superexp}}{\Longrightarrow}} \theta.$$

We may now refine this result by proving deviation inequality.

Proposition 2.4.9. *For all $\delta > 0$ and for all*

$$\gamma < \min\left(\frac{c_1 b}{1+\delta}, \frac{c_1 b}{1+\sqrt{\delta}}, \frac{c_1 b}{1+\sqrt[4]{\delta}}\right)$$

where c_1 is a positive constant which depends on μ_1, and for

$$r_0 := \frac{\log\left(\gamma^q \delta^{1-p/2}/c_0\right)}{\log \alpha},$$

we have

$$\mathbb{P}\left(\left\|\widehat{\theta}^r - \theta\right\| > \delta\right) \leq \begin{cases} c_2 \exp\left(c''\gamma^q\delta^{1-p/2}\right)\exp\left(-c'\gamma^{2q}\delta^{2-p}|\mathbb{T}_r|\right), \forall r \in \mathbb{N}, \text{ if } \alpha < \frac{1}{2} \\ c_2 \exp\left(c'\gamma^q\delta^{1-p/2}(r+1) - c'\gamma^{2q}\delta^{2-p}|\mathbb{T}_r|\right), \forall r \in \mathbb{N}, \text{ if } \alpha = \frac{1}{2} \\ c_2 \exp\left(-c'\gamma^{2q}\delta^{2-p}|\mathbb{T}_r|\right), \quad \forall r > r_0, \text{ if } \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2} \\ c_2 \exp\left(-c'\gamma^q\delta^{2-p}\frac{|\mathbb{T}_r|}{r+1}\right), \quad \forall r > r_0, \text{ if } \alpha = \frac{\sqrt{2}}{2} \\ c_2 \exp\left(-c'\gamma^{2q}\delta^{2-p}\frac{1}{\alpha^{2(r+1)}}\right), \quad \forall r > r_0, \text{ if } \alpha > \frac{\sqrt{2}}{2}, \end{cases} \tag{2.4.10}$$

where c_2 is a positive constant, c' and c'' depend on α, and c and may differ line by line, c_0 depends on α, c and γ, and may differ line by line, $p \in \{0, 1, 3/2\}$ and $q \in \{0, 1\}$.

We have now to consider super exponential convergence of the estimators of the other parameters.

Proposition 2.4.10. *Let* (b_n) *a sequence of real numbers satisfying the Assumption 1. Then we have*

$$(\widehat{\sigma}_r^2, \widehat{\rho}_r) \underset{\frac{b^2_{|\mathbb{T}_r|}}{|\mathbb{T}_r|}}{\overset{\text{superexp}}{\Longrightarrow}} (\sigma^2, \rho).$$

As previously we may now prove MDP for the estimator of θ.

Proposition 2.4.11. *Let* (b_n) *a sequence of real numbers satisfying the Assumption 1. Then* $\left(|\mathbb{T}_r|(\widehat{\theta}^r - \theta)/b_{|\mathbb{T}_r|}\right)$ *satisfies the MDP on* $\mathbb{R}^4$ *with the speed* $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ *and rate function*

$$I(x) = \frac{1}{2} x^t (\Sigma')^{-1} x,$$

where

$$\Sigma' = \sigma^2 \begin{pmatrix} K & \rho K \\ \rho K & K \end{pmatrix} \quad \textit{with} \quad K = \frac{1}{\mu_2(\theta, \sigma^2) - \mu_1(\theta)^2} \begin{pmatrix} 1 & -\mu_1(\theta) \\ -\mu_1(\theta) & \mu_2(\theta, \sigma^2) \end{pmatrix}.$$

Remark 2.4.12. *Notice that the proof of Proposition 2.4.11 does not need the cumulant method as in the proof of Proposition 2.4.6. Indeed, since we are in the bounded case, from MDP of martingale with bounded jumps (see [34]), we need only to prove the superexponential convergence of increasing process of the martingale. This convergence is easily obtained from Theorem 2.3.2.*

Let us give us our last result by considering a MDP for the test statistic.

Proposition 2.4.13. *Let* (b_n) *a sequence of real numbers satisfying the Assumption 1. Then under the null hypothesis* $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$,

$$\frac{|\mathbb{T}_r|^{1/2}}{b_{|\mathbb{T}_r|}} (\chi_r^{(1)})^{1/2}$$

satisfies a MDP on $\mathbb{R}$ *with speed* $b^2_{|\mathbb{T}_r|}/|\mathbb{T}_r|$ *and the rate function*

$$I'(y) = \begin{cases} \frac{y^2}{2} & \textit{if } y \in \mathbb{R}_+ \\ +\infty & \textit{otherwise.} \end{cases}$$

Under the alternative hypothesis H_1 *of* H_0, *we have for all* $A > 0$

$$\limsup_{r \to \infty} \frac{|\mathbb{T}_r|}{b^2_{|\mathbb{T}_r|}} \log \mathbb{P}\left(\chi_r^{(1)} < A\right) = -\infty.$$

Appendix 2.A Proof of the deviation inequalities

This section is devoted to the proof of the Theorem 2.2.11, Theorem 2.2.12, Theorem 2.3.1, Theorem 2.3.2 and Proposition 2.4.2.

2.A.1 Proof of Theorem 2.2.11

Let $f \in F$ such that $(\mu, f) = 0$. We shall study the three empirical averages $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_n^{\Pi}(f)$ and $\overline{M}_{\mathbb{T}_r}(f)$ successively.

Part 1. Let us first deal with $\overline{M}_{\mathbb{G}_r}(f)$. By Markov inequality, we get, for all $\delta > 0$

$$\mathbb{P}\Big(\big|\overline{M}_{\mathbb{G}_r}(f)\big| > \delta\Big) = \mathbb{P}\Big(\big|\overline{M}_{\mathbb{G}_r}(f)\big|^2 > \delta^2\Big) \leq \tfrac{1}{\delta^2}\mathbb{E}\Big[(\overline{M}_{\mathbb{G}_r}(f))^2\Big].$$

By Guyon (see [66]), we have

$$\mathbb{E}\Big[(\overline{M}_{\mathbb{G}_r}(f))^2\Big] = \sum_{p=0}^{r} 2^{-p-\mathbf{1}_{p<r}} \nu Q^p P\big(Q^{r-p-1}f \otimes Q^{r-p-1}f\big).$$

Hypothesis **(H1)** implies that there exists $g \in F$ and $\alpha \in (0,1)$ such that for all $p \in \big\{0, 1, \cdots, r\big\}$

$$\nu Q^p P(Q^{r-p-1}f \otimes Q^{r-p-1}f) \leq \alpha^{2(r-p-1)} \nu Q^p P(g \otimes g).$$

Next, hypothesis (iii), (v) and (vi) imply that there is a positive constant c such that for all $p \in \big\{0, 1, \cdots, r\big\}$

$$\alpha^{2(r-p-1)} \nu Q^p P(g \otimes g) \leq c\alpha^{2(r-p-1)}.$$

This leads us to

$$\begin{aligned} \mathbb{E}\Big[(\overline{M}_{\mathbb{G}_r}(f))^2\Big] &\leq c \sum_{p=0}^{r} 2^{-p-\mathbf{1}_{p<r}} \alpha^{2(r-p-1)} \\ &= \begin{cases} c\left(\frac{1}{2}\right)^r + c\frac{\alpha^{2r} - \left(\frac{1}{2}\right)^r}{2\alpha^2 - 1} & \text{if } \alpha^2 \neq \frac{1}{2} \\ cr\left(\frac{1}{2}\right)^r & \text{if } \alpha^2 = \frac{1}{2}, \end{cases} \end{aligned} \tag{2.A.1}$$

and therefore (2.2.14) follows.

Part 2. Let us now consider $\overline{M}_n^{\Pi}(f)$. By the Markov inequality and the triangle inequality, we get, for all $\delta > 0$

$$\begin{aligned} \mathbb{P}\Big(\big|\overline{M}_n^{\Pi}(f)\big| > \delta\Big) &= \mathbb{P}\Big(\big|\overline{M}_n^{\Pi}(f)\big|^2 > \delta^2\Big) \\ &\leq \tfrac{1}{\delta^2}\mathbb{E}\left[\Big(\overline{M}_n^{\Pi}(f)\Big)^2\right] \\ &\leq \tfrac{2}{\delta^2}\mathbb{E}\left[\left(\sum_{q=0}^{r_n-1} \tfrac{2^q}{n}\overline{M}_{\mathbb{G}_q}(f)\right)^2\right] + \tfrac{2}{\delta^2}\mathbb{E}\left[\left(\tfrac{1}{n}\sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)})\right)^2\right]. \end{aligned} \tag{2.A.2}$$

In the last inequality (2.A.2), we have used the decomposition

$$\overline{M}_n^{\Pi}(f) = \sum_{q=0}^{r_n-1} \frac{2^q}{n}\overline{M}_{\mathbb{G}_q}(f) + \frac{1}{n}\sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)}).$$

In what follows, the constant c may be slightly different from that of **Part 1** and may differ term by term. For the first term appearing in (2.A.2), we have

$$\mathbb{E}\left[\left(\sum_{q=0}^{r_n-1}\frac{2^q}{n}\overline{M}_{\mathbb{G}_q}(f)\right)^2\right] = \left\|\sum_{q=0}^{r_n-1}\frac{2^q}{n}\overline{M}_{\mathbb{G}_q}(f)\right\|_2^2 \le \left(\sum_{q=0}^{r_n-1}\frac{2^q}{n}\left\|\overline{M}_{\mathbb{G}_q}(f)\right\|_2\right)^2.$$

Using (2.A.1), we get that

$$\sum_{q=0}^{r_n-1}\frac{2^q}{n}\left\|\overline{M}_{\mathbb{G}_q}(f)\right\|_2 \le \begin{cases} \dfrac{c}{n}\displaystyle\sum_{q=0}^{r_n-1}(\sqrt{2})^q \le c\frac{\sqrt{2}^{r_n}}{n}, & \text{if } \alpha^2 < \dfrac{1}{2}, \\ \dfrac{c}{n}\displaystyle\sum_{q=0}^{r_n} q^{1/2}\sqrt{2}^q \le c\frac{r_n^{1/2}\sqrt{2}^{r_n}}{n}, & \text{if } \alpha^2 = \dfrac{1}{2}, \\ \dfrac{c}{n}\displaystyle\sum_{q=0}^{r_n-1}(2\alpha)^q \le c\alpha^{r_n}, & \text{if } \alpha^2 > \dfrac{1}{2}, \end{cases}$$

which implies that

$$\mathbb{E}\left[\left(\sum_{q=0}^{r_n-1}\frac{2^q}{n}\overline{M}_{\mathbb{G}_q}(f)\right)^2\right] \le \begin{cases} c\dfrac{2^{r_n}}{n^2} \le c\left(\dfrac{1}{2}\right)^{r_n+1}, & \text{if } \alpha^2 < \dfrac{1}{2}, \\ c\dfrac{r_n}{2^{r_n+1}}, & \text{if } \alpha^2 = \dfrac{1}{2}, \\ c\alpha^{2(r_n+1)}, & \text{if } \alpha^2 > \dfrac{1}{2}. \end{cases} \tag{2.A.3}$$

Now, we have to control the second term in (2.A.2). As in Guyon [66], we have that

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)})\right)^2\right] \le \frac{n-2^{r_n}+1}{n^2}\nu Q^{r_n} f^2$$
$$+ \frac{(n-2^{r_n})(n-2^{r_n}+1)}{n^2(1-2^{-r_n})}\sum_{p=0}^{r_n-1} 2^{-p-1}\nu Q^p P(Q^{r_n-p-1}f \otimes Q^{r_n-p-1}f)$$

$$\leq \frac{c}{n} + c\sum_{p=0}^{r_n-1} 2^{-p-1}\alpha^{2r_n-2p-2}.$$

Discussing following the value of α, we obtain that

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=2^{r_n}}^{n} f(X_{\Pi(i)})\right)^2\right] \leq \begin{cases} c\dfrac{1}{2^{r_n+1}}, & \text{if } \alpha^2 < \dfrac{1}{2}, \\ c\dfrac{r_n}{2^{r_n+1}}, & \text{if } \alpha^2 = \dfrac{1}{2}, \\ c\alpha^{2(r_n+1)}, & \text{if } \alpha^2 > \dfrac{1}{2}. \end{cases} \tag{2.A.4}$$

Inequality (2.2.15) then follows from (2.A.3) and (2.A.4).

Part 3. The case of $\overline{M}_{\mathbb{T}_r}(f)$ can be deduced from the previous by taking $n = |\mathbb{T}_r|$.

2.A.2 Proof of Theorem 2.2.12

Let $f \in \mathcal{B}(S^3)$ such that Pf and Pf^2 exist and belong to F and $(\mu, Pf) = 0$. We shall study the three empirical averages $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_n^{\Pi}(f)$ and $\overline{M}_{\mathbb{T}_r}(f)$ successively.

Part 1. Let us first deal with $\overline{M}_{\mathbb{G}_r}(f)$. By the Markov inequality, we get for all $\delta > 0$

$$\begin{aligned}\mathbb{P}\Big(\big|\overline{M}_{\mathbb{G}_r}(f)\big| > \delta\Big) &\leq \frac{1}{\delta^2}\mathbb{E}\Big[(\overline{M}_{\mathbb{G}_r}(f))^2\Big] \\ &= \frac{1}{\delta^2}\mathbb{E}\Big[(\overline{M}_{\mathbb{G}_r}(Pf))^2\Big] + \frac{1}{\delta^2}\frac{1}{|\mathbb{G}_r|}\mathbb{E}\Big[\overline{M}_{\mathbb{G}_r}\left(Pf^2 - (Pf)^2\right)\Big] \\ &\leq \frac{1}{\delta^2}\mathbb{E}\Big[(\overline{M}_{\mathbb{G}_r}(Pf))^2\Big] + \frac{c}{\delta^2}\left(\frac{1}{2}\right)^r.\end{aligned}$$

The last inequality follows from the convergence of the sequence (see [66])

$$\left(\mathbb{E}\Big[\overline{M}_{\mathbb{G}_r}\left(Pf^2 - (Pf)^2\right)\Big]\right)_r.$$

Now, using the **Part 1** of the proof of the Theorem 2.2.11 with Pf instead of f leads us to a similar inequality (2.2.14) in Theorem 2.2.12 for $f \in \mathcal{B}(S^3)$.

Part 2. Let us now treat $\overline{M}_n^{\Pi}(f)$. Using the two equalities

$$\overline{M}_n^{\Pi}(f) = \sum_{q=0}^{r_n-1}\frac{|\mathbb{G}_q|}{n}\overline{M}_{\mathbb{G}_q}(f) + \frac{1}{n}\sum_{i=2^{r_n}}^{n} f(\Delta_{\Pi(i)}),$$

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=2^r n}^{n} f(\Delta_{\Pi(i)})\right)^2\right] = \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=2^r n}^{n} Pf(X_{\Pi(i)})\right)^2\right]$$
$$+\frac{1}{n}\mathbb{E}\left[\frac{1}{n}\sum_{i=2^r n}^{n}\left(Pf^2-(Pf)^2\right)\left(X_{\Pi(i)}\right)\right],$$

and the **Part 2** of the proof of the Theorem 2.2.11 with Pf instead of f leads us to a similar inequality (2.2.15) in Theorem 2.2.12 for $f \in \mathcal{B}$.

Part 3. The case of $\overline{M}_{\mathbb{T}_r}(f)$ can be deduced from the previous by taking $n = |\mathbb{T}_r|$.

2.A.3 Proof of Theorem 2.3.1

Let $f \in \mathcal{B}_b(S)$ such that $(\mu, f) = 0$. We shall study the three empirical averages $\overline{M}_{\mathbb{G}_r}(f)$, $\overline{M}_n^{\Pi}(f)$ and $\overline{M}_{\mathbb{T}_r}(f)$ successively.

Part 1. Let us first deal with $\overline{M}_{\mathbb{G}_r}(f)$. We have for all $\lambda > 0$ and for all $\delta > 0$

$$\mathbb{P}(\overline{M}_{\mathbb{G}_r}(f) > \delta) \leq \exp(-\lambda\delta|\mathbb{G}_r|)\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_r} f(X_i)\right)\right], \tag{2.A.5}$$

By subtracting and adding terms, we get

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_r} f(X_i)\right)\right] = \mathbb{E}\left[\mathbb{E}\left[\prod_{i\in\mathbb{G}_{r-1}}\exp(\lambda(f(X_{2i})+f(X_{2i+1})-2Qf(X_i)))\right.\right.$$
$$\left.\left.\times\prod_{i\in\mathbb{G}_{r-1}}\exp(2\lambda Qf(X_i))\Big|\mathcal{F}_{r-1}\right]\right].$$

Now using the fact that conditionally to the $(r-1)$ first generations the sequence $\{\Delta_i, i \in \mathbb{G}_{r-1}\}$ is a sequence of independent random variables, we have that

$$\mathbb{E}\left[\mathbb{E}\left[\prod_{i\in\mathbb{G}_{r-1}}\exp(\lambda(f(X_{2i})+f(X_{2i+1})-2Qf(X_i)))\right.\right.$$
$$\left.\left.\times\prod_{i\in\mathbb{G}_{r-1}}\exp(2\lambda Qf(X_i))\Big|\mathcal{F}_{r-1}\right]\right]$$
$$=\mathbb{E}\left[\prod_{i\in\mathbb{G}_{r-1}}\exp\left(2\lambda Qf(X_i)\right)\right.$$
$$\left.\times\prod_{i\in\mathbb{G}_{r-1}}\mathbb{E}\left[\exp\left(\lambda(f(X_{2i})+f(X_{2i+1})-2Qf(X_i))\right)\Big|\mathcal{F}_{r-1}\right]\right].$$

Using the Azuma-Bennet-Hoeffding inequalities [6], [16], [70], we get according to (**H2**), for all $i \in \mathbb{G}_{r-1}$

$$\mathbb{E}\Big[\exp\Big(\lambda(f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i))\Big)\Big|\mathcal{F}_{r-1}\Big] \leq \exp\Big(2\lambda^2 c^2 (1+\alpha)^2\Big).$$

This leads us to

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{G}_r} f(X_i)\right)\right] \leq \exp\Big(\lambda^2 c^2 (1+\alpha)^2 |\mathbb{G}_r|\Big)\mathbb{E}\left[\prod_{i\in\mathbb{G}_{r-1}} \exp\left(2\lambda Qf(X_i)\right)\right].$$

Doing the same thing for $\mathbb{E}\left[\prod_{i\in\mathbb{G}_{r-1}} \exp\left(2\lambda Qf(X_i)\right)\right]$ with Qf replacing f, we get

$$\mathbb{E}\left[\prod_{i\in\mathbb{G}_{r-1}} \exp\left(2\lambda Qf(X_i)\right)\right] \leq \exp\Big(2\lambda^2 c^2 (\alpha+\alpha^2)^2 |\mathbb{G}_r|\Big) \times \mathbb{E}\left[\prod_{i\in\mathbb{G}_{r-2}} \exp\left(2^2\lambda Q^2 f(X_i)\right)\right].$$

Iterating this procedure, we get

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{G}_r} f(X_i)\right)\right] \leq \mathbb{E}\Big[\exp\Big(2^r \lambda Q^r f(X_1)\Big)\Big] \times \prod_{k=1}^{r} \exp\Big(2^{k-1}\lambda^2 c^2 (\alpha^{k-1}+\alpha^k)^2 |\mathbb{G}_r|\Big).$$

Once again, according to (**H2**), we have

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{G}_r} f(X_i)\right)\right] \leq \exp\Big(\lambda c \alpha^r |\mathbb{G}_r|\Big) \times \exp\Big(\lambda^2 c^2 (1+\alpha)^2 |\mathbb{G}_r| \sum_{k=1}^{r} (2\alpha^2)^{k-1}\Big).$$

Hence

- if $\alpha^2 \neq \frac{1}{2}$ then

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{G}_r} f(X_i)\right)\right] \leq \exp\Big(\lambda c \alpha^r |\mathbb{G}_r|\Big) \times \exp\Big(\lambda^2 c^2 (1+\alpha)^2 \frac{1-(2\alpha^2)^r}{1-2\alpha^2} |\mathbb{G}_r|\Big);$$

- if $\alpha^2 = \frac{1}{2}$ then

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_r} f(X_i)\right)\right] \leq \exp\left(\lambda c(\frac{\sqrt{2}}{2})^r|\mathbb{G}_r|\right) \times \exp\left(\lambda^2 c^2(1+\alpha)^2 r|\mathbb{G}_r|\right).$$

We then consider three cases.

(a) If $\alpha^2 < \frac{1}{2}$, then $\frac{1-(2\alpha^2)^r}{1-2\alpha^2} < \frac{1}{1-2\alpha^2}$ for all r. Taking $\lambda = \frac{(1-2\alpha^2)\delta}{2c^2(1+\alpha)^2}$ in (2.A.5) leads us to

$$\mathbb{P}\Big(\overline{M}_{\mathbb{G}_r}(f) > \delta\Big) \leq \exp\left(-\Big(\frac{(1-2\alpha^2)\delta^2}{4c^2(1+\alpha)^2} - \alpha^r\frac{(1-2\alpha^2)\delta}{2c(1+\alpha)^2}\Big)|\mathbb{G}_r|\right).$$

- If $\alpha \leq \frac{1}{2}$, then $(2\alpha)^r \leq 1$ for all $r \in \mathbb{N}$. We then have for all $r \in \mathbb{N}$.

$$\mathbb{P}\Big(\overline{M}_{\mathbb{G}_r}(f) > \delta\Big) \leq \exp\left(\frac{(1-2\alpha^2)\delta}{2c(1+\alpha)^2}\right)\exp\left(-\frac{(1-2\alpha^2)\delta^2|\mathbb{G}_r|}{4c^2(1+\alpha)^2}\right).$$

- If $\frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}$, then for all $r \in \mathbb{N}$ such that $r > \log\left(\frac{\delta}{4c}\right)/\log\alpha$, we have $(\delta - 2c\alpha^r) > \frac{\delta}{2}$ and it then follows that

$$\mathbb{P}\Big(\overline{M}_{\mathbb{G}_r}(f) > \delta\Big) \leq \exp\left(-\frac{(1-2\alpha^2)\delta^2|\mathbb{G}_r|}{8c^2(1+\alpha)^2}\right).$$

(b) If $\alpha^2 = \frac{1}{2}$. Then for all $\lambda > 0$,

$$\mathbb{P}\left(\overline{M}_{\mathbb{G}_r}(f) > \delta\right) \leq \exp\left(\left(-\delta\lambda + c^2(1+\alpha)^2 r\lambda^2\right)|\mathbb{G}_r|\right) \times \exp\left(\lambda c\left(\frac{\sqrt{2}}{2}\right)^r|\mathbb{G}_r|\right).$$

Taking $\lambda = \frac{\delta}{2c^2(1+\alpha)^2 r}$, we are led to

$$\mathbb{P}\left(\overline{M}_{\mathbb{G}_r}(f) > \delta\right) \leq \exp\left(-\frac{\delta|\mathbb{G}_r|}{4c^2(1+\alpha)^2 r}\left(\delta - 2c\left(\frac{\sqrt{2}}{2}\right)^r\right)\right).$$

For all $r \in \mathbb{N}$ such that $r > \log\left(\frac{\delta}{4c}\right)/\log\left(\frac{\sqrt{2}}{2}\right)$, we have $(\delta - 2c\left(\frac{\sqrt{2}}{2}\right)^r) > \frac{\delta}{2}$ and for such r, it follows that

$$\mathbb{P}\left(\overline{M}_{\mathbb{G}_r}(f) > \delta\right) \leq \exp\left(-\frac{\delta^2|\mathbb{G}_r|}{18c^2 r}\right).$$

(c) If $\alpha^2 > \frac{1}{2}$. Then for all $\lambda > 0$

$$\mathbb{P}\Big(\overline{M}_{\mathbb{G}_r}(f) > \delta\Big) \leq \exp\Big(-\lambda\delta|\mathbb{G}_r|\Big) \times \exp\left(\lambda^2 c^2 (1+\alpha)^2 \frac{(2\alpha^2)^r - 1}{2\alpha^2 - 1}|\mathbb{G}_r|\right)$$
$$\times \exp\Big(\lambda c\alpha^r|\mathbb{G}_r|\Big)$$
$$\leq \exp\left(-|\mathbb{G}_r|\left(\lambda\delta - \frac{\lambda^2 c^2(1+\alpha)^2}{2\alpha^2-1}(2\alpha^2)^r\right)\right) \times \exp\Big(\lambda c\alpha^r|\mathbb{G}_r|\Big).$$

Taking $\lambda = \frac{(2\alpha^2-1)\delta}{2c^2(1+\alpha)^2(2\alpha^2)^r}$ leads us to

$$\mathbb{P}\Big(\overline{M}_{\mathbb{G}_r}(f) > \delta\Big) \leq \exp\left(-\frac{(2\alpha^2-1)\delta}{4c^2(1+\alpha)^2\alpha^{2r}}\left(\delta - 2c\alpha^r\right)\right).$$

Now for all $r \in \mathbb{N}$ such that $r > \log\left(\frac{\delta}{4c}\right)/\log\alpha$, we have

$$\mathbb{P}\left(\overline{M}_{\mathbb{G}_r}(f)\right) \leq \exp\left(-\frac{(2\alpha^2-1)\delta^2}{8c^2(1+\alpha)^2\alpha^{2r}}\right).$$

Part 2. Let us now deal with $\overline{M}_{\mathbb{T}_r}(f)$. We have for all $\lambda > 0$ and all $\delta > 0$

$$\mathbb{P}\Big(\overline{M}_{\mathbb{T}_r}(f) > \delta\Big) \leq \exp\Big(-\lambda\delta|\mathbb{T}_r|\Big)\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_r} f(X_i)\right)\right]. \qquad (2.A.6)$$

By subtracting and adding terms, we get

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_r} f(X_i)\right)\right]$$
$$= \mathbb{E}\left[\mathbb{E}\left[\prod_{i\in\mathbb{G}_{r-1}}\exp\Big(\lambda\big(f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i)\big)\Big) \times \prod_{i\in\mathbb{G}_{r-1}}\exp\Big(2\lambda Qf(X_i)\Big)\right.\right.$$
$$\left.\left.\times \prod_{i\in\mathbb{T}_{r-1}}\exp\Big(\lambda f(X_i)\Big)\Big|\mathcal{F}_{r-1}\right]\right]$$
$$= \mathbb{E}\left[\mathbb{E}\left[\prod_{i\in\mathbb{G}_{r-1}}\exp\Big(\lambda\big(f(X_{2i}) + f(X_{2i+1}) - 2Qf(X_i)\big)\Big)\right.\right.$$
$$\left.\left.\times \prod_{i\in\mathbb{G}_{r-1}}\exp\Big(\lambda(f+2Qf)(X_i)\Big) \times \prod_{i\in\mathbb{T}_{r-2}}\exp\Big(\lambda f(X_i)\Big)\Big|\mathcal{F}_{r-1}\right]\right]$$

The fact that conditionally to the $(r-1)$ first generations the sequence $\{\Delta_i, i \in \mathbb{G}_{r-1}\}$ is a sequence of independent random variables and Azuma-Bennett-Hoeffding inequality (see lemma 2.B.1) lead us according to **(H2)** to

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_r}f(X_i)\right)\right]\leq\exp\Big(2\lambda^2c^2(1+\alpha)^2|\mathbb{G}_{r-1}|\Big)$$
$$\times\,\mathbb{E}\left[\prod_{i\in\mathbb{G}_{r-1}}\exp\Big(\lambda(f+2Qf)(X_i)\Big)\prod_{i\in\mathbb{T}_{r-2}}\exp\Big(\lambda f(X_i)\Big)\right].$$

Doing the same things for

$$\mathbb{E}\left[\prod_{i\in\mathbb{G}_{r-1}}\exp\Big(\lambda(f+2Qf)(X_i)\Big)\prod_{i\in\mathbb{T}_{r-2}}\exp\Big(\lambda f(X_i)\Big)\right]$$

with $f+2Qf$ replacing f we get

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_r}f(X_i)\right)\right]\leq\exp\Big(2\lambda^2c^2(1+\alpha)^2|\mathbb{G}_{r-1}|\Big)$$
$$\times\exp\Big(2\lambda^2c^2(1+3\alpha+2\alpha^2)^2|\mathbb{G}_{r-2}|\Big)$$
$$\times\,\mathbb{E}\left[\prod_{i\in\mathbb{G}_{r-2}}\exp\Big(\lambda(f+2Qf+2^2Q^2f)(X_i)\Big)\prod_{i\in\mathbb{T}_{r-3}}\exp\Big(\lambda f(X_i)\Big)\right].$$

Iterating this procedure leads us to

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_r}f(X_i)\right)\right]\leq\exp\left(2\lambda^2c^2(1+\alpha)^2\sum_{q=1}^{r}\left(\sum_{k=0}^{q-1}(2\alpha)^k\right)^2|\mathbb{G}_{r-q}|\right)$$
$$\times\,\mathbb{E}\left[\exp\left(\lambda\Big(f+2Qf+2^2Q^2f+\cdots+2^rQ^rf\Big)(X_1)\right)\right].$$

Using **(H2)** we get

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_r}f(X_i)\right)\right]$$
$$\leq\exp\left(\lambda c\sum_{k=0}^{r}(2\alpha)^k+2\lambda^2c^2(1+\alpha)^2\sum_{q=1}^{r}\left(\sum_{k=0}^{q-1}(2\alpha)^k\right)^2|\mathbb{G}_{r-q}|\right).$$

Now for $\alpha\neq\frac{1}{2}$ and $\alpha^2\neq\frac{1}{2}$ we have

$$\mathbb{P}\Big(\overline{M}_{\mathbb{T}_r}(f)>\delta\Big)\leq\exp\Big(-\lambda\delta|\mathbb{T}_r|\Big)\times\exp\left(\lambda c\frac{1-(2\alpha)^{r+1}}{1-2\alpha}\right)$$
$$\times\exp\left(2\lambda^2c^2(1+\alpha)^2\left(\frac{2^r-1}{(1-2\alpha)^2}-\frac{\alpha(1-\alpha^r)2^{r+1}}{(1-2\alpha)^2(1-\alpha)}+\frac{2\alpha^2(1-(2\alpha^2)^r)2^r}{(1-2\alpha)^2(1-2\alpha^2)}\right)\right)$$

$$\leq \exp\left(-|\mathbb{T}_r|\left(\lambda\delta - \frac{\lambda^2 c^2(1+\alpha)^2}{(1-2\alpha)^2}\left(1+\frac{4\alpha^2(1-(2\alpha^2)^r)}{1-2\alpha^2}\right)\right)\right)$$
$$\times \exp\left(\lambda c\frac{1-(2\alpha)^{r+1}}{1-2\alpha}\right).$$

Taking $\lambda = \dfrac{\delta}{\frac{2c^2(1+\alpha)^2}{(1-2\alpha)^2}\left(1+\frac{4\alpha^2(1-(2\alpha^2)^r)}{1-2\alpha^2}\right)}$ leads us to

$$\mathbb{P}\Big(\overline{M}_{\mathbb{T}_r}(f) > \delta\Big) \leq \exp\left(-|\mathbb{T}_r|\frac{(1-2\alpha)^2\delta^2}{4c^2(1+\alpha)^2\left(1+\frac{4\alpha^2(1-(2\alpha^2)^r)}{1-2\alpha^2}\right)}\right)$$
$$\times \exp\left(\frac{(1-2\alpha)^2\delta}{2c(1+\alpha)^2\left(1+\frac{4\alpha^2(1-(2\alpha^2)^r)}{1-2\alpha^2}\right)}\frac{1-(2\alpha)^{r+1}}{1-2\alpha}\right).$$

- If $\alpha < \frac{1}{2}$ then $\frac{1-(2\alpha^2)^r}{1-2\alpha^2} < \frac{1}{1-2\alpha^2}$ for all $r \in \mathbb{N}$.

$$\mathbb{P}\Big(\overline{M}_{\mathbb{T}_r}(f) > \delta\Big) \leq \exp\left(\frac{1-2\alpha}{2c(1+\alpha)^2}\delta\right) \times \exp\left(-\frac{(1-2\alpha^2)(1-2\alpha)^2\delta^2}{4c^2(1+\alpha)^2(1+2\alpha^2)}|\mathbb{T}_r|\right).$$

- If $\frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}$ then $\frac{1-(2\alpha^2)^r}{1-2\alpha^2} < \frac{1}{1-2\alpha^2}$ for all $r \in \mathbb{N}$.

$$\mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(f) > \delta\right) \leq \exp\left(-\frac{(1-2\alpha^2)(2\alpha-1)^2\delta|\mathbb{T}_r|}{4c^2(1+\alpha)^2(1+2\alpha^2)}\left(\delta - \frac{2c(1-2\alpha^2)\alpha^{r+1}}{(2\alpha-1)(1+2\alpha^2)}\right)\right),$$

Now for all $r \in \mathbb{N}$ such that $r+1 > \log\left(\frac{(2\alpha-1)(1+2\alpha^2)\delta}{4c(1-2\alpha^2)}\right)/\log\alpha$, we have $\delta - \frac{2c(1-2\alpha^2)\alpha^{r+1}}{(2\alpha-1)(1+2\alpha^2)} > \frac{\delta}{2}$ so that for such r, we have

$$\mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(f) > \delta\right) \leq \exp\left(-\frac{(1-2\alpha^2)(2\alpha-1)^2\delta^2|\mathbb{T}_r|}{8c^2(1+\alpha)^2(1+2\alpha^2)}\right).$$

- If $\alpha^2 > \frac{1}{2}$ then for all $r \geq 1$, we have

$$\mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(f) > \delta\right) \leq \exp\left(-\frac{(2\alpha-1)^2(2\alpha^2-1)\delta}{32c^2(1+\alpha)^2\alpha^{2(r+1)}}\left(\delta - \frac{16\alpha^2 c\alpha^{r+1}}{(2\alpha^2-1)(2\alpha-1)}\right)\right).$$

For all $r \in \mathbb{N}^*$ such that $r+3 > \log\left(\frac{(2\alpha^2-1)(2\alpha-1)\delta}{32c}\right)/\log\alpha$, we have $\delta - \frac{16\alpha^2 c\alpha^{r+1}}{(2\alpha^2-1)(2\alpha-1)} > \frac{\delta}{2}$ so that

$$\mathbb{P}\Big(\overline{M}_{\mathbb{T}_r}(f) > \delta\Big) \leq \exp\left(-\frac{(1-2\alpha)^2(2\alpha^2-1)\delta^2}{64c^2(1+\alpha)^2}\left(\frac{1}{\alpha^2}\right)^{r+1}\right).$$

Now if $\alpha = \frac{1}{2}$, then $\sum_{q=1}^{r} \frac{q^2}{2^q} < \sum_{q=1}^{\infty} \frac{q^2}{2^q} = 6$. Then for all $\lambda > 0$

$$\mathbb{P}\Big(\overline{M}_{\mathbb{T}_r}(f) > \delta\Big) \leq \exp\left(-(\lambda\delta - 27c^2\lambda^2)|\mathbb{T}_r|\right) \times \exp\left(\lambda c(r+1)\right).$$

Taking $\lambda = \frac{\delta}{54c^2}$ leads us to

$$\mathbb{P}\Big(\overline{M}_{\mathbb{T}_r}(f) > \delta\Big) \leq \exp\left(-\frac{\delta^2}{108c^2}|\mathbb{T}_r|\right) \times \exp\left(\frac{\delta}{54c}(r+1)\right).$$

Finally, if $\alpha^2 = \frac{1}{2}$, in the same way as previously, for all $r \in \mathbb{N}$ such that $r+1 > \log\left(\frac{(\sqrt{2}-1)\delta}{4c}\right) / \log\left(\frac{\sqrt{2}}{2}\right)$, we have

$$\mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(f) > \delta\right) \leq \exp\left(-\frac{(\sqrt{2}-1)^2\delta^2}{4c^2(1+\sqrt{2})^2}\frac{|\mathbb{T}_r|}{r+1}\right).$$

Part 3. Eventually, let us look at $\overline{M}_n^{\Pi}(f)$. We have for all $\delta > 0$

$$\mathbb{P}\left(\frac{1}{n}M_n^{\Pi}(f) > \delta\right) \leq \mathbb{P}\left(\frac{1}{n}\sum_{i\in\mathbb{T}_{r_n-1}} f\left(X_i\right) > \frac{\delta}{2}\right) + \mathbb{P}\left(\frac{1}{n}\sum_{i=2^{r_n}}^{n} f\left(X_{\Pi(i)}\right) > \frac{\delta}{2}\right).$$

On the one hand, (2.3.2) leads us to

$$\mathbb{P}\left(\frac{1}{n}\sum_{i\in\mathbb{T}_{r_n-1}} f\left(X_i\right) > \frac{\delta}{2}\right) \leq \begin{cases} \exp\left(c''\delta\right)\exp\left(-c'\delta^2 n\right), & \forall n\in\mathbb{N}, \quad \text{if } \alpha < \frac{1}{2} \\ \exp\left(2c'\delta(r_n+1)\right)\exp\left(-c'\delta^2 n\right), & \forall n\in\mathbb{N}, \quad \text{if } \alpha = \frac{1}{2} \\ \exp\left(-c'\delta^2 n\right), \forall r_n > r_0, \text{if } \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}, \\ \exp\left(-c'\delta^2\frac{n}{r_n+1}\right), \forall r_n > r_0, \text{ if } \alpha = \frac{\sqrt{2}}{2}, \\ \exp\left(-c'\delta^2\frac{1}{\alpha^{2(r_n+1)}}\right), \forall r_n > r_0 - 2, \text{if } \alpha > \frac{\sqrt{2}}{2}, \end{cases} \tag{2.A.7}$$

where $r_0 := \log\left(\frac{\delta}{c_0}\right) / \log\alpha$ and c_0, c' and c'' are positive constants which depend on α, $\|f\|_\infty$ and c. c_0, $c'$and c'' differ line by line. On the other hand, for all $\lambda > 0$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=2^{r_n}}^{n} f\left(X_{\Pi(i)}\right) > \frac{\delta}{2}\right) \leq \exp\left(-\frac{\lambda\delta}{2}n\right)\mathbb{E}\left[\exp\left(\lambda\sum_{i=2^{r_n}}^{n} f\left(X_{\Pi(i)}\right)\right)\right].$$

Now let

- $\mathcal{O}_{r_n} = \{\Pi(2^{r_n}), \Pi(2^{r_n}+1), \cdots, \Pi(n)\}$,
- $\mathcal{O}^1_{r_n-1}$ the set of individuals of generation $\mathbb{G}_{r_n-1}$ which are ancestors of one individual in $\mathcal{O}_{r_n}$,
- $\mathcal{O}^2_{r_n-1}$ the set of individuals of generation $\mathbb{G}_{r_n-1}$ which are ancestors of two individuals in $\mathcal{O}_{r_n}$,
- $\mathcal{O}^{'}_{r_n}$ the set of individuals of $\mathcal{O}_{r_n}$ whose parents belong to $\mathcal{O}^1_{r_n-1}$.
- $\mathcal{O}_{r_n-1} = \mathcal{O}^1_{r_n-1} \cup \mathcal{O}^2_{r_n-1}$.

We introduce the following filtration $\tilde{\mathcal{F}}_r := \sigma(\mathcal{F}_r, \Pi(i), i \in \mathbb{T})$. Then we have

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda \sum_{i=2^{r_n}}^{n} f\left(X_{\Pi(i)}\right)\right)\right] = \mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathcal{O}^2_{r_n-1}} 2Qf\left(X_i\right) + \lambda \sum_{i\in\mathcal{O}^1_{r_n-1}} Qf\left(X_i\right)\right)\right.\\ \times \mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathcal{O}^{'}_{r_n}} f\left(X_i\right) - Qf\left(X_{[\frac{i}{2}]}\right)\right)\Big|\tilde{\mathcal{F}}_{r_n-1}\right]\\ \left.\times \mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathcal{O}^2_{r_n-1}} f\left(X_{2i}\right) + f\left(X_{2i+1}\right) - 2Qf\left(X_i\right)\right)\Big|\tilde{\mathcal{F}}_{r_n-1}\right]\right]\end{aligned}$$

Using Azuma-Bennett-Hoeffding inequality, as in **Part 1**, we get

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathcal{O}^{'}_{r_n}} f\left(X_i\right) - Qf\left(X_{[\frac{i}{2}]}\right)\right)\Big|\tilde{\mathcal{F}}_{r_n-1}\right] \leq \exp\left(\frac{\lambda^2 c^2 (1+\alpha)^2}{2}|\mathcal{O}^{'}_{r_n}|\right),$$

and

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathcal{O}^2_{r_n-1}} f\left(X_{2i}\right) + f\left(X_{2i+1}\right) - 2Qf\left(X_i\right)\right)\Big|\tilde{\mathcal{F}}_{r_n-1}\right]\\ \leq \exp\left(2\lambda^2 c^2 (1+\alpha)^2 |\mathcal{O}^2_{r_n-1}|\right).\end{aligned}$$

Now, we have

$$\begin{aligned}\exp\left(\tfrac{\lambda^2 c^2 (1+\alpha)^2}{2}|\mathcal{O}^{'}_{r_n}|\right) &+ \exp\left(2\lambda^2 c^2 (1+\alpha)^2 |\mathcal{O}^2_{r_n-1}|\right)\\ &= \exp\left(\lambda^2 c^2 (1+\alpha)^2 \left(2|\mathcal{O}^2_{r_n-1}| + \tfrac{|\mathcal{O}^{'}_{r_n}|}{2}\right)\right)\\ &\leq \exp\left(\lambda^2 c^2 (1+\alpha)^2 n\right).\end{aligned}$$

This leads us to

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i=2^{r_n}}^{n} f\left(X_{\Pi(i)}\right)\right)\right]$$

$$\leq \exp\left(\lambda^2c^2(1+\alpha)^2n\right)\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathcal{O}^2_{r_n-1}} 2Qf\left(X_i\right)+\lambda\sum_{i\in\mathcal{O}^1_{r_n-1}} Qf\left(X_i\right)\right)\right].$$

Now let

- $\mathcal{O}^{1,1}_{r_n-2}$ the set of individuals of $\mathbb{G}_{r_n-2}$ which are ancestors of one individual in $\mathcal{O}_{r_n-1}$ and one individual in $\mathcal{O}_{r_n}$,
- $\mathcal{O}^{1,2}_{r_n-2}$ the set of individuals of $\mathbb{G}_{r_n-2}$ which are ancestors of one individual in $\mathcal{O}_{r_n-1}$ and two individuals in $\mathcal{O}_{r_n}$,
- $\mathcal{O}^{2,2}_{r_n-2}$ the set of individuals of $\mathbb{G}_{r_n-2}$ which are ancestors of two individuals in $\mathcal{O}_{r_n-1}$ and two individuals in $\mathcal{O}_{r_n}$,
- $\mathcal{O}^{2,3}_{r_n-2}$ the set of individuals of $\mathbb{G}_{r_n-2}$ which are ancestors of two individuals in $\mathcal{O}_{r_n-1}$ and three individuals in $\mathcal{O}_{r_n}$,
- $\mathcal{O}^{2,4}_{r_n-2}$ the set of individuals of $\mathbb{G}_{r_n-2}$ which are ancestors of two individuals in $\mathcal{O}_{r_n-1}$ and four individuals in $\mathcal{O}_{r_n}$,
- $\mathcal{O}'_{r_n-1}$ the set of individuals of $\mathcal{O}_{r_n-1}$ whose parents belong to $\mathcal{O}^{1,1}_{r_n-2}$,
- $\mathcal{O}''_{r_n-1}$ the set of individuals of $\mathcal{O}_{r_n-1}$ whose parents belong to $\mathcal{O}^{1,2}_{r_n-2}$.

Then we have

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathcal{O}^2_{r_n-1}} 2Qf\left(X_i\right)+\lambda\sum_{i\in\mathcal{O}^1_{r_n-1}} Qf\left(X_i\right)\right)\right]=\mathbb{E}\Big[I_1\times I_2\times I_3\times I_4\times I_5\times I_6\times I_7\Big]$$

where

$$\begin{aligned} I_1 = \quad & \exp\left(\lambda\sum_{i\in\mathcal{O}^{1,1}_{r_n-2}} Q^2f(X_i)+\lambda\sum_{i\in\mathcal{O}^{1,2}_{r_n-2}} 2Q^2f(X_i)+\lambda\sum_{i\in\mathcal{O}^{2,2}_{r_n-2}} 2Q^2f(X_i)\right. \\ & \left.+\lambda\sum_{i\in\mathcal{O}^{2,3}_{r_n-2}} 3Q^2f(X_i)+\lambda\sum_{i\in\mathcal{O}^{2,4}_{r_n-2}} 4Q^2f(X_i)\right), \end{aligned}$$

$$I_2=\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathcal{O}'_{r_n-1}} Qf(X_i)-Q^2f(X_{[\frac{i}{2}]})\right)\Bigg|\tilde{\mathcal{F}}_{r_n-2}\right],$$

$$I_3=\mathbb{E}\left[\exp\left(2\lambda\sum_{i\in\mathcal{O}''_{r_n-1}} Qf(X_i)-Q^2f(X_{[\frac{i}{2}]})\right)\Bigg|\tilde{\mathcal{F}}_{r_n-2}\right],$$

$$I_4 = \mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathcal{O}^{2,2}_{r_n-1}} Qf(X_{2i}) + Qf(X_{2i+1}) - 2Q^2f(X_i)\right) \Big| \tilde{\mathcal{F}}_{r_n-2}\right],$$

$$I_5 = \mathbb{E}\left[\exp\left(\tfrac{\lambda}{2} \sum_{i\in\mathcal{O}^{2,3}_{r_n-1}} 2Qf(X_{2i}) + Qf(X_{2i+1}) - 3Q^2f(X_i)\right) \Big| \tilde{\mathcal{F}}_{r_n-2}\right],$$

$$I_6 = \mathbb{E}\left[\exp\left(\tfrac{\lambda}{2} \sum_{i\in\mathcal{O}^{2,3}_{r_n-1}} Qf(X_{2i}) + 2Qf(X_{2i+1}) - 3Q^2f(X_i)\right) \Big| \tilde{\mathcal{F}}_{r_n-2}\right],$$

$$I_7 = \mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathcal{O}^{2,4}_{r_n-1}} 2Qf(X_{2i}) + 2Qf(X_{2i+1}) - 4Q^2f(X_i)\right) \Big| \tilde{\mathcal{F}}_{r_n-2}\right].$$

Using Azuma-Bennett-Hoeffding inequality, we get

$$I_2 \times I_3 \times I_4 \times I_5 \times I_6 \times I_7$$

$$\leq \exp\left(\lambda^2 c^2 (\alpha+\alpha^2)^2 \left(\tfrac{|\mathcal{O}'_{r_n-1}|}{2} + 2|\mathcal{O}''_{r_n-1}| + 2|\mathcal{O}^{2,2}_{r_n-1}| + \tfrac{9|\mathcal{O}^{2,3}_{r_n-1}|}{2} + 8|\mathcal{O}^{2,4}_{r_n-1}|\right)\right)$$

$$\leq \exp\left(2\lambda^2 c^2(\alpha+\alpha^2)^2 n\right),$$

hence

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i=2^{r_n}}^{n} f\left(X_{\Pi(i)}\right)\right)\right] \leq \exp\left(\lambda^2 c^2 (1+\alpha)^2 n\right) \exp\left(2\lambda^2 c^2 (\alpha+\alpha^2)^2 n\right) \mathbb{E}\Big[I_1\Big].$$

Now, iterating this procedure we get

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i=2^{r_n}}^{n} f\left(X_{\Pi(i)}\right)\right)\right] \leq \exp\left(\lambda^2 c^2 (1+\alpha)^2 n \sum_{p=0}^{r_n} (2\alpha^2)^p\right) \exp\Big(\lambda c \alpha^{r_n} n\Big).$$

Then it follows as in **Part 1** that

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=2^{r_n}}^{n} f\left(X_{\Pi(i)}\right) > \frac{\delta}{2}\right) \leq$$

$$\begin{cases} \exp\left(c''\delta\right)\exp\left(-c'\delta^2 n\right), \quad \forall n \in \mathbb{N}, & \text{if } \alpha \leq \frac{1}{2}, \\ \exp\left(-c'\delta^2 n\right), \forall n \in \mathbb{N} \text{ such that } r_n > r_0, & \text{if } \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}, \\ \exp\left(-c'\delta^2 \frac{n}{r_n}\right), \forall n \in \mathbb{N} \text{ such that } r_n > r_0, & \text{if } \alpha^2 = \frac{1}{2}, \\ \exp\left(-c'\delta^2 \left(\frac{1}{\alpha}\right)^{2r_n}\right), \forall n \in \mathbb{N} \text{ such that } r_n > r_0, & \text{if } \alpha^2 > \frac{1}{2}, \end{cases} \tag{2.A.8}$$

where $r_0 := \log(\frac{\delta}{c_0})/\log(\alpha)$ and the positive constants c_0, c' and c'' depend on α, δ, c and differ line to line. Finally (2.A.7) and (2.A.8) lead us to (2.3.3).

2.A.4 Proof of Theorem 2.3.2

Let $f \in \mathcal{B}_b\big(S^3\big)$ such that $(\mu, Pf) = 0$.

Part 1. Let us first deal with $\overline{M}_{\mathbb{G}_r}(f)$. We have for all $\delta > 0$ and $\lambda > 0$,

$$\mathbb{P}\Big(\overline{M}_{\mathbb{G}_r}(f) > \delta\Big) \le \exp\Big(-\lambda\delta|\mathbb{G}_r|\Big)\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_r} f(\Delta_i)\right)\right].$$

Conditioning and using Bennet-Hoeffding inequality give us

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_r} f(\Delta_i)\right)\right] \le \exp\Big(2\lambda^2\|f\|_\infty|\mathbb{G}_r|\Big)\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_r} Pf(X_i)\right)\right].$$

Now, applying the **Part 1** of the proof of the Theorem 2.3.1 to Pf, we get (2.3.1) for $f \in \mathcal{B}_b\big(S^3\big)$.

Part 2. Let us now treat $\overline{M}_{\mathbb{T}_r}(f)$. We have for all $\delta > 0$

$$\mathbb{P}\Big(\overline{M}_{\mathbb{T}_r}(f) > \delta\Big) \le \mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(f - Pf) > \frac{\delta}{2}\right) + \mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(Pf) > \frac{\delta}{2}\right) \tag{2.A.9}$$

Now, since $\Big(M_n^\Pi(f - Pf)\Big)_{n\ge1}$ is a $\mathcal{H}_n$-martingale with bounded jumps, Azuma inequality [6], gives us for some positive constant c'

$$\mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(f - Pf) > \frac{\delta}{2}\right) \le \exp\Big(-c'\delta^2|\mathbb{T}_r|\Big).$$

For the second term in the right hand side of (2.A.9), we use the inequalities (2.3.2) with Pf instead of f. Gathering these inequalities, we get (2.3.2) for all r large enough.

Part 3. The proof for the case $\overline{M}_n^\Pi(f)$ follows the same lines as the proof of **Part 2.**

2.A.5 Proof of Proposition 2.4.2

We will prove the deviation inequality for $|\widehat{\alpha_0^r} - \alpha_0|$. The other deviation inequalities for $|\widehat{\beta}_0^r - \beta_0|, |\widehat{\alpha}_1^r - \alpha_1|$ and $|\widehat{\beta}_1^r - \beta_1|$ may be treated in a similar way.

One easily checks that

$$\widehat{\alpha_0^r} - \alpha_0 = \frac{\big(\overline{M}_{\mathbb{T}_r}(\mathbf{xy}) - \overline{M}_{\mathbb{T}_r}(P(\mathbf{xy}))\big) - \big(\overline{M}_{\mathbb{T}_r}(\mathbf{x})\big)\big(\overline{M}_{\mathbb{T}_r}(\mathbf{y}) - \overline{M}_{\mathbb{T}_r}(P(\mathbf{y}))\big)}{B_r}.$$

We have for all $\delta > 0$

$$\mathbb{P}\left(\left|\widehat{\alpha_0^r}-\alpha_0\right|>\delta\right)\leq\mathbb{P}\left(\frac{\left|\overline{M}_{\mathbb{T}_r}(\mathbf{xy}-P(\mathbf{xy}))\right|}{B_r}>\frac{\delta}{2}\right)+$$
$$\mathbb{P}\left(\frac{\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x})\right|\left|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))\right|}{B_r}>\frac{\delta}{2}\right).$$

On the one hand, for all $\gamma_1>0$ we have,

$$\mathbb{P}\left(\frac{\left|\overline{M}_{\mathbb{T}_r}(\mathbf{xy}-P(\mathbf{xy}))\right|}{B_r}>\frac{\delta}{2}\right)\leq\mathbb{P}\Big(B_r<\gamma_1\Big)+\mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{xy}-P(\mathbf{xy}))\right|>\frac{\delta\gamma_1}{2}\right).\tag{2.A.10}$$

Now, for $b=\mu_2(\theta,\sigma^2)-\mu_1(\theta)^2$, where μ_1 and μ_2 are given in (2.4.5), we have

$$\mathbb{P}\Big(B_r<\gamma_1\Big)\leq\mathbb{P}\left(-\overline{M}_{\mathbb{T}_r}(\mathbf{x}^2-\mu_2)>\frac{b-\gamma_1}{3}\right)+\mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)\right|>\frac{\sqrt{b-\gamma_1}}{\sqrt{3}}\right)$$
$$+\mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)>\frac{b-\gamma_1}{6|\mu_1|}\right).$$

We choose

$$\gamma_1<\min\left\{\frac{2b}{2+3\delta},\frac{-4+\sqrt{48b\delta^2+16}}{6\delta^2},\frac{b}{1+3\delta|\mu_1|}\right\},$$

so that

$$\frac{\delta\gamma_1}{2}<\max\left\{\frac{b-\gamma_1}{3},\frac{\sqrt{b-\gamma_1}}{\sqrt{3}},\frac{b-\gamma_1}{6|\mu_1|}\right\}.$$

Then we have

$$\mathbb{P}\left(B_r<\gamma_1\right)\leq\mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(\mu_2-\mathbf{x}^2)>\frac{\delta\gamma_1}{2}\right)+2\mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)\right|>\frac{\delta\gamma_1}{2}\right),$$

and therefore we get

$$\mathbb{P}\left(\frac{\left|\overline{M}_{\mathbb{T}_r}(\mathbf{xy}-P(\mathbf{xy}))\right|}{B_r}>\frac{\delta}{2}\right)\leq 2\mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)\right|>\frac{\delta\gamma_1}{2}\right)$$
$$+\mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(\mu_2-\mathbf{x}^2)>\frac{\delta\gamma_1}{2}\right)+\mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{xy}-P(\mathbf{xy}))\right|>\frac{\delta\gamma_1}{2}\right).$$

On the other hand, we have

$$\mathbb{P}\left(\frac{\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x})\right|\left|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))\right|}{B_r}>\frac{\delta}{2}\right)$$
$$\leq\mathbb{P}\left(\frac{\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)\right|\left|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))\right|}{B_r}>\frac{\delta}{4}\right)+\mathbb{P}\left(\frac{\left|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))\right|}{B_r}>\frac{\delta}{4|\mu_1|}\right).$$

The last term of the previous inequality can be dealt with in the same way that inequality (2.A.10), using $\gamma_3 > 0$ such that

$$\gamma_3 < \min\left\{\frac{4b|\mu_1|}{4|\mu_1|+3\delta}, \frac{2|\mu_1|\left(-4+\sqrt{\frac{24b\delta^2}{|\mu_1|}+16}\right)}{3\delta^2}, \frac{2b}{2+3\delta}\right\}.$$

For the second term, we have

$$\begin{aligned}\mathbb{P}\left(\frac{\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)\right|\left|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))\right|}{B_r} > \frac{\delta}{4}\right) &\leq \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)\right| > \frac{\sqrt{\delta}}{2}\right) \\ &+\mathbb{P}\left(\frac{\left|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))\right|}{B_r} > \frac{\sqrt{\delta}}{2}\right).\end{aligned}$$

Let $\gamma_2 > 0$ such that $\gamma_2 < \min\left\{\frac{2b}{2+3\sqrt{\delta}}, \frac{-4+\sqrt{48b\delta+16}}{b\delta}, \frac{b}{1+3\sqrt{\delta}|\mu_1|}\right\}$, in such a way that we obtain $\frac{\gamma_2\sqrt{\delta}}{2} < \max\left\{\frac{b-\gamma_2}{3}, \frac{\sqrt{b-\gamma_2}}{\sqrt{3}}, \frac{b-\gamma_2}{6|\mu_1|}\right\}$. We thus have

$$\begin{aligned}\mathbb{P}\left(\frac{\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)\right|\left|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))\right|}{B_r} > \frac{\delta}{4}\right) &\leq \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)\right| > \frac{\sqrt{\delta}}{2}\right) \\ + \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x}^2-\mu_2)\right| > \frac{\gamma_2\sqrt{\delta}}{2}\right) &+ \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))\right| > \frac{\gamma_2\sqrt{\delta}}{2}\right) \\ &+ 2\mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)\right| > \frac{\gamma_2\sqrt{\delta}}{2}\right).\end{aligned}$$

From the foregoing, we deduce that for all $\gamma > 0$ such that $\gamma < \min(\gamma_1, \gamma_2, \gamma_3)$,

$$\begin{aligned}\mathbb{P}\left(|\widehat{\alpha}_0^{(r)} - \alpha_0| > \delta\right) &\leq 2\mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)\right| > \tfrac{\delta\gamma}{2}\right) + \mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(\mu_2-\mathbf{x}^2) > \tfrac{\delta\gamma}{2}\right) \\ &+ \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{xy}-P(\mathbf{xy}))\right| > \tfrac{\delta\gamma}{2}\right) + \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)\right| > \tfrac{\sqrt{\delta}}{2}\right) \\ &+ \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x}^2-\mu_2)\right| > \tfrac{\gamma\sqrt{\delta}}{2}\right) + \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))\right| > \tfrac{\gamma\sqrt{\delta}}{2}\right) \\ &+ 2\mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)\right| > \tfrac{\gamma\sqrt{\delta}}{2}\right) + 2\mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)\right| > \tfrac{\delta\gamma}{4|\mu_1|}\right) \\ &+ \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mu_2-\mathbf{x}^2) > \tfrac{\delta\gamma}{4|\mu_1|}\right|\right) + \mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))\right| > \tfrac{\delta\gamma}{4|\mu_1|}\right).\end{aligned}$$

Now, using (2.2.7) we get

$$\mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{xy}-P(\mathbf{xy}))\right| > \frac{\delta\gamma}{2}\right) \leq \frac{c}{\delta^4\gamma^4}\left(\frac{1}{4}\right)^{r+1},$$

$$\mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))\right| > \frac{\delta\gamma}{4|\mu_1|}\right) \leq \frac{c\mu_1^4}{\delta^4\gamma^4}\left(\frac{1}{4}\right)^{r+1},$$

and

$$\mathbb{P}\left(\left|\overline{M}_{\mathbb{T}_r}(\mathbf{y}-P(\mathbf{y}))\right| > \frac{\gamma\sqrt{\delta}}{2}\right) \leq \frac{c}{\delta^2\gamma^4}\left(\frac{1}{4}\right)^{r+1}.$$

where the constant c can be found as in remark 2.2.4

Finally, the other terms, that is the terms related to $\overline{M}_{\mathbb{T}_r}(\mathbf{x}^2-\mu_2)$ and $\overline{M}_{\mathbb{T}_r}(\mathbf{x}-\mu_1)$, can be bounded as in Corollary 2.2.2 and this ends the proof.

Appendix 2.B Some useful results

Let us gather here for the convenience of the readers various Theorems useful to establish LIL, ASFCLT or MDP.

First, let us enunciate Azuma-Bennett-Hoeffding inequality [6, 16, 70].

Lemma 2.B.1. *Let X be a real-valued and centered random variable such that $a \leq X \leq b$ a.s., with $a < b$. Then for all $\lambda > 0$, we have*

$$\mathbb{E}\left[\exp\left(\lambda X\right)\right] \leq \exp\left(\frac{\lambda^2(b-a)^2}{8}\right).$$

Lemma 2.B.2. *Let (E, d) a metric space. Let (Z_n) a sequence of random variables values in E, (v_n) a rate and $g : \mathcal{D}_E \subset E \to \mathbb{R}$ continuous. Let $z \in E$ a deterministic value.*

$$If \quad Z_n \underset{v_n}{\overset{\text{superexp}}{\Longrightarrow}} z, \quad then \quad g(Z_n) \underset{v_n}{\overset{\text{superexp}}{\Longrightarrow}} g(z).$$

Proof. For all $\delta > 0$, there exists (see e.g [106], proof of Theorem 2.3) $\alpha_0(\delta) > 0$

$$\mathbb{P}\Big(\left|g(Z_n) - g(z)\right| > \delta\Big) \leq \mathbb{P}\Big(d(Z_n, z) > \alpha_0(\delta)\Big). \tag{2.B.1}$$

Indeed, since g is continuous, for all $\delta > 0$, there exists $\alpha_0(\delta) > 0$ such that

$$|g(x) - g(z)| \leq \delta \quad \text{whenever} \quad d(x, z) \leq \alpha_0(\delta).$$

We then have

$$\{\omega : d(Z_n(\omega), z) \leq \alpha_0(\delta)\} \subset \{\omega : |g(Z_n(\omega)) - g(z)| \leq \delta\}$$

and therefore inequality (2.B.1). Now, the result of the lemma follows since $Z_n \underset{v_n}{\overset{\text{superexp}}{\Longrightarrow}} z$. ∎

Let $M = (M_n, \mathcal{H}_n, n \geq 0)$ be a centered square integrable martingale defined on a probability space $(\Omega, \mathcal{H}, \mathbb{P})$ and $(\langle M\rangle_n)$ its bracket. We recall some limit theorems for martingale used intensively in this chapter.

We recall the following result due to W. F. Stout (Theorem 3 in [102]).

Theorem 2.B.3. *Let (M_n) such that $M_0 = 0$. If $\langle M\rangle_n \to \infty$ a.s. and*

$$\sum_{n=1}^{\infty} \frac{2\log\log\langle M\rangle_n}{K_n^2\langle M\rangle_n}\mathbb{E}\left[(M_n - M_{n-1})^2\,\mathbf{1}_{\left\{(M_n-M_{n-1})^2 > \frac{K_n^2\langle M_n\rangle}{2\log\log\langle M\rangle_n}\right\}}\Big|\mathcal{H}_{n-1}\right] < \infty \quad a.s.$$

where K_n are $\mathcal{H}_{n-1}$ measurable and $K_n \to 0$ a.s., then

$$\limsup \frac{M_n}{\sqrt{2\langle M\rangle_n \log\log\langle M\rangle_n}} = 1 \quad a.s..$$

We recall the following result due to F. Chaabane (Corollary 2.2 in [26]).

Theorem 2.B.4. *Let (V_n) be a $(\mathcal{H}_n)$-predictable increasing process such that*

H-1 $V_n^{-2}\langle M\rangle_n \underset{n\to\infty}{\longrightarrow} 1$, *a.s.*

H-2 for all $\varepsilon > 0$, $\sum_{n\geq 1} V_n^{-2}\mathbb{E}\Big[(M_n - M_{n-1})^2\,\mathbf{1}_{|M_n-M_{n-1}|>\varepsilon V_n}\Big|\mathcal{H}_{n-1}\Big] < \infty;\quad a.s.$

H-3 for some $a > 1$, $\sum_{n\geq 1} V_n^{-2a}\mathbb{E}\Big[(M_n - M_{n-1})^{2a}\,\mathbf{1}_{|M_n-M_{n-1}|\leq V_n}\Big|\mathcal{H}_{n-1}\Big] < \infty$, *a.s.*

Then M_n satisfies an ASFCLT, that is, for almost all ω, the weighted random measures

$$W_N(\omega, \bullet) = (\log V_N^2)^{-1}\sum_{n=1}^{N}\left(1 - \frac{V_n^2}{V_{n+1}^2}\right)\delta_{\{\psi_n(\omega)\in\bullet\}}$$

associated to the continuous processes $\Psi_n(\omega) = \{\Psi_n(\omega, t), 0 \leq t \leq 1\}$ defined by

$$\Psi_n(\omega, t) = V_n^{-1}\{M_k + (V_{k+1}^2 - V_k^2)^{-1}(tV_n^2 - V_k^2)(M_{k+1} - M_k)\},$$

when $V_k^2 \leq tV_n^2 < V_{k+1}^2$, $0 \leq k \leq n-1$, weakly converge to the Wiener measure on $\mathcal{C}([0,1],\mathbb{R})$.

Let us enunciate the following which corresponds to the unidimensional case of Theorem 1 in [38].

Proposition 2.B.5. *Let (b_n) a sequence satisfying*

$$b_n \quad \textit{is increasing,} \quad \frac{b_n}{\sqrt{n}} \longrightarrow +\infty, \quad \frac{b_n}{n} \longrightarrow 0,$$

such that $c(n) := n/b_n$ is non-decreasing, and define the reciprocal function $c^{-1}(t)$ by

$$c^{-1}(t) := \inf\{n \in \mathbb{N} : c(n) \geq t\}.$$

Under the following conditions:

(C1) *there exists* $Q \in \mathbb{R}_+^*$ *such that* $\dfrac{\langle M \rangle_n}{n} \overset{\text{superexp}}{\underset{b_n^2/n}{\Longrightarrow}} Q$;

(C2) $\displaystyle \limsup_{n \to +\infty} \frac{n}{b_n^2} \log \left(n \operatorname*{ess\,sup}_{1 \le k \le c^{-1}(b_{n+1})} \mathbb{P}(|M_k - M_{k-1}| > b_n | \mathcal{H}_{k-1}) \right) = -\infty$;

(C3) *for all* $a > 0$ $\displaystyle \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left(|M_k - M_{k-1}|^2 \mathbf{1}_{\{|M_k - M_{k-1}| \ge a \frac{n}{b_n}\}} | \mathcal{H}_{k-1} \right) \overset{\text{superexp}}{\underset{b_n^2/n}{\Longrightarrow}} 0$;

$(M_n/b_n)_{n \in \mathbb{N}}$ *satisfies the MDP in* $\mathbb{R}$ *with the speed* b_n^2/n *and the rate function* $I(x) = \dfrac{x^2}{2Q}$.

Chapitre 3

Deviation inequalities and moderate deviations for estimators of parameters in bifurcating autoregressive models

3.1 Motivation and context

Bifurcating autoregressive processes (BAR, for short) are an adaptation of autoregressive processes, when the data have a binary tree structure. They were first introduced by Cowan and Staudte [28] for cell lineage data where each individual in one generation gives rise to two offspring in the next generation.

In their paper, the original BAR process was defined as follows. The initial cell is labelled 1, and the two offspring of cell k are labelled $2k$ and $2k+1$. If X_k denotes an observation of some characteristic of individual k then the first order BAR process is given, for all $k \geq 1$, by

$$\begin{cases} X_{2k} = a + bX_k + \varepsilon_{2k} \\ X_{2k+1} = a + bX_k + \varepsilon_{2k+1}. \end{cases}$$

The noise sequence $(\varepsilon_{2k}, \varepsilon_{2k+1})$ represents environmental effects, while a, b are unknown real parameters, with $|b| < 1$, related to inherited effects. The driven noise $(\varepsilon_{2k}, \varepsilon_{2k+1})$ was originally supposed to be independent and identically distributed with normal distribution. But since two sister cells are in the same environment at their birth, ε_{2k} and ε_{2k+1} are allowed to be correlated, inducing a correlation between sister cells, distinct from the correlation inherited from their mother.

Several extensions of the model have been proposed and various estimators are studied in the literature for the unknown parameters, see for instance [13],[10], [11], [12], [14], [15]. See [18] for relevant references.

Recently, there have been many studies of the asymmetric BAR process, that is when the quantitative characteristics of the even and odd sisters are allowed to depend from their mother's through different sets of parameters.

Guyon [66] proposes an interpretation of the asymmetric BAR process as a bifurcating Markov chain, which allows him to derive laws of large numbers and central limit theorems for the least squares estimators of the unknown parameters of the process. This Markov chain approach was further developed by Delmas and Marsalle [33], where the cells are allowed to die. They defined the genealogy of the cells through a Galton-Watson process, studying the same model on the Galton Watson tree instead of a binary tree.

Another approach based on martingales theory was proposed by Bercu, de Saporta and Gégout-Petit [18], to sharpen the asymptotic analysis of Guyon under weaker assumptions. It must be pointed out that missing data are not dealt with in this work. To take into account possibly missing data in the estimation procedure de Saporta et al. [32] use a two-type Galton-Watson process to model the genealogy.

Our objective in this chapter is to go a step further by

- studying the moderate deviation principle (MDP, for short) of the least squares estimators of the unknown parameters of a general asymmetric pth-order bifurcating autoregressive processes. More precisely we are interested in the asymptotic estimations of

$$\mathbb{P}\left(\frac{\sqrt{n}}{b_n}(\Theta_n - \Theta) \in A\right)$$

  where Θ_n denotes the estimator of the unknown parameter of interest Θ, A is a given domain of deviation, $(b_n > 0)$ is some sequence denoting the scale of deviation. When $b_n = 1$ this is exactly the estimation of the central limit theorem. When $b_n = \sqrt{n}$, it becomes the *large deviation*. And when $1 \ll b_n \ll \sqrt{n}$, this is the so called *moderate deviations*. Usually, MDP has a simpler rate function inherited from the approximated Gaussian process, and holds for a larger class of dependent random variables than the LDP.

  Let us notice that we have not found studies exactly on this question in the literatures, except the recent work of Biteski et al. [22]. However, technically, we are much inspired from two lines of studies:

  1. the work of Bercu et al. [18] on the almost sure convergence of the estimators with the quadratic strong law and the central limit theorem;
  2. the works of Dembo [34], and Worms [114], [115], [113] on the one hand, and of the papers of Puhalskii [91] and Djellout [38] on the other hand, about the MDP for martingales.

- giving deviation inequalities for the estimator of bifurcating autoregressive processes, which are important for a rigorous non asymptotic statistical study, i.e. for all $x > 0$
$$\mathbb{P}\left(||\Theta_n - \Theta|| \geq x\right) \leq e^{-C_n(x)},$$
where $C_n(x)$ will crucially depends on our set of assumptions. The upper bounds in this inequality hold for arbitrary n and x (not a limit relation, unlike the MDP results), hence they are much more practical (in statistics). Deviation inequalities for estimators of the parameters associated with linear regression, autoregressive and branching processes are investigated by Bercu and Touati [20]. In the martingale case, deviation inequalities for self normalized martingale have been developed by de la Peña et al. [29]. We also refer to the work of Ledoux [74] for precise credit and references. This type of inequalities are equally well motivated by theoretical question as by numerous applications in different fields including the analysis of algorithms, mathematical physics and empirical processes. For some applications in non asymptotic model selection problems we refer to Massart [78].

This chapter is organized as follows. First of all, in Section 3.2, we introduce the BAR(p) (that is p-order bifurcating autoregressive) model as well as the least squares estimators for the parameters of observed BAR(p) process and some related notation and hypothesis. In Section 3.3, we state our main results on the deviation inequalities and MDP of our estimators. Section 3.4 is dedicated to the superexponential convergence of the quadratic variation of the martingale, this section contains exponential inequalities which are crucial for the proof of the deviation inequalities. The proofs of the main results are postponed in section 3.5.

3.2 Notation and Hypothesis

In all the sequel, let $p \in \mathbb{N}^*$. We consider the asymmetric BAR(p) (p-order bifurcating autoregressive) process given, for all $n \geq 2^{p-1}$, by

$$\begin{cases} X_{2n} = a_0 + \sum_{k=1}^{p} a_k X_{[\frac{n}{2^{k-1}}]} + \varepsilon_{2n} \\ X_{2n+1} = b_0 + \sum_{k=1}^{p} b_k X_{[\frac{n}{2^{k-1}}]} + \varepsilon_{2n+1}, \end{cases} \tag{3.2.1}$$

where the notation $[x]$ stands for the largest integer less than or equal to the real x. The initial states $\{X_k, 1 \leq k \leq 2^{p-1} - 1\}$ are the ancestors while $(\varepsilon_{2n}, \varepsilon_{2n+1})$ is the driven noise of the process. The parameters $(a_0, a_1, \cdots, a_p)$ and $(b_0, b_1, \cdots, b_p)$ are unknown real vectors.

The BAR(p) process can be rewritten in the abbreviated vector form given, for all $n \geq 2^{p-1}$, by

$$\begin{cases} \mathbb{X}_{2n} = A\mathbb{X}_n + \eta_{2n} \\ \mathbb{X}_{2n+1} = B\mathbb{X}_n + \eta_{2n+1} \end{cases} \tag{3.2.2}$$

where the regression vector $\mathbb{X}_n = \left(X_n, X_{[\frac{n}{2}]}, \cdots, X_{[\frac{n}{2^{p-1}}]}\right)^t$, $\eta_{2n} = (a_0 + \varepsilon_{2n})e_1$, $\eta_{2n+1} = (b_0 + \varepsilon_{2n+1})e_1$, with $e_1 = (1, 0, \cdots, 0)^t \in \mathbb{R}^p$. Moreover, A and B are the $p \times p$ companion matrices

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_p \\ 1 & 0 & \cdots & 0 \\ 0 & . & . & . \\ 0 & . & 1 & . \end{pmatrix} \qquad \text{and} \quad B = \begin{pmatrix} b_1 & b_2 & \cdots & b_p \\ 1 & 0 & \cdots & 0 \\ 0 & . & . & . \\ 0 & . & 1 & . \end{pmatrix}.$$

In the sequel, we shall assume that the matrices A and B satisfy the contraction property

$$\beta = \max(||A||, ||B||) < 1, \tag{3.2.3}$$

where for any matrix M the notation M^t, $\|M\|$ and $\mathrm{Tr}(M)$ stand for the transpose, the Euclidean norm (that is $\|M\| = \sqrt{\lambda_{max}(M^t M)}$) and the trace of M, respectively.

One can see this BAR(p) process as a pth-order autoregressive process on a binary tree, where each vertex represents an individual or cell, vertex 1 being the original ancestor. For all $n \geq 1$, denote the n-th generation by $\mathbb{G}_n = \{2^n, 2^n + 1, \cdots, 2^{n+1} - 1\}$.

In particular, $\mathbb{G}_0 = \{1\}$ is the initial generation and $\mathbb{G}_1 = \{2, 3\}$ is the first generation of offspring from the first ancestor. Let $\mathbb{G}_{r_n}$ be the generation of individual n, which means that $r_n = [\log_2(n)]$. Recall that the two offspring of individual n are labelled $2n$ and $2n + 1$, or conversely, the mother of the individual n is $[n/2]$. More generally, the ancestors of individual n are $[n/2], [n/2^2], \cdots, [n/2^{r_n}]$. Furthermore, denote by

$$\mathbb{T}_n = \bigcup_{k=0}^{n} \mathbb{G}_k$$

the subtree of all individuals from the original individual up to the n-th generation. We denote by $\mathbb{T}_{n,p} = \{k \in \mathbb{T}_n, k \geq 2^p\}$ the subtree of all individuals up to the nth generation without $\mathbb{T}_{p-1}$. One can observe that , for all $n \geq 1$, $\mathbb{T}_{n,0} = \mathbb{T}_n$ and for all $p \geq 1$, $\mathbb{T}_{p,p} = \mathbb{G}_p$.

The BAR(p) process can be rewritten, for all $n \geq 2^{p-1}$, in the matrix form

$$Z_n = \theta^t Y_n + V_n$$

where

$$Z_n = \begin{pmatrix} X_{2n} \\ X_{2n+1} \end{pmatrix}, \quad Y_n = \begin{pmatrix} 1 \\ \mathbb{X}_n \end{pmatrix}, \quad V_n = \begin{pmatrix} \varepsilon_{2n} \\ \varepsilon_{2n+1} \end{pmatrix},$$

and the $(p+1)\times 2$ matrix parameter θ is given by

$$\theta = \begin{pmatrix} a_0 & b_0 \\ a_1 & b_1 \\ . & . \\ . & . \\ a_p & b_p \end{pmatrix}.$$

As in Bercu et al.[18], we introduce the least squares estimator $\widehat{\theta}_n$ of θ, from the observation of all individuals up to the n-th generation that is the complete sub-tree $\mathbb{T}_n$, for all $n \geq p$

$$\widehat{\theta}_n = S_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1,p-1}} Y_k Z_k^t, \tag{3.2.4}$$

where the $(p+1)\times(p+1)$ matrix S_n is defined as

$$S_n = \sum_{k \in \mathbb{T}_{n,p-1}} Y_k Y_k^t = \sum_{k \in \mathbb{T}_{n,p-1}} \begin{pmatrix} 1 & \mathbb{X}_k^t \\ \mathbb{X}_k & \mathbb{X}_k \mathbb{X}_k^t \end{pmatrix}. \tag{3.2.5}$$

We assume, without loss of generality, that for all $n \geq p-1$, S_n is invertible. In all that follows, we shall make a slight abuse of notation by identifying θ as well as $\widehat{\theta}_n$ to

$$\operatorname{vec}(\theta) = \begin{pmatrix} a_0 \\ \vdots \\ a_p \\ b_0 \\ \vdots \\ b_p \end{pmatrix} \quad \text{and} \quad \operatorname{vec}(\widehat{\theta}_n) = \begin{pmatrix} \widehat{a}_{0,n} \\ \vdots \\ \widehat{a}_{p,n} \\ \widehat{b}_{0,n} \\ \vdots \\ \widehat{b}_{p,n} \end{pmatrix}.$$

Let $\Sigma_n = I_2 \otimes S_n$, where $\otimes$ stands for the matrix Kronecker product. Therefore, we deduce from (3.2.4) that

$$\widehat{\theta}_n = \Sigma_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1,p-1}} \operatorname{vec}(Y_k Z_k^t) = \Sigma_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1,p-1}} \begin{pmatrix} X_{2k} \\ X_k \mathbb{X}_{2k} \\ X_{2k+1} \\ X_k \mathbb{X}_{2k+1} \end{pmatrix}. \tag{3.2.6}$$

Consequently, (3.2.2) yields to

$$\widehat{\theta}_n - \theta = \Sigma_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1,p-1}} \begin{pmatrix} \varepsilon_{2k} \\ \varepsilon_{2k} \mathbb{X}_k \\ \varepsilon_{2k+1} \\ \varepsilon_{2k+1} \mathbb{X}_k \end{pmatrix}. \tag{3.2.7}$$

Denote by $\mathbb{F} = (\mathcal{F}_n)$ the natural filtration associated with the BAR(p) process, which means that $\mathcal{F}_n$ is the $\sigma-$algebra generated by the individuals up to the n-th generation, in other words $\mathcal{F}_n = \sigma\{X_k, k \in \mathbb{T}_n\}$.

For the initial states, if we denote by $\overline{X}_1 = \max\left\{\|\mathbb{X}_k\|, k \leq 2^{p-1}\right\}$ with the convention that $X_0 = 0$, we introduce the following hypothesis.

(Xa) For some $a > 2$, there exists $\tau > 0$ such that

$$\mathbb{E}\left[\exp\left(\tau \overline{X}_1^a\right)\right] < \infty.$$

This assumption implies the weaker Gaussian integrability condition

(X2) There is $\tau > 0$ such that

$$\mathbb{E}\left[\exp\left(\tau \overline{X}_1^2\right)\right] < \infty.$$

For the noise $(\varepsilon_{2n}, \varepsilon_{2n+1})$ the assumption may be of two types.

1. In the first case we will assume the independence of the noise which allows us to impose less restrictive conditions on the exponential integrability of the noise.

   Case 1: We shall assume that $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$ forms a sequence of independent and identically distributed bi-variate centered random variables with covariance matrix Γ associated with $(\varepsilon_{2n}, \varepsilon_{2n+1})$, given by

$$\Gamma = \begin{pmatrix} \sigma^2 & \rho \\ \rho & \sigma^2 \end{pmatrix}, \quad \text{where } \sigma^2 > 0 \text{ and } |\rho| < \sigma^2. \tag{3.2.8}$$

   For all $n \geq p-1$ and for all $k \in \mathbb{G}_n$, we denote

$$\mathbb{E}[\varepsilon_k^2] = \sigma^2, \ \ \mathbb{E}[\varepsilon_k^4] = \tau^4, \ \ \mathbb{E}[\varepsilon_{2k}\varepsilon_{2k+1}] = \rho, \ \ \mathbb{E}[\varepsilon_{2k}^2\varepsilon_{2k+1}^2] = \nu^2,$$

   where $\tau^4 > 0$, $\nu^2 < \tau^4$. In addition, we assume that the condition **(X2)** on the initial state is satisfied and

   (G2) one can find $\gamma > 0$ and $c > 0$ such that for all $n \geq p-1$, for all $k \in \mathbb{G}_n$ and for all $|t| \leq c$

$$\mathbb{E}\left[\exp t\left(\varepsilon_k^2 - \sigma^2\right)\right] \leq \exp\left(\frac{\gamma t^2}{2}\right).$$

   In this case, we impose the following hypothesis on the scale of the deviation

(V1) (b_n) will denote an increasing sequence of positive real numbers such that

$$b_n \uparrow +\infty$$

and for β given by (3.2.3)

- if $\beta \leq \frac{1}{2}$, the sequence (b_n) is such that $\dfrac{b_n \log n}{\sqrt{n}} \downarrow 0$,
- if $\beta > \frac{1}{2}$, the sequence (b_n) is such that $(b_n\sqrt{\log n})\beta^{\frac{r_n+1}{2}} \downarrow 0$.

2. In contrast with the first case, in the second case, we will not assume that the sequence $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$ is i.i.d. The price to pay for giving up this i.i.d. assumption is higher exponential moments. Indeed we need them to make use of the MDP for martingale, especially to prove the Lindeberg condition via Lyapunov one.

   Case 2: We shall assume that for all $n \geq p-1$ and for all $j \in \mathbb{G}_{n+1}$, $\mathbb{E}[\varepsilon_j|\mathcal{F}_n] = 0$ and for all different $k, l \in \mathbb{G}_{n+1}$ with $[\frac{k}{2}] \neq [\frac{l}{2}]$, ε_k and ε_l are conditionally independent given $\mathcal{F}_n$. And we will use the same notation as in the case 1: for all $n \geq p-1$ and for all $k \in \mathbb{G}_{n+1}$

   $$\mathbb{E}[\varepsilon_k^2|\mathcal{F}_n] = \sigma^2,\ \mathbb{E}[\varepsilon_k^4|\mathcal{F}_n] = \tau^4,\ \mathbb{E}[\varepsilon_{2k}\varepsilon_{2k+1}|\mathcal{F}_n] = \rho,\ \mathbb{E}[\varepsilon_{2k}^2\varepsilon_{2k+1}^2|\mathcal{F}_n] = \nu^2\ a.s.$$

   where $\tau^4 > 0$, $\nu^2 < \tau^4$ and we use also Γ for the conditional covariance matrix associated with $(\varepsilon_{2n}, \varepsilon_{2n+1})$. In this case, we assume that the condition **(Xa)** on the initial state is satisfied, and we shall make use of the following hypotheses:

   (Ea) for some $a > 2$, there exist $t > 0$ and $E > 0$ such that for all $n \geq p-1$ and for all $k \in \mathbb{G}_{n+1}$,

   $$\mathbb{E}\left[\exp\left(t|\varepsilon_k|^{2a}\right)|\mathcal{F}_n\right] \leq E < \infty \quad a.s.$$

   Throughout this case, we introduce the following hypothesis on the scale of the deviation

   (V2) (b_n) will denote an increasing sequence of positive real numbers such that

   $$b_n \uparrow +\infty,$$

   and for β given by (3.2.3)

   - if $\beta^2 < \frac{1}{2}$, the sequence (b_n) is such that $\dfrac{b_n \log n}{\sqrt{n}} \downarrow 0$,
   - if $\beta^2 = \frac{1}{2}$, the sequence (b_n) is such that $\dfrac{b_n (\log n)^{3/2}}{\sqrt{n}} \downarrow 0$,

- if $\beta^2 > \frac{1}{2}$, the sequence (b_n) is such that $(b_n \log n)\beta^{r_n+1} \downarrow 0$.

Remark 3.2.1. *The conditions on the scale of the deviation in case 2 is less restrictive than in case 1, since we assume more integrability conditions. These conditions on the scale of the deviation naturally appear from the calculations (see the proof of Proposition 3.4.1). Specifically, the* log *term comes from the crossing of the probability of a sum to the sum of probability.*

Remark 3.2.2. *From [42] or [74], we deduce with* **(Ea)** *that*

(N1) *there is $\phi > 0$ such that for all $n \geq p-1$, for all $k \in \mathbb{G}_{n+1}$ and for all $t \in \mathbb{R}$,*

$$\mathbb{E}\Big[\exp\left(t\varepsilon_k\right)|\mathcal{F}_n\Big] < \exp\left(\frac{\phi t^2}{2}\right), \qquad a.s.$$

We have the same conclusion in case 1 without the conditioning, i.e.

(G1) *there is $\phi > 0$ such that for all $n \geq p-1$, for all $k \in \mathbb{G}_n$ and for all $t \in \mathbb{R}$,*

$$\mathbb{E}\Big[\exp(t\varepsilon_k)\Big] < \exp\left(\frac{\phi t^2}{2}\right).$$

Remark 3.2.3. *Armed with the recent development in the theory of transportation inequalities, exponential integrability and functional inequalities (see Ledoux [74], Gozlan [60] and Gozlan and Leonard [61]), we can prove that a sufficient condition for hypothesis* **(G2)** *to hold is existence of $t_0 > 0$ such that for all $n \geq p-1$ and for all $k \in \mathbb{G}_n$, $\mathbb{E}\left[\exp(t_0\varepsilon_k^2)\right] < \infty$.*

We now turn to the estimation of the parameters σ^2 and ρ. On the one hand, we propose to estimate the conditional variance σ^2 by

$$\widehat{\sigma}_n^2 = \frac{1}{2|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}} ||\widehat{V}_k||^2 = \frac{1}{2|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}} (\widehat{\varepsilon}_{2k}^2 + \widehat{\varepsilon}_{2k+1}^2) \tag{3.2.9}$$

where for all $n \geq p-1$ and all $k \in \mathbb{G}_n$, $\widehat{V}_k^t = (\widehat{\varepsilon}_{2k}, \widehat{\varepsilon}_{2k+1})^t$ with

$$\begin{cases} \widehat{\varepsilon}_{2k} = X_{2k} - \widehat{a}_{0,n} - \sum_{i=1}^p \widehat{a}_{i,n} X_{[\frac{k}{2^{i-1}}]} \\ \widehat{\varepsilon}_{2k+1} = X_{2k+1} - \widehat{b}_{0,n} - \sum_{i=1}^p \widehat{b}_{i,n} X_{[\frac{k}{2^{i-1}}]} \end{cases}$$

We also introduce the following

$$\sigma_n^2 = \frac{1}{2|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p}} (\varepsilon_{2k}^2 + \varepsilon_{2k+1}^2). \tag{3.2.10}$$

On the other hand, we estimate the conditional covariance ρ by

$$\widehat{\rho}_n = \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \widehat{\varepsilon}_{2k} \widehat{\varepsilon}_{2k+1} \tag{3.2.11}$$

We also introduce the following

$$\rho_n = \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p}} \varepsilon_{2k} \varepsilon_{2k+1}. \tag{3.2.12}$$

In order to establish the MDP results of our estimators, we shall make use of a martingale approach. For all $n \geq p$, denote

$$M_n = \sum_{k \in \mathbb{T}_{n-1,p-1}} \begin{pmatrix} \varepsilon_{2k} \\ \varepsilon_{2k} \mathbb{X}_k \\ \varepsilon_{2k+1} \\ \varepsilon_{2k+1} \mathbb{X}_k \end{pmatrix} \in \mathbb{R}^{2(p+1)}. \tag{3.2.13}$$

We can clearly rewrite (3.2.7) as

$$\widehat{\theta}_n - \theta = \Sigma_{n-1}^{-1} M_n. \tag{3.2.14}$$

We know from Bercu & al. [18] that (M_n) is a square integrable martingale adapted to the filtration $\mathbb{F} = (\mathcal{F}_n)$. Its increasing process is given for all $n \geq p$ by

$$\langle M \rangle_n = \Gamma \otimes S_{n-1}$$

where S_n is given in (3.2.5) and Γ is given in (3.2.8).

We recall that for a sequence of random variables $(Z_n)_n$ on $\mathbb{R}^{d \times p}$, we say that $(Z_n)_n$ converges $(b_n^2)-$superexponentially fast in probability to some random variable Z if, for all $\delta > 0$,

$$\limsup_{n \to \infty} \frac{1}{b_n^2} \log \mathbb{P}\Big(\|Z_n - Z\| > \delta \Big) = -\infty.$$

This exponential convergence with speed b_n^2 will be shortened as

$$Z_n \underset{b_n^2}{\overset{\text{superexp}}{\Longrightarrow}} Z.$$

We follow Dembo and Zeitouni [35] for the language of the large deviations, throughout this chapter. For completeness, let us recall the definition of a MDP: let (b_n) be an increasing sequence of positive real numbers such that

$$b_n \uparrow \infty \qquad \text{and} \qquad \frac{b_n}{\sqrt{n}} \downarrow 0. \tag{3.2.15}$$

We say that a sequence of centered random variables $(M_n)_n$ with topological state space $(S, \mathcal{S})$ satisfies a MDP with speed b_n^2 and rate function $I : S \to \mathbb{R}_+^*$ if for each $A \in \mathcal{S}$,

$$-\inf_{x\in A^o} I(x) \leq \liminf_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\sqrt{n}}{b_n} M_n \in A\right)$$
$$\leq \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\sqrt{n}}{b_n} M_n \in A\right) \leq -\inf_{x\in\overline{A}} I(x),$$

here A^o and $\overline{A}$ denote the interior and closure of A respectively.

Before the presentation of the main results, let us fix some more notation. Let $\overline{a} = \dfrac{a_0 + b_0}{2}$, $\overline{a^2} = \dfrac{a_0^2 + b_0^2}{2}$, $\overline{A} = \dfrac{A+B}{2}$ and $e_1 = (1, 0, \cdots, 0)^t \in \mathbb{R}^p$. We denote

$$\Xi = \overline{a}(I_p - \overline{A})^{-1} e_1, \tag{3.2.16}$$

and Λ the unique solution of the equation

$$\Lambda = T + \frac{1}{2}(A\Lambda A^t + B\Lambda B^t) \tag{3.2.17}$$

where

$$T = \left(\sigma^2 + \overline{a^2}\right) e_1 e_1^t + \frac{1}{2}\left(a_0\left(A\Xi e_1^t + e_1 \Xi^t A^t\right) + b_0\left(B\Xi e_1^t + e_1\Xi^t B^t\right)\right), \tag{3.2.18}$$

We also introduce the following matrices L and Σ given by

$$L = \begin{pmatrix} 1 & \Xi \\ \Xi & \Lambda \end{pmatrix} \quad \text{and} \quad \Sigma = I_2 \otimes L. \tag{3.2.19}$$

Remark 3.2.4. *In the special case $p = 1$, we have*

$$\Xi = \frac{\overline{a}}{1 - \overline{b}}, \quad \textit{and} \quad \Lambda = \frac{\overline{a^2} + \sigma^2 + 2\Xi\overline{ab}}{1 - \overline{b^2}},$$

where $\overline{ab} = \dfrac{a_0 a_1 + b_0 b_1}{2}$, $\overline{b} = \dfrac{a_1 + b_1}{2}$, $\overline{b^2} = \dfrac{a_1^2 + b_1^2}{2}$.

3.3 Main results

Let us present now the main results of this chapter. In the following theorem, we will give the deviation inequalities of the estimator of the parameters, useful for non asymptotic statistical studies.

Theorem 3.3.1.

(i) *In case 1 we have for all $\delta > 0$, for all $b > 0$ such that $b < \|\Sigma\|/(1+\delta)$ and for all $n > c_0 \frac{\log((1-\beta)b\delta)}{\log(\beta)} + p - 1$*

$$\mathbb{P}\left(\|\widehat{\theta}_n - \theta\| > \delta\right) \leq \begin{cases} c_1 \exp\left(-\frac{c_2(\delta b)^2}{c_3+(\delta b)} \frac{2^n}{(n-1)^2}\right) & \text{if } \beta < \frac{1}{2} \\ c_1(n-1) \exp\left(\frac{-c_2(\delta b)^2}{c_3+(\delta b)} \frac{2^n}{(n-1)^2}\right) & \text{if } \beta = \frac{1}{2} \\ c_1(n-1) \exp\left(\frac{-c_2(\delta b)^2}{c_3+(\delta b)} \frac{1}{(n-1)\beta^n}\right) & \text{if } \beta > \frac{1}{2}, \end{cases} \tag{3.3.1}$$

where the constants c_0, c_1, c_2 and c_3 depend on σ^2, β, γ and ϕ and are such that $c_0 \in \mathbb{R}^*$, $c_1, c_2 > 0$, $c_3 \geq 0$.

(ii) *In case 2 we have for all $\delta > 0$, for all $b > 0$ such that $b < \|\Sigma\|/(1+\delta)$ and for all $n > c_0 \frac{\log((1-\beta)b\delta)}{\log(\beta)} + p - 1$*

$$\mathbb{P}\left(\|\widehat{\theta}_n - \theta\| > \delta\right) \leq \begin{cases} c_1 \exp\left(-\frac{c_2(\delta b)^2}{c_3+c_4(\delta b)} \frac{2^n}{(n-1)^2}\right) & \text{if } \beta < \frac{\sqrt{2}}{2} \\ c_1 \exp\left(-\frac{c_2(\delta b)^2}{c_3+c_4(\delta b)} \frac{2^n}{(n-1)^3}\right) & \text{if } \beta = \frac{\sqrt{2}}{2} \\ c_1 \exp\left(-\frac{c_2(\delta b)^2}{c_3+c_4(\delta b)} \frac{1}{(n-1)^2\beta^{2n}}\right) & \text{if } \beta > \frac{\sqrt{2}}{2}, \end{cases} \tag{3.3.2}$$

where the constants c_0, c_1, c_2, c_3, and c_4 depend on σ^2, β, γ and ϕ and are such that $c_0 \in \mathbb{R}^*$, $c_1, c_2 > 0$, $c_3, c_4 \geq 0$, $(c_3, c_4) \neq (0,0)$.

Remark 3.3.2. *One can notice that the estimate (3.3.2) is stronger than estimate (3.3.1). This is due to the fact that the integrability condition in case 2 is stronger than integrability condition in case 1.*

Remark 3.3.3. *The upper bounds in previous theorem hold for arbitrary $n \geq n_0 + p - 1$, where n_0 can be computed explicitly (not a limit relation, unlike the below results). Hence, these estimates are much more practical (in non asymptotic statistics).*

In the next result, we will present the MDP of the estimator $\widehat{\theta}_n$.

Theorem 3.3.4. *In case 1 or in case 2, the sequence*

$$\left(\sqrt{|\mathbb{T}_{n-1}|}(\widehat{\theta}_n - \theta)/b_{|\mathbb{T}_{n-1}|}\right)_{n\geq 1}$$

satisfies the MDP on $\mathbb{R}^{2(p+1)}$ with speed $b^2_{|\mathbb{T}_{n-1}|}$ and rate function

$$I_\theta(x) = \sup_{\lambda \in \mathbb{R}^{2(p+1)}} \{\lambda^t x - \lambda(\Gamma \otimes L^{-1})\lambda^t\} = \frac{1}{2} x^t (\Gamma \otimes L^{-1})^{-1} x, \tag{3.3.3}$$

where L and Γ are given in (3.2.19) and (3.2.8) respectively.

Remark 3.3.5. *Similar results about deviation inequalities and MDP, are already obtained in [22], in a restrictive case of bounded or Gaussian noise and when $p = 1$, but results therein hold for general Markov models also.*

Let us consider now the estimation of the parameter in the noise process.

Theorem 3.3.6. *Let (b_n) an increasing sequence of positive real numbers which satisfy (3.2.15). In case 1 or in case 2,*

(1) the sequence

$$\left(\sqrt{|\mathbb{T}_{n-1}|}(\sigma_n^2 - \sigma^2)/b_{|\mathbb{T}_{n-1}|}\right)_{n\geq 1}$$

satisfies the MDP on $\mathbb{R}$ with speed $b^2_{|\mathbb{T}_{n-1}|}$ and rate function

$$I_{\sigma^2}(x) = \frac{x^2}{\tau^4 - 2\sigma^4 + \nu^2}. \tag{3.3.4}$$

(2) the sequence

$$\left(\sqrt{|\mathbb{T}_{n-1}|}(\rho_n - \rho)/b_{|\mathbb{T}_{n-1}|}\right)_{n\geq 1}$$

satisfies the MDP on $\mathbb{R}$ with speed $b^2_{|\mathbb{T}_{n-1}|}$ and rate function

$$I_{\rho}(x) = \frac{x^2}{2(\nu^2 - \rho^2)}. \tag{3.3.5}$$

Remark 3.3.7. *Note that in this case the MDP holds for all the scales (b_n) verifying (3.2.15) without other restriction.*

Remark 3.3.8. *It will be more interesting to prove the MDP for*

$$\left(\sqrt{|\mathbb{T}_{n-1}|}(\widehat{\sigma}_n^2 - \sigma^2)/b_{|\mathbb{T}_{n-1}|}\right)_{n\geq 1},$$

which will be the case if one proves for example that

$$\left(\sqrt{|\mathbb{T}_{n-1}|}(\widehat{\sigma}_n^2 - \sigma^2)/b_{|\mathbb{T}_{n-1}|}\right)_{n\geq 1} \quad \text{and} \quad \left(\sqrt{|\mathbb{T}_{n-1}|}(\sigma_n^2 - \sigma^2)/b_{|\mathbb{T}_{n-1}|}\right)_{n\geq 1}$$

are exponentially equivalent in the sense of the MDP. This is described by the following convergence

$$\frac{\sqrt{|\mathbb{T}_{n-1}|}}{b_{|\mathbb{T}_{n-1}|}}(\widehat{\sigma}_n^2 - \sigma_n^2) \underset{b^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

The proof is very technical and very restrictive for the scale of the deviation. Actually we are only able to prove that

$$\widehat{\sigma}_n^2 - \sigma_n^2 \underset{b^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0,$$

this superexponential convergence will be proved in Theorem 3.3.9.

In the following theorem we will state the superexponential convergence.

Theorem 3.3.9. *In case 1 or in case 2, we have*

$$\widehat{\sigma}_n^2 \overset{\text{superexp}}{\underset{b^2_{|\mathbb{T}_{n-1}|}}{\Longrightarrow}} \sigma^2.$$

In case 1 instead of **(G2)**, *if we assume that*

(G2') *one can find $\gamma' > 0$ such that for all $n \geq p-1$, for all $k, l \in \mathbb{G}_{n+1}$ with $[\frac{k}{2}] = [\frac{l}{2}]$ and for all $t \in]-c, c[$ for some $c > 0$,*

$$\mathbb{E}\left[\exp t\left(\varepsilon_k \varepsilon_l - \rho\right)\right] \leq \exp\left(\frac{\gamma' t^2}{2}\right),$$

and in case 2 instead of **(Ea)**, *if we assume that*

(E2') *one can find $\gamma' > 0$ such that for all $n \geq p-1$, for all $k, l \in \mathbb{G}_{n+1}$ with $[\frac{k}{2}] = [\frac{l}{2}]$ and for all $t \in \mathbb{R}$*

$$\mathbb{E}\left[\exp t\left(\varepsilon_k \varepsilon_l - \rho\right) | \mathcal{F}_n\right] \leq \exp\left(\frac{\gamma' t^2}{2}\right), \qquad a.s.$$

then in case 1 or in case 2, we have

$$\widehat{\rho}_n \overset{\text{superexp}}{\underset{b^2_{|\mathbb{T}_{n-1}|}}{\Longrightarrow}} \rho.$$

Before going to the proofs, let us gather here for the convenience of the readers two theorems useful to establish MDP for martingales and used intensively in this chapter. From these two theorems, we will be able to give a strategy for the proof.

Let $M = (M_n, \mathcal{H}_n, n \geq 0)$ be a centered square integrable martingale defined on a probability space $(\Omega, \mathcal{H}, \mathbb{P})$ and $(\langle M \rangle_n)$ its bracket. Let (b_n) be an increasing sequence of real numbers satisfying (3.2.15). The following proposition corresponds to the unidimensional case of Theorem 1 in [38].

Proposition 3.3.10. *Let $c(n) := \frac{\sqrt{n}}{b_n}$ be non-decreasing, and define the reciprocal function $c^{-1}(t)$ by*

$$c^{-1}(t) := \inf\{n \in \mathbb{N} : c(n) \geq t\}.$$

Under the following conditions:

(D1) *there exists $Q \in \mathbb{R}_+^*$ such that* $\dfrac{\langle M \rangle_n}{n} \overset{\text{superexp}}{\underset{b_n^2/n}{\Longrightarrow}} Q$;

(D2) $\displaystyle \limsup_{n \to +\infty} \frac{n}{b_n^2} \log\left(n \operatorname*{ess\,sup}_{1 \leq k \leq c^{-1}(\sqrt{n+1}\, b_{n+1})} \mathbb{P}(|M_k - M_{k-1}| > b_n \sqrt{n} | \mathcal{H}_{k-1})\right) = -\infty;$

(D3) *for all* $a > 0$ $\displaystyle\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\left(|M_k - M_{k-1}|^2\mathbf{1}_{\{|M_k-M_{k-1}|\geq a\frac{\sqrt{n}}{b_n}\}}|\mathcal{H}_{k-1}\right) \overset{\text{superexp}}{\underset{b_n^2/n}{\Longrightarrow}} 0;$

$(M_n/b_n\sqrt{n})_{n\in\mathbb{N}}$ *satisfies the MDP in* $\mathbb{R}$ *with the speed* b_n^2 *and the rate function* $I(x) = \dfrac{x^2}{2Q}$.

Let us introduce a simplified version of Puhalskii's result [91] applied to a sequence of martingale differences.

Theorem 3.3.11. *Let* $(m_j^n)_{1\leq j\leq n}$ *be a triangular array of martingale differences with values in* $\mathbb{R}^d$*, with respect to the filtration* $(\mathcal{H}_n)_{n\geq 1}$*. Under the following conditions*

(P1) *there exists a symmetric positive semi-definite matrix* Q *such that*

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\Big[m_k^n(m_k^n)'\big|\mathcal{H}_{k-1}\Big] \overset{\text{superexp}}{\underset{b_n^2/n}{\Longrightarrow}} Q,$$

(P2) *there exists a constant* $c > 0$ *such that, for each* $1 \leq k \leq n$, $|m_k^n| \leq c\frac{\sqrt{n}}{b_n}$ a.s.,

(P3) *for all* $a > 0$*, we have the exponential Lindeberg's condition*

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\Big[|m_k^n|^2\mathrm{I}_{\left\{|m_k^n|\geq a\frac{\sqrt{n}}{b_n}\right\}}\big|\mathcal{H}_{k-1}\Big] \overset{\text{superexp}}{\underset{b_n^2/n}{\Longrightarrow}} 0.$$

$(\sum_{k=1}^{n} m_k^n/(b_n\sqrt{n}))_{n\geq 1}$ *satisfies an MDP on* $\mathbb{R}^d$ *with speed* b_n^2 *and rate function*

$$\Lambda^*(v) = \sup_{\lambda\in\mathbb{R}^d}\left(\lambda' v - \frac{1}{2}\lambda' Q\lambda\right).$$

In particular, if Q *is invertible,* $\Lambda^*(v) = \frac{1}{2}v'Q^{-1}v$.

As the reader can imagine naturally now, the strategy of the proof of the MDP consists in the following steps :

- the superexponential convergence of the quadratic variation of the martingale (M_n). This step is very crucial and the key for the rest of the chapter. It will be realized by means of powerful exponential inequalities. This allows us to obtain the deviation inequalities for the estimator of the parameters,
- introduce a truncated martingale which satisfies the MDP, thanks to the classical Theorems 3.3.11,
- the truncated martingale is an exponentially good approximation of (M_n), in the sense of the moderate deviation.

3.4 Superexponential convergence of the quadratic variation of the martingale

At first, it is necessary to establish the superexponential convergence of the quadratic variation of the martingale (M_n), properly normalized in order to prove the MDP of the estimators. Its proof is very technical, but crucial for the rest of the chapter. This section contains also some deviation inequalities for some quantities needed in the proof later.

Proposition 3.4.1. *In case 1 or case 2, we have*

$$\frac{S_n}{|\mathbb{T}_n|} \underset{b^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} L, \tag{3.4.1}$$

where S_n is given in (3.2.5) and L is given in (3.2.19).

For the proof we focus on case 2. The Proposition 3.4.1 will follow from Proposition 3.4.3 and Proposition 3.4.4 below, where we assume that the sequence (b_n) satisfies the condition **(V2)**. Proposition 3.4.10 gives some ideas of the proof in case 1.

Remark 3.4.2. *Using [42], we infer from* **(Ea)** *that:*

(N2) *one can find $\gamma > 0$ such that for all $n \geq p-1$, for all $k \in \mathbb{G}_{n+1}$ and for all $t \in \mathbb{R}$*

$$\mathbb{E}\left[\exp t\left(\varepsilon_k^2 - \sigma^2\right)|\mathcal{F}_n\right] \leq \exp\left(\frac{\gamma t^2}{2}\right) \qquad a.s.$$

Proposition 3.4.3. *Assume that hypothesis* **(N2)** *and* **(Xa)** *are satisfied. Then we have*

$$\frac{1}{|\mathbb{T}_n|}\sum_{k\in\mathbb{T}_{n,p}} \mathbb{X}_k \underset{b^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} \Xi,$$

where Ξ is given in (3.2.16).

Proof. Let

$$H_n = \sum_{k\in\mathbb{T}_{n,p-1}} \mathbb{X}_k \qquad \text{and} \qquad P_n = \sum_{k\in\mathbb{T}_{n,p}} \varepsilon_k.$$

From Bercu et al. [18], we have

$$\frac{H_n}{2^{n+1}} = \sum_{k=p-1}^{n} (\overline{A})^{n-k}\frac{H_{p-1}}{2^{k+1}} + \sum_{k=p}^{n} \overline{a}(\overline{A})^{n-k}\left(\frac{2^k - 2^{p-1}}{2^k}\right)e_1 + \sum_{k=p}^{n}\frac{P_k}{2^{k+1}}(\overline{A})^{n-k}e_1. \tag{3.4.2}$$

Since the second term in the right hand side of this equality is deterministic and converges to Ξ, this proposition will be proved if we show that

$$\sum_{k=p-1}^{n} \frac{(\overline{A})^{n-k}}{2^k} H_{p-1} \overset{\text{superexp}}{\underset{b^2_{|\mathbb{T}_n|}}{\Longrightarrow}} 0, \qquad \sum_{k=p}^{n} \frac{P_k}{2^{k+1}} (\overline{A})^{n-k} e_1 \overset{\text{superexp}}{\underset{b^2_{|\mathbb{T}_n|}}{\Longrightarrow}} 0, \tag{3.4.3}$$

which follows proceeding as in the proof of Proposition 3.4.4 (see the proof of Proposition 3.4.4 for more details). ☐

Proposition 3.4.4. *Assume that hypothesis* **(N2)** *and* **(Xa)** *are satisfied. Then we have*

$$\frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n,p}} \mathbb{X}_k \mathbb{X}_k^t \overset{\text{superexp}}{\underset{b^2_{|\mathbb{T}_n|}}{\Longrightarrow}} \Lambda,$$

where Λ is given in (3.2.17).

Proof. Let

$$K_n = \sum_{k \in \mathbb{T}_{n,p-1}} \mathbb{X}_k \mathbb{X}_k^t \qquad \text{and} \qquad L_n = \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k^2. \tag{3.4.4}$$

Then from (3.2.2), and after straightforward calculations (see [18] for more details), we get that

$$\frac{K_n}{2^{n+1}} = \frac{1}{2^{n-p+1}} \sum_{C \in \{A;B\}^{n-p+1}} C \frac{K_{p-1}}{2^p} C^t + \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C T_{n-k} C^t,$$

where the notation $\{A;B\}^k$ means the set of all products of A and B with exactly k terms. The cardinality of $\{A;B\}^k$ is obviously 2^k, and

$$T_k = \frac{L_k}{2^{k+1}} e_1 e_1^t + \overline{a^2} \left(\frac{2^k - 2^{p-1}}{2^k} \right) e_1 e_1^t + I_k^{(1)} + I_k^{(2)} + \frac{1}{2^{k+1}} U_k$$

with $\overline{a^2} = (a_0^2 + b_0^2)/2$ and

$$I_k^{(1)} = \frac{1}{2} \left(a_0 \left(A \frac{H_{k-1}}{2^k} e_1^t + e_1 \frac{H_{k-1}}{2^k} A^t \right) + b_0 \left(B \frac{H_{k-1}}{2^k} e_1^t + e_1 \frac{H_{k-1}}{2^k} B^t \right) \right), \tag{3.4.5}$$

$$I_k^{(2)} = \left(\frac{1}{2^k} \sum_{l \in \mathbb{T}_{k-1,p-1}} (a_0 \varepsilon_{2l} + b_0 \varepsilon_{2l+1}) \right) e_1 e_1^t, \tag{3.4.6}$$

$$U_k = \sum_{l \in \mathbb{T}_{k-1,p-1}} \varepsilon_{2l} \Big(A \mathbb{X}_l e_1^t + e_1 \mathbb{X}_l^t A^t \Big) + \varepsilon_{2l+1} \Big(B \mathbb{X}_l e_1^t + e_1 \mathbb{X}_l^t B^t \Big). \tag{3.4.7}$$

Then proposition will follow if we prove Lemmas 3.4.5, 3.4.6, 3.4.7, 3.4.8 and 3.4.9.

Lemma 3.4.5. *Assume that hypothesis* **(Xa)** *is satisfied. Then we have*

$$\frac{1}{2^{n-p+1}} \sum_{C \in \{A;B\}^{n-p+1}} C \frac{K_{p-1}}{2^p} C^t \underset{b^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} 0, \tag{3.4.8}$$

where K_p is given in (3.4.4).

Proof. We get easily

$$\left\| \frac{1}{2^{n-p+1}} \sum_{C \in \{A;B\}^{n-p+1}} C \frac{K_{p-1}}{2^p} C^t \right\| \leq c\beta^{2n} \overline{X}_1^2,$$

where β is given in (3.2.3), $\overline{X}_1$ is introduced in **(Xa)** and c is a positive constant which depends on p. Next, Chernoff inequality and hypothesis **(X2)** lead us easily to (3.4.8). □

Lemma 3.4.6. *Assume that hypothesis* **(N2)** *and* **(Xa)** *are satisfied. Then we have*

$$\sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C \frac{L_{n-k}}{2^{n-k}} e_1 e_1^t C^t \underset{b^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} l, \tag{3.4.9}$$

where L_k is given in the second part of (3.4.4) and

$$l = \sum_{k=0}^{+\infty} \frac{1}{2^k} \sum_{C \in \{A,B\}^k} C(\sigma^2 e_1 e_1^t) C^t$$

is the unique solution of the equation

$$l = \sigma^2 e_1 e_1^t + \frac{1}{2}\left(AlA^t + BlB^t\right).$$

Proof. First, since we have for all $k \geq p$ the following decomposition on odd and even part

$$\sum_{i \in \mathbb{T}_{k,p}} (\varepsilon_i^2 - \sigma^2) = \sum_{i \in \mathbb{T}_{k-1,p-1}} (\varepsilon_{2i}^2 - \sigma^2) + (\varepsilon_{2i+1}^2 - \sigma^2),$$

we obtain for all $\delta > 0$ that

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_k| + 1} \sum_{i \in \mathbb{T}_{k,p}} (\varepsilon_i^2 - \sigma^2) > \delta \right) \leq \sum_{\eta=0}^{1} \mathbb{P}\left(\frac{1}{|\mathbb{T}_k| + 1} \sum_{i \in \mathbb{T}_{k-1,p-1}} (\varepsilon_{2i+\eta}^2 - \sigma^2) > \frac{\delta}{2} \right).$$

We will treat only the case $\eta = 0$. Chernoff inequality gives us for all $\lambda > 0$

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_k|+1}\sum_{i\in\mathbb{T}_{k-1,p-1}}(\varepsilon_{2i}^2-\sigma^2)>\frac{\delta}{2}\right)\leq \exp\left(-\lambda\frac{\delta}{2}2^{k+1}\right)$$
$$\times\,\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{k-1,p-1}}(\varepsilon_{2i}^2-\sigma^2)\right)\right].$$

We obtain from hypothesis **(N2)**, after conditioning by $\mathcal{F}_{k-1}$

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{k-1,p-1}}(\varepsilon_{2i}^2-\sigma^2)\right)\right]\leq \exp\left(\lambda^2\gamma|\mathbb{G}_{k-1}|\right)$$
$$\times\,\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{k-2,p-1}}(\varepsilon_{2i}^2-\sigma^2)\right)\right].$$

Iterating this, we deduce that

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{k-1,p-1}}(\varepsilon_{2i}^2-\sigma^2)\right)\right]\leq \exp\left(\gamma\lambda^2\sum_{l=p-1}^{k-1}|\mathbb{G}_l|\right)\leq\exp\left(\gamma\lambda^2 2^{k+1}\right).$$

Next, optimizing on λ, we get

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_k|+1}\sum_{i\in\mathbb{T}_{k-1,p-1}}(\varepsilon_{2i}^2-\sigma^2)>\frac{\delta}{2}\right)\leq \exp\left(-c\delta^2|\mathbb{T}_k|\right)$$

for some positive constant c which depends on γ. Applying the foregoing to the random variables $-(\varepsilon_i^2-\sigma^2)$, we obtain

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_k|+1}\left|\sum_{i\in\mathbb{T}_{k,p}}(\varepsilon_i^2-\sigma^2)\right|>\delta\right)\leq 4\exp\left(-c\delta^2|\mathbb{T}_k|\right). \tag{3.4.10}$$

Now we have

$$\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A,B\}^k}C\frac{L_{n-k}}{2^{n-k}}e_1e_1^tC^t-l=\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A,B\}^k}C\left(\frac{L_{n-k}}{2^{n-k}}-\sigma^2\right)e_1e_1^tC^t$$
$$-\sum_{k=n-p+1}^{+\infty}\frac{1}{2^k}\sum_{C\in\{A,B\}^k}C(\sigma^2e_1e_1^t)C^t$$

and since the second term of right hand side of the last equality is deterministic and tends to 0, to prove lemma 3.4.6, it suffices to show that

$$\sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A,B\}^k} C \left(\frac{L_{n-k}}{2^{n-k}} - \sigma^2 \right) e_1 e_1^t C^t \overset{\text{superexp}}{\underset{b^2_{|\mathbb{T}_n|}}{\Longrightarrow}} 0.$$

From the following inequalities

$$\begin{aligned}
\left\| \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C \left(\frac{L_{n-k}}{2^{n-k}} - \sigma^2 \right) e_1 e_1^t C^t \right\| & \\
&\leq \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} \left| \frac{L_{n-k}}{2^{n-k}} - \sigma^2 \right| \left\| C e_1 e_1^t C^t \right\| \\
&\leq \sum_{k=p}^{n} \beta^{2(n-k)} \left| \frac{L_k}{|\mathbb{T}_k| + 1} - \sigma^2 \right|
\end{aligned}$$

and from (3.4.10) applied with $\delta / ((n-p+1)\beta^{2(n-k)})$ instead of δ, we get

$$\begin{aligned}
\mathbb{P}\left(\left\| \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C \left(\frac{L_{n-k}}{2^{n-k}} - \sigma^2 \right) e_1 e_1^t C^t \right\| > \delta \right) & \\
&\leq \mathbb{P}\left(\sum_{k=p}^{n} \beta^{2(n-k)} \left| \frac{L_k}{|\mathbb{T}_k| + 1} - \sigma^2 \right| > \delta \right) \\
&\leq \sum_{k=p}^{n} \mathbb{P}\left(\left| \frac{L_k}{|\mathbb{T}_k| + 1} - \sigma^2 \right| > \frac{\delta}{(n-p+1)\beta^{2(n-k)}} \right) \\
&\leq c_1 \sum_{k=p}^{n} \exp\left(-c_2 \delta^2 \frac{(2\beta^4)^{k+1}}{n^2 \beta^{4n}} \right).
\end{aligned}$$

Now, following the same lines as in the proof of (3.4.17) we obtain

$$\mathbb{P}\left(\left\| \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C \left(\frac{L_{n-k}}{2^{n-k}} - \sigma^2 \right) e_1 e_1^t C^t \right\| > \delta \right) \leq \begin{cases} c_1 \exp\left(-c_2 \delta^2 \frac{2^{n+1}}{n^2} \right) & \text{if } \beta^4 < \frac{1}{2}, \\ c_1 n \exp\left(-c_2 \delta^2 \frac{2^{n+1}}{n^2} \right) & \text{if } \beta^4 = \frac{1}{2}, \\ c_1 \exp\left(-c_2 \delta^2 \frac{1}{n^2 \beta^{4n}} \right) & \text{if } \beta^4 > \frac{1}{2}, \end{cases} \tag{3.4.11}$$

for some positive constants c_1 and c_2. From (3.4.11), we infer that (3.4.9) holds. □

Lemma 3.4.7. *Assume that hypothesis* **(N1)** *is satisfied. Then we have*

$$\sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C\in\{A;B\}^k} C I_{n-k}^{(2)} C^t \underset{b^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} 0, \tag{3.4.12}$$

where $I_k^{(2)}$ *is given in (3.4.6).*

Proof. This proof follows the same lines as that of (3.4.9) and uses hypothesis **(N1)** instead of **(N2)**. ☐

Lemma 3.4.8. *Assume that hypothesis* **(N2)** *and* **(Xa)** *are satisfied. Then we have*

$$\sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C\in\{A;B\}^k} C I_{n-k}^{(1)} C^t \underset{b^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} \Lambda', \tag{3.4.13}$$

where

$$\Lambda' = \sum_{k=0}^{+\infty} \frac{1}{2^k} \sum_{C\in\{A,B\}^k} C\left(T - (\sigma^2 + \overline{a^2}) e_1 e_1^t\right) C^t$$

is the unique solution of equation

$$\Lambda' = T - (\sigma^2 + \overline{a^2}) e_1 e_1^t + \frac{1}{2}(A\Lambda' A^t + B\Lambda' B),$$

T is given in (3.2.18) and $I_k^{(1)}$ *is given in (3.4.5).*

Proof. Since in the definition of $I_n^{(1)}$ given by (3.4.5), there are four terms we will focus only on the first term

$$\frac{a_0}{2} A \frac{H_{k-1}}{2^k} e_1^t,$$

the other terms will be treated in the same way. Using (3.4.2), we obtain the following decomposition:

$$\frac{a_0}{2} \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C\in\{A;B\}^k} CA \frac{H_{n-k-1}}{2^{n-k}} e_1^t C^t = T_n^{(1)} + T_n^{(2)} + T_n^{(3)}$$

where

$$T_n^{(1)} = \frac{a_0}{2} \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C\in\{A;B\}^k} CA \left\{ \overline{A}^{n-k-p} \frac{H_{p-1}}{2^p} + \sum_{l=p}^{n-k-1} \overline{A}^{n-k-l-1} \frac{H_{p-1}}{2^{l+1}} \right\} e_1^t C^t,$$

$$T_n^{(2)} = \frac{a_0}{2} \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C\in\{A;B\}^k} CA \left\{ \sum_{l=p}^{n-k-1} \overline{A}^{n-k-l-1} \overline{a} \left(\frac{2^l - 2^{p-1}}{2^l} \right) e_1 e_1^t \right\} C^t,$$

and

$$T_n^{(3)} = \frac{a_0}{2} \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} CA \sum_{l=p}^{n-k-1} \overline{A}^{n-k-l-1} \frac{P_l}{2^{l+1}} e_1 e_1^t C^t.$$

On the one hand we have

$$\|T_n^{(3)}\| \leq c \sum_{k=p}^{n} \beta^{n-k} \frac{|P_k|}{2^{k+1}}$$

where c is a positive constant such that $c > |a_0| \frac{1-\beta^{n-l}}{1-\beta}$ for all $n \geq l$, so that

$$\mathbb{P}\Big(\|T_n^{(3)}\| > \delta\Big) \leq \sum_{k=p}^{n} \mathbb{P}\left(\frac{|P_k|}{|\mathbb{T}_k| + 1} > \frac{2\delta}{cn\beta^{n-k}}\right).$$

We deduce again from hypothesis **(N1)** and in the same way we have obtained (3.4.10) that

$$\mathbb{P}\left(\frac{P_k}{|\mathbb{T}_k| + 1} > \frac{2\delta}{cn\beta^{n-k}}\right) \leq \exp\left(-c_1 \delta^2 \frac{(2\beta^2)^{k+1}}{n^2 \beta^{2n}}\right) \quad \forall k \geq p,$$

for some positive constant c_1. It then follows as in the proof of (3.4.17) that

$$\mathbb{P}\Big(\|T_n^{(3)}\| > \delta\Big) \leq \begin{cases} \exp\left(-c_1 \delta^2 \frac{2^{n+1}}{n^2}\right) & \text{if } \beta^2 < \frac{1}{2}, \\ n \exp\left(-c_1 \delta^2 \frac{2^{n+1}}{n^2}\right) & \text{if } \beta^2 = \frac{1}{2}, \\ \exp\left(-c_1 \delta^2 \frac{1}{n^2 \beta^{2n}}\right) & \text{if } \beta^2 > \frac{1}{2}, \end{cases}$$

so that

$$T_n^{(3)} \underset{b^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{3.4.14}$$

On the other hand, we have after studious calculations

$$\|T_n^{(1)}\| \leq \begin{cases} c \frac{\overline{X}_1}{2^{n+1}} & \text{if } \beta < \frac{1}{2}, \\ c \frac{\overline{X}_1}{\sqrt{|\mathbb{T}_n| + 1}} & \text{if } \beta = \frac{1}{2}, \\ c\beta^n \overline{X}_1 & \text{if } \beta > \frac{1}{2}, \end{cases}$$

where c is a positive constant which depends on p and $|a_0|$. Next, from hypothesis **(X2)** and Chernoff inequality we conclude that

$$T_n^{(1)} \underset{b^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{3.4.15}$$

Furthermore, since $(T_n^{(2)})$ is a deterministic sequence, we have (see [18], lemma A.4)

$$T_n^{(2)} \overset{\text{superexp}}{\underset{b^2_{|\mathbb{T}_n|}}{\Longrightarrow}} \Lambda'', \tag{3.4.16}$$

where

$$\Lambda'' = \sum_{k=0}^{+\infty} \frac{1}{2^k} \sum_{C \in \{A,B\}^k} C \left(\frac{1}{2} a_0 A \Xi e_1^t \right) C^t$$

is the unique solution of

$$\Lambda'' = \frac{1}{2} a_0 A \Xi e_1^t + \frac{1}{2} \left(A \Lambda'' A^t + B \Lambda'' B \right).$$

It then follows that

$$\frac{a_0}{2} \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} CA \frac{H_{n-k-1}}{2^{n-k}} e_1^t C^t \overset{\text{superexp}}{\underset{b^2_{|\mathbb{T}_n|}}{\Longrightarrow}} \Lambda''.$$

Doing the same for the three other terms of $I_k^{(1)}$, we end the proof of Lemma (3.4.8). □

Lemma 3.4.9. *Assume that hypothesis* **(N2)** *and* **(Xa)** *are satisfied. Then we have*

$$\sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C \frac{U_{n-k}}{2^{n-k+1}} C^t \overset{\text{superexp}}{\underset{b^2_{|\mathbb{T}_n|}}{\Longrightarrow}} 0, \tag{3.4.17}$$

where U_k is given by (3.4.7).

Proof. Let $V_n = \sum_{k=2^{p-1}}^{n} \varepsilon_{2k} X_k$. Then (V_n) is a $\mathcal{G}_n$-martingale and its increasing process verifies that

$$\langle V \rangle_n = \sigma^2 \sum_{k=2^{p-1}}^{n} X_k^2 \leq \sigma^2 \sum_{k=2^{p-1}}^{n} \|\mathbb{X}_k\|^2 \leq \sigma^2 \sum_{k \in \mathbb{T}_{r_n, p-1}} \|\mathbb{X}_k\|^2$$

From [18], with $\alpha = \max(|a_0|, |b_0|)$, we have

$$\sum_{k \in \mathbb{T}_{r_n, p-1}} \|\mathbb{X}_k\|^2 \leq \frac{4}{1-\beta} P_{r_n} + \frac{4\alpha^2}{1-\beta} Q_{r_n} + 2\overline{X}_1^2 R_{r_n}, \tag{3.4.18}$$

where

$$P_{r_n} = \sum_{k \in \mathbb{T}_{r_n, p}} \sum_{i=0}^{r_k - p} \beta^i \varepsilon^2_{[\frac{k}{2^i}]}, \quad Q_{r_n} = \sum_{k \in \mathbb{T}_{r_n, p}} \sum_{i=0}^{r_k - p} \beta^i, \quad R_{r_n} = \sum_{k \in \mathbb{T}_{r_n, p-1}} \beta^{2(r_k - p + 1)}.$$

For $\lambda > 0$, we infer from hypothesis **(N1)** that $(Y_k)_{2^{p-1} \le k \le n}$ given by

$$Y_n = \exp\left(\lambda V_n - \frac{\lambda^2 \phi}{2} \sum_{k=2^{p-1}}^{n} X_k^2\right),$$

is a $\mathcal{G}_k$-supermartingale and moreover $\mathbb{E}\Big[Y_{2^{p-1}}\Big] \le 1$.

For $B > 0$ and $\delta > 0$, we have

$$\begin{aligned}\mathbb{P}\left(\frac{V_n}{2n} > \delta\right) &\le \mathbb{P}\Big(\frac{\phi}{2n} \sum_{i=2^{p-1}}^{n} X_k^2 > B\ \Big) + \mathbb{P}\left(Y_n > \exp\left(\lambda\delta - \frac{\lambda^2 B}{2}\right) 2n\right) \\ &\le \mathbb{P}\left(\frac{\phi}{2n} \sum_{k=2^{p-1}}^{n} X_k^2 > B\right) + \exp\left(\left(-\lambda\delta + \frac{\lambda^2 B}{2}\right) 2n\right).\end{aligned}$$

Optimizing on λ , we get

$$\mathbb{P}\left(\frac{V_n}{2n} > \delta\right) \le \mathbb{P}\left(\frac{\phi}{2n} \sum_{k \in \mathbb{T}_{r_n, p-1}} \|\mathbb{X}_k\|^2 > B\right) + \exp\left(-\frac{\delta^2}{B} 2n\right).$$

Since the same thing works for $-V_n$ instead of V_n, using $|\mathbb{T}_{n-1}|$ instead of n in the previous inequality, we have in particular

$$\mathbb{P}\left(\frac{|V_{|\mathbb{T}_{n-1}|}|}{|\mathbb{T}_n| + 1} > \delta\right) \le \mathbb{P}\left(\frac{\phi}{|\mathbb{T}_n| + 1} \sum_{k \in \mathbb{T}_{n-1, p-1}} \|\mathbb{X}_k\|^2 > B\right) + \exp\left(-\frac{\delta^2}{B} 2^{n+1}\right). \tag{3.4.19}$$

Now, to control the first term in the right hand side of the last inequality, we will use the decomposition given by (3.4.18). From the convergence of $\frac{4\phi}{(1-\beta)(|\mathbb{T}_n|+1)} P_n$ and $\frac{4\phi\alpha^2}{(1-\beta)(|\mathbb{T}_n|+1)} Q_n$ (see [18] for more details) let l_1 and l_2 such that $\forall n \ge p-1$

$$\frac{4\phi P_{n-1}}{(1-\beta)(|\mathbb{T}_n| + 1)} \to l_1 \quad \text{and} \quad \frac{4\phi\alpha^2 Q_{n-1}}{(1-\beta)(|\mathbb{T}_n| + 1)} < l_2.$$

For $\delta > 0$, we choose $B = \delta + l_1 + l_2$, using (3.4.18), we then have

$$\begin{aligned}\mathbb{P}\left(\tfrac{\phi}{|\mathbb{T}_n|+1} \sum_{k \in \mathbb{T}_{n-1, p-1}} \|\mathbb{X}_k\|^2 > B\right) &\le \mathbb{P}\left(\tfrac{P_{n-1}}{|\mathbb{T}_n|+1} - l_1' > \delta_1\right) \\ &+ \mathbb{P}\left(\tfrac{Q_{n-1}}{|\mathbb{T}_n|+1} - l_2' > \delta_2\right) + \mathbb{P}\left(\tfrac{R_{n-1}\overline{X}_1^2}{|\mathbb{T}_n|+1} > \delta_3\right)\end{aligned} \tag{3.4.20}$$

where

$$\delta_1 = \frac{(1-\beta)\delta}{12\phi}, \quad l_1' = \frac{(1-\beta) l_1}{4\phi}, \quad \delta_2 = \frac{(1-\beta)\delta}{12\alpha^2\phi}, \quad l_2' = \frac{(1-\beta) l_2}{4\alpha^2\phi}, \ \text{ and } \delta_3 = \frac{\delta}{6\phi}.$$

First, by the choice of l_2, we have

$$\mathbb{P}\left(\frac{Q_{n-1}}{|\mathbb{T}_n|+1} - l_2' > \delta_2\right) = 0. \tag{3.4.21}$$

Next, from Chernoff inequality and hypothesis **(X2)** we get easily

$$\mathbb{P}\left(\frac{R_{n-1}\overline{X}_1^2}{|\mathbb{T}_n|+1} > \delta_3\right) \leq \begin{cases} c_1 \exp\Big(-c_2\delta 2^{n+1}\Big) & \text{if } \beta < \frac{\sqrt{2}}{2} \\ c_1 \exp\left(-c_2\delta \frac{2^{n+1}}{n+1}\right) & \text{if } \beta = \frac{\sqrt{2}}{2} \\ c_1 \exp\left(-c_2\delta \left(\frac{1}{\beta^2}\right)^{n+1}\right) & \text{if } \beta > \frac{\sqrt{2}}{2}, \end{cases} \tag{3.4.22}$$

for some positive constants c_1 and c_2. Let us now control the first term of the right hand side of (3.4.20).

First case. If $\beta = \frac{1}{2}$, from [18]

$$P_{n-1} = \sum_{k=p}^{n-1}(n-k)\sum_{i\in\mathbb{G}_k}\varepsilon_i^2 \quad \text{and} \quad l_1' = \sigma^2.$$

We thus have

$$\frac{P_{n-1}}{|\mathbb{T}_n|+1} - \sigma^2 = \frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}(n-k)\sum_{i\in\mathbb{G}_k}(\varepsilon_i^2-\sigma^2) + \sigma^2\left(\sum_{k=p}^{n-1}\frac{n-k}{2^{n+1-k}} - 1\right).$$

In addition, we also have

$$\sigma^2\left(\sum_{k=p}^{n-1}\frac{n-k}{2^{n+1-k}} - 1\right) \leq 0.$$

We thus deduce that

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l_1' > \delta_1\right) \leq \mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}(n-k)\sum_{i\in\mathbb{G}_k}(\varepsilon_i^2-\sigma^2) > \delta_1\right).$$

On the one hand we have

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}(n-k)\sum_{i\in\mathbb{G}_k}(\varepsilon_i^2-\sigma^2) > \delta_1\right)$$
$$\leq \sum_{\eta=0}^{1}\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}(\varepsilon_{2i+\eta}^2-\sigma^2) > \delta_1/2\right). \tag{3.4.23}$$

On the other hand, for all $\lambda > 0$, an application of Chernoff inequality yields

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}(\varepsilon_{2i}^2-\sigma^2) > \delta_1/2\right)\\
&\leq \exp\left(\frac{-\delta_1\lambda 2^{n+1}}{2}\right)\times\mathbb{E}\left[\exp\left(\lambda\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}(\varepsilon_{2i}^2-\sigma^2)\right)\right].
\end{aligned}$$

From hypothesis **(N2)** we get

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}(\varepsilon_{2i}^2-\sigma^2)\right)\right]\\
&\qquad = \mathbb{E}\left[\mathbb{E}\left[\exp\left(\lambda\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}(\varepsilon_{2i}^2-\sigma^2)\right)\Big|\mathcal{F}_n\right]\right]\\
&= \mathbb{E}\left[\exp\left(\lambda\sum_{k=p-1}^{n-3}(n-k-1)\sum_{i\in\mathbb{G}_k}(\varepsilon_{2i}^2-\sigma^2)\right)\prod_{i\in\mathbb{G}_{n-2}}\mathbb{E}\left[\exp\left(\lambda(\varepsilon_{2i}^2-\sigma^2)\right)\Big|\mathcal{F}_n\right]\right]\\
&\leq \exp\left(\lambda^2\gamma|\mathbb{G}_{n-2}|\right)\mathbb{E}\left[\exp\left(\lambda\sum_{k=p-1}^{n-3}(n-k-1)\sum_{i\in\mathbb{G}_k}(\varepsilon_{2i}^2-\sigma^2)\right)\right].
\end{aligned}$$

Iterating this procedure, we obtain

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\lambda\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}(\varepsilon_{2i}^2-\sigma^2)\right)\right] &\leq \exp\left(\gamma\lambda^2\sum_{k=2}^{n-p+1}k^2|\mathbb{G}_{n-k}|\right)\\
&\leq \exp\left(c\gamma\lambda^2 2^{n+1}\right),
\end{aligned}$$

where $c = \sum_{k=1}^{\infty}\frac{k^2}{2^{k+2}}$. Optimizing on λ, we are led, for some positive constant c_1 to

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p-1}^{n-2}(n-k-1)\sum_{i\in\mathbb{G}_k}(\varepsilon_{2i}^2-\sigma^2) > \delta_1/2\right) \leq \exp\left(-c_1\delta^2|\mathbb{T}_n|\right).$$

Following the same lines, we obtain the same inequality for the second term in (3.4.23). It then follows that

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l_1' > \delta_1\right) \leq c_1\exp\left(-c_2\delta^2|\mathbb{T}_n|\right), \qquad (3.4.24)$$

for some positive constants c_1 and c_2.

Second case. If $\beta \neq \frac{1}{2}$, then from [18], we have $l_1' = \frac{\sigma^2}{2(1-\beta)}$. Since

$$\sigma^2\left(\sum_{k=p}^{n-1}\frac{1-(2\beta)^{n-k}}{(1-2\beta)2^{n-k+1}}\right) \leq \frac{\sigma^2}{2(1-\beta)},$$

we deduce that

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l'_1 > \delta_1\right) \leq \mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}\frac{1-(2\beta)^{n-k}}{1-2\beta}\sum_{i\in\mathbb{G}_k}(\varepsilon_i^2-\sigma^2) > \delta_1\right).$$

- If $\beta < \frac{1}{2}$, then for some positive constant c we have

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l'_1 > \delta_1\right) \leq \mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}\sum_{i\in\mathbb{G}_k}(\varepsilon_i^2-\sigma^2) > c\delta_1\right).$$

  Performing now as in the proof of (3.4.9), we get

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l'_1 > \delta_1\right) \leq c_1 \exp\left(-c_2\delta^2|\mathbb{T}_n|\right), \tag{3.4.25}$$

  for some positive constants c_1 and c_2.

- If $\beta > \frac{1}{2}$, then for some positive constant c, we have

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l'_1 > \delta_1\right) \leq \mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}(2\beta)^{n-k}\sum_{i\in\mathbb{G}_k}(\varepsilon_i^2-\sigma^2) > c\delta_1\right).$$

  Now, from Chernoff inequality, hypothesis **(N2)** and after several successive conditioning, we get for all $\lambda > 0$

$$\begin{aligned}&\mathbb{P}\left(\frac{1}{|\mathbb{T}_n|+1}\sum_{k=p}^{n-1}(2\beta)^{n-k}\sum_{i\in\mathbb{G}_k}(\varepsilon_i^2-\sigma^2) > c\delta_1\right)\\ \leq\ &\exp\Big(-c\delta_1\lambda 2^{n+1}\Big)\exp\left(\gamma\lambda^2 2^{n+1}\sum_{k=2}^{n-p+1}(2\beta^2)^k\right).\end{aligned}$$

  Next, optimizing over λ, we are led, for some positive constant c to

$$\mathbb{P}\left(\frac{P_{n-1}}{|\mathbb{T}_n|+1} - l'_1 > \delta_1\right) \leq \begin{cases}\exp\Big(-c\delta^2|\mathbb{T}_n|\Big) & \text{if } \frac{1}{2} < \beta < \frac{\sqrt{2}}{2},\\[2ex] \exp\Big(-c\delta^2\frac{|\mathbb{T}_n|}{n}\Big) & \text{if } \beta = \frac{\sqrt{2}}{2},\\[2ex] \exp\left(-c\delta^2\left(\frac{1}{\beta^2}\right)^{n+1}\right) & \text{if } \beta > \frac{\sqrt{2}}{2}.\end{cases} \tag{3.4.26}$$

Now combining (3.4.19), (3.4.20), (3.4.21), (3.4.22), (3.4.24),(3.4.25) and (3.4.26), we have thus showed that

$$\mathbb{P}\left(\tfrac{1}{|\mathbb{T}_n|+1}\left|V_{|\mathbb{T}_{n-1}|}\right| > \delta\right)$$

$$\leq \begin{cases} c_1\exp\left(-c_2\delta^2 2^{n+1}\right) + c_1\exp\left(-c_2\delta 2^{n+1}\right) \\ \qquad\qquad + \exp\left(\frac{-\delta^2}{\delta+l_1+l_2}2^{n+1}\right) \text{ if } \beta < \frac{\sqrt{2}}{2}, \\ c_1\exp\left(-c_2\delta^2\frac{2^{n+1}}{n+1}\right) + c_1\exp\left(-c_2\delta\frac{2^{n+1}}{n+1}\right) \\ \qquad\qquad + \exp\left(\frac{-\delta^2}{\delta+l_1+l_2}2^{n+1}\right) \text{ if } \beta = \frac{\sqrt{2}}{2}, \\ c_1\exp\left(-c_2\delta^2\left(\frac{1}{\beta^2}\right)^{n+1}\right) + c_1\exp\left(-c_2\delta\left(\frac{1}{\beta^2}\right)^{n+1}\right) \\ \qquad\qquad + \exp\left(\frac{-\delta^2}{\delta+l_1+l_2}2^{n+1}\right) \text{ if } \beta > \frac{\sqrt{2}}{2}, \end{cases} \tag{3.4.27}$$

where the positive constants c_1 and c_2 may differ term by term.

One can easily check that the coefficients of the matrix U_n are linear combinations of terms similar to $V_{|\mathbb{T}_{n-1}|}$, so that performing similar calculations as before for each of them, we deduce the same deviation inequalities for U_n as in (3.4.27).

Now we have

$$\mathbb{P}\left(\sum_{k=0}^{n-p}\frac{1}{2^k}\left\|\sum_{C\in\{A;B\}^k} C\frac{U_{n-k}}{2^{n-k+1}}C^t\right\| > \delta\right)$$

$$\leq \mathbb{P}\left(\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}\frac{1}{2^{n-k+1}}\left\|CU_{n-k}C^t\right\| > \delta\right)$$

$$\leq \mathbb{P}\left(\sum_{k=p}^{n}\beta^{2(n-k)}\frac{1}{|\mathbb{T}_k|+1}\|U_k\| > \delta\right)$$

$$\leq \sum_{k=p}^{n}\mathbb{P}\left(\frac{\|U_k\|}{|\mathbb{T}_k|+1} > \frac{\delta}{(n-p+1)\beta^{2(n-k)}}\right).$$

From (3.4.27), we infer the following

$$\mathbb{P}\left(\sum_{k=0}^{n-p}\frac{1}{2^k}\left\|\sum_{C\in\{A;B\}^k} C\frac{U_{n-k}}{2^{n-k+1}}C^t\right\|>\delta\right)$$

$$\leq \begin{cases} c_1\sum\limits_{k=p}^{n}\exp\left(-c_2\frac{\delta^2(2\beta^4)^{k+1}}{n^2\beta^{4n}}\right)+c_1\sum\limits_{k=p}^{n}\exp\left(-c_2\frac{\delta(2\beta^2)^{k+1}}{n\beta^{2n}}\right) \\ \qquad +c_1\sum\limits_{k=p}^{n}\exp\left(-c_2\frac{\delta^2 2^{k+1}}{(\delta+nl\beta^{2(n-k-1)})n\beta^{2(n-k-1)}}\right) \text{ if } \beta<\frac{\sqrt{2}}{2}, \\ c_1\sum\limits_{k=p}^{n}\exp\left(-c_2\frac{\delta^2 4^n}{n^2(k+1)2^{k+1}}\right)+c_1\sum\limits_{k=p}^{n}\exp\left(-c_2\frac{\delta 2^n}{(k+1)n}\right) \\ \qquad +c_1\sum\limits_{k=p}^{n}\exp\left(-c_2\frac{\delta^2 2^{k+1}}{(\delta+nl2^{-(n-k-1)})n2^{-(n-k-1)}}\right) \text{ if } \beta=\frac{\sqrt{2}}{2}, \\ c_1\sum\limits_{k=p}^{n}\exp\left(-c_2\frac{\delta^2(2\beta^2)^{k+1}}{n^2\beta^{4n}}\right)+c_1\sum\limits_{k=p}^{n}\exp\left(-c_2\frac{\delta}{n\beta^{2n}}\right) \\ \qquad +c_1\sum\limits_{k=p}^{n}\exp\left(-c_2\frac{\delta^2 2^{k+1}}{(\delta+nl\beta^{2(n-k-1)})n\beta^{2(n-k-1)}}\right) \text{ if } \beta>\frac{\sqrt{2}}{2}, \end{cases}$$

where $l=l_1+l_2$ and the positive constants c_1 and c_2 may differ term by term.

Now

- If $\beta<\frac{\sqrt{2}}{2}$, then on the one hand,

$$\begin{aligned}\sum_{k=p}^{n}\exp\left(-c\frac{\delta^2(2\beta^4)^{k+1}}{n^2\beta^{4n}}\right) &= \exp\left(-c\delta^2\beta^4\frac{2^{n+1}}{n^2}\right)\left(1+\sum_{k=p}^{n-1}\left(\exp\left(\frac{-c\delta^2}{n^2}\right)\right)^{(2\beta^4)^{k+1}\beta^{-4n}(1-(2\beta^4)^{n-k})}\right)\\ &\leq \exp\left(-c\delta^2\beta^4\frac{2^{n+1}}{n^2}\right)\Big(1+o(1)\Big),\end{aligned}$$

where the last inequality follows from the fact that for some positive constant c_1,

$$(2\beta^4)^{k+1}\beta^{-4n}(1-(2\beta^4)^{n-k})\propto c_1(2\beta^4)^{k+1}\beta^{-4n}.$$

On the other hand, following the same lines as before, we obtain

$$\begin{aligned}\sum_{k=p}^{n}\exp\left(-\frac{\delta^2 2^{k+1}}{(\delta+ln\beta^{2(n-k-1)})n\beta^{2(n-k-1)}}\right) &\leq \sum_{k=p}^{n}\exp\left(-c\delta^2\frac{2^{k+1}}{n^2\beta^{2(n-k-1)}}\right)\\ &\leq \exp\left(-c\frac{\delta^2 2^{n+1}}{(\delta+l)n^2}\right)\Big(1+o(1)\Big),\end{aligned}$$

and

$$\sum_{k=p}^{n}\exp\left(-c\frac{\delta(2\beta^2)^{k+1}}{n\beta^{2n}}\right) \leq \sum_{k=p}^{n}\exp\left(-c\frac{\delta(2\beta^2)^{k+1}}{n^2\beta^{2n}}\right) \leq \exp\left(-c\delta\frac{2^{n+1}}{n^2}\right)\Big(1+o(1)\Big).$$

We thus deduce that

$$\mathbb{P}\left(\sum_{k=0}^{n-p}\frac{1}{2^k}\left\|\sum_{C\in\{A;B\}^k}C\frac{U_{n-k}}{2^{n-k+1}}C^t\right\|>\delta\right) \leq c_1\exp\left(-c_2\delta^2\frac{2^{n+1}}{n^2}\right) + c_1\exp\left(-c_2\delta\frac{2^{n+1}}{n^2}\right), \tag{3.4.28}$$

for some positive constants c_1 and c_2.

- If $\beta=\frac{\sqrt{2}}{2}$, then following the same lines as before, we show that

$$\sum_{k=p}^{n}\exp\left(-c\delta^2\frac{4^n}{n^2(k+1)2^{k+1}}\right) \leq \exp\left(-c\delta^2\frac{2^{n+1}}{n^3}\right)\Big(1+o(1)\Big),$$

$$\sum_{k=p}^{n}\exp\left(-\frac{\delta^2 2^{k+1}}{(\delta+ln2^{-(n-k-1)})n2^{-(n-k-1)}}\right) \leq \exp\left(-c\frac{\delta^2 2^{n+1}}{n^2(\delta+l)}\right)\Big(1+o(1)\Big),$$

$$\sum_{k=p}^{n}\exp\left(-c\delta\frac{2^n}{n(k+1)}\right) \leq \exp\left(-c\delta\frac{2^{n+1}}{n^3}\right)\Big(1+o(1)\Big).$$

It then follows that

$$\mathbb{P}\left(\sum_{k=0}^{n-p}\frac{1}{2^k}\left\|\sum_{C\in\{A;B\}^k}C\frac{U_{n-k}}{2^{n-k+1}}C^t\right\|>\delta\right) \leq c_1\exp\left(-c_2\delta^2\frac{2^{n+1}}{n^3}\right) + c_1\exp\left(-c_2\frac{\delta^2 2^{n+1}}{n^2(\delta+l)}\right)+c_1\exp\left(-c_2\delta\frac{2^{n+1}}{n^3}\right), \tag{3.4.29}$$

for some positive constants c_1 and c_2.

- If $\beta>\frac{\sqrt{2}}{2}$, once again following the previous lines, we get

$$\mathbb{P}\left(\sum_{k=0}^{n-p}\frac{1}{2^k}\left\|\sum_{C\in\{A;B\}^k}C\frac{U_{n-k}}{2^{n-k+1}}C^t\right\|>\delta\right) \leq c_1\exp\left(-c_2\delta^2\frac{1}{n^2\beta^{2n}}\right) + c_1\exp\left(-c_2\frac{\delta^2}{(\delta+l)n^2\beta^{2n}}\right)+c_1 n\exp\left(-c_2\frac{\delta}{n^2\beta^{2n}}\right) \tag{3.4.30}$$

for some positive constants c_1 and c_2.

We infer from the inequalities (3.4.28), (3.4.29)and (3.4.30) that

$$\sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C \frac{U_{n-k}}{2^{n-k+1}} C^t \underset{b^2_{|\mathbb{T}_n|}}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

□

This ends the proof of the Proposition 3.4.4. □

We now explain the modification in the last proofs in case 1.

Proposition 3.4.10. *Within the framework 1, we have the same conclusions as the Proposition 3.4.3 and 3.4.4 with the sequence* (b_n) *which satisfies condition* **(V1)**.

Proof. The proof follows exactly the same lines as the proof of the Propositions 3.4.3 and 3.4.4, and uses the fact that if a superexponential convergence holds with a sequence (b_n) which satisfies condition **(V2)**, then it also holds with a sequence (b_n) which satisfies condition **(V1)**. We thus obtain the first convergence of (3.4.3), the convergences (3.4.8), (3.4.15), (3.4.16) and (3.4.12) within the framework 1 with (b_n) which satisfies condition **(V1)**. Next, following the same approach as which used to obtain (3.4.10), we get

$$\mathbb{P}\left(\frac{1}{|\mathbb{T}_k|+1}\left|\sum_{i \in \mathbb{T}_{k,p}} (\varepsilon_i^2 - \sigma^2)\right| > \delta\right) \leq \begin{cases} c_1 \exp\left(-c_2 \delta^2 |\mathbb{T}_k|\right) & \text{if } \delta \text{ is small enough} \\ c_1 \exp\left(-c_2 \delta |\mathbb{T}_k|\right) & \text{if } \delta \text{ is large enough,} \end{cases} \tag{3.4.31}$$

where c_1 and c_2 are positive constants which do not depend on δ. The first inequality holds for example if $\delta/\gamma < \varepsilon$ and the second holds for example if $\delta/\gamma > \varepsilon$, where ε is some positive constant. On the other hand, for n large enough, let n_0 such that for all $k < n_0$, $n\beta^{2(n-k)}$ is small enough so that $\delta/(n-p+1)\gamma\beta^{2(n-k)} > \varepsilon$. We have

$$\begin{aligned} \mathbb{P}\left(\left\|\sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C\left(\frac{L_{n-k}}{2^{n-k}} - \sigma^2\right) e_1 e_1^t C^t\right\| > \delta\right) \\ \leq \sum_{k=p}^{n_0-1} \mathbb{P}\left(\left|\frac{L_k}{|\mathbb{T}_k|+1} - \sigma^2\right| > \frac{\delta}{(n-p+1)\beta^{2(n-k)}}\right) \\ + \sum_{k=n_0}^{n} \mathbb{P}\left(\left|\frac{L_k}{|\mathbb{T}_k|+1} - \sigma^2\right| > \frac{\delta}{(n-p+1)\beta^{2(n-k)}}\right). \end{aligned}$$

Now, using (3.4.31) with $\delta/(n-p+1)\beta^{2(n-k)}$ instead of δ and following the same approach used to obtain (3.4.28)-(3.4.30) in the two sums of the right hand side of the above inequality, we are led to

$$\mathbb{P}\left(\left\|\sum_{k=0}^{n-p}\frac{1}{2^k}\sum_{C\in\{A;B\}^k}C\left(\frac{L_{n-k}}{2^{n-k}}-\sigma^2\right)e_1e_1^tC^t\right\|>\delta\right)$$
$$\leq\begin{cases}c_1\exp\left(-\frac{c_2\delta^2 2^{n+1}}{n^2}\right)+c_1\exp\left(-\frac{c_2\delta 2^{n+1}}{n}\right) & \text{if } \beta\leq\frac{1}{2}\\ c_1 n\exp\left(-\frac{c_2\delta^2}{n^2\beta^{4n}}\right)+c_1\exp\left(-\frac{c_2\delta}{n\beta^{2n}}\right) & \text{if } \beta>\frac{1}{2},\end{cases}$$

and we thus obtain convergence (3.4.9) with (b_n) which satisfies condition **(V1)**. In the same way we obtain

$$\mathbb{P}\left(\|T_n^{(3)}\|>\delta\right)\leq\begin{cases}c_1\exp\left(-\frac{c_2\delta^2 2^{n+1}}{n^2}\right)+c_1\exp\left(-\frac{c_2\delta 2^{n+1}}{n}\right) & \text{if } \beta<\frac{1}{2},\\ c_1 n\exp\left(-\frac{c_2\delta 2^{n+1}}{n}\right) & \text{if } \beta=\frac{1}{2},\\ c_1\exp\left(-\frac{c_2\delta^2}{n^2\beta^{2n}}\right)+c_1\exp\left(-\frac{c_2\delta}{n\beta^{n}}\right) & \text{if } \beta>\frac{1}{2},\end{cases}$$

so that (3.4.14) and then (3.4.13) hold for (b_n) which satisfies condition **(V1)**. To reach the convergence (3.4.17) and the second convergence of (3.4.3) with (b_n) which satisfies condition **(V1)**, we follow the same procedure as before and the proof of the proposition is then complete. □

Remark 3.4.11. *Let us note that we can actually prove that*

$$\frac{1}{n}\sum_{k=2^p}^{n}\mathbb{X}_k\overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}}\Xi\qquad\textit{and}\qquad\frac{1}{n}\sum_{k=2^p}^{n}\mathbb{X}_k\mathbb{X}_k^t\overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}}\Lambda,$$

where Ξ and Λ are defined in (3.2.16) and (3.2.17). Indeed, let $H_n=\sum_{k=2^{p-1}}^{n}\mathbb{X}_k$ and $P_l^{(n)}=\sum_{k=2^{r_n-l}}^{[\frac{n}{2^l}]}\varepsilon_k$. We have the following decomposition

$$\frac{H_n}{n}-\Xi=\frac{1}{n}\sum_{k\in\mathbb{T}_{r_n-1,p-1}}(\mathbb{X}_k-\Xi)+\frac{1}{n}\sum_{k=2^{r_n}}^{n}(\mathbb{X}_k-\Xi)+\frac{2^{p-1}-1}{n}\Xi.$$

On the one hand, observing that $b_n/b_{|\mathbb{T}_{r_n-1}|}<2$, we infer from Proposition 3.4.3 that

$$\frac{1}{n}\sum_{k\in\mathbb{T}_{r_n-1,p-1}}(\mathbb{X}_k-\Xi)\overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}}0.$$

The sequence $(\frac{2^{p-1}-1}{n}\Xi)$ being deterministic and converging to 0, we deduce that

$$\frac{2^{p-1}-1}{n}\Xi\overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}}0.$$

On the other hand, from (3.2.2) we have the following decomposition:

$$\sum_{k=2^{r_n}}^{n} \mathbb{X}_k = 2^{r_n-p+1}\left(\overline{A}\right)^{r_n-p+1} \sum_{k=2^{p-1}}^{[\frac{n}{2^{r_n-p+1}}]} \mathbb{X}_k + 2\overline{a}\sum_{k=0}^{r_n-p}\left([\frac{n}{2^k}] - 2^{r_n-k}+1\right)2^k\left(\overline{A}\right)^k e_1$$
$$+\sum_{k=0}^{r_n-p} 2^k\left(\overline{A}\right)^k P_k^{(n)} e_1 - \sum_{k=1}^{r_n-p+1} s_k 2^{k-1}\left(\overline{A}\right)^{k-1}\left(B\mathbb{X}_{[\frac{n}{2^k}]} + \eta_{[\frac{n}{2^{k-1}}]+1}\right),$$

where

$$s_k = \begin{cases} 1 & \textit{if } [\frac{n}{2^{k-1}}] \textit{ is even} \\ 0 & \textit{if } [\frac{n}{2^{k-1}}] \textit{ is odd.}\end{cases}$$

Performing now as in the proof of Proposition 3.4.3, tedious but straightforward calculations lead us to

$$\frac{1}{n}\sum_{k=2^{r_n}}^{n}(\mathbb{X}_k - \Xi) \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0.$$

It then follows that

$$\frac{1}{n}\sum_{k=2^{p}}^{n}\mathbb{X}_k \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} \Xi.$$

The term $\frac{1}{n}\sum_{k=2^p}^{n}\mathbb{X}_k\mathbb{X}_k^t$ *can be dealt with in the same way.*

The rest of the chapter is dedicated to the proof of our main results. We focus on the proof in case 2, and some explanations are given on how to obtain the results in case 1.

3.5 Proof of the main results

We start with the proof of the deviation inequalities.

3.5.1 Proof of Theorem 3.3.1

We begin the proof with the case 2. Let $\delta > 0$ and $b > 0$ such that $b < \|\Sigma\|/(1+\delta)$. We have from (3.2.14)

$$\begin{aligned}\mathbb{P}\left(\|\widehat{\theta}_n - \theta\| > \delta\right) &= \mathbb{P}\left(\frac{\|M_n\|}{\|\Sigma_{n-1}\|} > \delta, \frac{\|\Sigma_{n-1}\|}{|\mathbb{T}_{n-1}|} \geq b\right) \\ &\qquad + \mathbb{P}\left(\frac{\|M_n\|}{\|\Sigma_{n-1}\|} > \delta, \frac{\|\Sigma_{n-1}\|}{|\mathbb{T}_{n-1}|} < b\right) \\ &\leq \mathbb{P}\left(\frac{\|M_n\|}{|\mathbb{T}_{n-1}|} > \delta b\right) + \mathbb{P}\left(\left\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\right\| > \|\Sigma\| - b\right).\end{aligned}$$

Since $b < \|\Sigma\|/(1+\delta)$, then,

$$\mathbb{P}\left(\left\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\right\| > \|\Sigma\| - b\right) \leq \mathbb{P}\left(\left\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\right\| > \delta b\right).$$

It then follows that

$$\mathbb{P}\left(\|\widehat{\theta}_n - \theta\| > \delta\right) \leq 2\max\left\{\mathbb{P}\left(\frac{\|M_n\|}{|\mathbb{T}_{n-1}|} > \delta b\right), \mathbb{P}\left(\left\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\right\| > \delta b\right)\right\}.$$

On the one hand, we have

$$\mathbb{P}\left(\frac{\|M_n\|}{|\mathbb{T}_{n-1}|} > \delta b\right) \leq \sum_{\eta=0}^{1}\left\{\mathbb{P}\left(\left|\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}} \varepsilon_{2k+\eta}\right| > \frac{\delta b}{4}\right) + \mathbb{P}\left(\left\|\frac{1}{|\mathbb{T}_{n-1}|}\sum_{k\in\mathbb{T}_{n-1,p-1}} \varepsilon_{2k+\eta}\mathbb{X}_k\right\| > \frac{\delta b}{4}\right)\right\}.$$

Now, by carrying out the same calculations as those which have permitted us to obtain Lemma 3.4.7 and equation (3.4.27), we are led to

$$\mathbb{P}\left(\frac{\|M_n\|}{|\mathbb{T}_{n-1}|} > \delta b\right) \leq \begin{cases} c_1 \exp\left(-\frac{c_2(\delta b)^2}{c_3+c_4(\delta b)} 2^n\right) & \text{if } \beta < \frac{\sqrt{2}}{2}, \\ c_1 \exp\left(-\frac{c_2(\delta b)^2}{c_3+c_4(\delta b)} \frac{2^n}{n}\right) & \text{if } \beta = \frac{\sqrt{2}}{2}, \\ c_1 \exp\left(-\frac{c_2(\delta b)^2}{c_3+c_4(\delta b)} \left(\frac{1}{\beta^2}\right)^n\right) & \text{if } \beta > \frac{\sqrt{2}}{2}, \end{cases} \tag{3.5.1}$$

where positive constants c_1, c_2, c_3 and c_4 depend on σ, β, γ and ϕ and $(c_3, c_4) \neq (0,0)$.

On the other hand, noticing that $\Sigma_{n-1} = I_2 \otimes S_{n-1}$, we have

$$\mathbb{P}\left(\left\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\right\| > \delta b\right) \leq 2\mathbb{P}\left(\left\|\frac{S_{n-1}}{|\mathbb{T}_{n-1}|} - L\right\| > \frac{\delta b}{2}\right).$$

Next, from the proofs of Propositions 3.4.3 and 3.4.4, we deduce that

$$\mathbb{P}\left(\left\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\right\| > \frac{b\delta}{2}\right) \leq \begin{cases} c_1 \exp\left(-\frac{c_2(\delta b)^2}{c_3+c_4(\delta b)} \frac{2^n}{(n-1)^2}\right) & \text{if } \beta < \frac{\sqrt{2}}{2} \\ c_1 \exp\left(-\frac{c_2(\delta b)^2}{c_3+c_4(\delta b)} \frac{2^n}{(n-1)^3}\right) & \text{if } \beta = \frac{\sqrt{2}}{2} \\ c_1 \exp\left(-\frac{c_2(\delta b)^2}{c_3+c_4(\delta b)} \left(\frac{1}{(n-1)^2\beta^{2n}}\right)\right) & \text{if } \beta > \frac{\sqrt{2}}{2}, \end{cases} \tag{3.5.2}$$

for all $n > c_0 \frac{\log(\delta)}{\log(\beta)}$, where the constant $c_0 \in \mathbb{R}^*$ and the positive constants c_1, c_2, c_3 and c_4 depend on σ, β, γ and ϕ and $(c_3, c_4) \neq (0, 0)$. Now, (3.3.1) follows from (3.5.1) and (3.5.2).

In case 1, the proof follows exactly the same lines as before and uses the same ideas as the proof of Proposition 3.4.10. In particular, we have in this case

$$\mathbb{P}\left(\left\|\frac{\Sigma_{n-1}}{|\mathbb{T}_{n-1}|} - \Sigma\right\| > \frac{b\delta}{2}\right) \leq \begin{cases} c_1 \exp\left(-\frac{c_2(\delta b)^2}{c_3+(\delta b)} \frac{2^n}{(n-1)^2}\right) & \text{if } \beta < \frac{1}{2} \\ \\ c_1(n-1) \exp\left(-\frac{c_2(\delta b)^2}{c_3+(\delta b)} \frac{2^n}{(n-1)^2}\right) & \text{if } \beta = \frac{1}{2} \\ \\ c_1(n-1) \exp\left(-\frac{c_2(\delta b)^2}{c_3+(\delta b)} \left(\frac{1}{(n-1)\beta^n}\right)\right) & \text{if } \beta > \frac{1}{2} \end{cases}$$

for all $n > c_0 \frac{\log(\delta)}{\log(\beta)}$, where the constant $c_0 \in \mathbb{R}^*$ and the positive constants c_1, c_2 and c_3 depend on σ, β, γ and ϕ. (3.3.1) then follows in this case, and this ends the proof of Theorem. 3.3.1.

3.5.2 Proof of Theorem 3.3.4

At first we need to prove the following

Theorem 3.5.1. *In case 1 or in case 2, the sequence*

$$\left(M_n / \left(b_{|\mathbb{T}_{n-1}|} \sqrt{|\mathbb{T}_{n-1}|}\right)\right)_{n \geq 1}$$

satisfies the MDP on $\mathbb{R}^{2(p+1)}$ with speed $b^2_{|\mathbb{T}_{n-1}|}$ and rate function

$$I_M(x) = \sup_{\lambda \in \mathbb{R}^{2(p+1)}} \{\lambda^t x - \lambda^t (\Gamma \otimes L) \lambda\} = \frac{1}{2} x^t (\Gamma \otimes L)^{-1} x. \tag{3.5.3}$$

Proof of Theorem 3.5.1

Now, as in Bercu et al. [18], denote by $(\mathcal{G}_n)_{n \geq 1}$ the sister pair-wise filtration, that is $\mathcal{G}_n = \sigma\{X_1, (X_{2k}, X_{2k+1}), 1 \leq k \leq n\}$. We introduce the following $(\mathcal{G}_n)$ martingale difference sequence (D_n), given by

$$D_n = V_n \otimes Y_n = \begin{pmatrix} \varepsilon_{2n} \\ \varepsilon_{2n} \mathbb{X}_n \\ \varepsilon_{2n+1} \\ \varepsilon_{2n+1} \mathbb{X}_n \end{pmatrix}.$$

We clearly have

$$D_n D_n^t = V_n V_n^t \otimes Y_n Y_n^t.$$

So we obtain that the quadratic variation of the $(\mathcal{G}_n)$ martingale $(N_n)_{n\geq 2^{p-1}}$ given by

$$N_n = \sum_{k=2^{p-1}}^{n} D_k$$

is

$$\langle N \rangle_n = \sum_{k=2^{p-1}}^{n} \mathbb{E}(D_k D_k^t | \mathcal{G}_{k-1}) = \Gamma \otimes \sum_{k=2^{p-1}}^{n} Y_k Y_k^t.$$

Now we clearly have $M_n = N_{|\mathbb{T}_{n-1}|}$ and $\langle M \rangle_n = \langle N \rangle_{|\mathbb{T}_{n-1}|} = \Gamma \otimes S_{n-1}$. From Proposition 3.4.1, and since $\langle M \rangle_n = \Gamma \otimes S_{n-1}$, we have

$$\frac{\langle M \rangle_n}{|\mathbb{T}_n|} \underset{b^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} \Gamma \otimes L. \tag{3.5.4}$$

Before going to the proof of the MDP results, we state the exponential Lyapounov condition for $(N_n)_{n\geq 2^{p-1}}$, which implies exponential Lindeberg condition, that is

$$\limsup \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{1}{n} \sum_{k=2^{p-1}}^{n} \mathbb{E}\left[\|D_k\|^2 \mathbf{1}_{\left\{ \|D_k\| \geq r \frac{\sqrt{n}}{b_n} \right\}} \right] \geq \delta \right) = -\infty,$$

(see e.g [113] for more details on this implication).

Remark 3.5.2. *By [42], we infer from the condition* **(Ea)** *that*

(Na) *one can find $\gamma_a > 0$ such that for all $n \geq p-1$, for all $k \in \mathbb{G}_{n+1}$ and for all $t \in \mathbb{R}$, with $\mu_a = \mathbb{E}(|\varepsilon_k|^a | \mathcal{F}_n)$ a.s.*

$$\mathbb{E}\left[\exp t \left(|\varepsilon_k|^a - \mu_a \right) | \mathcal{F}_n \right] \leq \exp\left(\frac{\gamma_a t^2}{2} \right) \quad a.s.$$

Proposition 3.5.3. *Let (b_n) a sequence satisfying Assumption* **(V2)**. *Assume that hypotheses* **(Na)** *and* **(Xa)** *are satisfied. Then there exists $B > 0$ such that*

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{1}{n} \sum_{j=2^{p-1}}^{n} \mathbb{E}\left[\|D_j\|^a | \mathcal{G}_{j-1} \right] > B \right) = -\infty.$$

Proof. We are going to prove that

$$\limsup_{n\to\infty} \frac{1}{b^2_{|\mathbb{T}_n|}} \log \mathbb{P}\left(\frac{1}{|\mathbb{T}_n|} \sum_{j=2^p}^{|\mathbb{T}_n|} \mathbb{E}\left[\|D_j\|^a | \mathcal{G}_{j-1} \right] > B \right) = -\infty, \tag{3.5.5}$$

and Proposition (3.5.3) will follow proceeding as in Remark 3.4.11. We have

$$\sum_{j\in\mathbb{T}_{n,p}} \mathbb{E}\left[\|D_j\|^a | \mathcal{G}_{j-1} \right] \leq c\mu^a \sum_{j\in\mathbb{T}_{n,p}} \left(1 + \|\mathbb{X}_j\|^a \right),$$

where c is a positive constant which depends on a. From (3.2.2), we deduce that

$$\sum_{j\in\mathbb{T}_{n,p}} \|\mathbb{X}_j\|^a \leq \frac{c^2}{(1-\beta)^{a-1}} P_n + \frac{c^2\alpha^a Q_n}{(1-\beta)^{a-1}} + 2cR_n\overline{X}_1^a,$$

where

$$P_n = \sum_{j\in\mathbb{T}_{n,p}} \sum_{i=0}^{r_j-p} \beta^i |\varepsilon_{[\frac{j}{2^i}]}|^a, \quad Q_n = \sum_{j\in\mathbb{T}_{n,p}} \sum_{i=0}^{r_j-p} \beta^i, \quad R_n = \sum_{j\in\mathbb{T}_{n,p}} \beta^{a(r_j-p+1)},$$

and c is a positive constant. Now, proceeding as in the proof of Proposition 3.4.4, using hypothesis **(Na)** and **(Xa)** instead of **(N2)** and **(X2)** we get for B large enough

$$\limsup_{n\to\infty} \frac{1}{b^2_{|\mathbb{T}_n|}} \log \mathbb{P}\left(\frac{1}{|\mathbb{T}_n|}\sum_{j\in\mathbb{T}_{n,p}} \|\mathbb{X}_j\|^a > B\right) = -\infty. \tag{3.5.6}$$

Now (3.5.6) leads us to (3.5.5) and arguing as in Remark 3.4.11, we obtain Proposition 3.5.3. □

Remark 3.5.4. *In case 1, we clearly have that $(\mathbb{X}_n, n \in \mathbb{T}_{\cdot,p-1})$, where*

$$\mathbb{T}_{\cdot,p-1} = \bigcup_{r=p-1}^{\infty} \mathbb{G}_r,$$

is a bifurcating Markov chain with initial state $\mathbb{X}_{2^{p-1}} = (X_{2^{p-1}}, X_{2^{p-2}}, \cdots, X_1)^t$. Let ν be the law of $\mathbb{X}_{2^{p-1}}$. From hypothesis **(X2)**, *we deduce that ν has finite moments of all orders. We denote by P the transition probability kernel associated to $(\mathbb{X}_n, n \in \mathbb{T}_{\cdot,p-1})$. Let $(\mathbb{Y}_r, r \in \mathbb{N})$ the ergodic stable Markov chain associated to $(\mathbb{X}_n, n \in \mathbb{T}_{\cdot,p-1})$. This Markov chain is defined as follows, starting from the root $\mathbb{Y}_0 = \mathbb{X}_{2^{p-1}}$ and if $\mathbb{Y}_r = \mathbb{X}_n$ then $\mathbb{Y}_{r+1} = \mathbb{X}_{2n+\zeta_{r+1}}$ for a sequence of independent Bernoulli r.v. $(\zeta_q, q \in \mathbb{N}^*)$ such that $\mathbb{P}(\zeta_q = 0) = \mathbb{P}(\zeta_q = 1) = 1/2$. Let μ the stationary distribution associated to $(\mathbb{Y}_r, r \in \mathbb{N})$. For more details on bifurcating Markov chain and the associated ergodic stable Markov chain, we refer to [66] (see also [22]).*

From [22], we deduce that for all real bounded function f defined on $(\mathbb{R}^p)^3$,

$$\frac{1}{b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} \sum_{k\in\mathbb{T}_{n-1,p-1}} f(\mathbb{X}_k, \mathbb{X}_{2k}, \mathbb{X}_{2k+1})$$

satisfies a MDP on $\mathbb{R}$ with speed $b^2_{|\mathbb{T}_{n-1}|}$ and the rate function $I(x) = \frac{x^2}{2S^2(f)}$, where $S^2(f) =< \mu, P(f^2) - (Pf)^2 >$.

Now, let f be the function defined on $(\mathbb{R}^p)^3$ by $f(x,y,z) = \|x\|^2 + \|y\|^2 + \|z\|^2$. Then, using the relation (3.4.1) in Proposition 3.4.1, the above MDP for *real bounded functionals of the bifurcating Markov chain* $(\mathbb{X}_n, n \in \mathbb{T}_{\cdot,p-1})$ *and the truncation of the function* f*, we prove (in the same manner as the proof of lemma 3 in Worms [112]) that for all* $r > 0$

$$\limsup_{R\to\infty}\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{1}{n}\sum_{j=2^{p-1}}^{n}\left(\|\mathbb{X}_j\|^2+\|\mathbb{X}_{2j}\|^2+\|\mathbb{X}_{2j+1}\|^2\right)\right.$$
$$\left.\times\,\mathbf{1}_{\{||\mathbb{X}_j||+||\mathbb{X}_{2j}||+||\mathbb{X}_{2j+1}||>R\}} > r\right) = -\infty,$$

which implies the following Lindeberg condition (for more detail one can see Proposition 2 in Worms [112])

$$\limsup_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{P}\left(\frac{1}{n}\sum_{j=2^{p-1}}^{n}\left(\|\mathbb{X}_j\|^2+\|\mathbb{X}_{2j}\|^2+\|\mathbb{X}_{2j+1}\|^2\right)\right.$$
$$\left.\times\,\mathbf{1}_{\left\{||\mathbb{X}_j||+||\mathbb{X}_{2j}||+||\mathbb{X}_{2j+1}||>r\frac{\sqrt{n}}{b_n}\right\}} > \delta\right) = -\infty,$$

for all $\delta > 0$ *and for all* $r > 0$*. Notice that the above Lindeberg condition implies particularly the Lindeberg condition on the sequence* $(\mathbb{X}_n)$*.*

Now, we come back to the proof of Theorem 3.5.1. We divide the proof into four steps. In the first one, we introduce a truncation of the martingale $(M_n)_{n\geq 0}$ and prove that the truncated martingale satisfies some MDP thanks to Puhalskii's Theorem 3.3.11. In the second part, we show that the truncated martingale is an exponentially good approximation of (M_n), see e.g. Definition 4.2.14 in [35]. We conclude by the identification of the rate function.

Proof in case 2

Step 1. From now on, in order to apply Puhalskii's result [91] (Puhalskii's Theorem 3.3.11) for the MDP for martingales, we introduce the following truncation of the martingale $(M_n)_{n\geq 0}$. For $r > 0$ and $R > 0$,

$$M_n^{(r,R)} = \sum_{k\in\mathbb{T}_{n-1,p-1}} D_{k,n}^{(r,R)}.$$

where, for all $1 \leq k \leq n$, $D_{k,n}^{(r,R)} = V_k^{(R)} \otimes Y_{k,n}^{(r)}$, with

$$V_n^{(R)} = \left(\varepsilon_{2n}^{(R)}, \varepsilon_{2n+1}^{(R)}\right)^t \qquad \text{and} \qquad Y_{k,n}^{(r)} = \left(1, \mathbb{X}_{k,n}^{(r)}\right)^t,$$

where

$$\varepsilon_k^{(R)} = \varepsilon_k\mathbf{1}_{\{|\varepsilon_k|\leq R\}} - \mathbb{E}\left[\varepsilon_k\mathbf{1}_{\{|\varepsilon_k|\leq R\}}\right], \qquad \mathbb{X}_{k,n}^{(r)} = \mathbb{X}_k\mathbf{1}_{\left\{||\mathbb{X}_k||\leq r\frac{\sqrt{|\mathbb{T}_{n-1}|}}{b_{|\mathbb{T}_{n-1}|}}\right\}}.$$

We introduce $\Gamma^{(R)}$ the conditional covariance matrix associated with $(\epsilon_{2k}^{(R)}, \epsilon_{2k+1}^{(R)})^t$ and the truncated matrix associated with S_n :

$$\Gamma^{(R)} = \begin{pmatrix} \sigma_R^2 & \rho_R \\ \rho_R & \sigma_R^2 \end{pmatrix} \qquad \text{and} \qquad S_n^{(r)} = \sum_{k \in \mathbb{T}_{n,p-1}} \begin{pmatrix} 1 & (\mathbb{X}_{k,n}^{(r)})^t \\ \mathbb{X}_{k,n}^{(r)} & \mathbb{X}_{k,n}^{(r)} (\mathbb{X}_{k,n}^{(r)})^t \end{pmatrix}.$$

The condition **(P2)** in Puhalskii's Theorem 3.3.11 is verified by the construction of the truncated martingale, that is for some positive constant c, we have that for all $k \in \mathbb{T}_{n-1}$

$$||D_{k,n}^{(r,R)}|| \leq c \frac{\sqrt{|\mathbb{T}_{n-1}|}}{b_{|\mathbb{T}_{n-1}|}}.$$

From Proposition 3.5.3, we also have for all $r > 0$,

$$\frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \mathbb{X}_k \mathrm{I}_{\left\{||\mathbb{X}_k|| > r \frac{\sqrt{|\mathbb{T}_{n-1}|}}{b_{|\mathbb{T}_{n-1}|}}\right\}} \underset{b^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0; \tag{3.5.7}$$

and

$$\frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \mathbb{X}_k \mathbb{X}_k^t \mathrm{I}_{\left\{||\mathbb{X}_k|| > r \frac{\sqrt{|\mathbb{T}_{n-1}|}}{b_{|\mathbb{T}_{n-1}|}}\right\}} \underset{b^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{3.5.8}$$

From (5.4.80) and (3.5.8), we deduce that for all $r > 0$

$$\frac{1}{|\mathbb{T}_{n-1}|} \left(S_{n-1} - S_{n-1}^{(r)} \right) \underset{b^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{3.5.9}$$

Then, we easily transfer the properties (5.4.79) to the truncated martingale $(M_n^{(r,R)})_{n \geq 0}$. We have for all $R > 0$ and all $r > 0$,

$$\begin{aligned} \frac{\langle M^{(r,R)} \rangle_n}{|\mathbb{T}_{n-1}|} &= \Gamma^{(R)} \otimes \frac{S_{n-1}^{(r)}}{|\mathbb{T}_{n-1}|} \\ &= -\Gamma^{(R)} \otimes \left(\frac{S_{n-1} - S_{n-1}^{(r)}}{|\mathbb{T}_{n-1}|} \right) + \Gamma^{(R)} \otimes \frac{S_{n-1}}{|\mathbb{T}_{n-1}|} \underset{b^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} \Gamma^{(R)} \otimes L \end{aligned}$$

That is condition **(P1)** in Puhalskii's Theorem 3.3.11.

Note also that Proposition 3.5.3 works for the truncated martingale $(M_n^{(r,R)})_{n \geq 0}$, which ensures the Lindeberg's condition and thus condition **(P3)** to $(M_n^{(r,R)})_{n \geq 0}$. By Theorem 3.3.11, we deduce that $(M_n^{(r,R)} / (b_{|\mathbb{T}_{n-1}|} \sqrt{|\mathbb{T}_{n-1}|}))_{n \geq 0}$ satisfies a MDP on $\mathbb{R}^{2(p+1)}$ with speed $b^2_{|\mathbb{T}_{n-1}|}$ and good rate function given by

$$I_R(x) = \frac{1}{2} x^t (\Gamma^{(R)} \otimes L)^{-1} x. \tag{3.5.10}$$

Step 2. At first, we infer from the hypothesis **(Ea)** that:

(N1R) there is a sequence $(\kappa_R)_{R>0}$ with $\kappa_R \longrightarrow 0$ when R goes to infinity, such that for all $n \geq p-1$, for all $k \in \mathbb{G}_{n+1}$, for all $t \in \mathbb{R}$ and for R large enough

$$\mathbb{E}\left[\exp t\left(\varepsilon_k - \varepsilon_k^R\right)|\mathcal{F}_n\right] \leq \exp\left(\frac{\kappa_R t^2}{2}\right), \ \ a.s.$$

The approximation, in the sense of the moderate deviation, is described by the following convergence, for all $r > 0$ and all $\delta > 0$,

$$\limsup_{R\to\infty}\limsup_{n\to\infty}\frac{1}{b^2_{|\mathbb{T}_{n-1}|}}\log\mathbb{P}\left(\frac{\|M_n - M_n^{(r,R)}\|}{\sqrt{|\mathbb{T}_{n-1}|}b_{|\mathbb{T}_{n-1}|}} > \delta\right) = -\infty.$$

For that, we shall prove that for $\eta \in \{0,1\}$

$$I_1 = \frac{1}{\sqrt{|\mathbb{T}_{n-1}|}b_{|\mathbb{T}_{n-1}|}}\sum_{k\in\mathbb{T}_{n-1,p-1}}\left(\varepsilon_{2k+\eta} - \varepsilon_{2k+\eta}^{(R)}\right) \overset{\text{superexp}}{\underset{b^2_{|\mathbb{T}_{n-1}|}}{\Longrightarrow}} 0, \tag{3.5.11}$$

$$I_2 = \frac{1}{\sqrt{|\mathbb{T}_{n-1}|}b_{|\mathbb{T}_{n-1}|}}\sum_{k\in\mathbb{T}_{n-1,p-1}}\left(\varepsilon_{2k+\eta}\mathbb{X}_k - \varepsilon_{2k+\eta}^{(R)}\mathbb{X}_{k,n}^{(r)}\right) \overset{\text{superexp}}{\underset{b^2_{|\mathbb{T}_{n-1}|}}{\Longrightarrow}} 0. \tag{3.5.12}$$

To prove (3.5.11) and (3.5.12), we have to do it only for $\eta = 0$ the same proof works for $\eta = 1$.

Proof of (3.5.11) We have for all $\alpha > 0$ and R large enough

$$\begin{aligned}
&\mathbb{E}\left(\exp\left(\alpha\sum_{k\in\mathbb{T}_{n-1,p-1}}(\varepsilon_{2k} - \varepsilon_{2k}^{(R)})\right)\right)\\
= \ &\mathbb{E}\left[\prod_{k\in\mathbb{T}_{n-2,p-1}}\exp\left(\alpha(\varepsilon_{2k} - \varepsilon_{2k}^{(R)})\right)\times\mathbb{E}\left[\prod_{k\in\mathbb{G}_{n-1}}\exp\left(\alpha(\varepsilon_{2k} - \varepsilon_{2k}^{(R)})\right)\Big|\mathcal{F}_{n-1}\right]\right]\\
= \ &\mathbb{E}\left[\prod_{k\in\mathbb{T}_{n-2,p-1}}\exp\left(\alpha(\varepsilon_{2k} - \varepsilon_{2k}^{(R)})\right)\times\prod_{k\in\mathbb{G}_{n-1}}\mathbb{E}\left[\exp\left(\alpha(\varepsilon_{2k} - \varepsilon_{2k}^{(R)})\right)\Big|\mathcal{F}_{n-1}\right]\right]\\
\leq \ &\mathbb{E}\left[\prod_{k\in\mathbb{T}_{n-2,p-1}}\exp\left(\alpha(\varepsilon_{2k} - \varepsilon_{2k}^{(R)})\right)\exp\left(|\mathbb{G}_{n-1}|\alpha^2\kappa_R\right)\right]\\
\leq \ &\exp\left(|\mathbb{T}_{n-1}|\alpha^2\kappa_R\right).
\end{aligned}$$

where hypothesis **(N1R)** was used to get the first inequality, and the second was obtained by induction. By Chebyshev inequality and the previous calculation applied to $\alpha = \lambda b_{|\mathbb{T}_{n-1}|}/|\mathbb{T}_{n-1}|$, we obtain for all $\delta > 0$

$$\mathbb{P}\left(\frac{1}{\sqrt{|\mathbb{T}_{n-1}|}b_{|\mathbb{T}_{n-1}|}}\sum_{k\in\mathbb{T}_{n-1,p-1}}(\varepsilon_{2k}-\varepsilon_{2k}^{(R)})\geq\delta\right)\leq\exp\left(-b_{|\mathbb{T}_{n-1}|}^2\Big(\delta\lambda-\kappa_R\lambda^2\Big)\right).$$

Optimizing on λ, we obtain

$$\frac{1}{b_{|\mathbb{T}_{n-1}|}^2}\log\mathbb{P}\left(\frac{1}{\sqrt{|\mathbb{T}_{n-1}|}b_{|\mathbb{T}_{n-1}|}}\sum_{k\in\mathbb{T}_{n-1,p-1}}\Big(\varepsilon_{2k}-\varepsilon_{2k}^{(R)}\Big)\geq\delta\right)\leq-\frac{\delta^2}{4\kappa_R}.$$

Letting n goes to infinity and than R goes to infinity, we obtain the negligibility in (3.5.11).

Proof of (3.5.12) Now, since we have the decomposition

$$\varepsilon_{2k}\mathbb{X}_k-\varepsilon_{2k}^{(R)}\mathbb{X}_{k,n}^{(r)}=\Big(\varepsilon_{2k}-\varepsilon_{2k}^{(R)}\Big)\mathbb{X}_{k,n}^{(r)}+\varepsilon_{2k}\Big(\mathbb{X}_k-\mathbb{X}_{k,n}^{(r)}\Big),$$

we introduce the following notation

$$L_n^{(r)}=\sum_{k\in\mathbb{T}_{n-1,p-1}}\varepsilon_{2k}\Big(\mathbb{X}_k-\mathbb{X}_{k,n}^{(r)}\Big)\qquad\text{and}\qquad F_n^{(r,R)}=\sum_{k\in\mathbb{T}_{n-1,p-1}}\Big(\varepsilon_{2k}-\varepsilon_{2k}^{(R)}\Big)\mathbb{X}_{k,n}^{(r)}.$$

To prove (3.5.12), we will show that for all $r>0$

$$\frac{L_n^{(r)}}{\sqrt{|\mathbb{T}_{n-1}|}b_{|\mathbb{T}_{n-1}|}}\overset{\text{superexp}}{\underset{b_{|\mathbb{T}_{n-1}|}^2}{\Longrightarrow}}0,\tag{3.5.13}$$

and for all $r>0$ and all $\delta>0$

$$\limsup_{R\to\infty}\limsup_{n\to\infty}\frac{1}{b_{|\mathbb{T}_{n-1}|}^2}\log\mathbb{P}\left(\frac{\|F_n^{(r,R)}\|}{b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}}>\delta\right)=-\infty.\tag{3.5.14}$$

Let us first deal with $(L_n^{(r)})$. Let its first component be

$$L_{n,1}^{(r)}=\sum_{k\in\mathbb{T}_{n-1,p-1}}\varepsilon_{2k}\Big(X_k-X_{k,n}^{(r)}\Big).$$

For $\lambda\in\mathbb{R}$, we consider the random sequence $(Z_{n,1}^{(r)})_{n\geq p-1}$ defined by

$$Z_{n,1}^{(r)}=\exp\left(\lambda L_{n,1}^{(r)}-\frac{\lambda^2\phi}{2}\sum_{k\in\mathbb{T}_{n-1,p-1}}X_k^2\mathbf{1}_{\left\{\|\mathbb{X}_k\|>r\frac{\sqrt{|\mathbb{T}_{n-1}|}}{b_{|\mathbb{T}_{n-1}|}}\right\}}\right)$$

where ϕ appears in **(N1)**.

For $b > 0$, we introduce the following event

$$A_{n,1}^{(r)}(b) = \left\{ \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} X_k^2 \mathbf{1}_{\left\{ \|\mathbb{X}_k\| > r \frac{\sqrt{|\mathbb{T}_{n-1}|}}{b_{|\mathbb{T}_{n-1}|}} \right\}} > b \right\}.$$

Using **(N1)**, we have for all $\delta > 0$

$$\begin{aligned}
& \mathbb{P}\left(\frac{1}{b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} L_{n,1}^{(r)} > \delta \right) \\
\leq\ & \mathbb{P}\Big(A_{n,1}^{(r)}(b) \Big) + \mathbb{P}\left(Z_{n,1}^{(r)} > \exp\left(\delta\lambda b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|} - \frac{\lambda^2\phi}{2} b |\mathbb{T}_{n-1}| \right) \right) \\
\leq\ & \mathbb{P}\Big(A_{n,1}^{(r)}(b) \Big) + \exp\left(-b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|} \left(\delta\lambda - \frac{b\phi\sqrt{|\mathbb{T}_{n-1}|}}{2 b_{|\mathbb{T}_{n-1}|}} \lambda^2 \right) \right), \quad (3.5.15)
\end{aligned}$$

where the second term in (3.5.15) is obtained by conditioning successively on $(\mathcal{G}_i)_{2^{p-1} \leq i \leq |\mathbb{T}_{n-1}|-1}$ and using the fact that

$$\mathbb{E}\left[\exp\left(\lambda\varepsilon_{2^p}\left(X_{2^{p-1}} - X_{2^{p-1}}^{(r)} \right) - \frac{\lambda^2\phi}{2} X_{2^{p-1}}^2 \mathbf{1}_{\left\{ \|\mathbb{X}_{2^{p-1}}\| > r \frac{\sqrt{2^{p-1}}}{b_{2^{p-1}}} \right\}} \right) \right] \leq 1,$$

which follows from **(N1)**.

From Proposition 3.5.3, we have for all $b > 0$

$$\limsup_{n\to\infty} \frac{1}{b_{|\mathbb{T}_{n-1}|}^2} \log \mathbb{P}\Big(A_{n,1}^{(r)}(b) \Big) = -\infty,$$

so that taking $\lambda = \delta b_{|\mathbb{T}_{n-1}|} / (b\phi\sqrt{|\mathbb{T}_{n-1}|})$ in (3.5.15), we are led to

$$\limsup_{n\to\infty} \frac{1}{b_{|\mathbb{T}_{n-1}|}^2} \log \mathbb{P}\left(\frac{L_{n,1}^{(r)}}{b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} > \delta \right) \leq -\frac{\delta^2}{2b\phi}.$$

Letting $b \to 0$, we obtain that the right hand of the last inequality goes to $-\infty$. Proceeding in the same way for $-L_{n,1}^{(r)}$, we deduce that for all $r > 0$

$$\frac{L_{n,1}^{(r)}}{b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} \overset{\text{superexp}}{\underset{b_{|\mathbb{T}_{n-1}|}^2}{\Longrightarrow}} 0.$$

Now, it is easy to check that the same proof works for the others components of $L_n^{(r)}$. We thus conclude the proof of (3.5.13).

Eventually, let us treat the term $(F_n^{(r,R)})$. We follow the same approach as in the proof of (3.5.13). Let its first component be

$$F_{n,1}^{(r,R)} = \sum_{k \in \mathbb{T}_{n-1,p-1}} (\varepsilon_{2k} - \varepsilon_{2k}^{(R)}) X_{k,n}^{(r)}$$

For $\lambda \in \mathbb{R}$, we consider the random sequence $\left(W_{n,1}^{(r,R)}\right)_{n \geq p-1}$ defined by

$$W_{n,1}^{(r,R)} = \exp\left(\lambda \sum_{k \in \mathbb{T}_{n-1,p-1}} (\varepsilon_{2k} - \varepsilon_{2k}^{(R)}) X_{k,n}^{(r)} - \frac{\lambda^2 \kappa_R}{2} \sum_{k \in \mathbb{T}_{n-1,p-1}} (X_{k,n}^{(r)})^2\right)$$

where κ_R appears in **(N1R)**.

Let $b > 0$. Consider the following event $B_{n,1}^{(r)}(b) = \left\{\frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} (X_{k,n}^{(r)})^2 > b\right\}$.

We have for all $\delta > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\frac{F_{n,1}^{(r,R)}}{b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} > \delta\right) \\
\leq\ & \mathbb{P}\left(B_{n,1}^{(r)}(b)\right) + \mathbb{P}\left(W_{n,1}^{(r,R)} > \exp\left(\delta\lambda b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|} - \frac{\lambda^2\kappa_R}{2}|\mathbb{T}_{n-1}|b\right)\right) \\
\leq\ & \mathbb{P}\left(B_{n,1}^{(r)}(b)\right) + \exp\left(-b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}\left(\delta\lambda - \frac{b\kappa_R\sqrt{|\mathbb{T}_{n-1}|}}{2b_{|\mathbb{T}_{n-1}|}}\lambda^2\right)\right) \qquad (3.5.16)
\end{aligned}$$

where the second term in (3.5.16) is obtained by conditioning successively on $(\mathcal{G}_i)_{2^{p-1} \leq i \leq |\mathbb{T}_{n-1}|-1}$ and using the fact that

$$\mathbb{E}\left[\exp\left(\lambda\left(\varepsilon_{2^p} - \varepsilon_{2^p}^{(R)}\right) X_{2^{p-1}}^{(r)} - \frac{\lambda^2\kappa_R}{2}\left(X_{2^{p-1}}^{(r)}\right)^2\right)\right] \leq 1,$$

Since $B_{n,1}^{(r)}(b) \subset \left\{\frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} X_k^2 > b\right\}$, from Proposition 3.4.4, we deduce that for b large enough

$$\limsup_{n\to\infty} \frac{1}{b_{|\mathbb{T}_{n-1}|}^2} \log \mathbb{P}\left(B_{n,1}^{(r)}(b)\right) = -\infty,$$

so that choosing $\lambda = \delta b_{|\mathbb{T}_{n-1}|}/(\kappa_R b\sqrt{|\mathbb{T}_{n-1}|})$, we get for all $\delta > 0$

$$\limsup_{n\to\infty} \frac{1}{b_{|\mathbb{T}_{n-1}|}^2} \log \mathbb{P}\left(\frac{F_{n,1}^{(r,R)}}{b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} > \delta\right) \leq -\frac{\delta^2}{2\kappa_R b}.$$

Letting R to infinity, we obtain that

$$\limsup_{R\to\infty}\limsup_{n\to\infty}\frac{1}{b^2_{|\mathbb{T}_{n-1}|}}\log\mathbb{P}\left(\frac{F^{(r,R)}_{n,1}}{b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}}>\delta\right)=-\infty.$$

Now it is easy to check that the same works for $-F^{(r,R)}_{n,1}$ and for the others components of $F^{(r,R)}_n$. We thus conclude (3.5.14) holds for all $r>0$.

Step 3. By application of Theorem 4.2.16 in [35], we find that

$$(M_n/(b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}))$$

satisfies an MDP on $\mathbb{R}^{2(p+1)}$ with speed $b^2_{|\mathbb{T}_{n-1}|}$ and rate function

$$\widetilde{I}(x)=\sup_{\delta>0}\liminf_{R\to\infty}\inf_{z\in B_{x,\delta}}I_R(z),$$

where I_R is given in (5.4.81) and $B_{x,\delta}$ denotes the ball $\{z:|z-x|<\delta\}$. The identification of the rate function $\widetilde{I}=I_M$, where I_M is given in (3.5.3) is done easily (see for example [39]), which concludes the proof of Theorem 3.5.1.

Proof in case 1.
For the proof in case 1, there are no changes in Step 1, and Step 3, instead of (5.4.80), (3.5.8), and **(N1)**, we use Remark 3.5.4 and **(G1)**. In Step 2, the negligibility in (3.5.11), comes from the MDP of the i.i.d. sequences $(\varepsilon_{2k}-\varepsilon^{(R)}_{2k})$ since it verifies the condition, for $\lambda>0$ and all $R>0$

$$\mathbb{E}(\exp(\lambda(\varepsilon_{2k}-\varepsilon^{(R)}_{2k}))<\infty.$$

The negligibility of $(L^{(r)}_n)$ works in the same way. For $(F^{(r,R)}_n)$ we will use the MDP for martingale, see Proposition 3.3.10. For R large enough, we have

$$\begin{aligned}\mathbb{P}\left(\left|X^{(r)}_{k,n}\left(\varepsilon_{2k}-\varepsilon^{(R)}_{2k}\right)\right|>b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}\,\Big|\mathcal{F}_{k-1}\right)&\leq\mathbb{P}\left(\left|\varepsilon_{2k}-\varepsilon^{(R)}_{2k}\right|>\frac{b^2_{|\mathbb{T}_{n-1}|}}{r}\right)\\&=\mathbb{P}\left(\left|\varepsilon_{2}-\varepsilon^{(R)}_{2}\right|>\frac{b^2_{|\mathbb{T}_{n-1}|}}{r}\right)\\&=0.\end{aligned}$$

This implies that

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{b^2_{|\mathbb{T}_{n-1}|}}\log\Bigg(|\mathbb{T}_{n-1}|\ \operatorname*{ess\,sup}_{k\geq1}\mathbb{P}\Bigg(&\left|X^{(r)}_{k,n}\left(\varepsilon_{2k}-\varepsilon^{(R)}_{2k}\right)\right|\\&>b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}\,\Big|\mathcal{F}_{k-1}\Bigg)\Bigg)=-\infty.\end{aligned}$$

That is condition **(D2)** in Proposition 3.3.10.

For all $\gamma > 0$ and all $\delta > 0$, we obtain from Remark 3.5.4 , that

$$\limsup_{n\to\infty} \frac{1}{b^2_{|\mathbb{T}_{n-1}|}} \log \mathbb{P}\left(\frac{1}{|\mathbb{T}_{n-1}|} \sum_{k\in\mathbb{T}_{n-1,p-1}} \left(X^{(r)}_{k,n}\right)^2 \mathrm{I}_{\left\{|X^{(r)}_{k,n}|>\gamma\frac{\sqrt{|\mathbb{T}_{n-1}|}}{b_{|\mathbb{T}_{n-1}|}}\right\}} > \delta \right)$$
$$\leq \limsup_{n\to\infty} \frac{1}{b^2_{|\mathbb{T}_{n-1}|}} \log \mathbb{P}\left(\frac{1}{|\mathbb{T}_{n-1}|} \sum_{k\in\mathbb{T}_{n-1,p-1}} X_k^2 \mathbf{1}_{\left\{|X_k|>\gamma\frac{\sqrt{|\mathbb{T}_{n-1}|}}{b_{|\mathbb{T}_{n-1}|}}\right\}} > \delta \right) = -\infty.$$

That is condition **(D3)** in Proposition 3.3.10. Finally, from Remark 3.5.4 and in the same way as in (3.5.9), it follows that

$$\frac{\langle F^{(r,R)}\rangle_{n,1}}{|\mathbb{T}_{n-1}|} = Q_R \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k\in\mathbb{T}_{n-1,p-1}} (X^{(r)}_{k,n})^2 \overset{\text{superexp}}{\underset{b^2_{|\mathbb{T}_{n-1}|}}{\Longrightarrow}} Q_R \ell$$

for some positive constant ℓ, where $Q_R = \mathbb{E}\left[\left(\varepsilon_2 - \varepsilon_2^{(R)}\right)^2\right]$. That is condition **(D1)** in Proposition 3.3.10. Moreover, it is clear that Q_R converges to 0 as R goes to infinity. In light of foregoing, we infer from Proposition 3.3.10, that $(F^{(r,R)}_{n,1}/(b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}))$ satisfies an MDP on $\mathbb{R}$ of speed $b^2_{|\mathbb{T}_{n-1}|}$ and rate function $I_R(x) = x^2/(2Q_R\ell)$. In particular, this implies that for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{b^2_{|\mathbb{T}_{n-1}|}} \log \mathbb{P}\left(\frac{|F^{(r,R)}_{n,1}|}{b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}} > \delta \right) \leq -\frac{\delta^2}{2Q_R\ell},$$

and letting R go to infinity clearly leads to the result.

Proof of Theorem 3.3.4

The proof works in case 1 and in case 2. From (3.2.14), we have

$$\frac{\sqrt{|\mathbb{T}_{n-1}|}}{b_{|\mathbb{T}_{n-1}|}}(\widehat{\theta}_n - \theta) = |\mathbb{T}_{n-1}|\Sigma^{-1}_{n-1} \frac{M_n}{b_{|\mathbb{T}_{n-1}|}|\mathbb{T}_{n-1}|}$$

From Proposition 3.4.1, we obtain that

$$\frac{\Sigma_n}{|\mathbb{T}_n|} = I_2 \otimes \frac{S_n}{|\mathbb{T}_n|} \overset{\text{superexp}}{\underset{b^2_{|\mathbb{T}_n|}}{\Longrightarrow}} I_2 \otimes L. \tag{3.5.17}$$

According to Lemma 4.1 of [113], together with (3.5.17), we deduce that

$$|\mathbb{T}_{n-1}|\Sigma^{-1}_{n-1} \overset{\text{superexp}}{\underset{b^2_{|\mathbb{T}_{n-1}|}}{\Longrightarrow}} I_2 \otimes L^{-1}. \tag{3.5.18}$$

From Theorem 3.5.1, (3.5.18) and the contraction principle [35], we deduce that the sequence $\left(\sqrt{|\mathbb{T}_{n-1}|}(\widehat{\theta}_n - \theta)/b_{|\mathbb{T}_{n-1}|}\right)_{n\geq 1}$ satisfies the MDP with rate function I_θ given by (3.3.3).

3.5.3 Proof of Theorem 3.3.6

For the proof of Theorem 3.3.6, the case 1 is an easy consequence of the classical MDP for i.i.d.r.v. applied to the sequence $(\varepsilon_{2k}^2 + \varepsilon_{2k+1}^2)$, for the case 2, we will use Proposition 3.3.10, rather than Puhalskii's Theorem 3.3.11.

We will prove that the sequence$\left(\sqrt{|\mathbb{T}_{n-1}|}(\sigma_n^2 - \sigma^2)/b_{|\mathbb{T}_{n-1}|}\right)$ satisfies the MDP. For that,we will prove that conditions **(D1), (D2)** and **(D3)** of Proposition 3.3.10 are verified. Let us consider the $\mathcal{G}_n$-martingale $(N_n)_{n\geq 2^{p-1}}$ given by

$$N_n = \sum_{k=2^{p-1}}^{n} \nu_k, \quad \text{where } \nu_k = \varepsilon_{2k}^2 + \varepsilon_{2k+1}^2 - 2\sigma^2.$$

It is easy to see that its predictable quadratic variation is given by

$$\langle N \rangle_n = \sum_{k=2^{p-1}}^{n} \mathbb{E}\left[\nu_k^2 | \mathcal{G}_{k-1}\right] = (n - 2^{p-1} + 1)(2\tau^4 - 4\sigma^4 + 2\nu^2),$$

which immediately implies that

$$\frac{\langle N \rangle_n}{n} \underset{b_n^2/n}{\overset{\text{superexp}}{\Longrightarrow}} 2\tau^4 - 4\sigma^4 + 2\nu^2,$$

ensuring condition **(D1)** in Proposition 3.3.10.

Next, for $B > 0$ large enough, we have for $a > 2$ (in **(Ea)**), and some positive constant c

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=2^{p-1}}^{n} |\nu_k|^a > B\right) \leq 3 \max_{\eta\in\{0,1\}} \left\{\mathbb{P}\left(\frac{1}{n}\sum_{k=2^{p-1}}^{n} |\varepsilon_{2k+\eta}|^{2a} > \frac{B}{3c}\right)\right\}.$$

From hypothesis **(Ea)** and since B is large enough, we obtain, for a suitable $t > 0$ via the Chernoff inequality and several successive conditioning on $(\mathcal{G}_n)$, for $\eta \in \{0, 1\}$

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=2^{p-1}}^{n} |\varepsilon_{2k+\eta}|^{2a} > \frac{B}{3c}\right) \leq \exp\left(-tn\left(\frac{B}{3c} - \log E\right)\right) \leq \exp\left(-tc'n\right),$$

where c, c' are a positive generic constant. Therefore, for $B > 0$ large enough, we deduce that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}\sum_{k=2^{p-1}}^{n} |\nu_k|^a > B\right) < 0,$$

and this implies (see e.g [113]) exponential Lindeberg condition, that is for all $r > 0$

$$\frac{1}{n}\sum_{k=2^{p-1}}^{n}\nu_k^2\mathbf{1}_{\left\{|\nu_k|>r\frac{\sqrt{n}}{b_n}\right\}}\overset{\text{superexp}}{\underset{b_n^2/n}{\Longrightarrow}}0.$$

That is condition **(D3)** in Proposition 3.3.10.

Now, for all $k\in\mathbb{N}$ and a suitable $t>0$ we have

$$\begin{aligned}\mathbb{P}\left(|\nu_k|>b_n\sqrt{n}|\mathcal{G}_{k-1}\right) &\leq \sum_{\eta=0}^{1}\mathbb{P}\left(|\varepsilon_{2k+\eta}^2-\sigma^2|>\frac{b_n\sqrt{n}}{2}|\mathcal{G}_{k-1}\right)\\ &\leq \exp\left(\frac{-tb_n\sqrt{n}}{2}\right)\sum_{\eta=0}^{1}\mathbb{E}\Big[\exp\left(t|\varepsilon_{2k+\eta}^2-\sigma^2|\right)|\mathcal{G}_{k-1}\Big]\\ &\leq 2E'\exp\left(\frac{-tb_n\sqrt{n}}{2}\right),\end{aligned}$$

where from hypothesis **(Na)**, E' is finite and positive. We are thus led to

$$\frac{1}{b_n^2}\log\left(n\ \operatorname*{ess\,sup}_{k\in\mathbb{N}^*}\mathbb{P}\left(|\nu_k|>b_n\sqrt{n}|\mathcal{G}_{k-1}\right)\right)\leq\frac{\log(2E'n)}{b_n^2}-\frac{t\sqrt{n}}{b_n},$$

and consequently, letting n goes to infinity, we get the condition **(D2)** in Proposition 3.3.10.

Now, applying Proposition 3.3.10, we conclude that $(N_n/(b_n\sqrt{n}))_{n\geq 0}$ satisfies the MDP with speed b_n^2 and rate function

$$I_N(x)=\frac{x^2}{4(\tau^4-2\sigma^4+2\nu^2)}.$$

Applying the foregoing to $|\mathbb{T}_{n-1}|$ and using contraction principle (see e.g [35]), we deduce that the sequence

$$\frac{\sqrt{|\mathbb{T}_{n-1}|}}{b_{|\mathbb{T}_{n-1}|}}(\sigma_n^2-\sigma^2)=\frac{N_{|\mathbb{T}_{n-1}|}}{2b_{|\mathbb{T}_{n-1}|}\sqrt{|\mathbb{T}_{n-1}|}}$$

satisfies a MDP with speed $b_{|\mathbb{T}_{n-1}|}^2$ and rate function I_{σ^2} given by (3.3.4).

We obtain as in the proof of the first part, with a slight modification that the sequence $(|\mathbb{T}_{n-1}|(\rho_n-\rho)/b_{|\mathbb{T}_{n-1}|})$ satisfies a MDP with speed $b_{|\mathbb{T}_{n-1}|}^2$ and rate function I_ρ given by (3.3.5).

3.5.4 Proof of Theorem 3.3.9

Here also the proof works for the two cases. Let us first deal with $\widehat{\sigma}_n$. We have

$$\widehat{\sigma}_n^2-\sigma^2=(\widehat{\sigma}_n^2-\sigma_n^2)+(\sigma_n^2-\sigma^2).$$

From (3.4.10) and (3.4.31), we easily deduce that $\sigma_n^2 \underset{b^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} \sigma^2$ in case 1 and in case 2. Thus, it is enough to prove that $\widehat{\sigma_n^2} - \sigma_n^2 \underset{b^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0$. Let $\theta^{(0)} = (a_0, a_1, \cdots, a_p)^t$, $\theta^{(1)} = (b_0, b_1, \cdots, b_p)^t$, $\widehat{\theta}_n^{(0)} = (\widehat{a}_{0,n}, \widehat{a}_{1,n}, \cdots, \widehat{a}_{p,n})$, $\widehat{\theta}_n^{(1)} = (\widehat{b}_{0,n}, \widehat{b}_{1,n}, \cdots, \widehat{b}_{p,n})$.

Let us introduce the following function f defined for x and z in $\mathbb{R}^{p+1}$ by

$$f(x,z) = \left(x_1 - z_1 - \sum_{i=2}^{p+1} z_i x_i\right)^2,$$

where x_i and z_i denote respectively the i-th component of x and z. One can observe that

$$\begin{aligned}\widehat{\sigma}_n^2 - \sigma_n^2 = &\frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \left\{ f\left(\mathbb{X}_{2k}, \widehat{\theta}_n^{(0)}\right) - f\left(\mathbb{X}_{2k}, \theta^{(0)}\right)\right\} \\ &+ \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \left\{ f\left(\mathbb{X}_{2k+1}, \widehat{\theta}_n^{(1)}\right) - f\left(\mathbb{X}_{2k+1}, \theta^{(1)}\right)\right\}.\end{aligned}$$

By the Taylor-Lagrange formula, $\forall x \in \mathbb{R}^{p+1}$ and $\forall z, z' \in \mathbb{R}^{p+1}$, one can find $\lambda \in (0,1)$ such that

$$f(x,z') - f(x,z) = \sum_{j=1}^{p+1} (z'_j - z_j) \partial_{z_j} f\left(x, z + \lambda(z'-z)\right).$$

Let the function g defined by

$$g(x,z) = x_1 - z_1 - \sum_{j=2}^{p+1} z_j x_j.$$

Observing that

$$\begin{cases} \frac{\partial f}{\partial z_1}(x,z) = -2g(x,z) \\ \frac{\partial f}{\partial z_j}(x,z) = -2x_j g(x,z) \quad \forall j \geq 2, \end{cases}$$

we get easily that $\left|\frac{\partial f}{\partial z_j}(x,z)\right| \leq 4(1+\|z\|)(1+\|x\|^2)$ for all $j \geq 1$, and this implies

$$|f(x,z') - f(x,z)| \leq c\|z'-z\| \left(1 + \|z\| + \|z'-z\|\right)\left(1+\|x\|^2\right),$$

for some positive constant c. Now, applying the foregoing to

$$f\left(\mathbb{X}_{2k}, \widehat{\theta}_n^{(0)}\right) - f\left(\mathbb{X}_{2k}, \theta^{(0)}\right) \quad \text{and} \quad f\left(\mathbb{X}_{2k+1}, \widehat{\theta}_n^{(1)}\right) - f\left(\mathbb{X}_{2k+1}, \theta^{(1)}\right),$$

we deduce easily that

$$|\widehat{\sigma}_n^2 - \sigma_n^2| \leq c\|\widehat{\theta}_n - \theta\| \left(1 + \|\theta\| + \|\widehat{\theta}_n - \theta\|\right) \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \left(1 + \|\mathbb{X}_k\|^2\right),$$

for some positive constant c. From the MDP of $\widehat{\theta}_n - \theta$, we infer that

$$\|\widehat{\theta}_n - \theta\| \underset{b^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{3.5.19}$$

Form Proposition 3.4.4 we deduce that

$$\frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \left(1 + \|\mathbb{X}_k\|^2\right) \underset{b^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 1 + \operatorname{Tr}(\Lambda). \tag{3.5.20}$$

We thus conclude via (3.5.19) and (3.5.20) that

$$\widehat{\sigma_n^2} - \sigma_n^2 \underset{b^2_{|\mathbb{T}_{n-1}|}}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

This ends the proof for $\widehat{\sigma}_n$. The proof for $\widehat{\rho}_n$ is very similar and uses hypothesis **(G2')** and **(N2')** to get inequalities similar to (3.4.10) and (3.4.31).

Chapitre 4

Deviation inequalities for bifurcating Markov chains on Galton-Watson tree

4.1 Introduction

Bifurcating Markov chains (BMC) on Galton-Watson (GW) tree are an extension of BMC to GW tree data. They were introduced by Delmas and Marsalle [33] in order to take into account the death of individuals in the Escherichia coli's (E.coli) reproduction model. E.coli is a rod-shaped bacterium which reproduces by dividing in the middle, thus producing two cells. One which has the new pole of the mother and that we call new pole progeny cell, and the other which has the old pole of the mother and that we call old pole progeny cell. In fact, each daughter cell has two poles. One which is new (new pole) and the other which already existed (old pole). The age of a cell is given by the age of its old pole (i.e the number of generations in the past of the cell before the old pole was produced).

Guyon & Al [67] proposed the following linear Gaussian model to describe the evolution of the growth rate of the population of cells derived from an initial individual

$$\mathcal{L}(X_1) = \nu, \qquad \text{and} \qquad \forall n \geq 1, \quad \begin{cases} X_{2n} = \alpha_0 X_n + \beta_0 + \varepsilon_{2n} \\ \\ X_{2n+1} = \alpha_1 X_n + \beta_1 + \varepsilon_{2n+1}, \end{cases} \tag{4.1.1}$$

where X_n is the growth rate of individual n, n is the mother of $2n$ (the new pole progeny cell) and $2n+1$ (the old pole progeny cell), ν is a distribution probability on $\mathbb{R}$, $\alpha_0, \alpha_1 \in (-1,1)$; $\beta_0, \beta_1 \in \mathbb{R}$ and $\big((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1\big)$ forms a sequence of i.i.d bivariate random variables with law $\mathcal{N}_2(0, \Gamma)$, where

$$\Gamma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \sigma^2 > 0, \quad \rho \in (-1,1). \tag{4.1.2}$$

The processes (X_n) defined by (4.1.1) are typical example of BMC which are called the first order bifurcating autoregressive processes (BAR(1)). The BAR(1) processes are an adaptation of autoregressive processes, when the data have a binary tree structure (see Figure 1). They were first introduced by Cowan and Staudte [28] for cell lineage data where each individual in one generation gives rise to two offspring in the next generation.

In [66], Guyon, using the theory of BMC, gave laws of large numbers and central limit theorem for the least-squares estimator $\widehat{\theta}^r = (\widehat{\alpha}_0^r, \widehat{\beta}_0^r, \widehat{\alpha}_1^r, \widehat{\beta}_1^r)$ of the 4-dimensional parameter $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$. He has also built some statistical tests which allow to test if the model is symmetric or not, and if the new pole and the old pole populations are even distinct in mean. This allowed him to conclude a statistical evidence in aging in E. Coli. Let us also mention [18], where Bercu & Al. using the martingale approach give asymptotic analysis of the least squares estimator of the unknown parameters of a general asymmetric pth-order BAR processes.

However, in the BMC model presented by Guyon, cells are assumed to never die (a death corresponds to no more division). To take into account cells's death, Delmas and Marsalle [33], instead of a regular binary tree, used a binary GW tree to label cells. In the sequel, we will introduce the model which allowed them to study the behavior of the growth rate of cells, taking into account their possible death.

4.1.1 The model

Let $\mathbb{T}$ be a binary regular tree in which each vertex is seen as a positive integer different from 0, see Figure 4.1. For $r \in \mathbb{N}$, let

$$\mathbb{G}_r = \Big\{2^r, 2^r + 1, \cdots, 2^{r+1} - 1\Big\}, \quad \mathbb{T}_r = \bigcup_{q=0}^{r} \mathbb{G}_q,$$

which denote respectively the r-th column and the first $(r+1)$ columns of the tree. Then, the cardinality $|\mathbb{G}_r|$ of $\mathbb{G}_r$ is 2^r and that of $\mathbb{T}_r$ is $|\mathbb{T}_r| = 2^{r+1} - 1$. A column of a given integer n is $\mathbb{G}_{r_n}$ with $r_n = \lfloor \log_2 n \rfloor$, where $\lfloor x \rfloor$ denotes the integer part of the real number x.

The genealogy of the cells is described by this tree. In the sequel we will thus see $\mathbb{T}$ as a given population. Then the vertex n, the column $\mathbb{G}_r$ and the first $(r+1)$ columns $\mathbb{T}_r$ designate respectively individual n, the r-th generation and the first $(r+1)$ generations. The initial individual is denoted 1. The model proposed by Delmas and Marsalle [33] is defined as follows. The growth rate of cell n is X_n.

- With probability $p_{1,0}$, n gives birth to two cells $2n$ and $2n+1$ with both divide. The growth rate of the daughters X_{2n} and X_{2n+1} are then linked to the mother's one through auto-regressive equations (4.1.1).

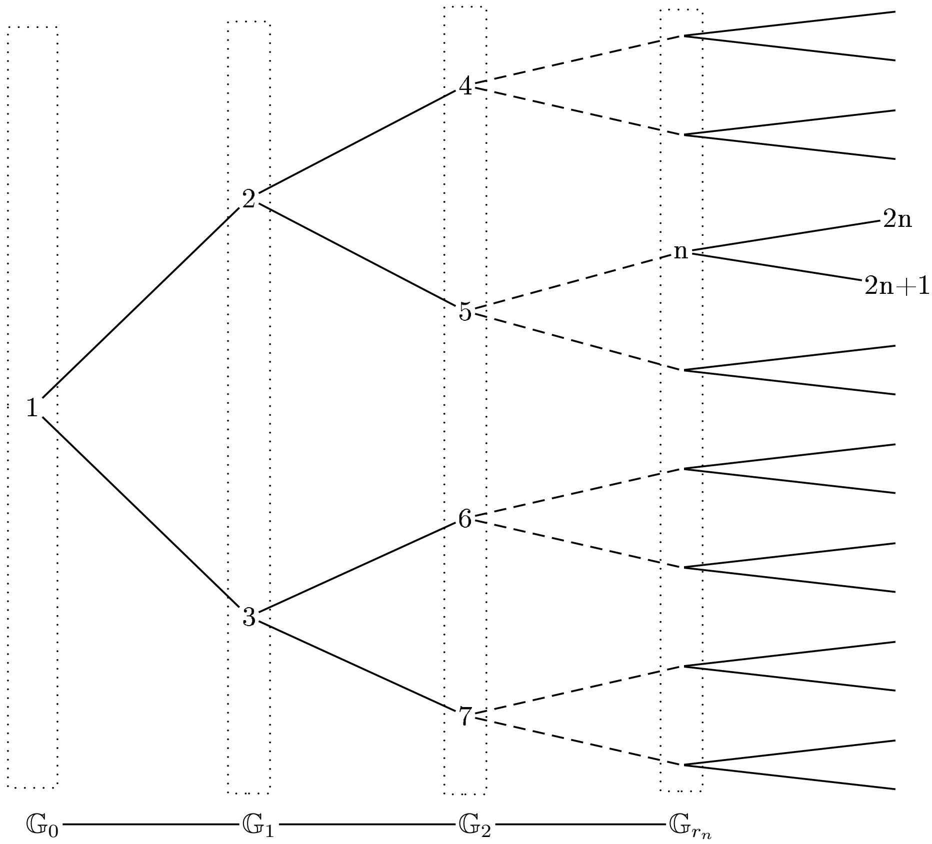

Figure 4.1: The binary tree $\mathbb{T}$

- With probability p_0, only the new pole $2n$ divides. Its growth rate X_{2n} is linked to its mother's one X_n through the relation

$$X_{2n} = \alpha'_0 X_n + \beta'_0 + \varepsilon'_{2n}, \tag{4.1.3}$$

where $\alpha'_0 \in (-1,1)$, $\beta'_0 \in \mathbb{R}$ and $(\varepsilon'_{2n}, n \in \mathbb{T})$ is a sequence of independent centered Gaussian random variables with variance $\sigma_0^2 > 0$.

- With probability p_1, only the old pole $2n+1$ divides. Its growth rate X_{2n+1} is linked to its mother's one X_n through the relation

$$X_{2n+1} = \alpha'_1 X_n + \beta'_1 + \varepsilon'_{2n+1}, \tag{4.1.4}$$

where $\alpha'_1 \in (-1,1)$, $\beta'_1 \in \mathbb{R}$ and $(\varepsilon'_{2n+1}, n \in \mathbb{T})$ is a sequence of independent centered Gaussian random variables with variance $\sigma_1^2 > 0$.

- With probability $1 - p_{1,0} - p_1 - p_0$, which is non-negative, n gives birth to two cells which do not divide.

- The sequences $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \in \mathbb{T})$, $(\varepsilon'_{2n}, n \in \mathbb{T})$ and $(\varepsilon'_{2n+1}, n \in \mathbb{T})$ are independent.

The process (X_n) described above is a typical example of BMC on GW tree. In [32], this process is called bifurcating autoregressive process (BAR) with missing data. It is an extension of bifurcating autoregressive process when the data have a binary GW tree structure, see figure 2 for example of binary GW tree. Indeed, one can assume that the cells which do not divide and those which do not exist are missing or dead.

In [33], Delmas and Marsalle using their results for BMC on GW tree, gave laws of large numbers and central limit theorem for the maximum likelihood estimator of the parameter

$$\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1, \alpha'_0, \beta'_0, \alpha'_1, \beta'_1). \tag{4.1.5}$$

In this paper, we will give deviation inequalities for the least squares estimator of the parameter θ, in case the noise sequence and the initial state X_1 valued in a compact set. Note that this implies that the BAR process with missing data describes above also valued in compact set.

We are now going to give a rigorous definition of BMC on GW tree. We refer to [33] for more details.

4.1.2 Definitions

For an individual $n \in \mathbb{T}$, we are interested in the quantity X_n (it may be the weight, the growth rate,$\cdots$) with values in the metric space S endowed with its Borel σ-field $\mathcal{S}$.

Definition 4.1.1 ($\mathbb{T}$-transition probability, see ([66])). *We call $\mathbb{T}$-transition probability any mappings $P : S \times \mathcal{S}^2 \to [0,1]$ such that*

- *$P(.,A)$ is measurable for all $A \in \mathcal{S}^2$,*
- *$P(x,.)$ is a probability measure on $(S^2, \mathcal{S}^2)$ for all $x \in S$.*

For $p \geq 1$, we denote by $\mathcal{B}(S^p)$(resp. $\mathcal{B}_b(S^p)$, $\mathcal{C}(S^p)$, $\mathcal{C}_b(S^p)$) the set of all $\mathcal{S}^p$-measurable (resp. $\mathcal{S}^p$-measurable and bounded, continuous, continuous and bounded) mapping $f : S^p \to \mathbb{R}$. For $f \in \mathcal{B}(S^3)$, when it is defined, we denote by $Pf \in \mathcal{B}(S)$ the function

$$x \mapsto Pf(x) = \int_{S^2} f(x,y,z)P(x,dy,dz).$$

Definition 4.1.2 (Bifurcating Markov Chains, see ([66])). *Let $(X_n, n \in \mathbb{T})$ be a family of S-valued random variables defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_r, r \in \mathbb{N}), \mathbb{P})$. Let ν be a probability on $(S, \mathcal{S})$ and P be a $\mathbb{T}$-transition probability. We say that $(X_n, n \in \mathbb{T})$ is a $(\mathcal{F}_r)$-bifurcating Markov chain with initial distribution ν and $\mathbb{T}$-transition probability P if*

- *X_n is $\mathcal{F}_{r_n}$-measurable for all $n \in \mathbb{T}$,*
- *$\mathcal{L}(X_1) = \nu$,*
- *for all $r \in \mathbb{N}$ and for all family $(f_n, n \in \mathbb{G}_r) \subseteq \mathcal{B}_b(S^3)$*

$$\mathbb{E}\left[\prod_{n\in\mathbb{G}_r} f_n(X_n, X_{2n}, X_{2n+1})\Big|\mathcal{F}_r\right] = \prod_{n\in\mathbb{G}_r} Pf_n(X_n).$$

Now, we add a cemetery point to S, ∂. Let $\bar{S} = S \cup \{\partial\}$, and $\bar{\mathcal{S}}$ be the $\sigma-$field generated by $\mathcal{S}$ and $\{\partial\}$. In the previous biological framework, S corresponds to the state space of the quantities related to living cells, and ∂ is the default value for dead cells. Let P^* be a $\mathbb{T}$-transition probability defined on $\bar{S} \times \bar{\mathcal{S}}$ such that

$$P^*(\partial, \{(\partial, \partial)\}) = 1. \tag{4.1.6}$$

In the previous biological framework, (4.1.6) means that no dead cell can give birth to a living cell. We denote by P_0^* and P_1^* the restriction of the first and the second marginal of P^* to S, that is:

$$P_0^* = P^*\left(\cdot, \left(\cdot \bigcap S\right) \times \bar{S}\right) \quad \text{and} \quad P_1^* = P^*\left(\cdot, \bar{S} \times \left(\cdot \bigcap S\right)\right).$$

Definition 4.1.3 (BMC on GW tree, see [33]). *Let $X = (X_n, n \in \mathbb{T})$ be a P^*-BMC on $(\bar{S}, \bar{\mathcal{S}})$, with P^* satisfying (4.1.6). We call $(X_n, n \in \mathbb{T}^*)$, with $\mathbb{T}^* = \{n \in \mathbb{T} : X_n \neq \partial\}$, a BMC on GW tree. The P^*-BMC is said spatially homogeneous if $p_{1,0} = P^*(x, S \times S)$, $p_0 = P^*(x, S \times \{\partial\})$, and $p_1 = P^*(x, \{\partial\} \times S)$ do not depend on $x \in S$. A spatially homogeneous P^*-BMC is said super-critical if $m > 1$, where $m = 2p_{1,0} + p_1 + p_0$.*

We denote by $(Y_n, n \in \mathbb{N})$ the Markov chain on S with $Y_0 = X_1$ and transition probability $Q = \frac{1}{m}(P_0^* + P_1^*)$.

Remark 4.1.4.
- *The name BMC on GW tree comes from the fact that condition (4.1.6) and spatial homogeneity imply that $\mathbb{T}^*$ is a GW tree.*
- *All through this work, we shall assume that the P^*-BMC is super-critical.*

Now, for any subset $J \subset \mathbb{T}$, let

$$J^* = J \cap \mathbb{T}^* = \{j \in J : X_j \neq \partial\}$$

be the subset of living cells among J, and $|J|$ be the cardinal of J. The process $(|\mathbb{G}_k^*|, k \in \mathbb{N})$, is a GW process with the reproduction generating function

$$\psi(z) = (1 - p_0 - p_1 - p_{1,0}) + (p_0 + p_1)z + p_{1,0}z^2,$$

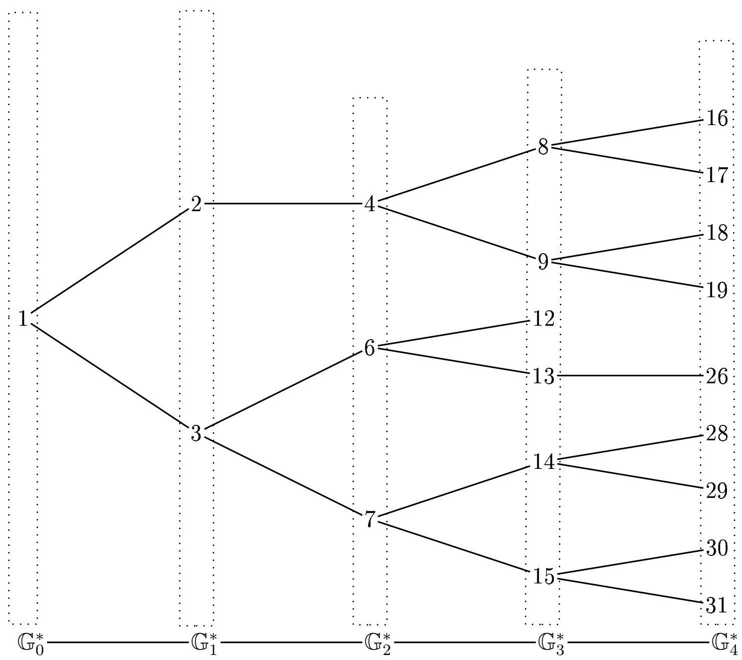

Figure 4.2: A binary GW tree up to the 4 th generation. *In this tree, individual 1 gives birth to two individuals which both divide, this happen with probability* $p_{1,0}$. *Individual 2 gives birth to two individuals which only one (the new pole) divides, this happen which probability* p_0. *Individual 12 gives birth to two individuals which do not divide, this happen with probability* $1 - p_{1,0} - p_0 - p_1$.

and the average number of daughters alive is m. It is known, see e.g [5], that $m^{-k}|\mathbb{G}_k^*|$ converges in probability to a non-negative random variable W. Moreover, $\mathbb{P}(W > 0) = 1$ iff there is no extinction. We have for all $r \geq 0$,

$$\mathbb{E}\left[|\mathbb{G}_r^*|\right] = m^r \quad \text{and} \quad \mathbb{E}\left[|\mathbb{T}_r^*|\right] = \sum_{q=0}^{r} \mathbb{E}\left[|\mathbb{G}_q^*|\right] = \frac{m^{r+1}-1}{m-1} := t_r. \tag{4.1.7}$$

It is known, see [33], that $t_r^{-1}|\mathbb{T}_r^*|$ converges in probability to W as well.

For $i \in \mathbb{T}$, set $\Delta_i = (X_i, X_{2i}, X_{2i+1})$ the mother-daughters quantities of interest. For a finite subset $J \subset \mathbb{T}$, we set

$$M_J(f) = \begin{cases} \sum_{i \in J} f(X_i) & \text{for } f \in \mathcal{B}(\bar{S}), \\ \sum_{i \in J} f(\Delta_i) & \text{for } f \in \mathcal{B}(\bar{S}^3), \end{cases} \tag{4.1.8}$$

with the convention that a sum over an empty set is null. We also define the following two averages of f over J

$$\overline{M}_J(f) = \frac{1}{|J|} M_J(f) \quad \text{if } |J| > 0 \quad \text{and} \quad \widetilde{M}_J(f) = \frac{1}{\mathbb{E}\left[|J|\right]} M_J(f) \quad \text{if } \mathbb{E}\left[|J|\right] > 0. \tag{4.1.9}$$

Limit theorems for averages (4.1.9) have been studied in [33] for $J = \mathbb{G}_n^*$ and $J = \mathbb{T}_n^*$, as n goes to infinity. Under uniform geometric ergodicity assumption for Q, we will establish in this paper deviation inequalities for this averages. Notice that deviation inequalities were already studied in the no death case [22], that is $m = 2$. We will follow essentially the same approach that the latter paper for the proofs of our results. Let us also mention [21], where the authors establish deviation inequalities for estimators of parameters of the p-order bifurcating autoregressive process.

The rest of this chapter is organized as follows. In section 4.2, we states our main results, that is deviation inequalities for averages (4.1.9), for $J = \mathbb{G}_n^*$ and $J = \mathbb{T}_n^*$. This will be done under uniform geometric ergodicity assumption for Q, and suitable assumptions on the binary GW tree. In section 4.3, we will interest particularly to the first order bifurcating autoregressive process with missing data described in section 4.1.1. Section 4.4 is dedicated to the proofs of our results.

4.2 Main results

We consider the following hypothesis:

(H1): There exists a probability measure μ on $(S, \mathcal{S})$ such that for all $f \in \mathcal{B}_b(S)$ with $\langle \mu, f \rangle = 0$, there is $c > 0$ such that for all $k \in \mathbb{N}$ and for all $x \in S$, $|Q^k f(x)| \leq c\alpha^k$.

(H2): $m > \sqrt{2}$.

(H3): $p_{1,0} + p_0 + p_1 = 1$, where $p_{1,0}$, p_0 and p_1 are defined in section 4.1.1.

Remark 4.2.1. *Hypothesis* **(H1)** *implies that the Markov chain Y is ergodic, that is for all $f \in \mathcal{C}_b(S)$ and for all $x \in S$, $\lim_{k\to\infty} \mathbb{E}_x[f(Y_k)] = \langle \mu, f \rangle$. Assuming hypothesis* ***(H3)*** *means that we work conditionally to the non-extinction.*

In the sequel, $\mathbb{H}_r$ will denote one of the set $\mathbb{G}_r$ or $\mathbb{T}_r$. We set $h_r = (m^2/2)^r$ if $\mathbb{H}_r = \mathbb{G}_r$ and $h_r = (m^2/2)^{r+1}$ if $\mathbb{H}_r = \mathbb{T}_r$. We can now state our main results. Notice that any function f defined on S is extended to $\bar{S}$ by setting $f(\partial) = 0$.

Theorem 4.2.2. *Under hypothesis* **(H1)** *and* **(H2)**, *let $f \in \mathcal{B}_b(S)$ such that*

$\langle \mu, f\rangle = 0$. *Then we have for all* $\delta > 0$

$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) > \delta\Big) \leq \begin{cases} \exp\left(c''\delta\right)\exp\left(-c'\delta^2 h_r\right), & \forall r \in \mathbb{N}, \quad \textit{if } m\alpha < 1, \\ \exp\left(c''\delta\right)\exp\left(-c'\delta^2 (m^2/2)^r\right), \forall r \in \mathbb{N}, \textit{if } \alpha m = 1 \textit{ and } \mathbb{H}_r = \mathbb{G}_r, \\ \exp\left(c''\delta(r+1)\right)\exp\left(-c'\delta^2 (m^2/2)^{r+1}\right), \forall r \in \mathbb{N}, \textit{if } \alpha m = 1 \textit{ and } \mathbb{H}_r = \mathbb{T}_r \\ \exp\left(-c'\delta^2 h_r\right), \forall r \in \mathbb{N} \textit{ such that } r > r_0, \textit{ if } 1 < m\alpha < \sqrt{2}, \\ \exp\left(-\frac{c'\delta^2 h_r}{r}\right), \forall r \in \mathbb{N} \textit{ such that } r > r_0, \textit{ if } m\alpha = \sqrt{2}, \\ \exp\left(-\frac{c'\delta^2}{\alpha^{2r}}\right), \forall r \in \mathbb{N}^* \textit{ such that } r > r_0, \textit{ if } m\alpha > \sqrt{2}. \end{cases} \tag{4.2.1}$$

Furthermore, under additional hypothesis **(H3)**, *we have for all* $f \in \mathcal{B}_b(S)$

$$\mathbb{P}\Big(\widetilde{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f\rangle W > \delta\Big) \leq \begin{cases} \exp\left(c''\delta\right)\exp\left(-c'\delta^2 h_r\right) + A_r, & \forall r \in \mathbb{N}, \quad \textit{if } m\alpha < 1, \\ \exp\left(c''\delta\right)\exp\left(-c'\delta^2 (m^2/2)^r\right) + A_r, \forall r \in \mathbb{N}, \textit{if } \alpha m = 1 \\ \qquad \textit{and } \mathbb{H}_r = \mathbb{G}_r, \\ \exp\left(c''\delta(r+1)\right)\exp\left(-c'\delta^2 (m^2/2)^{r+1}\right) + A_r, \forall r \in \mathbb{N}, \textit{if } \alpha m = 1 \\ \qquad \textit{and } \mathbb{H}_r = \mathbb{T}_r \\ \exp\left(-c'\delta^2 h_r\right) + A_r, \forall r \in \mathbb{N} \textit{ such that } r > r_0, \textit{ if } 1 < m\alpha < \sqrt{2}, \\ \exp\left(-\frac{c'\delta^2 h_r}{r}\right) + A_r, \forall r \in \mathbb{N} \textit{ such that } r > r_0, \textit{ if } m\alpha = \sqrt{2}, \\ \exp\left(-\frac{c'\delta^2}{\alpha^{2r}}\right) + A_r, \forall r \in \mathbb{N}^* \textit{ such that } r > r_0, \textit{ if } m\alpha > \sqrt{2}, \end{cases} \tag{4.2.2}$$

where,

- *for all* $r \in \mathbb{N}$,

$$A_r = \begin{cases} c' \exp\left(-c''\delta^{2/3}(m^{1/3})^r\right) & \textit{if } \mathbb{H}_r = \mathbb{G}_r \\ \exp\left(c'\delta^{2/3}\right)\exp\left(-c''\delta^{2/3}\left(t_r/(r+1)^2\right)^{1/3}\right) & \textit{if } \mathbb{H}_r = \mathbb{T}_r, \end{cases}$$

- $r_0 := \log\left(\frac{\delta}{c_0}\right)/\log(\alpha) - k_0$, *with* $k_0 \in \{0, 1\}$,

- c_0, c' *and* c'' *are positive constants which depend on* α, m, *and* c *and may differ line by line.*

The next results can be seen as a consequence of the previous results.

Theorem 4.2.3. *We assume that hypothesis* **(H1)-(H3)** *are satisfied. Let* $f \in \mathcal{B}_b(S)$. *For all* $\delta > 0$, *for all* $a > 0$ *and for all* $b > 0$ *such that* $b < a/(\delta+1)$, *we have*

$$\mathbb{P}\Big(\overline{M}_{\mathbb{H}_r^*}(f) - \langle \mu, f\rangle > \delta \Big| W \geq a\Big)$$
$$\leq \begin{cases} \exp\left(c''\delta b\right)\exp\left(-c'(\delta b)^2 h_r\right) + A_r, & \forall r \in \mathbb{N}, \quad \text{if } m\alpha < 1, \\ \exp\left(c''\delta b\right)\exp\left(-c'(\delta b)^2 (m^2/2)^r\right) + A_r, \forall r \in \mathbb{N}, \text{if } \alpha m = 1 & \\ \qquad \text{and } \mathbb{H}_r = \mathbb{G}_r, & \\ \exp\left(2c'\delta b(r+1)\right)\exp\left(-c'(\delta b)^2 (m^2/2)^{r+1}\right) + A_r, \forall r \in \mathbb{N}, & \\ \qquad \text{if } \alpha m = 1 \text{ and } \mathbb{H}_r = \mathbb{T}_r & \\ \exp\left(-c'(\delta b)^2 h_r\right) + A_r, \forall r \in \mathbb{N} \text{ such that } r > r_0, & \\ \qquad \text{if } 1 < m\alpha < \sqrt{2}, & \\ \exp\left(-\frac{c'(\delta b)^2 h_r}{r}\right) + A_r, \forall r \in \mathbb{N} \text{ such that } r > r_0, \text{ if } m\alpha = \sqrt{2}, & \\ \exp\left(-\frac{c'(\delta b)^2}{\alpha^{2r}}\right) + A_r, \forall r \in \mathbb{N}^* \text{ such that } r > r_0, \text{ if } m\alpha > \sqrt{2}, & \end{cases} \tag{4.2.3}$$

where,

- *for all* $r \in \mathbb{N}$,

$$A_r = \begin{cases} c' \exp\left(-c''(\delta b)^{2/3}(m^{1/3})^r\right) & \text{if } \mathbb{H}_r = \mathbb{G}_r \\ \exp\left(c'(\delta b)^{2/3}\right)\exp\left(-c''(\delta b)^{2/3}\left(t_r/(r+1)^2\right)^{1/3}\right) & \text{if } \mathbb{H}_r = \mathbb{T}_r, \end{cases}$$

- $r_0 := \log\left(\frac{\delta b}{c_0}\right)/\log(\alpha) - k_0$, *with* $k_0 \in \{0, 1\}$,
- c_0, c' *and* c'' *are positive constants which depend on* α, m, a, *and* c, *and may differ line by line.*

We have the following extension of above theorems when f does not only depend on an individual X_i, but on the mother-daughters triangle Δ_i.

Theorem 4.2.4. *Let* $f \in \mathcal{B}_b(S^3)$. *If* $\langle \mu, P^* f\rangle = 0$, *then, under hypothesis* ***(H1)*** *and* ***(H2)***, *we have deviation inequalities (4.2.1) for* $\widetilde{M}_{\mathbb{H}_r^*}(f)$. *If* $\langle \mu, P^* f\rangle \neq 0$, *under additional hypothesis* ***(H3)***, *we have deviation inequalities (4.2.2) for* $\widetilde{M}_{\mathbb{H}_r^*}(f) - \langle \mu, P^* f\rangle W$ *and (4.2.3) for* $\overline{M}_{\mathbb{H}_r^*}(f) - \langle \mu, P^* f\rangle$.

4.3 Application:First order bifurcating autoregressive processes with missing data

We consider the asymmetric auto-regressive processes given in section 4.1.1. Notice that the process $(X_i, i \in \mathbb{T})$ defined in section 4.1.1, with the convention that $X_i = \partial$ if the cell i is missing, is a spatially homogeneous BMC on a GW tree. We will assume $2p_{1,0}+p_1+p_0 > \sqrt{2}$. This implies particularly that the BMC on GW is super-critical. We will also assume that the noise sequences $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \in \mathbb{T})$, $(\varepsilon'_{2n}, n \in \mathbb{T})$ and $(\varepsilon'_{2n+1}, n \in \mathbb{T})$, and the initial state X_1 valued in a compact set. The latter implies that the process $(X_i, i \in \mathbb{T})$ is bounded. We denote by S the state space of $(X_i, i \in \mathbb{T})$. We assume without loss of generality that S is a compact subset of $\mathbb{R}$.

Let $\mathbb{T}_n^{0,1}$ be the subset of cells in $\mathbb{T}_n^*$ with two living daughters, $\mathbb{T}_n^0$ (resp.$\mathbb{T}_n^1$) be the set of cells of $\mathbb{T}_n^*$ with only the new (resp. old) pole daughter alive:

$$\mathbb{T}_n^{1,0} = \{i \in \mathbb{T}_n^* : \Delta_i \in S^3\}, \qquad \mathbb{T}_n^0 = \{i \in \mathbb{T}_n^* : \Delta_i \in S^2 \times \{\partial\}\}$$

and

$$\mathbb{T}_n^1 = \{i \in \mathbb{T}_n^* : \Delta_i \in S \times \{\partial\} \times S\}.$$

We compute the least-squares estimator (LSE)

$$\widehat{\theta}_n = (\widehat{\alpha}_0^n, \widehat{\beta}_0^n, \widehat{\alpha}_1^n, \widehat{\beta}_1^n, \widehat{\alpha}_0'^n, \widehat{\beta}_0'^n, \widehat{\alpha}_1'^n, \widehat{\beta}_1'^n)$$

of θ given by (4.1.5), based on the observation of a sub-tree $\mathbb{T}_{n+1}^*$. Consequently, we obviously have for $\eta \in \{0,1\}$,

$$\widehat{\alpha}_\eta^n = \frac{|\mathbb{T}_n^{1,0}|^{-1} \sum_{i \in \mathbb{T}_n^{1,0}} X_i X_{2i+\eta} - \left(|\mathbb{T}_n^{1,0}|^{-1} \sum_{i \in \mathbb{T}_n^{1,0}} X_i\right)\left(|\mathbb{T}_n^{1,0}|^{-1} \sum_{i \in \mathbb{T}_n^{1,0}} X_{2i+\eta}\right)}{|\mathbb{T}_n^{1,0}|^{-1} \sum_{i \in \mathbb{T}_n^{1,0}} X_i^2 - \left(|\mathbb{T}_n^{1,0}|^{-1} \sum_{i \in \mathbb{T}_n^{1,0}} X_i\right)^2},$$

$$\widehat{\beta}_\eta^n = |\mathbb{T}_n^{1,0}|^{-1} \sum_{i \in \mathbb{T}_n^{1,0}} X_{2i+\eta} - \widehat{\alpha}_\eta^n |\mathbb{T}_n^{1,0}|^{-1} \sum_{i \in \mathbb{T}_n^{1,0}} X_i,$$

$$\widehat{\alpha}_\eta'^n = \frac{|\mathbb{T}_n^\eta|^{-1} \sum_{i \in \mathbb{T}_n^\eta} X_i X_{2i+\eta} - \left(|\mathbb{T}_n^\eta|^{-1} \sum_{i \in \mathbb{T}_n^\eta} X_i\right)\left(|\mathbb{T}_n^\eta|^{-1} \sum_{i \in \mathbb{T}_n^\eta} X_{2i+\eta}\right)}{|\mathbb{T}_n^\eta|^{-1} \sum_{i \in \mathbb{T}_n^\eta} X_i^2 - \left(|\mathbb{T}_n^\eta|^{-1} \sum_{i \in \mathbb{T}_n^\eta} X_i\right)^2},$$

$$\widehat{\beta}_\eta'^n = |\mathbb{T}_n^\eta|^{-1} \sum_{i \in \mathbb{T}_n^\eta} X_{2i+\eta} - \widehat{\alpha}_\eta'^n |\mathbb{T}_n^\eta|^{-1} \sum_{i \in \mathbb{T}_n^\eta} X_i.$$

Notice that those LSE are based on polynomial functions of the observations. So, since the latter are bounded, we are in the functional setting of the results of section 4.2. Recalling the Markov chain $(Y_n, n \in \mathbb{N})$, notice that Y_n is distributed as $Z_n = a_1 a_2 \cdots a_{n-1} a_n Y_0 + \sum_{k=1}^{n} a_1 a_2 \cdots a_{k-1} b_k$, where $b_n = b'_n + s_n e_n$, $((a_n, b'_n, s_n), n \geq 1)$ is a sequence of independent identically distributed random variables, whose common distribution is given by, for $\eta \in \{0,1\}$,

$$\mathbb{P}(a_1 = \alpha_\eta, b'_1 = \beta_\eta, s_1 = \sigma) = \frac{p_{1,0}}{m} \quad \text{and} \quad \mathbb{P}(a_1 = \alpha'_\eta, b'_1 = \beta'_\eta, s_1 = \sigma_\eta) = \frac{p_\eta}{m},$$

$(e_n, n \geq 1)$ is a sequence of independent $\mathcal{N}(0,1)$ random variables, and is independent of $((a_n, b'_n, s_n), n \geq 1)$, and both sequences are independent of Y_0. Moreover, it is easy to check that the sequence $(Z_n, n \in \mathbb{N})$ converge a.s. to a limit Z, which implies that the Markov chain $(Y_n, n \in \mathbb{N})$ converge in distribution to Z. We refer to [33], section 6, for more details. Following the proof of Proposition 28, step 1 in [66], we check hypothesis **(H1)** with $\alpha = \max(|\alpha_0|, |\alpha_1|, |\alpha'_0|, |\alpha'_1|) < 1$ and with μ the distribution of Z. Let $\mu_1 = \mathbb{E}[Z]$ and $\mu_2 = \mathbb{E}[Z^2]$. We have (see [33])

$$\mu_1 = \frac{\overline{\beta}}{1 - \overline{\alpha}} \quad \text{and} \quad \mu_2 = \frac{2\overline{\alpha\beta}\overline{\beta}/(1-\overline{\alpha}) + \overline{\beta^2} + \overline{\alpha^2}}{1 - \overline{\alpha^2}},$$

where $\overline{\alpha} = \mathbb{E}[a_1]$, $\overline{\alpha^2} = \mathbb{E}\left[a_1^2\right]$, $\overline{\beta} = \mathbb{E}[b_1]$, $\overline{\beta^2} = \mathbb{E}\left[b_1^2\right]$, $\overline{\alpha\beta} = \mathbb{E}[a_1 b_1]$ and $\overline{\sigma^2} = \mathbb{E}\left[s_1^2\right]$.

We then have the following deviation inequality for $\widehat{\theta}_n - \theta$.

Proposition 4.3.1. *For all $\delta > 0$, for all $a > 0$, for all $b > 0$ and for all $\gamma > 0$ such that $b < a/(\delta + 1)$ and $\gamma < \min\left\{c_1/(1+\delta), c_1/\left(1+\sqrt{\delta}\right)\right\}$, where c_1 is a positive constant which depends on $p_{1,0}$, p_0, p_1, μ_1 and μ_2, and for $n_0 := \left(\log\left(\gamma^q \delta^p b / c_0\right) / \log \alpha\right) - 1$, we have*

$$\mathbb{P}\left(\|\widehat{\theta}_n - \theta\| > \delta | W \geq a\right) \leq$$
$$\begin{cases} c_2 \exp\left(c'' \gamma^q \delta^p b\right) \exp\left(-c' \left(\gamma^q \delta^p b\right)^2 \left(m^2/2\right)^{n+1}\right) + A_n + C_n & \forall n \in \mathbb{N} \text{ if } m\alpha < 1, \\ c_2 \exp\left(c'' \gamma^q \delta^p b (n+1)\right) \exp\left(-c' \left(\gamma^q \delta^p b\right)^2 \left(m^2/2\right)^{n+1}\right) + A_n + C_n & \forall n \in \mathbb{N} \text{ if } m\alpha = 1, \\ c_2 \exp\left(-c' \left(\gamma^q \delta^p b\right)^2 \left(m^2/2\right)^{n+1}\right) + A_n + C_n & \forall n > n_0 \text{ if } 1 < m\alpha < \sqrt{2}, \\ c_2 \exp\left(-c' \left(\gamma^q \delta^p b\right)^2 (1/n) \left(m^2/2\right)^{n+1}\right) + A_n + C_n & \forall n > n_0 \text{ if } m\alpha = \sqrt{2}, \\ c_2 \exp\left(-c' \left(\gamma^q \delta^p b\right)^2 \alpha^{-2n}\right) + A_n + C_n & \forall n > n_0 \text{ if } m\alpha > \sqrt{2}, \end{cases} \tag{4.3.1}$$

where

$$C_n = c_3\mathbb{P}\left(\left|\frac{|\mathbb{T}_n^{1,0}|}{|\mathbb{T}_n^*|} - p_{1,0}\right| > c_4\gamma^q\delta^p b \Big| W \geq a\right)$$
$$+ \sum_{\eta=0}^{1} c_3\mathbb{P}\left(\left|\frac{|\mathbb{T}_n^{\eta}|}{|\mathbb{T}_n^*|} - p_{\eta}\right| > c_4\gamma^q\delta^p b \Big| W \geq a\right),$$
$$A_n = c_3 \exp\left(c' \left(\gamma^q\delta^p b\right)^{2/3}\right) \exp\left(-c'' \left(\gamma^q\delta^p b\right)^{2/3} \left(t_n/(n+1)^2\right)^{1/3}\right),$$
$p \in \{1/2, 1\}$, $q \in \{0, 1/2, 1\}$,
c_2, c_3, c_4, c' and c'' are positive constants which depend on c, m, α $p_{1,0}$, p_0, p_1, μ_1 and μ_2.

Remark 4.3.2. *Notice that the constants c_2, c_3, c_4, c' and c'' which appear in Proposition 4.3.1 may differ term by term. The values of p and q depend on the magnitude of δ and γ. For example, for δ and γ small enough, we have $p = 1$ and $q = 1$.*

4.4 Proofs of the main results

4.4.1 Proof of Theorem 4.2.2

Let $f \in \mathcal{B}_b(S)$. We are going to study successively $\widetilde{M}_{\mathbb{H}_r^*}(f)$ for $\mathbb{H}_r = \mathbb{G}_r$ and $\mathbb{H}_r = \mathbb{T}_r$.

Step 1. Let us first deal with $\widetilde{M}_{\mathbb{G}_r^*}(f)$. We first assume that $\langle \mu, f \rangle = 0$. By Chernoff inequality, we have for all $\delta > 0$ and for all $\lambda > 0$

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp\left(-\lambda\delta m^r\right) \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathbb{G}_r^*} f(X_i)\right)\right]. \tag{4.4.1}$$

Recall that for all $i \in \mathbb{G}_{r-1}^*$,

$$\mathbb{E}\left[f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}} + f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}} | \mathcal{F}_{r-1}\right] = mQf(X_i).$$

By subtracting and adding terms in expectation of the right hand of (4.4.1), and conditioning with respect to $\mathcal{F}_{r-1}$, we get

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathbb{G}_r^*} f(X_i)\right)\right] = \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathbb{G}_{r-1}^*} mQf(X_i)\right)\right. \tag{4.4.2}$$
$$\left.\times\, \mathbb{E}\left[\exp\left(\sum_{i \in \mathbb{G}_{r-1}^*} \lambda\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}} + f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}} - mQf(X_i)\right)\right) \Big| \mathcal{F}_{r-1}\right]\right].$$

Observing that $\mathbb{G}_{r-1}^*$ is $\mathcal{F}_{r-1}$ measurable, and using the fact that conditionally to $\mathcal{F}_{r-1}$, the triplets $\{(\Delta_i), i \in \mathbb{G}_{r-1}\}$ are independent (this is due to the Markov property), we have

$$\mathbb{E}\left[\exp\left(\sum_{i\in\mathbb{G}^*_{r-1}}\lambda\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}}+f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}}-mQf(X_i)\right)\right)\middle|\mathcal{F}_{r-1}\right]$$

$$=\prod_{i\in\mathbb{G}^*_{r-1}}\mathbb{E}\left[\exp\left(\lambda\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}}+f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}}-mQf(X_i)\right)\right)\middle|\mathcal{F}_{r-1}\right]. \tag{4.4.3}$$

Using Azuma-Bennet-Hoeffding inequality [6], [16], [70], we get according to **(H1)**, for all $i\in\mathbb{G}^*_{r-1}$,

$$\mathbb{E}\left[\exp\left(\lambda\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}}+f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}}-mQf(X_i)\right)\right)\middle|\mathcal{F}_{r-1}\right] \leq \exp\left(\frac{c^2\lambda^2(2+m\alpha)^2}{2}\right).$$

From (4.4.3), this implies that

$$\mathbb{E}\left[\exp\left(\sum_{i\in\mathbb{G}^*_{r-1}}\lambda\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}}+f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}}-mQf(X_i)\right)\right)\middle|\mathcal{F}_{r-1}\right]$$
$$\leq \exp\left(\frac{c^2\lambda^2(2+m\alpha)^2|\mathbb{G}^*_{r-1}|}{2}\right) \leq \exp\left(\frac{c^2\lambda^2(2+m\alpha)^2|\mathbb{G}_{r-1}|}{2}\right),$$

where we have used the fact that $|\mathbb{G}^*_{r-1}|\leq|\mathbb{G}_{r-1}|$ in the last inequality. Recalling (4.4.2), we are led to

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}^*_r}f(X_i)\right)\right]\leq\exp\left(\frac{c^2\lambda^2(2+m\alpha)^2|\mathbb{G}_{r-1}|}{2}\right) \times\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}^*_{r-1}}mQf(X_i)\right)\right].$$

Reproducing the same reasoning with Qf and $\mathbb{G}_{r-1}^*$ instead of f and $\mathbb{G}_r^*$, we get

$$\mathbb{E}\left[\exp\left(\lambda m \sum_{i\in\mathbb{G}_{r-1}^*} Qf(X_i)\right)\right] \leq \exp\left(\frac{c^2\lambda^2 m^2(2\alpha+m\alpha^2)^2|\mathbb{G}_{r-2}|}{2}\right) \times\mathbb{E}\left[\exp\left(\lambda m^2 \sum_{i\in\mathbb{G}_{r-2}^*} Q^2 f(X_i)\right)\right].$$

Iterating this procedure, we get

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{G}_r^*} f(X_i)\right)\right] &\leq \exp\left(\frac{c^2\lambda^2}{2}\sum_{q=0}^{r-1}\left(2\alpha^q+m\alpha^{q+1}\right)^2 m^{2q}2^{r-1-q}\right) \\ &\quad\times \mathbb{E}\Big[\exp\left(\lambda m^r Q^r f(X_1)\right)\Big] \\ &\leq \exp\left(\frac{c^2\lambda^2(2+m\alpha)^2 2^{r-1}}{2}\sum_{q=0}^{r-1}\left(\frac{\alpha^2 m^2}{2}\right)^q\right)\times\exp\left(\lambda c(\alpha m)^r\right),\end{aligned}$$

where the last inequality was obtained from **(H1)**. From the foregoing and from (4.4.1), we deduce that

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f)>\delta\right)\leq\begin{cases}\exp\left(-\lambda\delta m^r+\frac{c^2\lambda^2(2+m\alpha)^2\left(2^r-(\alpha^2m^2)^r\right)}{2(2-\alpha^2m^2)}\right) \\ \qquad\times\exp\left(\lambda c(\alpha m)^r\right)\quad\text{if } \alpha^2m^2\neq 2, \\ \exp\left(-\lambda\delta m^r+c^2\lambda^2(2+\sqrt{2})^2 r2^{r-2}\right)\exp\left(\lambda c(\sqrt{2})^r\right) \\ \qquad\text{if } \alpha^2m^2=2.\end{cases}$$

Now, the rest divides into four cases. In the sequel c_1 and c_2 will denote positive constants which depend on c, m, and α.

• If $m\alpha\leq 1$, then, for all $r\in\mathbb{N}$, $(m\alpha)^r<1$ and $2^r-(\alpha^2m^2)^r<2^r$. We then have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f)>\delta\right)\leq\exp\left(c\lambda\right)\exp\left(-\lambda\delta m^r+\lambda^2c_1 2^r\right).$$

Taking $\lambda=(\delta m^r)/(2^{r+1}c_1)$, we are led to

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f)>\delta\right)\leq\exp\left(c_1\delta\right)\exp\left(-\delta^2c_1\left(\frac{m^2}{2}\right)^r\right).$$

• If $1<m\alpha<\sqrt{2}$, then, since $2^r-(\alpha^2m^2)^r<2^r$, we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f)>\delta\right)\leq\exp\left(-\lambda\delta m^r+\lambda^2c_1 2^r\right)\exp\left(\lambda c(m\alpha)^r\right).$$

Taking $\lambda=(\delta m^r)/(2^{r+1}c_1)$, we are led to

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f)>\delta\right)\leq\exp\left(-c_2\delta(m^2/2)^r(\delta-2c\alpha^r)\right).$$

For all $r \in \mathbb{N}$ such that $r > \log(\delta/4c)/\log(\alpha)$, we have $\delta - 2c\alpha^r > \delta/2$ and it then follows that

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp\left(-c_2\delta^2(m^2/2)^r\right).$$

• If $m\alpha = \sqrt{2}$, then we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp\left(-\lambda\delta m^r + \lambda^2 c_1 r 2^{r-2})\right) \exp\left(\lambda c \left(\sqrt{2}\right)^r\right).$$

Taking $\lambda = (\delta m^r)/(c_1 r 2^{r-1})$, we have for all $r > \log(\delta/4c)/\log(\sqrt{2}/m)$,

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp\left(-c_2\delta^2(1/r)(m^2/2)^r\right).$$

• If $m\alpha > \sqrt{2}$, then we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp\left(-\lambda\delta m^r + \lambda^2 c_1(m^2\alpha^2)^r\right) \exp\left(\lambda c(m\alpha)^r\right).$$

Taking $\lambda = \delta/(2c_1(m\alpha^2)^r)$, we have for all $r > \log(\delta/4c)/\log\alpha$,

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) > \delta\right) \leq \exp\left(-c_3\delta^2\alpha^{-2r}\right).$$

This ends the proof when $\langle\mu, f\rangle = 0$ for $\mathbb{H}_r = \mathbb{G}_r$. Now, if $\langle\mu, f\rangle \neq 0$, set $g = f - \langle\mu, f\rangle$. Then, $\langle\mu, g\rangle = 0$ and $\widetilde{M}_{\mathbb{G}_r^*}(f) = \widetilde{M}_{\mathbb{G}_r^*}(g) + (|\mathbb{G}_r^*|/m^r)\langle\mu, f\rangle$. We have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) - \langle\mu, f\rangle W > \delta\right) \leq \mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(g) > \delta/2\right) + \mathbb{P}\left(\left|\frac{|\mathbb{G}_r^*|}{m^r} - W\right| > \frac{\delta}{2|\langle\mu, f\rangle|}\right). \tag{4.4.4}$$

As $\langle\mu, g\rangle = 0$, the previous computations give us a bound for the first term of right hand of (4.4.4), similar to (4.2.1). Now, under hypothesis **(H3)**, we deduce, from [4] Theorem 5, that

$$\mathbb{P}\left(\left|\frac{|\mathbb{G}_r^*|}{m^r} - W\right| > \frac{\delta}{2|\langle\mu, f\rangle|}\right) \leq c_2 \exp\left(-c_3\delta^{2/3}m^{r/3}\right),$$

and this ends the proof of (4.2.2) when $\mathbb{H}_r = \mathbb{G}_r$.

Step 2. Let us look at $\widetilde{M}_{\mathbb{T}_r^*}(f)$. First, we suppose that $\langle\mu, f\rangle = 0$. By Chernoff inequality, we have for all $\delta > 0$ and for all $\lambda > 0$

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(-\lambda\delta t_r\right) \mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{T}_r^*} f(X_i)\right)\right]. \tag{4.4.5}$$

Expectation which appears in the right hand of (4.4.5) can be written as

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_r^*} f(X_i)\right)\right] = \mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{r-2}^*} f(X_i)\right)\right. \tag{4.4.6}$$

$$\times\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-1}^*}(f+mQf)(X_i)\right)$$

$$\left.\times\,\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-1}^*}\big(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}}+f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}}-mQf(X_i)\big)\right)\Big|\mathcal{F}_{r-1}\right]\right].$$

Observing that $\mathbb{G}_{r-1}^*$ is $\mathcal{F}_{r-1}$ measurable, and using the fact that conditionally to $\mathcal{F}_{r-1}$, the triplets $\{(\Delta_i), i\in\mathbb{G}_{r-1}\}$ are independent and Azuma-Bennet-Hoeffding inequality, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-1}^*}\big(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}}+f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}}-mQf(X_i)\big)\right)\Big|\mathcal{F}_{r-1}\right]\\
&\quad=\prod_{i\mathbb{G}_{r-1}^*}\mathbb{E}\left[\exp\left(\lambda\left(f(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}}+f(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}}-mQf(X_i)\right)\right)\Big|\mathcal{F}_{r-1}\right]\\
&\qquad\le\exp\left(\frac{c^2\lambda^2(2+m\alpha)^2|\mathbb{G}_{r-1}^*|}{2}\right)\\
&\qquad\le\exp\left(\frac{c^2(2+m\alpha)^2|\mathbb{G}_{r-1}|}{2}\right),
\end{aligned}$$

where the last inequality was obtained using the fact that $|\mathbb{G}_{r-1}^*|\le|\mathbb{G}_{r-1}|$. From the foregoing and from (4.4.6), we deduce that

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_r^*}f(X_i)\right)\right]&\le\exp\left(\frac{c^2(2+m\alpha)^2|\mathbb{G}_{r-1}|}{2}\right)\\
&\quad\times\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{r-2}^*}f(X_i)\right)\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-1}^*}(f+mQf)(X_i)\right)\right]
\end{aligned}$$

Doing the same thing with $(f+mQf)$ and $\mathbb{G}_{r-1}^*$ instead of f and $\mathbb{G}_r^*$, we get

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{T}_{r-2}^*}f(X_i)\right)\exp\left(\lambda\sum_{i\in\mathbb{G}_{r-1}^*}(f+mQf)(X_i)\right)\right]$$

$$= \mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{T}^*_{r-3}} f(X_i)\right) \times \exp\left(\lambda \sum_{i\in\mathbb{G}^*_{r-2}} (f + mQf + m^2Q^2f)(X_i)\right)\right.$$

$$\times \mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{G}^*_{r-2}} \Big((f+mQf)(X_{2i})\mathbf{1}_{\{2i\in\mathbb{T}^*\}} + (f+mQf)(X_{2i+1})\mathbf{1}_{\{2i+1\in\mathbb{T}^*\}}\right.\right.$$

$$\left.\left.\left. - (mQf + m^2Q^2f)(X_i)\Big)\right)\middle|\mathcal{F}_{r-1}\right]\right]$$

$$\leq \exp\left(\frac{c^2\lambda^2(2+3m\alpha+m^2\alpha^2)^2|\mathbb{G}_{r-2}|}{2}\right)$$

$$\times \mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{T}^*_{r-3}} f(X_i)\right) \times \exp\left(\lambda \sum_{i\in\mathbb{G}^*_{r-2}} (f + mQf + m^2Q^2f)(X_i)\right)\right].$$

Iterating this procedure, we are led to

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i\in\mathbb{T}^*_r} f(X_i)\right)\right] \leq \exp\left(\frac{c^2(2+m\alpha)^2\lambda^2}{2}\sum_{q=1}^{r}\left(\sum_{k=0}^{q-1}(m\alpha)^k\right)^2 2^{r-q}\right)$$

$$\times \mathbb{E}\left[\exp\left(\lambda \sum_{q=0}^{r} m^qQ^qf(X_1)\right)\right]$$

$$\leq \exp\left(\frac{c^2(2+m\alpha)^2\lambda^2}{2}\sum_{q=1}^{r}\left(\sum_{k=0}^{q-1}(m\alpha)^k\right)^2 2^{r-q}\right)\exp\left(\lambda c\sum_{q=0}^{r}(m\alpha)^q\right),$$

where the last inequality was obtained using hypothesis **(H1)**. In the sequel, c_0, c_1 and c_2 will denote some positive constants which depend on α, m, and c. They may differ from one line to another. For $m\alpha \neq 1$ and $m\alpha \neq \sqrt{2}$, we deduce from the foregoing and from (4.4.5) that

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}^*_r}(f) > \delta\right) \leq \exp\left(-\lambda\delta t_r\right)\exp\left(\frac{\lambda c(1-(m\alpha)^{r+1})}{1-m\alpha}\right)$$

$$\times \exp\left(\frac{c^2(2+m\alpha)^2\lambda^2}{2(1-m\alpha)^2}\left((2^r-1) - \frac{2m\alpha(2^r-(m\alpha)^r)}{2-m\alpha}\right.\right.$$

$$\left.\left. + \frac{(m\alpha)^2(2^r-(m^2\alpha^2)^r)}{2-(m\alpha)^2}\right)\right)$$

$$\leq \exp\left(-\lambda\delta t_r + \frac{c^2(2+m\alpha)^2\lambda^2}{2(m\alpha-1)^2}\left((2^r-1) + \frac{(m\alpha)^2(2^r-(m^2\alpha^2)^r)}{2-(m\alpha)^2}\right)\right)$$

$$\times \exp\left(\frac{\lambda c(1-(m\alpha)^{r+1})}{1-m\alpha}\right).$$

Taking $\lambda = \dfrac{\delta t_r (m\alpha - 1)^2}{c^2(2+m\alpha)^2\left((2^r-1) + \frac{(m\alpha)^2(2^r-(m^2\alpha^2)^r)}{2-(m\alpha)^2}\right)}$, we are led to

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(-\frac{\delta^2(1-m\alpha)^2 t_r^2}{2c^2(2+m\alpha)^2\left(2^r - 1 + \frac{(m\alpha)^2(2^r-(m^2\alpha^2)^r)}{2-(m\alpha)2}\right)}\right)$$
$$\times \exp\left(\frac{\delta(1-m\alpha)^2 t_r}{c(2+m\alpha)^2\left(2^r - 1 + \frac{(m\alpha)^2(2^r-(m^2\alpha^2)^r)}{2-(m\alpha)^2}\right)} \times \frac{1-(m\alpha)^{r+1}}{1-m\alpha}\right).$$

Now, the rest of the proof divides into five cases.

• If $m\alpha < 1$, then, for all $r \in \mathbb{N}$, $(m\alpha)^{r+1} - 1 \leq m\alpha - 1$ and $2^r - (m\alpha)^{2r} < 2^r$. We then deduce that

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp(c_2\delta)\exp\left(-c_2\delta^2(m^2/2)^{r+1}\right).$$

• If $1 < m\alpha < \sqrt{2}$, then have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(-c_1\delta^2(m^2/2)^{r+1}\right)\exp\left(c_2\delta\frac{(m\alpha)^{r+1}-1}{m\alpha - 1}\right)$$
$$\leq \exp\left(-\delta c_2(m^2/2)^{r+1}(\delta - c_0\alpha^{r+1})\right).$$

Now, for all $r \in \mathbb{N}$ such that $r+1 > \log(\delta/2c_0)/\log(\alpha)$, we have $\delta - c_0\alpha^{r+1} > \delta/2$, in such a way that

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(\delta^2 c_2(m^2/2)^{r+1}\right).$$

• If $m\alpha > \sqrt{2}$, then for all $r \in \mathbb{N}$, $(m^2\alpha^2)^r > 2^r$. We then have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(-c_2\delta\alpha^{-2r}(\delta - c_0\alpha^{r+1})\right).$$

Now for all $r \in \mathbb{N}$ such that $r+1 > \log(\delta/c_0)/\log(\alpha)$, we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(-\frac{c_2\delta^2}{\alpha^{2r}}\right).$$

• If $m\alpha = 1$, then

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(-\lambda\delta t_r + c_1 2^r\lambda^2\right)\exp\left(\lambda c(r+1)\right)$$

Taking $\lambda = \delta t_r / c_1 2^{r+1}$, we are led to

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \exp\left(c_1\delta\frac{(r+1)t_r}{2^{r+1}}\right)\exp\left(-c_2\delta^2(m^2/2)^{r+1}\right).$$

• If $m\alpha = \sqrt{2}$, then

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \le \exp\left(-\lambda\delta t_r + \lambda^2 c_1 (r+1)2^r\right) \exp\left(\lambda c_1 (\sqrt{2})^{r+1}\right).$$

Taking $\lambda = \delta t_r/(2c_1(r+1)2^r)$, we are led to

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \le \exp\left(-\frac{c_2\delta}{r+1}\left(\frac{m^2}{2}\right)^{r+1}\left(\delta - c_0\left(\frac{\sqrt{2}}{m}\right)^{r+1}\right)\right).$$

Now, for all $r \in \mathbb{N}$ such that $r+1 > \log(\delta/c_0)/\log(\sqrt{2}/m)$, we get

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \le \exp\left(-\frac{c_2\delta^2}{r+1}\left(\frac{m^2}{2}\right)^{r+1}\right).$$

This ends the proof when $\langle \mu, f\rangle = 0$ for $\mathbb{H}_r = \mathbb{T}_r$. Now, for $f \in \mathcal{B}_b(S)$, set $g = f - \langle \mu, f\rangle$. Then, $\langle \mu, g\rangle = 0$ and $\widetilde{M}_{\mathbb{T}_r^*}(f) = \widetilde{M}_{\mathbb{T}_r^*}(g) + (|\mathbb{T}_r^*|/t_r)\langle \mu, f\rangle$. We have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) - \langle \mu, f\rangle W > \delta\right) \le \mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(g) > \delta/2\right) + \mathbb{P}\left(\left|\frac{|\mathbb{T}_r^*|}{t_r} - W\right| > \frac{\delta}{2|\langle \mu, f\rangle|}\right). \tag{4.4.7}$$

Since $\langle \mu, g\rangle = 0$, the first term of the right hand of (4.4.7) can be bounded as in the previous computations. Under additional hypothesis **(H3)**, we have, from [4] Theorem 5,

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{|\mathbb{T}_r^*|}{t_r} - W\right| > \frac{\delta}{2|\langle \mu, f\rangle|}\right) &\le \sum_{q=0}^{r} \mathbb{P}\left(\frac{m^q}{t_r}\left|\frac{|\mathbb{G}_q^*|}{m^q} - W\right| > \frac{\delta}{2(r+1)|\langle \mu, f\rangle|}\right)\\
&= \sum_{q=0}^{r} \mathbb{P}\left(\left|\frac{|\mathbb{G}_q^*|}{m^q} - W\right| > \frac{\delta t_r}{2(r+1)|\langle \mu, f\rangle| m^q}\right)\\
&\le \sum_{q=0}^{r} c_2 \exp\left(-c_3\delta^{2/3}\left(\frac{t_r^2}{(r+1)m^q}\right)^{1/3}\right)\\
&\le c_2 \exp\left(-c_3\delta^{2/3}\left(\frac{t_r}{(r+1)^2}\right)^{1/3}\right)\left(1 + o(1)\right),
\end{aligned}$$

and this ends the proof of (4.2.2) when $\mathbb{H}_r = \mathbb{T}_r$.

4.4.2 Proof of Theorem 4.2.3

Let $f \in \mathcal{B}_b(S)$. Without loss of generality, we assume that $\langle \mu, f\rangle = 0$. For all $\delta > 0$, for all $a > 0$ and for all $b > 0$ such that $b < a/(\delta+1)$, we have

$$\mathbb{P}\left(\overline{M}_{\mathbb{H}_r^*}(f) > \delta | W \ge a\right) = \mathbb{P}\left(\overline{M}_{\mathbb{H}_r^*}(f) > \delta, \frac{|\mathbb{H}_r^*|}{h_r} > b | W \ge a\right)$$

$$
\begin{aligned}
&+\mathbb{P}\left(\overline{M}_{\mathbb{H}_r^*}(f)>\delta, \frac{|\mathbb{H}_r^*|}{h_r}\leq b|W\geq a\right)\\
&=\frac{1}{\mathbb{P}\left(W\geq a\right)}\left(\mathbb{P}\left(\overline{M}_{\mathbb{H}_r^*}(f)>\delta, \frac{|\mathbb{H}_r^*|}{h_r}>b, W\geq a\right)\right.\\
&\left.+\mathbb{P}\left(\overline{M}_{\mathbb{H}_r^*}(f)>\delta, \frac{|\mathbb{H}_r^*|}{h_r}\leq b, W\geq a\right)\right)\\
&\leq p_a\mathbb{P}\left(\widetilde{M}_{\mathbb{H}_r^*}(f)>\delta b\right)+p_a\mathbb{P}\left(\left|\frac{|H_r^*|}{h_r}-W\right|>W-b, W\geq a\right)\\
&\leq p_a\mathbb{P}\left(\widetilde{M}_{\mathbb{H}_r^*}(f)>\delta b\right)+p_a\mathbb{P}\left(\left|\frac{|H_r^*|}{h_r}-W\right|>\delta b\right),
\end{aligned}
$$

where $p_a = \mathbb{P}\left(W \geq a\right)$. Now, the first term of the last inequality can be bounded as in Theorem 4.2.2, and the second term is bounded as in the **step 1** and and **step 2** of the proof of Theorem 4.2.2. This ends the proof.

4.4.3 Proof of Theorem 4.2.4

Let $f \in \mathcal{B}_b(S^3)$.

Step 1. Let us first deal with $\widetilde{M}_{\mathbb{G}_r^*}(f)$. Assume that $\langle \mu, P^* f\rangle = 0$. By Chernoff inequality, we have for all $\delta > 0$ and for all $\lambda > 0$,

$$
\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f)>\delta\right)\leq \exp\left(-\lambda\delta m^r\right)\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_r^*}f(\Delta_i)\right)\right].
$$

Conditioning by $\mathcal{F}_r$, and using, conditional independence of triplets $\{\Delta_i, i \in \mathbb{G}_r\}$ with respect to $\mathcal{F}_r$, Azuma-Bennet-Hoeffding inequality and **(H2)**, we obtain

$$
\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_r^*}f(\Delta_i)\right)\right]\\
&\quad=\mathbb{E}\left[\exp\left(\lambda\sum_{i\mathbb{G}_r^*}P^*f(X_i)\right)\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_r^*}(f(\Delta_i)-P^*f(X_i))\right)\Big|\mathcal{F}_r\right]\right]\\
&\quad=\mathbb{E}\left[\exp\left(\lambda\sum_{i\mathbb{G}_r^*}P^*f(X_i)\right)\prod_{i\in\mathbb{G}_r^*}\mathbb{E}\left[\exp\left(\lambda(f(\Delta_i)-P^*f(X_i))\right)\Big|\mathcal{F}_r\right]\right]\\
&\quad\leq\exp\left(2\lambda^2\|f\|_\infty c_1 m^r\right)\mathbb{E}\left[\exp\left(\lambda\sum_{i\in\mathbb{G}_r^*}P^*f(X_i)\right)\right].
\end{aligned}
$$

We control the last expectation as in the **Step 1** of the proof of Theorem 4.2.2, apply to P^*f. Next, we get the result discussing as in the proof of Theorem 4.2.2.

If $\langle \mu, P^*f \rangle \neq 0$, we set $g = f - \langle \mu, P^*f \rangle$. Then, we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(f) - \langle \mu, P^*f \rangle W > \delta\right) \leq \mathbb{P}\left(\widetilde{M}_{\mathbb{G}_r^*}(g) > \delta/2\right) + \mathbb{P}\left(\left|\frac{\mathbb{G}_r^*}{m^r} - W\right| > \delta/2|\langle \mu, P^*f \rangle|\right). \tag{4.4.8}$$

The first term of the right hand of (4.4.8) can be bounded as previously since $\langle \mu, P^*g \rangle = 0$. The second term can be bounded as in **Step 1** of the proof of Theorem 4.2.2. This ends the proof for $\widetilde{M}_{\mathbb{G}_r^*}(f)$.

Step 2. Let us now treat $\widetilde{M}_{\mathbb{T}_r^*}(f)$. First, we assume that $\langle \mu, P^*f \rangle = 0$. For all $\delta > 0$, we have

$$\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(f) > \delta\right) \leq \mathbb{P}\left(\frac{1}{t_r}\sum_{i \in \mathbb{T}_r^*}(f(\Delta_i) - P^*f(X_i)) > \delta/2\right) + \mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(P^*f) > \delta/2\right).$$

By chernoff inequality, we have for all $\lambda > 0$,

$$\mathbb{P}\left(\frac{1}{t_r}\sum_{i \in \mathbb{T}_r^*}(f(\Delta_i) - P^*f(X_i)) > \delta/2\right) \leq \exp\left(-\frac{\lambda \delta t_r}{2}\right) \times \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathbb{T}_r^*}(f(\Delta_i) - P^*f(X_i))\right)\right]$$

Conditioning successively with respect to $(\mathcal{F}_q)_{0 \leq q \leq r}$, using conditional independence of triplets $\{\Delta_i, i \in \mathbb{G}_q\}$ with respect to $\mathcal{F}_q$ and applying successively Azuma-Bennet-Hoeffding inequality and the fact that $|\mathbb{G}_q^*| \leq |\mathbb{G}_q|$ for all $q \in \{0, \cdots, r\}$, we get

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathbb{T}_r^*}(f(\Delta_i) - P^*f(X_i))\right)\right] \\
&= \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathbb{T}_{r-1}^*}(f(\Delta_i) - P^*f(X_i))\right) \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathbb{G}_r^*}(f(\Delta_i) - P^*f(X_i))\right) \Big| \mathcal{F}_r\right]\right] \\
&= \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathbb{T}_{r-1}^*}(f(\Delta_i) - P^*f(X_i))\right) \prod_{i \in \mathbb{G}_r^*} \mathbb{E}\left[\exp\left(\lambda(f(\Delta_i) - P^*f(X_i))\right) \Big| \mathcal{F}_r\right]\right] \\
&\leq \exp\left(2\lambda^2 \|f\|_\infty^2 |\mathbb{G}_r|\right) \mathbb{E}\left[\exp\left(\lambda \sum_{i \in \mathbb{T}_{r-1}^*}(f(\Delta_i) - P^*f(X_i))\right)\right]
\end{aligned}$$

$$\vdots$$
$$\leq \exp\left(2\lambda^2\|f\|_\infty^2|\mathbb{T}_r|\right).$$

Next, optimizing on λ, we obtain

$$\mathbb{P}\left(\frac{1}{t_r}\sum_{i\in\mathbb{T}_r^*}(f(\Delta_i)-P^*f(X_i))>\delta/2\right)\leq \exp\left(-c_1\delta^2\left(\frac{m^2}{2}\right)^{r+1}\right),$$

for some positive constant c_1. The term $\mathbb{P}\left(\widetilde{M}_{\mathbb{T}_r^*}(P^*f)>\delta/2\right)$ can be bounded as in the proof of Theorem 4.2.2, and this ends the proof when $\langle\mu,P^*f\rangle=0$. On the other hand, if $\langle\mu,P^*f\rangle\neq 0$, we have

$$\widetilde{M}_{\mathbb{T}_r^*}(f)-\langle\mu,P^*f\rangle W=\widetilde{M}_{\mathbb{T}_r^*}(g)+\left(\frac{|\mathbb{T}_r^*|}{t_r}-W\right)\langle\mu,P^*f\rangle.$$

We then proceed as for (4.4.8), and this ends the proof for $\widetilde{M}_{\mathbb{T}_r^*}(f)$.

Step 3. Eventually, we bound $\mathbb{P}\left(\overline{M}_{\mathbb{H}_r^*}(f)>\delta-\langle\mu,P^*f\rangle>\delta\right)$, using **Step 1** and **Step 2**, as in the proof of Theorem 4.2.3.

4.4.4 Proof of Proposition 4.3.1

We are going to treat $\widehat{\alpha}_0^n-\alpha_0$. Deviation inequalities for $\widehat{\alpha}_1^n-\alpha_1$, $\widehat{\beta}_\eta^n-\beta_\eta$, $\widehat{\alpha}_\eta^{\prime n}-\alpha_\eta^{'}$, $\widehat{\beta}_\eta^{\prime n}-\beta_\eta^{'}$, $\eta\in\{0,1\}$, can be treated in the same way. Recalling that the state space of the process $(X_i, i\in\mathbb{T}^*)$, denoted by S, is assumed to be a compact subset of $\mathbb{R}$.

Let g_1, g_2, h_1 and h_2 the functions defined on S^3 respectively by

$$g_1(x,y,z)=(xy-x(\alpha_0x+\beta_0))\,\mathbf{1}_{S^3}(x,y,z),$$
$$g_2(x,y,z)=(y-\alpha_0x-\beta_0))\,\mathbf{1}_{S^3}(x,y,z),$$
$$h_1(x,y,z)=x\mathbf{1}_{S^3}(x,y,z),$$
$$h_2(x,y,z)=x^2\mathbf{1}_{S^3}(x,y,z).$$

It is easy to see that $P^*g_1(x)=0$, $P^*g_2(x)=0$, $P^*h_1(x)=p_{1,0}x$ and $P^*h_2(x)=p_{1,0}x^2$ where P^* denote the transition kernel associated to the BAR(1) process with missing data. With this notation, we can rewrite $\widehat{\alpha}_0^n-\alpha_0$ as

$$\widehat{\alpha}_0^n-\alpha_0=\frac{|\mathbb{T}_n^*|^{-1}|\mathbb{T}_n^{1,0}|\left(|\mathbb{T}_n^*|^{-1}\sum_{i\in\mathbb{T}_n^*}g_1(\Delta_i)\right)}{B_n}-\frac{\left(|\mathbb{T}_n^*|^{-1}\sum_{i\in\mathbb{T}_n^*}h_1(\Delta_i)\right)\left(|\mathbb{T}_n^*|^{-1}\sum_{i\in\mathbb{T}_n^*}g_2(\Delta_i)\right)}{B_n},$$

where $B_n = |\mathbb{T}_n^*|^{-1}|\mathbb{T}_n^{1,0}|\left(|\mathbb{T}_n^*|^{-1}\sum_{i\in\mathbb{T}_n^*} h_2(\Delta_i)\right) - \left(|\mathbb{T}_n^*|^{-1}\sum_{i\in\mathbb{T}_n^*} h_1(\Delta_i)\right)^2.$

Recalling (4.1.8), we then have for all $\delta > 0$ and $a > 0$

$$\mathbb{P}\left(|\widehat{\alpha}_0^n - \alpha_0| > \delta | W \geq a\right) \leq \mathbb{P}\left(\frac{|\mathbb{T}_n^*|^{-1}|\mathbb{T}_n^{1,0}||\overline{M}_{\mathbb{T}_n^*}(g_1)|}{|B_n|} > \frac{\delta}{2}\Big| W \geq a\right) + \mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_n^*}(h_1)||\overline{M}_{\mathbb{T}_n^*}(g_2)|}{|B_n|} > \frac{\delta}{2}\Big| W \geq a\right). \tag{4.4.9}$$

For the first term of the right hand of (4.4.9), since $|\mathbb{T}_n^*|^{-1}|\mathbb{T}_n^{1,0}| \leq 1$, we have for all $\gamma > 0$

$$\mathbb{P}\left(\frac{|\mathbb{T}_n^*|^{-1}|\mathbb{T}_n^{1,0}||\overline{M}_{\mathbb{T}_n^*}(g_1)|}{|B_n|} > \frac{\delta}{2}\Big| W \geq a\right) \leq \mathbb{P}\left(|B_n| < \gamma | W \geq a\right) + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_n^*}(g_1)| > \frac{\delta\gamma}{2}\Big| W \geq a\right).$$

Notice that

$$B_n - (p_{1,0}^2\mu_2 - p_{1,0}^2\mu_1^2) = p_{1,0}\mu_2\left(\frac{|\mathbb{T}_n^{1,0}|}{|\mathbb{T}_n^*|} - p_{1,0}\right) + \frac{|\mathbb{T}_n^{1,0}|}{|\mathbb{T}_n^*|}\overline{M}_{\mathbb{T}_n^*}(h_2 - p_{1,0}\mu_2) - \left(\overline{M}_{\mathbb{T}_n^*}(h_1 - p_{1,0}\mu_1)\right)^2 - 2p_{1,0}\mu_1\overline{M}_{\mathbb{T}_n^*}(h_1 - p_{1,0}\mu_1)$$

and

$$\{|B_n| < \gamma\} \subset \left\{|B_n - (p_{1,0}^2\mu_2 - p_{1,0}^2\mu_1^2)| > |p_{1,0}^2\mu_2 - p_{1,0}^2\mu_1^2| - \gamma\right\}.$$

We then have for all $0 < \gamma < \dfrac{2|p_{1,0}^2\mu_2 - p_{1,0}^2\mu_1^2|}{2+\delta}$,

$$\begin{aligned}
&\mathbb{P}\left(\frac{|\mathbb{T}_n^*|^{-1}|\mathbb{T}_n^{1,0}||\overline{M}_{\mathbb{T}_n^*}(g_1)|}{|B_n|} > \frac{\delta}{2}\Big| W \geq a\right) \\
&\leq \mathbb{P}\left(|B_n| < \gamma | W \geq a\right) + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_n^*}(g_1)| > \frac{\delta\gamma}{2}\Big| W \geq a\right) \\
&\leq \mathbb{P}\left(|B_n - (p_{1,0}^2\mu_2 - p_{1,0}^2\mu_1^2)| > \frac{\gamma\delta}{2}\Big| W \geq a\right) + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_n^*}(g_1)| > \frac{\delta\gamma}{2}\Big| W \geq a\right) \\
&\leq \mathbb{P}\left(|\overline{M}_{\mathbb{T}_n^*}(g_1)| > \frac{\delta\gamma}{2}\Big| W \geq a\right) + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_n^*}(h_2 - p_{1,0}\mu_2)| > \frac{\gamma\delta}{8}\Big| W \geq a\right) \\
&+ \mathbb{P}\left(|\overline{M}_{\mathbb{T}_n^*}(h_1 - p_{1,0}\mu_1)| > \frac{\sqrt{\gamma\delta}}{2\sqrt{2}}\Big| W \geq a\right) + \mathbb{P}\left(\left|\frac{|\mathbb{T}_n^{1,0}|}{|\mathbb{T}_n^*|} - p_{1,0}\right| > \frac{\delta\gamma}{8p_{1,0}\mu_2}\Big| W \geq a\right) \\
&+ \mathbb{P}\left(|\overline{M}_{\mathbb{T}_n^*}(h_1 - p_{1,0}\mu_1)| > \frac{\gamma\delta}{16p_{1,0}\mu_1}\Big| W \geq a\right).
\end{aligned}$$

For the second term of the right hand of (4.4.9), we have

$$\mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_n^*}(h_1)||\overline{M}_{\mathbb{T}_n^*}(g_1)|}{|B_n|} > \frac{\delta}{2}\Big|W \geq a\right) \leq \mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_n^*}(g_2)|}{|B_n|} > \frac{\delta}{4p_{1,0}\mu_1}\Big|W \geq a\right)$$
$$+\,\mathbb{P}\left(\frac{|\overline{M}_{\mathbb{T}_n^*}(g_2)|}{|B_n|} > \frac{\sqrt{\delta}}{2}\Big|W \geq a\right) + \mathbb{P}\left(|\overline{M}_{\mathbb{T}_n^*}(h_1 - p_{1,0}\mu_1)| > \frac{\sqrt{\delta}}{2}\Big|W \geq a\right)$$

Now, the first and the second term of the right hand of the last inequality can be treated as the first term of the right hand of (4.4.9). Finally, to get the result, just apply Theorem 4.2.4 to functions g_1, g_2, h_1 and h_2.

Chapitre 5

Moderate deviations for the Durbin-Watson statistic related to the first-order autoregressive process

5.1 Introduction

This chapter is focused on the stable first-order autoregressive process where the driven noise is also given by a first-order autoregressive process. The purpose is to investigate moderate deviations for both the least squares estimator of the unknown parameter of the autoregressive process as well as for the serial correlation estimator associated with the driven noise. Our goal is to establish moderate deviations for the Durbin-Watson statistic [50], [51], [52], in a lagged dependent random variables framework. First of all, we shall assume that the driven noise is normally distributed. Then, we will extend our investigation to the more general framework where the driven noise satisfies a less restrictive Chen-Ledoux type condition [27], [73]. We are inspired by the recent paper of Bercu and Proïa [19], where the almost sure convergence and the central limit theorem are established for both the least squares estimators and the Durbin-Watson statistic. Our results are proved via an extensive use of the results of Dembo [34], Dembo and Zeitouni [35] and Worms [113], [114], [115] on the one hand, and of the paper of Puhalskii [91] and Djellout [38] on the other hand, about moderate deviations for martingales. In order to introduce the Durbin-Watson statistic, we shall focus our attention on the first-order autoregressive process given, for all $n \geq 1$, by

$$\begin{cases} X_n &= \theta X_{n-1} + \varepsilon_n \\ \varepsilon_n &= \rho \varepsilon_{n-1} + V_n \end{cases} \tag{5.1.1}$$

where we shall assume that the unknown parameters $|\theta| < 1$ and $|\rho| < 1$ to ensure the stability of the model. In all the sequel, we also assume that (V_n) is a sequence of independent and identically distributed random variables with zero mean, positive variance σ^2 and satisfying some suitable assumptions. The square-integrable initial values X_0 and ε_0 may be arbitrarily chosen. We have decided to estimate θ by the least squares estimator

$$\widehat{\theta}_n = \frac{\sum_{k=1}^n X_k X_{k-1}}{\sum_{k=1}^n X_{k-1}^2}. \tag{5.1.2}$$

Then, we also define a set of least squares residuals given, for all $1 \leq k \leq n$, by

$$\widehat{\varepsilon}_k = X_k - \widehat{\theta}_n X_{k-1}, \tag{5.1.3}$$

which leads to the estimator of ρ,

$$\widehat{\rho}_n = \frac{\sum_{k=1}^n \widehat{\varepsilon}_k \widehat{\varepsilon}_{k-1}}{\sum_{k=1}^n \widehat{\varepsilon}_{k-1}^{\,2}}. \tag{5.1.4}$$

Finally, the Durbin-Watson statistic is defined, for $n \geq 1$, as

$$\widehat{D}_n = \frac{\sum_{k=1}^n (\widehat{\varepsilon}_k - \widehat{\varepsilon}_{k-1})^2}{\sum_{k=0}^n \widehat{\varepsilon}_k^{\,2}}. \tag{5.1.5}$$

This well-known statistic was introduced by the pioneer work of Durbin and Watson [50], [51], [52], in the middle of last century, to test the presence of a significative first order serial correlation in the residuals of a regression analysis. A wide range of literature is available on the asymptotic behavior of the Durbin-Watson statistic, frequently used in Econometry. While it appeared to work pretty well in the classical independent framework, Malinvaud [76] and Nerlove and Wallis [86] observed that, for linear regression models containing lagged dependent random variables, the Durbin-Watson statistic may be asymptotically biased, potentially leading to inadequate conclusions. Durbin [49] proposed alternative tests to prevent this misuse, such as the *h-test* and the *t-test*, then substantial contributions were brought by Inder [71], King and Wu [72] and more recently Stocker [101]. Lately, a set of results have been established by Bercu and Proïa in [19], in particular a test procedure as powerful as the *h-test*, and they will be summarized thereafter as a basis for this chapter.

The chapter is organized as follows. First of all, we recall the results recently established by Bercu and Proïa [19]. In Section 5.2, we propose moderate deviation principles for the estimators of θ and ρ and for the Durbin-Watson statistic, given by (5.1.2), (5.1.4) and (5.1.5), under the normality assumption on the driven noise. Section 5.3 deals with the generalization of the latter results under a less restrictive Chen-Ledoux type condition on (V_n). Finally, all technical proofs are postponed to Section 5.4.

Lemma 5.1.1. *We have the almost sure convergence of the autoregressive estimator,*

$$\lim_{n\to\infty} \widehat{\theta}_n = \theta^* \quad \text{a.s.}$$

where the limiting value

$$\theta^* = \frac{\theta+\rho}{1+\theta\rho}. \tag{5.1.6}$$

In addition, as soon as $\mathbb{E}[V_1^4] < \infty$*, we also have the asymptotic normality,*

$$\sqrt{n}\left(\widehat{\theta}_n - \theta^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \sigma_\theta^2\big)$$

where the asymptotic variance

$$\sigma_\theta^2 = \frac{(1-\theta^2)(1-\theta\rho)(1-\rho^2)}{(1+\theta\rho)^3}. \tag{5.1.7}$$

Lemma 5.1.2. *We have the almost sure convergence of the serial correlation estimator,*

$$\lim_{n\to\infty} \widehat{\rho}_n = \rho^* \quad \text{a.s.}$$

where the limiting value

$$\rho^* = \theta\rho\theta^*. \tag{5.1.8}$$

Moreover, as soon as $\mathbb{E}[V_1^4] < \infty$*, we have the asymptotic normality,*

$$\sqrt{n}\Big(\widehat{\rho}_n - \rho^*\Big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \sigma_\rho^2\big)$$

with the asymptotic variance

$$\sigma_\rho^2 = \frac{(1-\theta\rho)}{(1+\theta\rho)^3}\big((\theta+\rho)^2(1+\theta\rho)^2 + (\theta\rho)^2(1-\theta^2)(1-\rho^2)\big). \tag{5.1.9}$$

On top of that, we have the joint asymptotic normality,

$$\sqrt{n}\begin{pmatrix} \widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^* \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, \Gamma\big)$$

where the covariance matrix

$$\Gamma = \begin{pmatrix} \sigma_\theta^2 & \theta\rho\sigma_\theta^2 \\ \theta\rho\sigma_\theta^2 & \sigma_\rho^2 \end{pmatrix}. \tag{5.1.10}$$

Lemma 5.1.3. *We have the almost sure convergence of the Durbin-Watson statistic,*

$$\lim_{n\to\infty} \widehat{D}_n = D^* \quad \text{a.s.}$$

where the limiting value

$$D^* = 2(1-\rho^*). \tag{5.1.11}$$

In addition, as soon as $\mathbb{E}[V_1^4] < \infty$, *we have the asymptotic normality,*

$$\sqrt{n}\left(\widehat{D}_n - D^*\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_D^2)$$

where the asymptotic variance

$$\sigma_D^2 = 4\sigma_\rho^2. \tag{5.1.12}$$

Proof. The proofs of Lemma 5.1.1, Lemma 5.1.2 and Lemma 5.1.3 may be found in [19]. □

Our objective is to establish a set of moderate deviation principles on these estimates in order to get a better asymptotic precision than the central limit theorem. In all the sequel, (b_n) will denote a sequence of increasing positive numbers satisfying $1 = o(b_n^2)$ and $b_n^2 = o(n)$, that is

$$b_n \longrightarrow \infty, \qquad \frac{b_n}{\sqrt{n}} \longrightarrow 0. \tag{5.1.13}$$

Remarks and Notation. *In the whole chapter, for any matrix* M, M' *and* $\|M\|$ *stand for the transpose and the euclidean norm of* M, *respectively. For any square matrix* M, $\det(M)$ *and* $\rho(M)$ *are the determinant and the spectral radius of* M, *respectively. Moreover, we will shorten* large deviation principle *by LDP. In addition, for a sequence of random variables* $(Z_n)_n$ *on* $\mathbb{R}^{d\times p}$, *we say that* $(Z_n)_n$ *converges* $(b_n^2)-$*superexponentially fast in probability to some random variable* Z *if, for all* $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\Big(\|Z_n - Z\| > \delta\Big) = -\infty.$$

This exponential convergence *with speed* b_n^2 *will be shortened as*

$$Z_n \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} Z.$$

The exponential equivalence *with speed* b_n^2 *between two sequences of random variables* $(Y_n)_n$ *and* $(Z_n)_n$, *whose precise definition is given in Definition 4.2.10 of [35], will be shortened as*

$$Y_n \overset{\text{superexp}}{\underset{b_n^2}{\sim}} Z_n.$$

5.2 On moderate deviations under the Gaussian condition

In this first part, we focus our attention on moderate deviations for the Durbin-Watson statistic in the easy case where the driven noise (V_n) is normally distributed. This restrictive assumption allows us to reduce the set of hypothesis to the existence of $t > 0$ such that

(G.1)

$$\mathbb{E}\Big[\exp(t\varepsilon_0^2)\Big] < \infty,$$

(G.2)

$$\mathbb{E}\Big[\exp(tX_0^2)\Big] < \infty.$$

Theorem 5.2.1. *Assume that there exists $t > 0$ such that* **(G.1)** *and* **(G.2)** *are satisfied. Then, the sequence*

$$\left(\frac{\sqrt{n}}{b_n}\left(\widehat{\theta}_n - \theta^*\right)\right)_{n\geq 1}$$

satisfies an LDP on $\mathbb{R}$ with speed b_n^2 and good rate function

$$I_\theta(x) = \frac{x^2}{2\sigma_\theta^2} \tag{5.2.1}$$

where σ_θ^2 is given by (5.1.7).

Theorem 5.2.2. *Assume that there exists $t > 0$ such that* **(G.1)** *and* **(G.2)** *are satisfied. Then, as soon as $\theta \neq -\rho$, the sequence*

$$\left(\frac{\sqrt{n}}{b_n}\begin{pmatrix}\widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^*\end{pmatrix}\right)_{n\geq 1}$$

satisfies an LDP on $\mathbb{R}^2$ with speed b_n^2 and good rate function

$$K(x) = \frac{1}{2}x'\Gamma^{-1}x \tag{5.2.2}$$

where Γ is given by (5.1.10). In particular, the sequence

$$\left(\frac{\sqrt{n}}{b_n}\left(\widehat{\rho}_n - \rho^*\right)\right)_{n\geq 1}$$

satisfies an LDP on $\mathbb{R}$ with speed b_n^2 and good rate function

$$I_\rho(x) = \frac{x^2}{2\sigma_\rho^2} \tag{5.2.3}$$

where σ_ρ^2 is given by (5.1.9).

Remark 5.2.1. *The covariance matrix Γ is invertible if and only if $\theta \neq -\rho$ since one can see by a straightforward calculation that*

$$\det(\Gamma) = \frac{\sigma_\theta^2(\theta+\rho)^2(1-\theta\rho)}{(1+\rho^2)}.$$

Moreover, in the particular case where $\theta = -\rho$, the sequences

$$\left(\frac{\sqrt{n}}{b_n}\left(\widehat{\theta}_n - \theta^*\right)\right)_{n\geq 1} \qquad \text{and} \qquad \left(\frac{\sqrt{n}}{b_n}\left(\widehat{\rho}_n - \rho^*\right)\right)_{n\geq 1}$$

satisfy LDP on $\mathbb{R}$ with speed b_n^2 and good rate functions respectively given by

$$I_\theta(x) = \frac{x^2(1-\theta^2)}{2(1+\theta^2)} \qquad \text{and} \qquad I_\rho(x) = \frac{x^2(1-\theta^2)}{2\theta^4(1+\theta^2)}.$$

Theorem 5.2.3. *Assume that there exists $t > 0$ such that* **(G.1)** *and* **(G.2)** *are satisfied. Then, the sequence*

$$\left(\frac{\sqrt{n}}{b_n}\left(\widehat{D}_n - D^*\right)\right)_{n\geq 1}$$

satisfies an LDP on $\mathbb{R}$ with speed b_n^2 and good rate function

$$I_D(x) = \frac{x^2}{2\sigma_D^2} \tag{5.2.4}$$

where σ_D^2 is given by (5.1.12).

Proof. Theorem 5.2.1, Theorem 5.2.2 and Theorem 5.2.3 are proved in Section 4. □

5.3 On moderate deviations under the Chen-Ledoux type condition

Via an extensive use of Puhalskii's result, we will now focus our attention on the more general framework where the driven noise (V_n) is assumed to satisfy the Chen-Ledoux type condition. Accordingly, one shall introduce the following hypothesis, for $a = 2$ and $a = 4$.

(CL.1) Chen-Ledoux.

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log n\mathbb{P}\Big(|V_1|^a > b_n\sqrt{n}\Big) = -\infty.$$

(CL.2)

$$\frac{|\varepsilon_0|^a}{b_n\sqrt{n}} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0.$$

(CL.3)

$$\frac{|X_0|^a}{b_n\sqrt{n}} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0.$$

Remark 5.3.1. *If the random variable V_1 satisfies* **(CL.1)** *with $a = 2$, then*

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log n\mathbb{P}\Big(\big|V_1^2 - \mathbb{E}[V_1^2]\big| > b_n\sqrt{n}\Big) = -\infty, \tag{5.3.1}$$

which implies in particular that $\mathrm{Var}(V_1^4) < \infty$. Moreover, if the random variable V_1 has exponential moments, i.e. if there exists $t > 0$ such that

$$\mathbb{E}\Big[\exp\left(tV_1^2\right)\Big] < \infty,$$

then **(CL.1)** *is satisfied for every increasing sequence (b_n). From [3], [53], condition (5.3.1) is equivalent to say that the sequence*

$$\left(\frac{1}{b_n\sqrt{n}}\sum_{k=1}^{n}\Big(V_k^2 - \mathbb{E}[V_k^2]\Big)\right)_{n\geq 1}$$

satisfies an LDP on $\mathbb{R}$ with speed b_n^2 and good rate function

$$I(x) = \frac{x^2}{2\mathrm{Var}(V_1^2)}.$$

Remark 5.3.2. *If we choose $b_n = n^\alpha$ with $0 < \alpha < 1/2$,* **(CL.1)** *is immediately satisfied if there exists $t > 0$ and $0 < \beta < 1$ such that*

$$\mathbb{E}\Big[\exp\left(tV_1^{2\beta}\right)\Big] < \infty,$$

which is clearly a weaker assumption than the existence of $t > 0$ such that

$$\mathbb{E}\Big[\exp\left(tV_1^2\right)\Big] < \infty,$$

imposed in the previous section.

Remark 5.3.3. *If* **(CL.1)** *is satisfied for $a = 4$, then it is also satisfied for all $0 < b < a$.*

Remark 5.3.4. *In the technical proofs that will follow, rather than* **(CL.1)** *with $a = 4$, the weakest assumption really needed could be summarized by the existence of a large constant C such that*

$$\limsup_{n\to\infty} \frac{1}{b_n^2}\log\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} V_k^4 > C\right) = -\infty.$$

Theorem 5.3.1. *Assume that* **(CL.1)**, **(CL.2)** *and* **(CL.3)** *are satisfied. Then, the sequence*

$$\left(\frac{\sqrt{n}}{b_n} \left(\widehat{\theta}_n - \theta^* \right) \right)_{n \geq 1}$$

satisfies the LDP on $\mathbb{R}$ *given in Theorem 5.2.1.*

Theorem 5.3.2. *Assume that* **(CL.1)**, **(CL.2)** *and* **(CL.3)** *are satisfied. Then, as soon as* $\theta \neq -\rho$, *the sequence*

$$\left(\frac{\sqrt{n}}{b_n} \begin{pmatrix} \widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^* \end{pmatrix} \right)_{n \geq 1}$$

satisfies the LDP on $\mathbb{R}^2$ *given in Theorem 5.2.2. In particular, the sequence*

$$\left(\frac{\sqrt{n}}{b_n} \left(\widehat{\rho}_n - \rho^* \right) \right)_{n \geq 1}$$

satisfies the LDP on $\mathbb{R}$ *also given in Theorem 5.2.2.*

Remark 5.3.5. *We have already seen in Remark 5.2.1 that the covariance matrix* Γ *is invertible if and only if* $\theta \neq -\rho$. *In the particular case where* $\theta = -\rho$, *the sequences*

$$\left(\frac{\sqrt{n}}{b_n} \left(\widehat{\theta}_n - \theta^* \right) \right)_{n \geq 1} \qquad \text{and} \qquad \left(\frac{\sqrt{n}}{b_n} \left(\widehat{\rho}_n - \rho^* \right) \right)_{n \geq 1}$$

satisfy the LDP on $\mathbb{R}$ *given in Remark 5.2.1.*

Theorem 5.3.3. *Assume that* **(CL.1)**, **(CL.2)** *and* **(CL.3)** *are satisfied. Then, the sequence*

$$\left(\frac{\sqrt{n}}{b_n} \left(\widehat{D}_n - D^* \right) \right)_{n \geq 1}$$

satisfies the LDP on $\mathbb{R}$ *given in Theorem 5.2.3.*

Proof. Theorem 5.3.1, Theorem 5.3.2 and Theorem 5.3.3 are proved in Section 4. □

5.4 Proof of the main results

For a matter of readability, some notation commonly used in the following proofs have to be introduced. First, for all $n \geq 1$, let

$$L_n = \sum_{k=1}^{n} V_k^2. \tag{5.4.1}$$

Then, let us define M_n, for all $n \geq 1$, as

$$M_n = \sum_{k=1}^{n} X_{k-1} V_k \tag{5.4.2}$$

where $M_0 = 0$. For all $n \geq 1$, denote by $\mathcal{F}_n$ the σ-algebra of the events occurring up to time n, $\mathcal{F}_n = \sigma(X_0, \varepsilon_0, V_1, \cdots, V_n)$. We infer from (5.4.2) that $(M_n)_{n\geq 0}$ is a locally square-integrable real martingale with respect to the filtration $\mathbb{F} = (\mathcal{F}_n)_{n\geq 0}$ with predictable quadratic variation given by $\langle M \rangle_0 = 0$ and for all $n \geq 1$, $\langle M \rangle_n = \sigma^2 S_{n-1}$, where

$$S_n = \sum_{k=0}^{n} X_k^2. \tag{5.4.3}$$

Moreover, $(N_n)_{n\geq 0}$ is defined, for all $n \geq 2$, as

$$N_n = \sum_{k=2}^{n} X_{k-2} V_k \tag{5.4.4}$$

and $N_0 = N_1 = 0$. It is not hard to see that $(N_n)_{n\geq 0}$ is also a locally square-integrable real martingale sharing the same properties than $(M_n)_{n\geq 0}$. More precisely, its predictable quadratic variation is given by $\langle N \rangle_n = \sigma^2 S_{n-2}$. To conclude, let $P_0 = 0$ and, for all $n \geq 1$,

$$P_n = \sum_{k=1}^{n} X_{k-1} X_k. \tag{5.4.5}$$

4.1. Proof of Theorem 5.2.1.

Before starting the proof of Theorem 5.2.1, we need to introduce some technical tools. Denote by ℓ the almost sure limit of S_n/n [19], given by

$$\ell = \frac{\sigma^2(1+\theta\rho)}{(1-\theta^2)(1-\theta\rho)(1-\rho^2)}. \tag{5.4.6}$$

Lemma 5.4.1. *Under the assumptions of Theorem 5.2.1, we have the exponential convergence*

$$\frac{S_n}{n} \xrightarrow[b_n^2]{\text{superexp}} \ell \tag{5.4.7}$$

where ℓ is given by (5.4.6).

Proof. After straightforward calculations, we get that for all $n \geq 2$,

$$\frac{S_n}{n} - \ell = \frac{\ell}{\sigma^2}\left[\left(\frac{L_n}{n} - \sigma^2\right) + 2\theta^* \frac{M_n}{n} - 2\theta\rho \frac{N_n}{n} + \frac{R_n}{n}\right] \tag{5.4.8}$$

where L_n, M_n, S_n and N_n are respectively given by (5.4.1), (5.4.2), (5.4.3) and (5.4.4),

$$R_n = [2(\theta+\rho)\rho^* - (\theta+\rho)^2 - (\theta\rho)^2]X_n^2 - (\theta\rho)^2 X_{n-1}^2 + 2\rho^* X_n X_{n-1} + \xi_1,$$

and where the remainder term

$$\xi_1 = (1 - 2\theta\rho - \rho^2)X_0^2 + \rho^2\varepsilon_0^2 + 2\theta\rho X_0\varepsilon_0 - 2\rho\rho^*(\varepsilon_0 - X_0)X_0 + 2\rho(\varepsilon_0 - X_0)V_1.$$

First of all, (V_n) is a sequence of independent and identically distributed Gaussian random variables with zero mean and variance $\sigma^2 > 0$. It immediately follows from Cramér-Chernoff's Theorem, expounded e.g. in [35], that for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left| \frac{L_n}{n} - \sigma^2 \right| > \delta \right) < 0. \tag{5.4.9}$$

Since $b_n^2 = o(n)$, the latter convergence leads to

$$\frac{L_n}{n} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} \sigma^2, \tag{5.4.10}$$

ensuring the exponential convergence of L_n/n to σ^2 with speed b_n^2. Moreover, for all $\delta > 0$ and a suitable $t > 0$, we clearly obtain from Markov's inequality that

$$\mathbb{P}\left(\frac{X_0^2}{n} > \delta \right) \leq \exp(-tn\delta)\, \mathbb{E}\Big[\exp(tX_0^2) \Big],$$

which immediately implies via **(G.2)**,

$$\frac{X_0^2}{n} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0, \tag{5.4.11}$$

and we get the exponential convergence of X_0^2/n to 0 with speed b_n^2. The same is true for V_1^2/n, ε_0^2/n and more generally for any isolated term of order 2 in relation (5.4.8) whose numerator do not depend on n. Let us now focus our attention on X_n^2/n. The model (5.1.1) can be rewritten in the vectorial form,

$$\Phi_n = A\Phi_{n-1} + W_n \tag{5.4.12}$$

where $\Phi_n = \begin{pmatrix} X_n & X_{n-1} \end{pmatrix}'$ stands for the lag vector of order 2, $W_n = \begin{pmatrix} V_n & 0 \end{pmatrix}'$ and

$$A = \begin{pmatrix} \theta+\rho & -\theta\rho \\ 1 & 0 \end{pmatrix}. \tag{5.4.13}$$

It is easy to show that $\rho(A) = \max(|\theta|, |\rho|) < 1$ under the stability conditions. According to Proposition 4.1 of [113],

$$\frac{\|\Phi_n\|^2}{n} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0,$$

which is clearly sufficient to deduce that

$$\frac{X_n^2}{n} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0. \tag{5.4.14}$$

The exponential convergence of R_n/n to 0 with speed b_n^2 is achieved following exactly the same lines. To conclude the proof of Lemma 5.4.1, it remains to study the exponential asymptotic behavior of M_n/n. For all $\delta > 0$ and a suitable $y > 0$,

$$\begin{aligned} \mathbb{P}\left(\frac{M_n}{n} > \delta\right) &= \mathbb{P}\left(\frac{M_n}{n} > \delta, \langle M\rangle_n \leq y\right) + \mathbb{P}\left(\frac{M_n}{n} > \delta, \langle M\rangle_n > y\right), \\ &\leq \exp\left(-\frac{n^2\delta^2}{2y}\right) + \mathbb{P}\Big(\langle M\rangle_n > y\Big), \end{aligned} \tag{5.4.15}$$

by application of Theorem 4.1 of [20] in case of a Gaussian martingale. Then, noting that we have the following inequality,

$$S_n \leq \alpha X_0^2 + \beta\varepsilon_0^2 + \beta L_n \quad \text{a.s.} \tag{5.4.16}$$

with $\alpha = 1 + (1 - |\theta|)^{-2}$ and $\beta = (1 - |\rho|)^{-2}\,(1 - |\theta|)^{-2}$, we get for a suitable $t > 0$,

$$\begin{aligned} \mathbb{P}\Big(\langle M\rangle_n > y\Big) &\leq \mathbb{P}\Big(X_0^2 > \frac{y}{3\alpha\sigma^2}\Big) + \mathbb{P}\left(\varepsilon_0^2 > \frac{y}{3\beta\sigma^2}\right) + \mathbb{P}\left(L_{n-1} > \frac{y}{3\beta\sigma^2}\right), \\ &\leq \exp\left(\frac{-yt}{3\alpha\sigma^2}\right)\mathbb{E}\Big[\exp(tX_0^2)\Big] + \exp\left(\frac{-yt}{3\beta\sigma^2}\right)\mathbb{E}\Big[\exp(t\varepsilon_0^2)\Big] \\ &\qquad + \mathbb{P}\left(L_{n-1} > \frac{y}{3\beta\sigma^2}\right), \\ &\leq 3\max\left(\exp\left(\frac{-yt}{3\alpha\sigma^2}\right)\mathbb{E}\Big[\exp(tX_0^2)\Big], \exp\left(\frac{-yt}{3\beta\sigma^2}\right)\mathbb{E}\Big[\exp(t\varepsilon_0^2)\Big],\right. \\ &\qquad \left.\mathbb{P}\left(L_{n-1} > \frac{y}{3\beta\sigma^2}\right)\right). \end{aligned}$$

Let us choose $y = nx$, assuming $x > 3\beta\sigma^4$. It follows that

$$\begin{aligned} \frac{1}{b_n^2}\log\mathbb{P}\Big(\langle M\rangle_n > nx\Big) &\leq \frac{\log 3}{b_n^2} + \frac{1}{b_n^2}\max\left(\frac{-nxt}{3\alpha\sigma^2} + \log\mathbb{E}\Big[\exp(tX_0^2)\Big],\right. \\ &\qquad \left.\frac{-nxt}{3\beta\sigma^2} + \log\mathbb{E}\Big[\exp(t\varepsilon_0^2)\Big], \log\mathbb{P}\left(L_{n-1} > \frac{nx}{3\beta\sigma^2}\right)\right). \end{aligned}$$

Since $b_n^2 = o(n)$ and by virtue of (5.4.10) with $\delta = x/(3\beta\sigma^2) - \sigma^2 > 0$, we obtain that

$$\limsup_{n\to\infty} \frac{1}{b_n^2}\log\mathbb{P}\Big(\langle M\rangle_n > nx\Big) = -\infty. \tag{5.4.17}$$

It enables us by (5.4.15) to deduce that for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{M_n}{n} > \delta\right) = -\infty. \tag{5.4.18}$$

The same result is also true replacing M_n by $-M_n$ in (5.4.18) since M_n and $-M_n$ share the same distribution. Therefore, we find that

$$\frac{M_n}{n} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0. \tag{5.4.19}$$

A similar reasoning leads to the exponential convergence of N_n/n to 0, with speed b_n^2. Finally, we obtain (5.4.7) from (5.4.8) together with (5.4.10), (5.4.11), (5.4.14) and (5.4.19) which achieves the proof of Lemma 5.4.1. □

Corollary 5.4.2. *By virtue of Lemma 5.4.1 and under the same assumptions, we have the exponential convergence*

$$\frac{P_n}{n} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} \ell_1 \tag{5.4.20}$$

where $\ell_1 = \theta^* \ell$.

Proof. The proof of Corollary 5.4.2 is immediately derived from the following inequality,

$$\begin{aligned} \left|\frac{P_n}{n} - \theta^* \frac{S_n}{n}\right| &= \left|\frac{1}{1+\theta\rho}\frac{M_n}{n} + \frac{1}{1+\theta\rho}\frac{R_n(\theta)}{n} - \theta^* \frac{X_n^2}{n}\right|, \\ &\leq \frac{1}{1+\theta\rho}\frac{|M_n|}{n} + \frac{1}{1+\theta\rho}\frac{|R_n(\theta)|}{n} + |\theta^*|\frac{X_n^2}{n} \end{aligned} \tag{5.4.21}$$

with $R_n(\theta) = \theta\rho X_n X_{n-1} + \rho X_0(\varepsilon_0 - X_0)$. □

We are now in the position to prove Theorem 5.2.1. We shall make use of the following deviation principle for martingales established by Worms [112].

Theorem 5.4.3 (Worms). *Let (Y_n) be an adapted sequence with values in $\mathbb{R}^p$, and (V_n) a Gaussian noise with variance $\sigma^2 > 0$. We suppose that (Y_n) satisfies, for some invertible square matrix C of order p and a speed sequence (b_n^2) such that $b_n^2 = o(n)$, the exponential convergence for any $\delta > 0$,*

$$\lim_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\left\|\frac{1}{n}\sum_{k=0}^{n-1} Y_k Y_k' - C\right\| > \delta\right) = -\infty. \tag{5.4.22}$$

Then, the sequence

$$\left(\frac{M_n}{b_n\sqrt{n}}\right)_{n\geq 1}$$

satisfies an LDP on $\mathbb{R}^p$ of speed b_n^2 and good rate function

$$I(x) = \frac{1}{2\sigma^2} x' C^{-1} x \tag{5.4.23}$$

where (M_n) is the martingale given by

$$M_n = \sum_{k=1}^{n} Y_{k-1} V_k.$$

Proof. The proof of Theorem 5.4.3 is contained in the one of Theorem 5 of [112] with $d = 1$. ☐

Proof of Theorem 5.2.1. Let us consider the decomposition

$$\frac{\sqrt{n}}{b_n}\left(\widehat{\theta}_n - \theta^*\right) = \frac{\sqrt{n}}{b_n}\left(\frac{\sigma^2}{1+\theta\rho}\right)\frac{M_n}{\langle M\rangle_n} + \frac{\sqrt{n}}{b_n}\left(\frac{1}{1+\theta\rho}\right)\frac{R_n(\theta)}{S_{n-1}}, \tag{5.4.24}$$

that can be obtained by a straightforward calculation, where the remainder term $R_n(\theta)$ is defined in (5.4.21). First, by using the same methodology as in convergence (5.4.11), we obtain that for all $\delta > 0$ and for a suitable $t > 0$,

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{X_0^2}{b_n\sqrt{n}} > \delta\right) &\leq \lim_{n\to\infty}\left(-t\delta\frac{\sqrt{n}}{b_n}\right) + \lim_{n\to\infty}\frac{1}{b_n^2}\log\mathbb{E}\Big[\exp(tX_0^2)\Big], \\ &= -\infty, \end{aligned} \tag{5.4.25}$$

since $b_n = o(\sqrt{n})$. Using inequality $X_0\varepsilon_0 \leq X_0^2/2 + \varepsilon_0^2/2$, we can prove in the same manner that $X_0\varepsilon_0/b_n\sqrt{n} \underset{b_n^2}{\overset{\text{superexp}}{\Longrightarrow}} 0$. Moreover, under the Gaussian assumption on the driven noise (V_n), it is not hard to see that

$$\frac{1}{b_n\sqrt{n}} \max_{1\leq k\leq n} V_k^2 \underset{b_n^2}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{5.4.26}$$

As a matter of fact, for all $\delta > 0$ and for all $t > 0$,

$$\begin{aligned}\mathbb{P}\left(\max_{1\leq k\leq n} V_k^2 \geq \delta b_n\sqrt{n}\right) &= \mathbb{P}\left(\bigcup_{k=1}^{n}\left\{V_k^2 \geq \delta b_n\sqrt{n}\right\}\right) \leq \sum_{k=1}^{n}\mathbb{P}\Big(V_k^2 \geq \delta b_n\sqrt{n}\Big), \\ &\leq n\exp\left(-t\delta b_n\sqrt{n}\right)\mathbb{E}\Big[\exp\left(tV_1^2\right)\Big].\end{aligned}$$

In addition, as soon as $0 < t < 1/(2\sigma^2)$, $\mathbb{E}\big[\exp(tV_1^2)\big] < \infty$. Consequently,

$$\begin{aligned}\frac{1}{b_n^2}\log\mathbb{P}\left(\max_{1\leq k\leq n} V_k^2 \geq \delta b_n\sqrt{n}\right) &\leq \frac{\log n}{b_n^2} - \frac{t\delta\sqrt{n}}{b_n} + \frac{\log\mathbb{E}\Big[\exp\left(tV_1^2\right)\Big]}{b_n^2}, \\ &\leq \frac{\sqrt{n}}{b_n}\left(\frac{\log n}{b_n\sqrt{n}} - t\delta + \frac{\log\mathbb{E}\Big[\exp\left(tV_1^2\right)\Big]}{b_n\sqrt{n}}\right)\end{aligned}$$

which clearly leads to (5.4.26). Furthermore, it follows from (5.1.1) that

$$\max_{1\leq k\leq n} X_k^2 \leq \frac{1}{1-|\theta|}X_0^2 + \left(\frac{1}{1-|\theta|}\right)^2 \max_{1\leq k\leq n} \varepsilon_k^2, \tag{5.4.27}$$

as well as

$$\max_{1\leq k\leq n} \varepsilon_k^2 \leq \frac{1}{1-|\rho|}\varepsilon_0^2 + \left(\frac{1}{1-|\rho|}\right)^2 \max_{1\leq k\leq n} V_k^2. \tag{5.4.28}$$

Then, we deduce from (5.4.25), (5.4.26), (5.4.27) and (5.4.28) that

$$\frac{1}{b_n\sqrt{n}} \max_{1\leq k\leq n} \varepsilon_k^2 \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0 \qquad \text{and} \qquad \frac{1}{b_n\sqrt{n}} \max_{1\leq k\leq n} X_k^2 \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0,$$

which of course imply the exponential convergence of $X_n^2/(b_n\sqrt{n})$ to 0, with speed b_n^2. Therefore, we obtain that

$$\frac{R_n(\theta)}{b_n\sqrt{n}} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0. \tag{5.4.29}$$

We infer from Lemma 5.4.1 together with Lemma 4.1 of [113] that the following convergence is satisfied,

$$\frac{n}{S_n} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} \frac{1}{\ell} \tag{5.4.30}$$

where $\ell > 0$ is given by (5.4.6). According to (5.4.29), the latter convergence and again Lemma 4.1 of [113], we deduce that

$$\frac{\sqrt{n}}{b_n}\left(\frac{1}{1+\theta\rho}\right)\frac{R_n(\theta)}{S_{n-1}} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0. \tag{5.4.31}$$

Hence, we obtain from (5.4.30) that the same is true for

$$\frac{\sigma^2}{1+\theta\rho}\frac{M_n}{b_n\sqrt{n}}\left(\frac{n}{\langle M\rangle_n} - \frac{1}{\sigma^2\ell}\right) \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0, \tag{5.4.32}$$

since Lemma 5.4.1 together with Theorem 5.4.3 with $p = 1$ directly show that $(M_n/(b_n\sqrt{n}))$ satisfies an LDP with speed b_n^2 and good rate function given, for all $x \in \mathbb{R}$, by

$$J(x) = \frac{x^2}{2\ell\sigma^2}. \tag{5.4.33}$$

As a consequence,

$$\frac{\sqrt{n}}{b_n}\left(\widehat{\theta}_n - \theta^*\right) \overset{\text{superexp}}{\underset{b_n^2}{\sim}} \frac{1}{\ell(1+\theta\rho)}\frac{M_n}{b_n\sqrt{n}}, \tag{5.4.34}$$

and this implies that both of them share the same LDP, see e.g. [35]. One shall now take advantage of the contraction principle [35] to establish that $(\sqrt{n}(\widehat{\theta}_n - \theta^*)/b_n)$

satisfies an LDP with speed b_n^2 and good rate function $I_\theta(x)$ given by (5.2.1). The contraction principle enables us to conclude that the good rate function of the LDP with speed b_n^2 associated with equivalence (5.4.34) is given by $I_\theta(x) = J(\ell(1+\theta\rho)x)$, that is

$$I_\theta(x) = \frac{x^2}{2\sigma_\theta^2},$$

which achieves the proof of Theorem 5.2.1. □

4.2. Proof of Theorem 5.2.2.

We need to introduce some more notation. For all $n \geq 2$, let

$$Q_n = \sum_{k=2}^{n} X_{k-2} V_k. \tag{5.4.35}$$

In addition, for all $n \geq 1$, denote

$$T_n = 1 + \theta^* \rho^* - \Big(1 + \rho^*(\widehat{\theta}_n + \theta^*)\Big) \frac{S_n}{S_{n-1}} + \Big(2\rho^* + \widehat{\theta}_n + \theta^*\Big) \frac{P_n}{S_{n-1}} - \frac{Q_n}{S_{n-1}}, \tag{5.4.36}$$

where S_n and P_n are respectively given by (5.4.3) and (5.4.5). Finally, for all $n \geq 0$, let

$$J_n = \sum_{k=0}^{n} \widehat{\varepsilon}_k^{\,2} \tag{5.4.37}$$

where the residual set $(\widehat{\varepsilon}_n)$ is given in (5.1.3). A set of additional technical tools has to be expounded to make the proof of Theorem 5.2.2 more tractable.

Corollary 5.4.4. *By virtue of Lemma 5.4.1 and under the same assumptions, we have the exponential convergence*

$$\frac{Q_n}{n} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} \ell_2$$

where $\ell_2 = ((\theta + \rho)\theta^* - \theta\rho)\ell$.

Proof. The proof of Corollary 5.4.4 immediately follows from the inequality,

$$\begin{aligned} \left| \frac{Q_n}{n} - ((\theta+\rho)\theta^* - \theta\rho)\frac{S_n}{n} \right| &= \left| \theta^* \frac{M_n}{n} + \frac{N_n}{n} + \frac{\xi_n^Q}{n} \right|, \\ &\leq |\theta^*| \frac{|M_n|}{n} + \frac{|N_n|}{n} + \frac{|\xi_n^Q|}{n} \end{aligned} \tag{5.4.38}$$

where ξ_n^Q is a residual made of isolated terms such that

$$\frac{\xi_n^Q}{n} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0,$$

see e.g. the proof of Theorem 3.2 in [19] where more details are given on ξ_n^Q. □

Lemma 5.4.5. *Under the assumptions of Theorem 5.2.2, we have the exponential convergence*

$$A_n \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} A$$

where

$$A_n = \frac{n}{1+\theta\rho}\begin{pmatrix} \dfrac{1}{S_{n-1}} & 0 \\ \dfrac{T_n}{J_{n-1}} & -\dfrac{(\theta+\rho)}{J_{n-1}} \end{pmatrix}, \tag{5.4.39}$$

and

$$A = \frac{1}{\ell(1+\theta\rho)(1-(\theta^*)^2)}\begin{pmatrix} 1-(\theta^*)^2 & 0 \\ \theta\rho+(\theta^*)^2 & -(\theta+\rho) \end{pmatrix}. \tag{5.4.40}$$

Proof. Via (5.4.30), we directly obtain the exponential convergence,

$$\frac{1}{(1+\theta\rho)}\frac{n}{S_{n-1}} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} \frac{1}{\ell(1+\theta\rho)}. \tag{5.4.41}$$

The combination of Lemma 5.4.1, Corollary 5.4.2, Corollary 5.4.4 and Lemma 4.1 of [113] shows, after a simple calculation, that

$$T_n \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} (\theta^*)^2 + \theta\rho. \tag{5.4.42}$$

Moreover, J_n given by (5.4.37) can be rewritten as

$$J_n = S_n - 2\widehat{\theta}_n P_n + \widehat{\theta}_n^{\,2} S_{n-1},$$

which leads, via Lemma 4.1 in [113], to

$$\frac{J_n}{n} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} \ell(1-(\theta^*)^2). \tag{5.4.43}$$

Convergences (5.4.42) and (5.4.43) imply

$$\left(\frac{n}{1+\theta\rho}\right)\frac{T_n}{J_{n-1}} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} \frac{(\theta^*)^2+\theta\rho}{\ell(1+\theta\rho)(1-(\theta^*)^2)}, \tag{5.4.44}$$

and finally,

$$\left(\frac{n}{1+\theta\rho}\right)\frac{\theta+\rho}{J_{n-1}} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} \frac{\theta+\rho}{\ell(1+\theta\rho)(1-(\theta^*)^2)}. \tag{5.4.45}$$

Finally, (5.4.41) together with (5.4.44) and (5.4.45) achieve the proof of Lemma 5.4.5. □

Proof of Theorem 5.2.2. We shall make use of the decomposition

$$\frac{\sqrt{n}}{b_n}\begin{pmatrix}\widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^*\end{pmatrix} = \frac{1}{b_n\sqrt{n}} A_n Z_n + B_n, \tag{5.4.46}$$

where A_n is given by (5.4.39), $(Z_n)_{n\geq 0}$ is the 2-dimensional vector martingale given by

$$Z_n = \begin{pmatrix} M_n \\ N_n \end{pmatrix}, \tag{5.4.47}$$

and where the remainder term

$$B_n = \frac{1}{(1+\theta\rho)}\frac{\sqrt{n}}{b_n}\begin{pmatrix} \dfrac{R_n(\theta)}{S_{n-1}} \\ \dfrac{R_n(\rho)}{J_{n-1}} \end{pmatrix}. \tag{5.4.48}$$

The first component $R_n(\theta)$ is given in (5.4.21) while $R_n(\rho)$, whose definition may be found in the proof of Theorem 3.2 in [19], is made of isolated terms. Consequently, (5.4.25) and (5.4.29) are sufficient to ensure that

$$\frac{R_n(\theta)}{b_n\sqrt{n}} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0 \qquad \text{and} \qquad \frac{R_n(\rho)}{b_n\sqrt{n}} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0.$$

Therefore, we obtain that

$$B_n \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0. \tag{5.4.49}$$

In addition, it follows from Lemma 5.4.5 and Theorem 5.4.3 with $p = 2$ that $(Z_n/(b_n\sqrt{n}))$ satisfies an LDP on $\mathbb{R}^2$ with speed b_n^2 and good rate function given, for all $x \in \mathbb{R}^2$, by

$$J(x) = \frac{1}{2\sigma^2}x'\Lambda^{-1}x, \tag{5.4.50}$$

where

$$\Lambda = \ell\begin{pmatrix} 1 & \theta^* \\ \theta^* & 1 \end{pmatrix}, \tag{5.4.51}$$

since we have the exponential convergence

$$\frac{\langle Z\rangle_n}{n} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} \sigma^2\Lambda \tag{5.4.52}$$

by application of Lemma 5.4.1 and Corollary 5.4.2. One observes that $\det(\Lambda) = \ell^2(1-(\theta^*)^2) > 0$ implying that Λ is invertible. As a consequence,

$$\frac{1}{b_n\sqrt{n}}(A_n - A)Z_n \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0, \tag{5.4.53}$$

and we deduce from (5.4.46) that

$$\frac{\sqrt{n}}{b_n}\begin{pmatrix}\widehat{\theta}_n-\theta^*\\ \widehat{\rho}_n-\rho^*\end{pmatrix} \underset{b_n^2}{\overset{\text{superexp}}{\sim}} \frac{1}{b_n\sqrt{n}}AZ_n. \tag{5.4.54}$$

This of course implies that both of them share the same LDP. The contraction principle [35] enables us to conclude that the rate function of the LDP on $\mathbb{R}^2$ with speed b_n^2 associated with equivalence (5.4.54) is given, for all $x \in \mathbb{R}^2$, by $K(x) = J(A^{-1}x)$, that is

$$K(x) = \frac{1}{2}x'\Gamma^{-1}x,$$

where $\Gamma = \sigma^2 A\Lambda A'$ is given by (5.1.10), and where we shall suppose that $\theta \neq -\rho$ to ensure that A is invertible. In particular, the latter result also implies that the good rate function of the LDP on $\mathbb{R}$ with speed b_n^2 associated with $(\sqrt{n}(\widehat{\rho}_n-\rho^*)/b_n)$ is given, for all $x \in \mathbb{R}$, by

$$I_\rho(x) = \frac{x^2}{2\sigma_\rho^2},$$

where σ_ρ^2 is the last element of the matrix Γ. This achieves the proof of Theorem 5.2.2. □

4.3. Proof of Theorem 5.2.3.

For all $n \geq 1$, denote by f_n the explosion coefficient associated with J_n given by (5.4.37), that is

$$f_n = \frac{J_n - J_{n-1}}{J_n} = \frac{\widehat{\varepsilon}_n^{\,2}}{J_n}. \tag{5.4.55}$$

It follows from decomposition (C.4) in [19] that

$$\frac{\sqrt{n}}{b_n}\Big(\widehat{D}_n - D^*\Big) = -2\frac{\sqrt{n}}{b_n}\Big(1-f_n\Big)\Big(\widehat{\rho}_n-\rho^*\Big) + \frac{\sqrt{n}}{b_n}\zeta_n, \tag{5.4.56}$$

where the remainder term ζ_n is made of isolated terms. As before, we clearly have

$$\frac{\sqrt{n}}{b_n}\zeta_n \underset{b_n^2}{\overset{\text{superexp}}{\Longrightarrow}} 0 \qquad \text{and} \qquad f_n \underset{b_n^2}{\overset{\text{superexp}}{\Longrightarrow}} 0.$$

As a consequence,

$$\frac{\sqrt{n}}{b_n}\Big(\widehat{D}_n - D^*\Big) \underset{b_n^2}{\overset{\text{superexp}}{\sim}} -2\frac{\sqrt{n}}{b_n}\Big(\widehat{\rho}_n-\rho^*\Big), \tag{5.4.57}$$

and this implies that both of them share the same LDP. The contraction principle [35] enables us to conclude that the rate function of the LDP on $\mathbb{R}$ with speed b_n^2

associated with equivalence (5.4.57) is given, for all $x \in \mathbb{R}$, by $I_D(x) = I_\rho(-x/2)$, that is

$$I_D(x) = \frac{x^2}{2\sigma_D^2},$$

which achieves the proof of Theorem 5.2.3. □

4.4. Proofs of Theorem 5.3.1, Theorem 5.3.2 and Theorem 5.3.3.

We shall now propose a technical lemma ensuring that all results already proved under the Gaussian assumption still hold under the Chen-Ledoux type condition.

Lemma 5.4.6. *Under* **(CL.1)**, **(CL.2)** *and* **(CL.3)**, *all exponential convergences of Lemma 5.4.1, Corollary 5.4.2, Corollary 5.4.4 and Lemma 5.4.5 still hold.*

Proof. Under **(CL.1)**, **(CL.2)** and **(CL.3)**, and following the same methodology as the one used to establish (5.4.29), we get

$$\frac{X_n^2}{b_n\sqrt{n}} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0, \tag{5.4.58}$$

and Cauchy-Schwarz inequality implies that this is also the case for any isolated term of order 2, such as $X_n X_{n-1}/(b_n\sqrt{n})$. This allows us to control each remainder term. Note that **(CL.2)**, **(CL.3)** and (5.4.58) are obviously true for ε_0^4/n, X_0^4/n, ε_0^2/n, X_0^2/n and X_n^2/n, since $b_n\sqrt{n} = o(n)$. Moreover, if follows from Theorem 2.2 of [53] under **(CL.1)** with $a = 2$, that

$$\frac{L_n}{n} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} \sigma^2. \tag{5.4.59}$$

Furthermore, since (M_n) is a locally square integrable martingale, we infer from Theorem 2.1 of [20] that for all $x, y > 0$,

$$\mathbb{P}\Big(|M_n| > x, \langle M\rangle_n + [M]_n \leq y\Big) \leq 2\exp\left(-\frac{x^2}{2y}\right), \tag{5.4.60}$$

where the predictable quadratic variation $\langle M\rangle_n = \sigma^2 S_{n-1}$ is described in (5.4.3) and the total quadratic variation is given by $[M]_0 = 0$ and, for all $n \geq 1$, by

$$[M]_n = \sum_{k=1}^{n} X_{k-1}^2 V_k^2. \tag{5.4.61}$$

According to (5.4.60), we have for all $\delta > 0$ and a suitable $b > 0$,

$$\begin{aligned}
\mathbb{P}\left(\frac{|M_n|}{n} > \delta\right) &\leq \mathbb{P}\Big(|M_n| > \delta n, \langle M\rangle_n + [M]_n \leq nb\Big) + \mathbb{P}\Big(\langle M\rangle_n + [M]_n > nb\Big), \\
&\leq 2\exp\left(-\frac{n\delta^2}{2b}\right) + \mathbb{P}\Big(\langle M\rangle_n + [M]_n > nb\Big), \\
&\leq 2\max\left(\mathbb{P}\Big(\langle M\rangle_n + [M]_n > nb\Big), 2\exp\left(-\frac{n\delta^2}{2b}\right)\right).
\end{aligned}$$

Consequently,

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{|M_n|}{n} > \delta\right) \leq \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\Big(\langle M\rangle_n + [M]_n > nb\Big). \quad (5.4.62)$$

We have for all $b > 0$,

$$\begin{aligned}
\mathbb{P}\Big(\langle M\rangle_n + [M]_n > nb\Big) &\leq \mathbb{P}\left(\langle M\rangle_n > \frac{nb}{2}\right) + \mathbb{P}\left([M]_n > \frac{nb}{2}\right) \\
&\leq 2\max\left(\mathbb{P}\left(\langle M\rangle_n > \frac{nb}{2}\right), \mathbb{P}\left([M]_n > \frac{nb}{2}\right)\right). \quad (5.4.63)
\end{aligned}$$

Moreover, for all $n \geq 1$, let us define

$$T_n = \sum_{k=0}^{n} X_k^4 \qquad \text{and} \qquad \Gamma_n = \sum_{k=1}^{n} V_k^4,$$

and note that we easily have the following inequality,

$$T_n \leq \alpha X_0^4 + \beta \varepsilon_0^4 + \beta \Gamma_n \quad \text{a.s.} \quad (5.4.64)$$

with $\alpha = 1 + (1 - |\theta|)^{-4}$ and $\beta = (1 - |\rho|)^{-4}(1 - |\theta|)^{-4}$. This implies that, for n large enough, one can find $\gamma > 0$ such that

$$T_n \leq \gamma \Gamma_n \quad \text{a.s.}$$

choosing for example $\gamma = 3\max(\alpha, \beta)$, under **(CL.2)** and **(CL.3)** for $a = 4$. According to Theorem 2.2 of [53] under **(CL.1)** with $a = 4$, we also have the exponential convergence,

$$\frac{\Gamma_n}{n} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} \tau^4, \quad (5.4.65)$$

where $\tau^4 = \mathbb{E}[V_1^4]$, leading, via Cauchy-Schwarz inequality and (5.4.64), to

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{[M]_n}{n} > \delta\right) &\leq \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\Gamma_n}{n} > \frac{\delta}{\sqrt{\gamma}}\right), \\
&= -\infty, \quad (5.4.66)
\end{aligned}$$

where $\delta > \tau^4\sqrt{\gamma}$. Exploiting (5.4.16) and (5.4.59), the same result can be achieved for $\langle M\rangle_n/n$ under **(CL.1)** with $a = 2$ and $\delta > \sigma^4\gamma$. As a consequence, it follows from (5.4.63), (5.4.66) and the latter remark that

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{\langle M\rangle_n + [M]_n}{n} > b\right) = -\infty, \tag{5.4.67}$$

as soon as $b > \sigma^4\gamma + \tau^4\sqrt{\gamma}$. Therefore, the exponential convergence of M_n/n to 0 with speed b_n^2 is obtained via (5.4.62) and (5.4.67), that is, for all $\delta > 0$ and $b > \sigma^4\gamma + \tau^4\sqrt{\gamma}$,

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{|M_n|}{n} > \delta\right) = -\infty. \tag{5.4.68}$$

The same obviously holds for N_n/n. Following the same lines as in the proofs of Lemma 5.4.1, Corollary 5.4.2, Corollary 5.4.4 and Lemma 5.4.5, hypothesis **(CL.2)** and **(CL.3)** with $a = 4$ together with exponential convergences (5.4.58), (5.4.59) and (5.4.68) are sufficient to achieve the proof of Lemma 5.4.6. □

Let us introduce a simplified version of Puhalskii's result [91] applied to a sequence of martingale differences, and two technical lemmas that shall help us to prove our results.

Theorem 5.4.7 (Puhalskii). *Let $(m_j^n)_{1\le j\le n}$ be a triangular array of martingale differences with values in $\mathbb{R}^d$, with respect to the filtration $(\mathcal{F}_n)_{n\ge1}$. Let (b_n) be a sequence of real numbers satisfying (5.1.13). Suppose that there exists a symmetric positive-semidefinite matrix Q such that*

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\Big[m_k^n (m_k^n)' \big| \mathcal{F}_{k-1}\Big] \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} Q. \tag{5.4.69}$$

Suppose that there exists a constant $c > 0$ such that, for each $1 \le k \le n$,

$$|m_k^n| \le c\frac{\sqrt{n}}{b_n} \quad \text{a.s.} \tag{5.4.70}$$

Suppose also that, for all $a > 0$, we have the exponential Lindeberg's condition

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\Big[|m_k^n|^2 \mathrm{I}_{\left\{|m_k^n| \ge a\frac{\sqrt{n}}{b_n}\right\}} \big| \mathcal{F}_{k-1}\Big] \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0. \tag{5.4.71}$$

Then, the sequence

$$\left(\frac{1}{b_n\sqrt{n}}\sum_{k=1}^{n} m_k^n\right)_{n\ge1}$$

satisfies an LDP on $\mathbb{R}^d$ with speed b_n^2 and good rate function

$$\Lambda^*(v) = \sup_{\lambda \in \mathbb{R}^d} \left(\lambda' v - \frac{1}{2} \lambda' Q \lambda \right).$$

In particular, if Q is invertible,

$$\Lambda^*(v) = \frac{1}{2} v' Q^{-1} v. \tag{5.4.72}$$

Proof. The proof of Theorem 5.4.7 is contained e.g. in the proof of Theorem 3.1 in [91]. □

Lemma 5.4.8. *Under* **(CL.1)**, **(CL.2)** *and* **(CL.3)** *with $a = 2$, we have for all $\delta > 0$,*

$$\limsup_{R\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{b_n^2} \sum_{k=1}^n X_k^2 \mathrm{I}_{\{|X_k|>R\}} > \delta \right) = -\infty.$$

Remark 5.4.1. *Lemma 5.4.8 implies that the exponential Lindeberg's condition given by (5.4.71) is satisfied.*

Proof. We introduce the empirical measure associated with the geometric ergodic Markov chain $(X_n)_{n\geq 0}$,

$$\Lambda_n = \frac{1}{n} \sum_{k=1}^n \delta_{X_k}, \tag{5.4.73}$$

with invariant probability measure denoted by μ. It is well-known that the sequence (Λ_n) satisfies the upper bound of the moderate deviations, see e.g. [40] for more details. Let us define, for $f(x) = x^2$, the following truncations,

$$f^{(R)}(x) = f(x) \min\Big(1, \big(f(x) - (R-1)\big)_+\Big) \quad \text{and} \quad \widetilde{f}^{(R)}(x) = \min\Big(f^{(R)}(x), R\Big).$$

Thus, we have

$$0 \leq f(x)\mathrm{I}_{\{f(x)\geq R\}} \leq f^{(R)}(x) \leq f(x),$$

and, as a consequence,

$$0 \leq \Lambda_n\Big(f\mathrm{I}_{\{f\geq R\}}\Big) \leq \Lambda_n\Big(f^{(R)} - \widetilde{f}^{(R)}\Big) + \Lambda_n\Big(\widetilde{f}^{(R)}\Big) - \mu\Big(\widetilde{f}^{(R)}\Big) + \mu\Big(\widetilde{f}^{(R)}\Big).$$

We also have

$$f^{(R)} - \widetilde{f}^{(R)} = \Big(f^{(R)} - R\Big)\mathrm{I}_{\{f^{(R)}\geq R\}} \leq \Big(f - R\Big)\mathrm{I}_{\{f\geq R\}} = f - \Big(f \wedge R\Big).$$

For $\delta > 0$, the functions $\widetilde{f}^{(R)}$ and $f - (f \wedge R)$ are continuous and bounded by f which is μ-integrable, and they converge to 0 as R goes to infinity. By Lebesgue's

Theorem, there exists $R > 0$ large enough such that $\mu(\widetilde{f}^{(R)})+\mu(f-(f\wedge R)) < \delta/4$. Thus,

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_k^2 \mathrm{I}_{\{X_k^2\geq R\}} > \delta\right) \leq \mathbb{P}\Big(\Lambda_n(f) - \mu(f) > \delta/4\Big)$$
$$+\mathbb{P}\Big(\Lambda_n(f\wedge R) - \mu(f\wedge R) > \delta/4\Big) + \mathbb{P}\Big(\Lambda_n(\widetilde{f}^{(R)}) - \mu(\widetilde{f}^{(R)}) > \delta/4\Big). \quad (5.4.74)$$

From Lemma 5.4.6, we have that for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\Big(\Lambda_n(f) - \mu(f) > \delta\Big) = -\infty.$$

By the upper bound of the moderate deviation principle for the sequence (Λ_n) given in [40], we obtain that

$$\limsup_{R\to\infty}\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\Big(\Lambda_n(f\wedge R) - \mu(f\wedge R) > \delta\Big) = -\infty,$$

and

$$\limsup_{R\to\infty}\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\Big(\Lambda_n(\widetilde{f}^{(R)}) - \mu(\widetilde{f}^{(R)}) > \delta\Big) = -\infty,$$

which, via inequality (5.4.74), achieves the proof of Lemma 5.4.8. Note that Remark 5.4.1 is immediately derived from the latter proof, see e.g. [113] for more details. □

Lemma 5.4.9. *Under* **(CL.1)**, **(CL.2)** *and* **(CL.3)**, *the sequence*

$$\left(\frac{M_n}{b_n\sqrt{n}}\right)_{n\geq 1}$$

satisfies an LDP on $\mathbb{R}$ *with speed* b_n^2 *and good rate function*

$$J(x) = \frac{x^2}{2\ell\sigma^2} \quad (5.4.75)$$

where ℓ *is given by (5.4.6).*

Proof. From now on, in order to apply Puhalskii's result for the moderate deviations for martingales, we introduce the following modification of the martingale $(M_n)_{n\geq 0}$, for $r > 0$ and $R > 0$,

$$M_n^{(r,R)} = \sum_{k=1}^{n} X_{k-1}^{(r)} V_k^{(R)} \quad (5.4.76)$$

where, for all $1 \leq k \leq n$,

$$X_k^{(r)} = X_k \mathrm{I}_{\left\{|X_k|\leq r\frac{\sqrt{n}}{b_n}\right\}} \quad \text{and} \quad V_k^{(R)} = V_k \mathrm{I}_{\left\{|V_k|\leq R\right\}} - \mathbb{E}\Big[V_k \mathrm{I}_{\left\{|V_k|\leq R\right\}}\Big]. \quad (5.4.77)$$

Then, we have to prove that for all $r > 0$ the sequence $(M_n^{(r,R)})$ is an exponentially good approximation of (M_n) as R goes to infinity, see e.g. Definition 4.2.14 in [35]. This approximation, in the sense of the large deviations, is described by the following convergence, for all $r > 0$ and all $\delta > 0$,

$$\limsup_{R\to\infty} \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{|M_n - M_n^{(r,R)}|}{b_n\sqrt{n}} > \delta \right) = -\infty. \tag{5.4.78}$$

From Lemma 5.4.6, and since $\langle M \rangle_n = \sigma^2 S_{n-1}$, we have

$$\frac{\langle M \rangle_n}{n} \underset{b_n^2}{\overset{\text{superexp}}{\Longrightarrow}} \sigma^2 \ell. \tag{5.4.79}$$

From Lemma 5.4.6 and Remark 5.4.1, we also have for all $r > 0$,

$$\frac{1}{n} \sum_{k=0}^{n} X_k^2 \mathrm{I}_{\left\{|X_k| > r\frac{\sqrt{n}}{b_n}\right\}} \underset{b_n^2}{\overset{\text{superexp}}{\Longrightarrow}} 0. \tag{5.4.80}$$

We introduce the following notation,

$$\sigma_R^2 = \mathbb{E}\left[(V_1^{(R)})^2\right] \qquad \text{and} \qquad S_n^{(r)} = \sum_{k=0}^{n} (X_k^{(r)})^2.$$

Then, we easily transfer properties (5.4.79) and (5.4.80) to the truncated martingale $(M_n^{(r,R)})_{n\geq 0}$. We have for all $R > 0$ and all $r > 0$,

$$\frac{\langle M^{(r,R)} \rangle_n}{n} = \sigma_R^2 \frac{S_{n-1}^{(r)}}{n} = -\sigma_R^2 \left(\frac{S_{n-1}}{n} - \frac{S_{n-1}^{(r)}}{n} \right) + \sigma_R^2 \frac{S_{n-1}}{n} \underset{b_n^2}{\overset{\text{superexp}}{\Longrightarrow}} \sigma_R^2 \ell$$

which ensures that (5.4.69) is satisfied for the martingale $(M_n^{(r,R)})_{n\geq 0}$. Note also that Lemma 5.4.6 and Remark 5.4.1 work for the martinagle $(M_n^{(r,R)})_{n\geq 0}$. So, for all $r > 0$, the exponential Lindeberg's condition and thus (5.4.71) are satisfied for $(M_n^{(r,R)})_{n\geq 0}$. By Theorem 5.4.7, we deduce that $(M_n^{(r,R)}/b_n\sqrt{n})$ satisfies an LDP on $\mathbb{R}$ with speed b_n^2 and good rate function

$$J_R(x) = \frac{x^2}{2\sigma_R^2 \ell}. \tag{5.4.81}$$

It will be possible to drive the moderate deviations result for the martingale $(M_n)_{n\geq 0}$ by proving relation (5.4.78). For that matter, let us now introduce the following decomposition,

$$M_n - M_n^{(r,R)} = L_n^{(r)} + F_n^{(r,R)}$$

where

$$L_n^{(r)} = \sum_{k=1}^{n} \Big(X_{k-1} - X_{k-1}^{(r)} \Big) V_k \quad \text{and} \quad F_n^{(r,R)} = \sum_{k=1}^{n} \Big(V_k - V_k^{(R)} \Big) X_{k-1}^{(r)}.$$

One has to show that for all $r > 0$,

$$\frac{L_n^{(r)}}{b_n \sqrt{n}} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0, \tag{5.4.82}$$

and, for all $r > 0$ and all $\delta > 0$, that

$$\limsup_{R \to \infty} \limsup_{n \to \infty} \frac{1}{b_n^2} \log \mathbb{P} \left(\frac{|F_n^{(r,R)}|}{b_n \sqrt{n}} > \delta \right) = -\infty. \tag{5.4.83}$$

On the one hand,
note that for any $\eta > 0$,

$$\sum_{k=0}^{n} |X_k|^{2+\eta} \leq \alpha |X_0|^{2+\eta} + \beta |\varepsilon_0|^{2+\eta} + \beta \sum_{k=1}^{n} |V_k|^{2+\eta} \quad \text{a.s.}$$

with $\alpha = 1 + (1 - |\theta|)^{-(2+\eta)}$ and $\beta = (1 - |\rho|)^{-(2+\eta)} (1 - |\theta|)^{-(2+\eta)}$. This implies that, for n large enough, one can find $\gamma > 0$ such that

$$\sum_{k=0}^{n} |X_k|^{2+\eta} \leq \gamma \sum_{k=1}^{n} |V_k|^{2+\eta} \quad \text{a.s.} \tag{5.4.84}$$

taking for example $\gamma = 3 \max(\alpha, \beta)$, under **(CL.2)** and **(CL.3)** for $a = 2 + \eta$.
Thus,

$$\begin{aligned}
\frac{|L_n^{(r)}|}{b_n \sqrt{n}} &= \frac{1}{b_n \sqrt{n}} \left| \sum_{k=1}^{n} X_{k-1} \mathrm{I}_{\left\{ |X_{k-1}| > r \frac{\sqrt{n}}{b_n} \right\}} V_k \right|, \\
&\leq \frac{1}{b_n \sqrt{n}} \left(r \frac{\sqrt{n}}{b_n} \right)^{-\eta} \left(\sum_{k=1}^{n} |X_{k-1}|^{2+\eta} \right)^{1/2} \left(\sum_{k=1}^{n} V_k^2 |X_{k-1}|^{\eta} \right)^{1/2}, \\
&\leq \lambda(r, \eta, \gamma) \left(\frac{b_n}{\sqrt{n}} \right)^{\eta - 1} \frac{1}{n} \sum_{k=1}^{n} |V_k|^{2+\eta} \quad \text{a.s.}
\end{aligned} \tag{5.4.85}$$

by virtue of (5.4.84) and Hölder's inequality, where $\lambda(r, \eta, \gamma) > 0$ can be evaluated under suitable assumptions of moment on (V_n). As a consequence, for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{|L_n^{(r)}|}{b_n\sqrt{n}} > \delta \right) \leq \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} |V_k|^{2+\eta} > \frac{\delta}{\lambda(r,\eta,\gamma)} \left(\frac{\sqrt{n}}{b_n}\right)^{\eta-1} \right) = -\infty, \tag{5.4.86}$$

as soon as $\eta > 1$, under **(CL.1)** with $a = 2+\eta$. We deduce that

$$\frac{L_n^{(r)}}{b_n\sqrt{n}} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0, \tag{5.4.87}$$

which achieves the proof of (5.4.82), under **(CL.1)**, **(CL.2)** and **(CL.3)** for $a > 3$. On the other hand, $(F_n^{(r,R)})_{n\geq 0}$ is a locally square-integrable real martingale whose predictable quadratic variation is given by $\langle F^{(r,R)}\rangle_0 = 0$ and, for all $n \geq 1$, by

$$\langle F^{(r,R)}\rangle_n = \mathbb{E}\left[\left(V_1 - V_1^{(R)}\right)^2 \right] S_{n-1}^{(r)}.$$

To prove (5.4.83), we will use Theorem 1 of [38]. For R large enough and all $k \geq 1$, we have

$$\begin{aligned} \mathbb{P}\left(\left| X_{k-1}^{(r)} \left(V_k - V_k^{(R)}\right) \right| > b_n\sqrt{n} \,\middle|\, \mathcal{F}_{k-1} \right) &\leq \mathbb{P}\left(\left| V_k - V_k^{(R)} \right| > \frac{b_n^2}{r} \right), \\ &= \mathbb{P}\left(\left| V_1 - V_1^{(R)} \right| > \frac{b_n^2}{r} \right) = 0. \end{aligned}$$

This implies that

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \left(n \operatorname*{ess\,sup}_{k\geq 1} \mathbb{P}\left(\left| X_{k-1}^{(r)} \left(V_k - V_k^{(R)}\right) \right| > b_n\sqrt{n} \,\middle|\, \mathcal{F}_{k-1} \right) \right) = -\infty. \tag{5.4.88}$$

For all $\gamma > 0$ and all $\delta > 0$, we obtain from Lemma 5.4.8 and Remark 5.4.1, that

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} \left(X_{k-1}^{(r)}\right)^2 \mathrm{I}_{\left\{|X_{k-1}^{(r)}| > \gamma \frac{\sqrt{n}}{b_n}\right\}} > \delta \right) \leq \limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_{k-1}^2 \mathrm{I}_{\left\{|X_{k-1}| > \gamma \frac{\sqrt{n}}{b_n}\right\}} > \delta \right) = -\infty.$$

Finally, from Lemma 5.4.6, Lemma 5.4.8 and Remark 5.4.1, it follows that

$$\frac{\langle F^{(r,R)}\rangle_n}{n} = Q_R \frac{S_{n-1}^{(r)}}{n} = -Q_R\left(\frac{S_{n-1}}{n} - \frac{S_{n-1}^{(r)}}{n} \right) + Q_R \frac{S_{n-1}}{n} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} Q_R \ell$$

where

$$Q_R = \mathbb{E}\left[\left(V_1 - V_1^{(R)}\right)^2\right],$$

and ℓ is given by (5.4.6). Moreover, it is clear that Q_R converges to 0 as R goes to infinity. In light of foregoing, we infer from Theorem 1 of [38] that $(F_n^{(r,R)}/(b_n\sqrt{n}))$ satisfies an LDP on $\mathbb{R}$ of speed b_n^2 and good rate function

$$I_R(x) = \frac{x^2}{2Q_R\ell}.$$

In particular, this implies that for all $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log \mathbb{P}\left(\frac{|F_n^{(r,R)}|}{b_n\sqrt{n}} > \delta\right) \leq -\frac{\delta^2}{2Q_R\ell}, \tag{5.4.89}$$

and letting R go to infinity clearly leads to the end of the proof of (5.4.83). We are able to conclude now that $(M_n^{(r,R)}/(b_n\sqrt{n}))$ is an exponentially good approximation of $(M_n/(b_n\sqrt{n}))$. By application of Theorem 4.2.16 in [35], we find that $(M_n/(b_n\sqrt{n}))$ satisfies an LDP on $\mathbb{R}$ with speed b_n^2 and good rate function

$$\widetilde{J}(x) = \sup_{\delta>0} \liminf_{R\to\infty} \inf_{z\in B_{x,\delta}} J_R(z),$$

where J_R is given in (5.4.81) and $B_{x,\delta}$ denotes the ball $\{z : |z-x| < \delta\}$. The identification of the rate function $\widetilde{J} = J$, where J is given in (5.4.75) is done easily, which concludes the proof of Lemma 5.4.9. □

Remark 5.4.2. *If we suppose that* **(CL.1)** *holds with $a > 2$, then the exponential Lindeberg's condition in Lemma 5.4.8 is easier to establish. Indeed, using (5.4.84), it follows that*

$$\left(r\frac{\sqrt{n}}{b_n}\right)^{\eta} \sum_{k=1}^{n} X_{k-1}^2 \mathrm{I}_{\left\{|X_{k-1}|>r\frac{\sqrt{n}}{b_n}\right\}} \leq \sum_{k=1}^{n} |X_{k-1}|^{2+\eta} \leq \gamma \sum_{k=1}^{n} |V_k|^{2+\eta},$$

for n large enough and $\eta > 0$, leading to

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} X_{k-1}^2 \mathrm{I}_{\left\{|X_{k-1}|>r\frac{\sqrt{n}}{b_n}\right\}} > \delta\right) \leq \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} |V_k|^{2+\eta} > \frac{\delta}{\gamma}\left(r\frac{\sqrt{n}}{b_n}\right)^{\eta}\right).$$

Lemma 5.4.10. *Under* **(CL.1)**, **(CL.2)** *and* **(CL.3)**, *the sequence*

$$\left(\frac{1}{b_n\sqrt{n}}\begin{pmatrix} M_n \\ N_n \end{pmatrix}\right)_{n\geq 1}$$

satisfies an LDP on $\mathbb{R}^2$ with speed b_n^2 and good rate function

$$J(x) = \frac{1}{2\sigma^2} x' \Lambda^{-1} x \tag{5.4.90}$$

where Λ is given by (5.4.51).

Proof. We follow the same approach as in the proof of Lemma 5.4.9. We shall consider the 2-dimensional vector martingale $(Z_n)_{n\geq 0}$ defined in (5.4.47). In order to apply Theorem 5.4.7, we introduce the following truncation of the martingale $(Z_n)_{n\geq 0}$, for $r > 0$ and $R > 0$,

$$Z_n^{(r,R)} = \begin{pmatrix} M_n^{(r,R)} \\ N_n^{(r,R)} \end{pmatrix}$$

where $M_n^{(r,R)}$ is given in (5.4.76) and where $N_n^{(r,R)}$ is defined in the same manner, that is, for all $n \geq 2$,

$$N_n^{(r,R)} = \sum_{k=2}^{n} X_{k-2}^{(r)} V_k^{(R)} \tag{5.4.91}$$

with $X_n^{(r)}$ and $V_n^{(R)}$ given by (5.4.77). The exponential convergence (5.4.52) still holds, by virtue of Lemma 5.4.6, which immediately implies hypothesis (5.4.69). On top of that, Lemma 5.4.8 ensures that, for all $r > 0$,

$$\frac{1}{n} \sum_{k=0}^{n} X_k^2 \mathrm{I}_{\left\{|X_k| > r \frac{\sqrt{n}}{b_n}\right\}} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0, \tag{5.4.92}$$

justifying hypothesis (5.4.71). Via Theorem 5.4.7, $(Z_n^{(r,R)}/(b_n\sqrt{n}))$ satisfies an LDP on $\mathbb{R}^2$ with speed b_n^2 and good rate function J_R given by

$$J_R(x) = \frac{1}{2\sigma_R^2} x' \Lambda^{-1} x. \tag{5.4.93}$$

Finally, it is straightforward to prove that $(Z_n^{(r,R)}/(b_n\sqrt{n}))$ is an exponentially good approximation of $(Z_n/(b_n\sqrt{n}))$. By application of Theorem 4.2.16 in [35], we deduce that $(Z_n/(b_n\sqrt{n}))$ satisfies an LDP on $\mathbb{R}^2$ with speed b_n^2 and good rate function given by

$$\widetilde{J}(x) = \sup_{\delta > 0} \liminf_{R \to \infty} \inf_{z \in B_{x,\delta}} J_R(z),$$

where J_R is given in (5.4.93) and $B_{x,\delta}$ denotes the ball $\{z : |z - x| < \delta\}$. The identification of the rate function $\widetilde{J} = J$ is done easily, which concludes the proof of Lemma 5.4.10. ☐

Proofs of Theorem 5.3.1, Theorem 5.3.2 and Theorem 5.3.3. The residuals appearing in the decompositions (5.4.24), (5.4.46) and (5.4.56) still converge exponentially to zero under **(CL.1)**, **(CL.2)** and **(CL.3)**, with speed b_n^2, as it was already proved. Therefore, for a better readability, we may skip the most accessible parts of these proofs whose development merely consists in following the same lines as those in the proofs of Theorem 5.2.1, Theorem 5.2.2 and Theorem 5.2.3, taking advantage of Lemma 5.4.9 and Lemma 5.4.10, and applying the contraction principle given e.g. in [35]. □

Appendix 5.A On the moderate deviations without serial correlation

The purpose of this section is to provide a moderate deviation principle for the empirical mean and the empirical covariance under the Chen-Ledoux type condition on the noise. We generalize the results of Mas and Menneteau [77], Miao and Shen [84], where the exponential integrability of the noise is imposed. One can see our results as an extension to the autoregressive case of the beautiful characterization of moderate deviations for i.i.d. case of Ledoux [73], see also Chen [27] and Eichelsbacher and Löwe [53]. Consider the linear autoregressive model in $\mathbb{R}$

$$X_n = \theta X_{n-1} + V_n$$

where θ is unknown, $(V_n)_{n\geq 1}$ is a sequence of $\mathbb{R}$-valued centered i.i.d. random variables representing the noise. This is a particular case of the model (5.1.1) when the serial correlation $\rho = 0$. Here, we are interested in the asymptotic behavior of the two empirical average

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k, \qquad C_n = \frac{1}{n}\sum_{k=1}^{n} X_k^2.$$

Proposition 5.A.1. *Suppose that the following Assumptions are satisfied :*

$$\limsup_{n\to\infty} \frac{1}{b_n^2} \log n\mathbb{P}\Big(V_1 > b_n\sqrt{n}\Big) = -\infty; \tag{5.A.1}$$

and

$$\frac{X_0}{b_n\sqrt{n}} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0. \tag{5.A.2}$$

Then $(\sqrt{n}\overline{X}_n/b_n)$ satisfies the large deviation principle on $\mathbb{R}$ with speed b_n^2 and rate function

$$I(x) = (1-\theta)^2 \frac{x^2}{2\mathbb{E}(V_1^2)}. \tag{5.A.3}$$

Remark 5.A.1. *When $b_n = n^\alpha$ with $0 < \alpha < 1/2$, the Assumption (5.A.1) is equivalent to there exists $\delta > 0$ and $0 < \kappa < 1$ such that $\mathbb{E}(e^{\delta V_1^\kappa}) < \infty$, which is weaker than the assumption $\mathbb{E}(e^{\delta V_1}) < \infty$ imposed in the previous works.*

Proof. We have the following decomposition

$$\frac{\sqrt{n}}{b_n}\overline{X}_n = (1-\theta)^{-1}\frac{\sqrt{n}}{b_n}\overline{V}_n + \theta(1-\theta)^{-1}\frac{1}{\sqrt{n}b_n}\left(X_0 - X_n\right),$$

where $\overline{V}_n = \frac{1}{n}\sum_{k=1}^n V_k$. By the moderate deviations result for means of i.i.d. random variables under the Assumption (5.A.1) (see for example Eichelsbacher and Löwe [53]) we deduce that $(\sqrt{n}\overline{V}_n/b_n)$ follows the large deviation principle on $\mathbb{R}$ with speed b_n^2 and rate function given by $I_1(x) = \frac{x^2}{2\mathbb{E}(V_1^2)}$. It is easy to see that under Assumption (5.A.1) and (5.A.2)

$$\frac{\sqrt{n}}{b_n}\overline{X}_n \underset{b_n^2}{\overset{\text{superexp}}{\sim}} (1-\theta)^{-1}\frac{\sqrt{n}}{b_n}\overline{V}_n.$$

By the contraction principle (see for example Dembo and Zeitouni [35]), we get that $(\sqrt{n}\overline{X}_n/b_n)$ follows the large deviation principle on $\mathbb{R}$ with speed b_n^2 and rate function given by (5.A.3). □

Proposition 5.A.2. *Assume that Assumptions* **(CL.1)** *and* **(CL.3)** *are satisfied. Then $(\sqrt{n}(C_n - \mathbb{E}(C_n))/b_n)$ satisfies the large deviation on $\mathbb{R}$ with speed b_n^2 and rate function*

$$J(x) = (1-\theta^2)^2\frac{x^2}{2\sigma^2}, \qquad \text{where} \quad \sigma^2 = \frac{4\theta^2\mathbb{E}(V_1^2)^2}{1-\theta^2} + \operatorname{Var}(V_1^2). \tag{5.A.4}$$

Proof. Since $X_k^2 = \theta^2 X_{k-1}^2 + 2\theta X_{k-1}V_k + V_k^2$, we deduce that

$$C_n = \frac{1}{1-\theta^2}\frac{1}{n}\sum_{k=1}^n(2\theta X_{k-1}V_k + V_k^2 - \mathbb{E}(V_1^2)) + \frac{1}{1-\theta^2}\mathbb{E}(V_1^2) + \frac{\theta^2}{1-\theta^2}\frac{1}{n}(X_0^2 - X_n^2)$$

We denote $u_k = 2\theta X_{k-1}V_k + V_k^2 - \mathbb{E}(V_1^2)$, so

$$\begin{aligned}\frac{\sqrt{n}}{b_n}(C_n - \mathbb{E}(C_n)) &= \frac{1}{1-\theta^2}\frac{1}{b_n\sqrt{n}}\sum_{k=1}^n u_k + \frac{\theta^2}{1-\theta^2}\frac{1}{b_n\sqrt{n}}(X_0^2 - X_n^2)\\ &\quad - \frac{\theta^2(1-\theta^{2n})}{1-\theta^2}\frac{1}{b_n\sqrt{n}}\mathbb{E}(X_0^2) + \frac{\theta^2(1-\theta^{2n})}{(1-\theta^2)^2}\frac{1}{b_n\sqrt{n}}\mathbb{E}(V_1^2).\end{aligned}$$

It is easy to see that, under the Assumptions **(CL.1)** and **(CL.3)**, we have that

$$\frac{\sqrt{n}}{b_n}(C_n - \mathbb{E}(C_n)) \underset{b_n^2}{\overset{\text{superexp}}{\sim}} \frac{1}{1-\theta^2}\frac{1}{b_n\sqrt{n}}\sum_{k=1}^n u_k.$$

The sequence $(G_n = \sum_{k=1}^{n} u_k)_{n\geq 1}$ is a square-integrable martingale, with predictable quadratic variation given by

$$\langle G\rangle_n = 4\theta^2\mathbb{E}(V_1^2)\sum_{k=1}^{n} X_{k-1}^2 + 4\theta\mathbb{E}(V_1^3)\sum_{k=1}^{n} X_{k-1} + n\mathrm{Var}(V_1^2).$$

We can see that

$$\frac{1}{n}\sum_{k=1}^{n} X_k \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} 0 \quad \text{and} \quad \frac{1}{n}\sum_{k=1}^{n} X_k^2 \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} \frac{\mathbb{E}(V_1^2)}{1-\theta^2}.$$

So we obtain

$$\frac{\langle G\rangle_n}{n} \overset{\text{superexp}}{\underset{b_n^2}{\Longrightarrow}} \sigma^2 = \frac{4\theta^2\mathbb{E}(V_1^2)^2}{1-\theta^2} + \mathrm{Var}(V_1^2).$$

We follow the same steps as in the previous section (by using truncation of the martinagle (G_n) and Puhalskii's result) to prove that $(G_n/b_n\sqrt{n})$ satisfies the large deviation principle on $\mathbb{R}$ with speed b_n^2 and rate function given by $J_1(x) = \frac{x^2}{2\sigma^2}$. See the previous section for more details. □

Remark 5.A.2. *Under the Assumptions* **(CL.1)** *and* **(CL.3)***, we obtain also the large deviation principle for the sequence* $(\sum_{k=1}^{n}(X_kX_{k+l} - \mathbb{E}(X_kX_{k+l}))/(b_n\sqrt{n}))_{n\geq 1}$*, for all* $l \geq 0$*.*

Chapitre 6

Explicit sub exponential rates of convergence for continuous time Markov processes

6.1 Introduction and definitions

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, (\mathbb{P}_x)_{x\in X})$ be a Markov family on a locally compact separable metric space X endowed with its Borel σ-field $\mathcal{B}(X)$: $(\Omega, \mathcal{F})$ is a measurable space, $(X_t)_{t\geq 0}$ is a Markov process with respect to the filtration $(\mathcal{F}_t)_{\geq 0}$ and $\mathbb{P}_x$ (resp. $\mathbb{E}_x$) denotes the canonical probability (resp. expectation) associated to the Markov process with initial distribution the point mass at x. Throughout this chapter, the process is assumed to be a time-homogeneous strong Markov process with cad-lag paths, and we denote by $(P_t)_{t\geq 0}$ the associated semigroup on $(X, \mathcal{B}(X))$.

There are two important notions in the study of long time behavior of Markov processes: small sets and hitting times and their integrability properties. It is largely due to the following heuristics: a small set is a set with a nice behavior, i.e. a local version of a Doeblin condition, and integrability of the hitting times of this small set measures the speed of return of the process to this nice set. Let us be more precise with the following definitions.

For any $t^* > 0$ and any closed set $C \in \mathcal{B}(X)$, let $\tau_C^{t^*} = \inf\{t \geq t^*, X_t \in C\}$ be the hitting time on C delayed by t^*.

We say that a non-empty subset $C \subseteq X$ is (t^*, ε)-small (or simply small) , for a positive time t^* and $\varepsilon > 0$, if there exists a probability measure $Q(\cdot)$ on X satisfying the minorization condition

$$P_{t^*}(x, \cdot) \geq \varepsilon Q(\cdot) \ \ \forall x \in C. \tag{6.1.1}$$

For completeness, let us give also the definition of a petite set. The subset C is ν_a-petite (or simply petite) if there exists a probability measure a on the Borel σ-field of $[0, +\infty)$ and a non-trivial σ-finite measure ν_a on $\mathcal{B}(X)$ such that

$$\int_0^\infty P_t(x,\cdot)a(dt) \geq \nu_a(\cdot) \ \ \forall x \in C.$$

One can thus see that every small set is a petite set.

Long time behavior of Markov process has been studied a lot through various set of techniques, but mainly for exponential speed of decay, such as coupling, control of hitting times and functional inequalities. It is of course a too long story to be exhaustive but let us refer to [87, 81, 85, 46] for Markov chains, [69, 82, 83, 95, 92, 48] for Markov processes via hitting times and Lyapunov conditions and [37, 2, 7, 8] for Markov processes via functional inequalities (Poincaré inequality, logarithmic Sobolev inequality). The case of sub exponential speed of decay with practical conditions has been considered only recently and only partial results exist, first for Markov chains [47] via Lyapunov type conditions, and extended to continuous time Markov process in [43]. See also [108] for a coupling approach and [25] for functional inequalities (weak Poincaré inequality). If usually functional inequalities usually provide quantitative estimates, even poor due to the difficulty to estimate precisely say the spectral gap constant, the hitting times approach of [43] furnishes speed of decay but not the constant multiplicative of this speed, so that the results are difficult to use in practice. It is the main goal of this chapter: give quantitative estimates under easy to verify conditions. Our main assumptions will be Lyapunov conditions and small set conditions.

Let us consider the following Lyapunov (or drift) condition towards a closed petite set C.

We will say that a Φ-Lyapunov condition holds if there exist a closed petite set C, a cad-lag function $V : X \longrightarrow [1,\infty)$, an increasing differentiable concave positive function $\Phi : [1,\infty) \longrightarrow (0,\infty)$ and a constant $b < \infty$ such that for any $s \geq 0$, $x \in X$,

$$\mathbb{E}_x\Big[V(X_s)\Big] + \mathbb{E}_x\left[\int_0^s \Phi \circ V(X_u)du\right] \leq V(x) + b\mathbb{E}_x\left[\int_0^s 1_C(X_u)du\right]. \qquad (6.1.2)$$

note that (6.1.2) is equivalent to the condition that the functional

$$s \longmapsto V(X_s) - V(X_0) + \int_0^s \Phi \circ V(X_u)du - b\int_0^s 1_C(X_u)du$$

is for all $x \in X$ a $\mathbb{P}_x$-supermartingale with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$.

Stability and subgeometric rates of convergence have been studied in [43] under the main hypothesis of a $\Phi-$Lyapunov condition. More precisely, they showed the following.

Proposition 6.1.1 ([43], Proposition 3.9 and Theorem 3.10).

Assume that a $\Phi-$Lyapunov condition holds and $\sup_C V < \infty$. Then the process is positive Harris recurrent with an invariant probability measure π such that $\pi(\Phi\circ$

$V) < \infty$. If moreover some skeleton chain is irreducible and $\lim_{+\infty} \Phi' = 0$, then there exists a finite constant c such that for all $t > 0$ and all $x \in X$,

$$\Phi \circ H_\Phi^{-1}(t) \|P_t(x, \cdot) - \pi(\cdot)\|_{TV} \le cV(x) \qquad (6.1.3)$$

where the function H_Φ is defined by

$$H_\Phi(u) = \int_1^u \frac{ds}{\Phi(s)} \quad \forall u \ge 1.$$

Note that the inequality obtained in [43] is more general than (6.1.3). More precisely, let Λ_0 denote the class of the measurable and non-decreasing functions $r : [0, +\infty) \to [2, +\infty)$ such that $\log r(t)/t \downarrow 0$ as $t \to +\infty$. Let Λ denote the class of positive measurable functions $\overline{r}$ such that for some $r \in \Lambda_0$,

$$0 < \liminf_t \frac{\overline{r}(t)}{r(t)} \le \limsup_t \frac{\overline{r}(t)}{r(t)} < \infty.$$

Then there exists a positive constant c such that for $r(t) \le \Phi \circ H_\Phi^{-1}(t)$

$$r(t)\|P_t(x, \cdot) - \pi(\cdot)\|_f \le cV(x),$$

where $r \in \Lambda$, for a signed measure μ, $\|\mu\|_f = \sup_{|g| \le f} \mu(g)$ and $f : X \to [0, \infty)$ a measurable function f linked to the choice of r (via inverse Young function).

Unfortunately, the constant c which appears in (6.1.3) is not explicit. Our main goal is thus to obtain quantitative bounds for the subgeometric convergence rates of the process to its stationary probability distribution π. More precisely, to determine a function $g : X \to [0, \infty)$, which can be computed explicitly, such that

$$r(t)\|P_t(x, \cdot) - \pi(\cdot)\|_{TV} \le g(x) \quad \forall x \in X, \qquad (6.1.4)$$

where $r \in \Lambda$.

Lyapunov conditions

As explained in [43], the $\Phi-$Lyapunov condition is not easy to derive. It is thus important to provide a sufficient condition which is more tractable. For this, we use the usual criteria based on extended generator. Let $\mathcal{D}(\mathcal{A})$ denote the set of measurable functions $f : X \to \mathbb{R}$ with the following property: there exists a measurable function $h : X \to \mathbb{R}$ such that the function $t \mapsto h(X_t)$ is integrable $\mathbb{P}_x-$a.s. for each $x \in X$ and the process

$$t \mapsto f(X_t) - f(X_0) - \int_0^t h(X_s) ds$$

is a $\mathbb{P}_x-$local martingale for all x. Then we write $h = \mathcal{A}f$, and f is said to be in the domain of the extended generator $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$. We then have.

Theorem 6.1.2 ([43], Theorem 3.11). *Assume that there exist a closed petite set C, a cad-lag function $V : X \to [1, +\infty)$ with $V \in \mathcal{D}(\mathcal{A})$, $\Phi \in \mathcal{C}$, and a constant $b < \infty$ such that for all $x \in X$,*

$$\mathcal{A}V(x) \leq -\Phi \circ V(x) + b1_C(x). \tag{6.1.5}$$

then a $\Phi-$Lyapunov condition holds.

The Lyapunov conditions are the most useful one to verify in practice, even if of course they can be tricky to exhibit (see for example the oscillator case [68]). In particular, we will be able to construct such a Lyapunov conditions for a coupling process which will be the key tool for our main result. Due to this last result we will use now the following formal definition even if the generator applied to the function makes no sense (remembering in this case that the first definition is in charge):

$$\mathcal{A}V(x) \leq -\Phi \circ V(x) + b1_C(x).$$

As explained before, the key tool to establish speed of convergence to equilibrium is the proof of integrability of hitting time, we thus recall the following result which will be useful for the proof of our results.

Theorem 6.1.3 ([43], Theorem 3.1). *Assume that a $\Phi-$Lyapunov condition holds.*
i)For all $x \in X$ and $t^* > 0$,

$$\mathbb{E}_x \left[\int_0^{\tau_C^{t^*}} \Phi \circ V(X_s) ds \right] \leq V(x) - 1 + bt^*.$$

ii)For all $x \in X$ and $t^* \geq 0$,

$$\mathbb{E}_x \left[\int_0^{\tau_C^{t^*}} \Phi \circ H_\Phi^{-1}(s) ds \right] \leq V(x) - 1 + \frac{b}{\Phi(1)} \int_0^{t^*} \Phi \circ H_\Phi^{-1}(s) ds.$$

Note that we will provide in Appendix a shorter (and we believe clearer) proof of a modified version of this result, however perfectly fitted for our use.

The chapter is then organized as follows. In section 6.2 we state our main results. This includes quantitative bounds of the distance of $\mathcal{L}(X_t)$ and ergodic average laws $\int_0^t P(X_s \in \cdot) ds$ to the stationarity. Section 6.3 is devoted to the proof of our main results. Section 6.4 focuses on some examples. The appendix is dedicated to a short proof of the integrability of Hitting times under Lyapunov conditions.

6.2 Main results

6.2.1 Notation and a preliminary lemma

Let us begin by some notation. Define

$$\mathcal{C} = \Big\{\Phi : [1,\infty) \to \mathbb{R}^+ : \Phi \text{ is increasing, differentiable, concave, } \lim_{t\to+\infty} \Phi(t) = +\infty, \quad \lim_{t\to+\infty} \Phi'(t) = 0, \log \Phi(t)/t \downarrow 0 \text{ as } t \to +\infty\Big\}.$$

Let us notice that the condition $\log \Phi(t)/t \downarrow 0$ as $t \to +\infty$ it is of importance, since we are working on sub-exponential ergodicity.

For $\Phi \in \mathcal{C}$, let

$$H_\Phi(u) = \int_1^u \frac{ds}{\Phi(s)}, \quad u \geq 1.$$

It's easy to see that

$$(H_\Phi^{-1})' = \Phi \circ H_\Phi^{-1}, (H_\Phi^{-1})'' = \Phi' \circ H_\Phi^{-1} \cdot \Phi \circ H_\Phi^{-1}, (\ln \Phi \circ H_\Phi^{-1})' = \Phi' \circ H_\Phi^{-1},$$
$$\lim_{s\to+\infty} \frac{\ln H_\Phi^{-1}(s)}{s} = \lim_{s\to+\infty} \frac{\Phi \circ H_\Phi^{-1}(s)}{H_\Phi^{-1}(s)} = \lim_{t\to+\infty} \Phi'(t) = 0.$$

Then H_Φ^{-1} is increasing, convex, sub-geometric. By the log-concavity of $\Phi \circ H_\Phi^{-1}$,

$$z \to \frac{\Phi \circ H_\Phi^{-1}(z+s)}{\Phi \circ H_\Phi^{-1}(z)} \quad \text{is non-increasing for } s > 0.$$

Let us first state the following elementary lemma.

Lemma 6.2.1. *$H_\Phi^{-1}(\lambda a) = H_{\lambda\Phi}^{-1}(a)$, $H_\Phi^{-1}(a+b) \leq H_\Phi^{-1}(a) H_\Phi^{-1}(b)$, for any $\lambda > 0$, $a \geq 0, b \geq 0$ and integer $n \geq 1$.*

6.2.2 Quantitative subexponential ergodicity

We are now in position to present our main results.

Theorem 6.2.2. *Given a Markov process (X_t) with semigroup (P_t), assumethat a $\Phi-$Lyapunov condition holds and $\sup_C V < \infty$, $\Phi \in \mathcal{C}$, $C \in X$ is (t^*, ϵ)-small, for some positive time t^*, and $\epsilon > 0$, stationary distribution is π, then if $\pi(V) < \infty$*

$$||\mathcal{L}(X_s) - \pi||_{TV} \leq \sum_{0 \leq n, s \geq nt^*} (1-\epsilon)^n \widetilde{A}_1 \frac{H_\Phi^{-1}(\lambda \widetilde{A}_0 n)}{H_\Phi^{-1}(\lambda(s-t^*))} \leq \frac{\widetilde{A}_1 \widetilde{A}_{\epsilon,\Phi}}{H_\Phi^{-1}(\lambda(s-t^*))}$$

where $\widetilde{A}_{\epsilon,\Phi}$ *will be given later and*

$$\widetilde{A}_0 =: \frac{2}{\lambda\Phi(1)} \sup_{x\in C}\left\{V(x) - 1 + \frac{b}{\Phi(1)}\int_0^{t^*} \Phi\circ H_\Phi^{-1}(\lambda s)ds)\right\} < \infty,$$

$$\widetilde{A}_1 = \max\left\{\frac{\mathbb{E}\Big[V(X_0)\Big] + \pi(V) - 2}{H_\Phi^{-1}(\lambda\widetilde{A}_0)}, 1\right\},$$

and

$$d_0 = \inf_{x\notin C} V(x), \quad 0 < \lambda \leq 1 - \frac{b}{\Phi(d_0)}, \quad \widetilde{A}_{\epsilon,\Phi} < \infty.$$

In general we have,

$$||\mathcal{L}(X_s^x) - \mathcal{L}(X_s^y)||_{TV} \leq \frac{A(V(x)+V(y))}{H_\Phi^{-1}(\lambda(s-t^*))}$$

for A *given by*

$$A = \max\left\{\frac{1}{V(x)+V(y)}, \frac{1}{H_\Phi(\lambda\widetilde{A}_0)}\right\} \times \widetilde{A}_{\varepsilon,\Phi}.$$

In particular, if $\pi(V)$ *is finite then*

$$||\mathcal{L}(X_s^x) - \pi||_{TV} \leq \frac{A(V(x)+\pi(V))}{H_\Phi^{-1}(\lambda(s-t^*))}.$$

Remark 6.2.3. *One has easily using the Lyapunov condition that* $\pi(\Phi(V)) < \infty$ *and* $\pi(V)$ *is not controlled. However, if we suppose that* $\pi(\psi(V)) < \infty$ *(eventually use* Φ*) for* $\Phi(x) \leq \psi(x) \leq x$ *for large* x *and* ψ *concave, one can get*

$$||\mathcal{L}(X_s^x) - \pi||_{TV} \leq \frac{A(V(x))}{H_\Phi^{-1}(\lambda(s-t^*))} + CR(s)$$

for some constant C *and* R *decaying to zero and being solution to the minimization in* R *of*

$$\frac{R}{\psi(R)H_\Phi^{-1}(t)} + \frac{1}{\psi(R)}.$$

Let us state now a result related to shift-coupling [93], which provides for bounds on the ergodic averages of distances to stationary distributions.

Theorem 6.2.4. *Given a Markov process* (X_t) *with semigroup* (P_t) *and stationary distribution* $\pi(.)$*, assume that a* $\Phi-$*Lyapunov condition holds and* $\sup_C V < \infty$*,* $C \in \mathcal{X}$ *is* (t^*, ε)*-small for some positive time* t^* *and* $\varepsilon > 0$*. Then for* $t > 0$

$$\left\|\frac{1}{t}\int_0^t \mathbb{P}(X_s \in .)ds - \pi(.)\right\| \leq \frac{1}{t}\left(2t^* + c + c'\right),$$

where

$$c = A_1 A_{\varepsilon,\Phi} \int_{t^*}^{+\infty} \frac{1}{H_\Phi^{-1}(s-t^*)} ds, \quad c' = A_1' A_{\varepsilon,\Phi} \int_{t^*}^{+\infty} \frac{1}{H_\Phi^{-1}(s-t^*)} ds,$$

with

$$A_1 = \max\left\{ \frac{\mathbb{E}\Big[V(X_0)\Big] - 1}{H_\Phi^{-1}(A_0)}, 1 \right\}, \quad A_1' = \max\left\{ \frac{\mathbb{E}_\pi\big[V\big] - 1}{H_\Phi^{-1}(A_0)}, 1 \right\} \quad \textit{and} \quad A_{\varepsilon,\Phi} < \infty.$$

Proof. Theorem 6.2.2 and 6.2.4 are proven in the next section. □

One interesting point, compared to the one of [92] where they only treated the exponential case, is that we obtain the same speed, i.e. t^{-1}, but valid as soon as $1/H_\Phi^{-1}$ is integrable at infinity. We may now a little bit further in the details of the constants appearing in our main results.

Explicit constant $A_{\epsilon,\Phi}$ **(**$\widetilde{A}_{\epsilon,\Phi}$**)**

For any $0 < a_\epsilon < \epsilon$, We choose α_c, such that

$$(1-\epsilon)^n H_\Phi^{-1}(cn) \le \alpha_c (1-a_\epsilon)^n \quad \text{for } n \ge 1.$$

In fact, we can choose $\alpha_c = \sup_{n\ge1} \frac{H_\Phi^{-1}(cn)}{(\frac{1-a_\epsilon}{1-\epsilon})^n} < \infty$. Then we have

$$\sum_{n=0}^{\infty} (1-\epsilon)^n H_\Phi^{-1}(cn) \le \alpha_c \sum_{n=0}^{\infty} (1-a_\epsilon)^n = \frac{\alpha_c}{a_\epsilon}.$$

So we can get more explicit expressions:

$$A_{\epsilon,\Phi} = \frac{\sup_{n\ge1} \frac{H_\Phi^{-1}(A_0 n)}{(\frac{1-a_\epsilon}{1-\epsilon})^n}}{a_\epsilon}, \qquad \widetilde{A}_{\epsilon,\Phi} = \frac{\sup_{n\ge1} \frac{H_\Phi^{-1}(\lambda \widetilde{A}_0 n)}{(\frac{1-a_\epsilon}{1-\epsilon})^n}}{a_\epsilon}$$

(I) $\Phi(x) = x^\alpha$, $0 < \alpha < 1$. $H_\Phi^{-1}(x) = ((1-\alpha)x + 1)^{\frac{1}{1-\alpha}}$,

$$A_{\epsilon,\Phi} = \frac{\sup_{n\ge1} \frac{((1-\alpha)A_0 n + 1)^{\frac{1}{1-\alpha}}}{(\frac{1-a_\epsilon}{1-\epsilon})^n}}{a_\epsilon}$$

$$\widetilde{A}_{\epsilon,\Phi} = \frac{\sup_{n\ge1} \frac{((1-\alpha)\lambda \widetilde{A}_0 n + 1)^{\frac{1}{1-\alpha}}}{(\frac{1-a_\epsilon}{1-\epsilon})^n}}{a_\epsilon}.$$

(II) $\Phi(x) = \frac{x}{(\ln x)^\alpha}$, $\alpha > 0$. $H_\Phi^{-1}(x) = \exp((1+\alpha)x)^{\frac{1}{1+\alpha}}$, and we get the similar expressions.

6.3 Useful results and proof of our main Theorems

6.3.1 Proof of Theorem 6.2.2

Our proof relies mainly on coupling method as presented for example in Roberts-Rosenthal [92]. Namely we build two Markov processes which will have some coupling property, which as usual give some upper bound on the total variation norm of the difference of the law of the two processes via a control of the tail of the coupling time.

To control this coupling time we need two tools. The minorization condition is needed to validate the coupling construction. The second important tool is a good control of expectation of subexponential moments of the coupling time. We will show that we are able to do so using the Lyapunov condition.

We will show first, to simplify the arguments, how to control the tail of regeneration time of some process with a minorization condition, and we hope that it will help the reader to understand the proof for the coupled process.

Trajectory construction and control of the tail of regeneration time

$C \in X$ is (t^*, ϵ)-small in $\mathbf{D}(\mathbf{C}, \mathbf{V}, \Phi, b)$, that means $P^{t^*}(x, \cdot) \geq \epsilon Q(\cdot)$ for $x \in C$. we shall use the following construction in this chapter, which is from Roberts and Rosenthal ([92], proof of Lemma 6). Given a sequence of i.i.d random variables $Z_1, Z_2, ..., Z_i \sim B(1, \epsilon)$, We construct X_t and the random stopping time $T = \inf\left\{\tau_i^{t^*} + t^*, X_{\tau_i^{t^*}} \in C, Z_i = 1\right\}$. let $\tau_1^{t^*} = \tau_C^0$, $\tau_i^{t^*} = \inf\left\{t \geq \tau_{i-1}^{t^*} + t^*, X_t \in C\right\}, i \geq 2$.

- 1. if $Z_1 = 1$, $\tau_1^{t^*} = \tau_C^0$ is the first time the process $\{X_t\}$ is in the set C, set $X_{\tau_1^{t^*}+t^*} \sim Q(\cdot)$;

  2. if $Z_1 = 0$, set $X_{\tau_1^{t^*}+t^*} \sim \frac{1}{1-\epsilon}(P_{t^*}(X_{\tau_1^{t^*}}, \cdot) - \epsilon Q(\cdot))$.

  in either case above, fill in X_t for $\tau_1^{t^*} < t < \tau_1^{t^*} + t^*$ from the appropriate conditional distributions.

- similarly, for $i \geq 2$, if $Z_i = 1$, set $X_{\tau_i^{t^*}+t^*} \sim Q(\cdot)$; otherwise $Z_i = 0$, set $X_{\tau_i^{t^*}+t^*} \sim \frac{1}{1-\epsilon}(P_{t^*}(X_{\tau_i^{t^*}}, \cdot) - \epsilon Q(\cdot))$.

Under this construction, we get the process X_t for all times $t > 0$ with the semigroup (P^t), and a random stopping time T with $X_T \sim Q(\cdot)$

Lemma 6.3.1. *Assume that a $\Phi-$Lyapunov condition holds and $\sup_C V < \infty$, $\Phi \in \mathcal{C}$, $C \in X$ is (t^*, ϵ)-small, for some positive time t^*, We construct the processes $\{X_t\}$ and a random stopping time T as above, then we have*

$$\mathbb{P}(T > s) \leq \sum_{0 \leq n, s \geq nt^*} (1-\epsilon)^n \frac{A_1 H_\Phi^{-1}(A_0 n)}{H_\Phi^{-1}(s - t*)} \leq \frac{A_1 A_{\epsilon,\Phi}}{H_\Phi^{-1}(s - t*)}, \qquad (6.3.1)$$

where $A_0 =: \frac{1}{\Phi(1)} \sup_{x\in C} \left\{ V(x) - 1 + \frac{b}{\Phi(1)} \int_0^{t^*} \Phi \circ H_\Phi^{-1}(s)ds) \right\} < \infty,$

$$A_1 = \max\left\{ \frac{\mathbb{E}[V(X_0)] - 1}{H_\Phi^{-1}(A_0)}, 1 \right\} \text{ and } A_{\epsilon,\Phi} < \infty$$

Proof. We define $N_s = \max\left\{ i : \tau_i^{t^*} \leq s \right\}$. By the construction above,

$$\begin{aligned}
\mathbb{P}\big(T > s\big) &\leq \sum_{0\leq n, s\geq nt^*} \mathbb{P}\Big(T > s, N_{s-t*} \leq n\Big) \\
&= \sum_{0\leq n, s\geq nt^*} \mathbb{P}\Big(T > s \big| N_{s-t*} \leq n\Big) \mathbb{P}\Big(N_{s-t*} \leq n\Big) \\
&\leq \sum_{0\leq n, s\geq nt^*} (1-\epsilon)^n P\Big(N_{s-t*} \leq n\Big)
\end{aligned}$$

Setting $D_1 = \tau_1^{t^*}$ and $D_i = \tau_i^{t^*} - \tau_{i-1}^{t^*}$ for $i \geq 2$, it's easy to see that,

$$\mathbb{E}\Big[H_\Phi^{-1}(D_1)\Big] = \mathbb{E}\Big[H_\Phi^{-1}(\tau_1^{t^*})\Big] \leq \mathbb{E}\Big[V(X_0)\Big] - 1 \quad (\text{ by Theorem 6.1.3 } (i))$$

for $i \geq 2$, let

$$A_0 =: \frac{1}{\Phi(1)} \sup_{x\in C} \left\{ V(x) - 1 + \frac{b}{\Phi(1)} \int_0^{t^*} \Phi \circ H_\Phi^{-1}(s)ds \right\} < \infty.$$

We have

$$\begin{aligned}
\mathbb{E}\Big[H_\Phi^{-1}(D_i)\Big] &= \mathbb{E}\left[\int_0^{\tau_i^{t^*} - \tau_{i-1}^{t^*}} \Phi \circ H_\Phi^{-1}(s)ds \right] \\
&= \mathbb{E}\left[\mathbb{E}^{X_{\tau_{i-1}^{t^*}}} \int_0^{\tau_C^{t^*}} \Phi \circ H_\Phi^{-1}(s)ds \right] \\
&\leq \mathbb{E}\left[V(X_{\tau_{i-1}^{t^*}}) - 1 + \frac{b}{\Phi(1)} \int_0^{t^*} \Phi \circ H_\Phi^{-1}(s)ds \right] \\
&\leq A_0 \Phi(1)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left[H_\Phi^{-1}\left(\tau_i^{t^*}\right)\right] - \mathbb{E}\left[H_\Phi^{-1}\left(\tau_{i-1}^{t^*}\right)\right] &= \mathbb{E}\left[\int_{\tau_{i-1}^{t^*}}^{\tau_i^{t^*}} \Phi \circ H_\Phi^{-1}(s)ds \right] \\
&= \mathbb{E}\left[\int_0^{\tau_i^{t^*} - \tau_{i-1}^{t^*}} \Phi \circ H_\Phi^{-1}\left(s + \tau_{i-1}^{t^*}\right) ds \right] \\
&\leq \mathbb{E}\left[\int_0^\infty \Phi \circ H_\Phi^{-1}(s) \left(\frac{\Phi \circ H_\Phi^{-1}(\tau_{i-1}^{t^*})}{\Phi(1)} \right) 1_{s \leq \tau_C^{t^*} \circ \theta_{\tau_{i-1}^{t^*}}} ds \right]
\end{aligned}$$

$$\left(z \rightarrow \frac{\Phi \circ H_\Phi^{-1}(z+s)}{\Phi \circ H_\Phi^{-1}(z)} \text{is non-increasing for } s>0\right)$$
$$\leq \mathbb{E}\left[\frac{\Phi \circ H_\Phi^{-1}\left(\tau_{i-1}^{t^*}\right)}{\Phi(1)} \mathbb{E}^{X_{\tau_{i-1}^{t^*}}} \int_0^{\tau_C^{t^*}} \Phi \circ H_\Phi^{-1}(s) ds\right]$$
$$\leq \mathbb{E}\left[\frac{\Phi \circ H_\Phi^{-1}\left(\tau_{i-1}^{t^*}\right)}{\Phi(1)} \left(V(X_{\tau_{i-1}^{t^*}}) - 1 + \frac{b}{\Phi(1)} \int_0^{t^*} \Phi \circ H_\Phi^{-1}(s) ds\right)\right]$$
(by Theorem 6.1.3 (ii))
$$\leq A_0 \mathbb{E}\left[\Phi \circ H_\Phi^{-1}\left(\tau_{i-1}^{t^*}\right)\right]$$
$$\leq A_0 \Phi \circ \mathbb{E}\left[H_\Phi^{-1}\left(\tau_{i-1}^{t^*}\right)\right] \quad \text{(by the concavity of } \Phi)$$

Then we get

$$\mathbb{E}\left[H_\Phi^{-1}\left(\tau_i^{t^*}\right)\right] \leq \mathbb{E}\left[H_\Phi^{-1}\left(\tau_{i-1}^{t^*}\right)\right] + A_0 \Phi \circ \mathbb{E}\left[H_\Phi^{-1}\left(\tau_{i-1}^{t^*}\right)\right].$$

Using the following lemma 6.3.2, note $H_{A_1\Phi}^{-1}(n) = H_\Phi^{-1}(A_1 n)$, and let

$$A' = A_1 = \max\left\{\frac{\mathbb{E}\Big[V(X_0)\Big] - 1}{H_\Phi^{-1}(A_0)}, 1\right\},$$

we get

$$\mathbb{E}\left[H_\Phi^{-1}(\tau_n^{t^*})\right] \leq \quad A_1 H_\Phi^{-1}(A_0 n), \text{for} \quad n \geq 2.$$

For any positive integer $n \geq$2, we can get the estimation from Markov's inequality,

$$\begin{aligned}
\mathbb{P}\Big(N_s \leq n\Big) &= \mathbb{P}\left(\sum_{i=1}^n D_i \geq s\right) \\
&= \mathbb{P}\left(H_\Phi^{-1}\left(\sum_{i=1}^n D_i\right) \geq H_\Phi^{-1}(s)\right) \\
&\leq \frac{\mathbb{E}\Big[H_\Phi^{-1}(\sum_{i=1}^n D_i)\Big]}{H_\Phi^{-1}(s)} \\
&= \frac{\mathbb{E}\Big[H_\Phi^{-1}\left(\tau_n^{t^*}\right)\Big]}{H_\Phi^{-1}(s)} \\
&\leq \frac{A_1 H_\Phi^{-1}(A_0 n)}{H_\Phi^{-1}(s)}
\end{aligned}$$

This estimation combined with formula 6.3.2, we get

$$\mathbb{P}(T > s) \leq \sum_{0 \leq n, s \geq nt^*} (1-\epsilon)^n \frac{A_1 H_\Phi^{-1}(A_0 n)}{H_\Phi^{-1}(s - t*)}.$$

Noting that $\lim_{s\to+\infty}\frac{\ln H_\Phi^{-1}(s)}{s}=0$, there exits a constant $A_{\epsilon,\Phi}$, such that

$$\sum_{0\le n,s\ge nt^*}(1-\epsilon)^n H_\Phi^{-1}(A_0 n)\le A_{\epsilon,\Phi}<+\infty,$$

then we finish the proof of the lemma. □

Lemma 6.3.2. *$\varphi\in\mathcal{C}$, if $a_{n+1}\le c\varphi(a_n)+a_n$, where $\{a_n\}$ is a non-negative increasing sequence, then there exists a constant $A'\ge 1$, such that*

$$a_n\le A'H_\Phi^{-1}(n)$$

where $\Phi(x)=c\varphi(x), H_\Phi(x)=\int_1^x\frac{1}{\Phi(s)}ds$

Remark 6.3.3. *Φ is also concave, $\Phi(A'x)\le A'\Phi(x)$ for $A'\ge 1$, $(H_\Phi^{-1})'=\Phi\circ H_\Phi^{-1}$ and $\Phi\circ H_\Phi^{-1}(n)\le H_\Phi^{-1}(n+1)-H_\Phi^{-1}(n)$*

Proof. . It is easy to choose a constant $A'\ge 1$ s.t $a_1\le A'H_\Phi^{-1}(1)$, by induction if $a_{n-1}\le A'H_\Phi^{-1}(n-1)$, then

$$\begin{aligned}a_n &\le c\varphi(a_{n-1})+a_{n-1}=\Phi(a_{n-1})+a_{n-1}\\ &\le \Phi(A'H_\Phi^{-1}(n-1))+A'H_\Phi^{-1}(n-1)\\ &\le A'\{\Phi(H_\Phi^{-1}(n-1))+H_\Phi^{-1}(n-1)\}\\ &\le A'H_\Phi^{-1}(n)\end{aligned}$$

□

The coupling construction

Recall $C\in X$ is (t^*,ϵ)-small in the $\Phi-$Lyapunov condition. Given a sequence of i.i.d random variables $Z_1,Z_2,...,\ Z_i\sim B(1,\epsilon)$, We construct (X_t,X'_t) and the random stopping time

$$\widetilde{T}=\inf\left\{\widetilde{\tau}_i^{t^*}+t^*,\left(X_{\widetilde{\tau}_i^{t^*}},X'_{\widetilde{\tau}_i^{t^*}}\right)\in C\times C, Z_i=1\right\}.$$

Let

$$\widetilde{\tau}_{C\times C}^{t^*}=\inf\left\{t\ge t^*,(X_t,X'_t)\in C\times C\right\},\quad \widetilde{\tau}_1^{t^*}=\widetilde{\tau}_{C\times C}^0,$$

and

$$\widetilde{\tau}_i^{t^*}=\inf\left\{t\ge\widetilde{\tau}_{i-1}^{t^*}+t^*,(X_t,X'_t)\in C\times C\right\},\quad i\ge 2.$$

For each time $\widetilde{\tau}_i^{t^*}$, if X_t,X'_t have not yet coupled, then we processes as follows:

1. if $Z_i=1$, set
$$X_{\widetilde{\tau}_i^{t^*}+t^*}=X'_{\widetilde{\tau}_i^{t^*}+t^*}\sim Q(\cdot),$$
and we declare the processes to have coupled, and from the coupling time, let $X_{\widetilde{\tau}_i^{t^*}+t^*}=X'_{\widetilde{\tau}_i^{t^*}+t^*}$,

2. if $Z_i = 0$, set

$$X_{\widetilde{\tau}_i^{t^*}+t^*} \sim \frac{1}{1-\epsilon}\left(P^{t^*}(X_{\widetilde{\tau}_i^{t^*}},\cdot) - \epsilon Q(\cdot)\right) \text{ and }$$
$$X'_{\widetilde{\tau}_i^{t^*}+t^*} \sim \frac{1}{1-\epsilon}\left(P^{t^*}(X'_{\widetilde{\tau}_i^{t^*}},\cdot) - \epsilon Q(\cdot)\right)$$

conditionally independently.

In either case above, we fill in X_t and X'_t for $\widetilde{\tau}_i^{t^*} < t < \widetilde{\tau}_i^{t^*} + t^*$ conditionally independently, using the correct conditional distributions given $X_{\widetilde{\tau}_i^{t^*}}$, $X'_{\widetilde{\tau}_i^{t^*}}$, $X_{\widetilde{\tau}_i^{t^*}+t^*}$, $X'_{\widetilde{\tau}_i^{t^*}+t*}$. $\widetilde{T}$ is the coupling time, and we still have similar estimation as Lemma 6.3.1. It is easily seen that X_t and X'_t each marginally has the transition probabilities $P^t(x,\cdot)$.

Lemma 6.3.4. *Assume that the $\Phi-$Lyapunov condition (6.1.5) is satisfied for the generator associated to the semigroup (P_t), we build a two-dimensional Markov process (X_t, X'_t) on $X \times X$ as above with semigroup $(\widetilde{P}_t)$, which satisfies the bivariate Lyapunov condition:*

$$\widetilde{\mathcal{A}}W(x,x') \le -\lambda\Phi \circ W(x,x') + 2b1_{C\times C}(x,x')$$

with $W(x,x') = V(x) + V(x') - 1$, $d_0 = inf_{x\notin C}V(x)$, $0 < \lambda \le 1 - \frac{b}{\Phi(d_0)}$

Proof. Note that $\Phi \circ (V(x) + V(x') - 1) - \Phi \circ V(x) \le \Phi \circ V(x') - \Phi(1)$, we have

$$\begin{aligned}
\widetilde{\mathcal{A}}W(x,x') &= \mathcal{A}V(x) + \mathcal{A}V(x') \\
&\le -\Phi \circ V(x) - \Phi \circ V(x') + b1_C(x) + b1_C(x') \\
&\le -\Phi \circ (V(x) + V(x') - 1) - \Phi(1) + b1_C(x) + b1_C(x').
\end{aligned}$$

Then we can get the result easily. □

Lemma 6.3.5. *Assume that the drift condition (6.1.5) is satisfied, then the $\Phi-$Lyapunov condition is satisfied for P_t and $(\widetilde{P}_t)$ satisfies a $2\Phi-$Lyapunov condition with function $W(x,x') = V(x)+V(x')-1$, $d_0 = \inf_{x\notin C} V(x)$, $0 < \lambda \le 1-\frac{b}{\Phi(d_0)}$.*

In the same way of the proof of Lemma 6.3.1, we get

Lemma 6.3.6. *Assume that a $\Phi-$Lyapunov condition is satisfied and $\sup_C V < \infty$, $\Phi \in \mathcal{C}$, $C \in X$ is (t^*, ϵ)-small, for some positive time t^*, and $\epsilon > 0$, we construct the process (X_t, X'_t) and the coupling time $\widetilde{T}$ as above, then we have*

$$\mathbb{P}(\widetilde{T} > s) \le \sum_{0\le n, s\ge nt^*} (1-\epsilon)^n \widetilde{A}_1 \frac{H_\Phi^{-1}(\lambda\widetilde{A}_0 n)}{H_\Phi^{-1}(\lambda(s-t^*))} \le \frac{\widetilde{A}_1\widetilde{A}_{\epsilon,\Phi}}{H_\Phi^{-1}(\lambda(s-t^*))},$$

where

$$\widetilde{A}_0 =: \frac{2}{\lambda\Phi(1)} \sup_{x\in C}\left\{V(x) - 1 + \frac{b}{\Phi(1)}\int_0^{t^*} \Phi \circ H_\Phi^{-1}(\lambda s) ds\right\} < \infty,$$

$$\widetilde{A}_1 = \max\left\{\frac{\mathbb{E}\Big[V(X_0)+V(X'_0)\Big]-2}{H_\Phi^{-1}(\lambda\widetilde{A}_0)}, 1\right\},$$

and

$$d_0 = \inf_{x\notin C} V(x), \quad 0<\lambda\le 1-\frac{b}{\Phi(d_0)}, \quad \widetilde{A}_{\epsilon,\Phi}<\infty$$

Proof of Theorem 6.2.2

We construct the processes X_t and X'_t jointly as in 6.3.1. By the coupling inequality $||\mathcal{L}(X_t)-\mathcal{L}(X'_t)||_{TV}\le P(\widetilde{T}>t)$, lemma 6.3.6 and setting $\mathcal{L}(X'_0)=\pi$, we can get the theorem directly.

6.3.2 Proof of Theorem 6.2.4

Our proof relies mainly on the shift-coupling method presented for example in [93] (see also [1], [103], [104]).

We construct a second process (Y_t) with the same marginal $P_t(x,\cdot)$. From 6.3.1, there are times T and T' such that $\mathcal{L}(X_T)=\mathcal{L}(Y_{T'})=Q(.)$

We define the joint process (X_t,X'_t) as follows

$$\begin{cases} X'_{T'} = X_T, \\ X'_{T'+t} = X_{T+t} & \text{for } t>0 \\ X'_t = Y_t & \text{for } t<T'. \end{cases}$$

It is easily seen that X'_t still follows the transition probabilities $P_t(x,\cdot)$. The times T and T' are the so-called shift-coupling epochs for (X_t) and (X'_t). The lemma 6.3.1 gives upper bounds on $P(T>s)$ and $P(T'>s)$.

By the shift-coupling inequality(see e.g [92], Proposition 5), we have

$$\begin{aligned}\left\|\frac{1}{t}\int_0^t \mathbb{P}(X_s\in .)ds-\frac{1}{t}\int_0^t \mathbb{P}(X'_s\in .)ds\right\| \\ &\le \frac{1}{t}\left(\int_0^{+\infty}\mathbb{P}(T>s)ds+\int_0^{+\infty}\mathbb{P}(T'>s)ds\right) \\ &\le \frac{1}{t}\left(2t^*+\int_{t^*}^{+\infty}\mathbb{P}(T>s)ds+\int_{t^*}^{+\infty}\mathbb{P}(T'>s)ds\right).\end{aligned}$$

The result thus follows by integrating the upper bounds of $\mathbb{P}(T>s)$ and $\mathbb{P}(T'>s)$ given by the lemma 6.3.1 and letting $\mathcal{L}(X'_0)=\pi(.)$.

6.4 Applications

6.4.1 The reversible case

Let $\{X_t\}$ be a n-dimensional diffusion process defined by

$$dX_t = dB_t + \frac{1}{2}\nabla \ln \pi(X_t)dt \tag{6.4.1}$$

where B_t is a standard n-dimensional Brownian motion.

We suppose that $\pi(x) = e^{-V(x)}$ where $V(x) = (1+|x|^2)^{\frac{\alpha}{2}}$, $0 < \alpha < 1$. It is know that π is up to a normalizing constant the density on $\mathbb{R}^n$ with respect to the Lebesgue measure, of the unique invariant probability distribution of the process $\{X_t\}$(see for example [8], [2] or [25]). In the sequel, π will denote at the same time the invariant probability and the density function. Our aim is to study the ergodicity of the solution of the stochastic differential equation (6.4.1). In what follows, we will study drift condition and minorization. At the end, we will compute the constant in order to apply Theorem 6.2.2. We recall that generator of the process is given by $L = \frac{1}{2}\Delta + \frac{1}{2}\nabla \ln \pi(x) \cdot \nabla$(where the dot denote scalar product).

Lyapunov function

Let $\iota \in]0,1[$. We define the function W by

$$W(x) = e^{\iota\left(1+|x|^2\right)^{\frac{\alpha}{2}}}$$

Then $W \in \mathcal{D}(\mathcal{A})$. The following is in part deduced from the the calculations of [43], section 4.1; see also [8] section 4.3.

$$\begin{aligned}\mathcal{A}W = LW = \frac{1}{2}\Big\{&\iota\alpha n\left(1+|x|^2\right)^{\frac{\alpha}{2}-1} + \iota\alpha(\alpha-2)|x|^2\left(1+|x|^2\right)^{\frac{\alpha}{2}-2} \\ &+ \iota^2\alpha^2|x|^2\left(1+|x|^2\right)^{\alpha-2} - \iota\alpha^2|x|^2\left(1+|x|^2\right)^{\alpha-2}\Big\} \times e^{\iota\left(1+|x|^2\right)^{\frac{\alpha}{2}}} \\ = -\frac{1}{2}&\underbrace{\left\{-\frac{\alpha n\iota^{\frac{2-\alpha}{\alpha}}}{(1+|x|^2)^{\frac{\alpha}{2}}} + \frac{\iota^{\frac{2-\alpha}{\alpha}}\alpha(2-\alpha)|x|^2}{(1+|x|^2)^{1+\frac{\alpha}{2}}} - \frac{\iota\iota^{\frac{2-\alpha}{\alpha}}\alpha^2|x|^2}{1+|x|^2} + \frac{\iota^{\frac{2-\alpha}{\alpha}}\alpha^2|x|^2}{1+|x|^2}\right\}}_{g(x)} \\ &\times \Big(\log W(x)\Big)^{-2\frac{1-\alpha}{\alpha}} \times \exp\left(\iota(1+x^2)^{\frac{\alpha}{2}}\right).\end{aligned}$$

Let M such that for all $|x| > M$, $g(x) > 0$. Such M exists since we can see that the graph of g has the shape of a bowl. let $C = \overline{\mathcal{B}}(0,M)$ the closed ball centered at 0 and radius M. Set $\kappa = \frac{1}{2}\inf_{x\notin C} g(x)$. For all $a > 0$, we have

$$\mathcal{A}W(x) \leq -\Phi \circ W(x) + b\mathbf{1}_C(x)$$

where

$$\Phi(x) = \kappa x \Big(a + \log x\Big)^{-2\frac{1-\alpha}{\alpha}}, \quad b = \sup_C \mathcal{A}W + \sup_C \Phi \circ W.$$

W is thus a Lyapunov function under the hypothesis that C verifies the minorization condition.

Now in order to be within the context of application of lemma 6.3.4 and lemma 6.3.5 we choose $d_0 > 0$ in such a way that $\Phi(d_0) > b$ and $M' = \sqrt{\left(\frac{\ln d_0}{\iota}\right)^{\frac{2}{\alpha}} - 1} > M$ (this is always possible since the function Φ is non-decreasing and $\lim_{x \to +\infty} \Phi(x) = +\infty$). We state again $C = \mathcal{B}(0, M')$. Then we have again

$$\mathcal{A}W(x) \leq -\Phi \circ W(x) + b\mathbf{1}_C(x)$$

and $d_0 = \inf_{x \notin C} W(x)$, $\Phi(d_0) > b$.

Minorization condition

We deal now with the minorization condition. In what follows, we will check that the hypothesis of Theorem 9 (or Theorem 7 in one dimensional case) of [92] holds. To this end let us recall the definition of a pseudo small set: a subset C is a $(t, \epsilon)-$pseudo small set if for all $x, y \in C$, there exists a probability measure Q_{xy} such that

$$P^t(x, \cdot) \geq \epsilon\, Q_{xy}(\cdot), \qquad P^t(x, \cdot) \geq \epsilon\, Q_{xy}(\cdot).$$

This definition is more convenient for the numerical calculus in high dimension than the usual definition of small set.

Let $C = \mathbf{B}(0, M') = \prod_{i=1}^{n} [-M', M']$ where M' is as above. We denote by D the diameter of C, i.e $D = 2\sqrt{n}M'$. Let $a > M'$, $S = \prod_{i=1}^{n} [-a, a]$. Set $\nabla \ln \pi(x) = (\mu_1(x), \cdots, \mu_n(x))$ where $\mu_i(x) = \frac{\alpha x_i}{(1+|x|^2)^{1-\frac{\alpha}{2}}}$. Let c and d such that $c \leq \mu_i(x) \leq d$ for all $x \in S$. Set $L = \sqrt{n}(d - c)$.

Then, given $t_0 > 0$, for all $t \geq t_0$, C is (t, ε)-pseudo-small(small in one dimensional case) with

$$\begin{aligned} \varepsilon &= \Psi(\tfrac{-D-t_0 L}{\sqrt{4t_0}}) + e^{-\frac{DL}{2}} \Psi(\tfrac{t_0 L - D}{\sqrt{4t_0}}) \\ &-2n\Psi(\tfrac{-(a-M')-t_0 c}{\sqrt{t_0}}) - 2ne^{-2(a-M')c} \Psi(\tfrac{t_0 c-(a-M')}{\sqrt{t_0}}) \\ &-2n\Psi(\tfrac{-(a-M')+t_0 d}{\sqrt{t_0}}) - 2ne^{2(a-M')d} \Psi(\tfrac{-t_0 d-(a-M')}{\sqrt{t_0}}) \end{aligned}$$

where $\Psi(s) = \int_{-\infty}^{s} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$.

Remark 6.4.1. *In the one dimensional case we have*

$$\begin{aligned}\varepsilon\ &= \Psi(\tfrac{-D-t_0L}{\sqrt{4t_0}}) + e^{-\frac{DL}{2}}\Psi(\tfrac{t_0L-D}{\sqrt{4t_0}})\\ &-\Psi(\tfrac{-(a-M')-t_0c}{\sqrt{t_0}}) - e^{-2(a-M')c}\Psi(\tfrac{t_0c-(a-M')}{\sqrt{t_0}})\\ &-\Psi(\tfrac{-(a-M')+t_0d}{\sqrt{t_0}}) - e^{2(a-M')d}\Psi(\tfrac{-t_0d-(a-M')}{\sqrt{t_0}})\end{aligned}$$

Proof. See ([92], Theorem 7 and Theorem 9). □

Moreover, for one-dimensional diffusions, Roberts and Tweedie [95] have provided the following result which gives a more accessible form to get minorization condition in certain special cases

Lemma 6.4.2 ([95], lemma 6.1). *Let X be a non-explosive one dimensional diffusion satisfying the SDE*

$$dX_t = dB_t + \zeta(X_t)dt,$$

where B_t is a standard Brownian motion. Suppose that ζ is a non-increasing function. Then for $-\infty < c < a < \infty$, $[c, a]$ is (t^*, ε)-small with

$$\varepsilon = 2\Phi\left(-\frac{a-c}{2\sqrt{t^*}}\right),$$

where Φ is the standard cumulative normal distribution function.

Bounds in case of pseudo-small sets

In the multi-dimensional case we have obtained pseudo-small sets instead of small-sets. The results we will give now will permit us to have the same bound as in Theorem 6.2.2 in case the minorization condition deal with pseudo-small sets.

Theorem 6.4.3. *Given a Markov process (X_t) with semigroup (P_t) and stationary distribution $\pi(.)$, suppose $C \subset \mathcal{X}$ is (t^*, ε) pseudo-small, for some positive time t^* and $\varepsilon > 0$. Suppose further that a $\Phi-$Lyapunov condition holds and* $\sup_C V$ *hold. Then*

$$\|\mathcal{L}(X_s) - \pi\|_{TV} \le \sum_{0\le n, s\ge nt^*} (1-\epsilon)^n \widetilde{A}_1 \frac{H_\Phi^{-1}(\lambda \widetilde{A}_0 n)}{H_\Phi^{-1}(\lambda(s-t^*))} \le \frac{\widetilde{A}_1 \widetilde{A}_{\epsilon,\Phi}}{H_\Phi^{-1}(\lambda(s-t^*))},$$

where $\widetilde{A}_0$, $\widetilde{A}_1$, λ, and $\widetilde{A}_{\varepsilon,\Phi}$ are as in Theorem 6.2.2.

Proof. We construct the process X_t and Y_t jointly and the coupling time T as in([92] Theorem 8). The result thus follow by applying the coupling inequality and setting $\mathcal{L}(Y_0) = \pi(.)$ as in Theorem 6.2.2. □

Computation of the constants

This section deals with the calculation of the constants which appear in Theorem 6.2.2 and Theorem 6.4.3 above.

First we have that

$$H_\Phi^{-1}(x) = \exp\left(\left(\kappa\left(\frac{2-\alpha}{\alpha}\right)x + a^{\frac{2-\alpha}{\alpha}}\right)^{\frac{\alpha}{2-\alpha}} - a\right)$$

where κ is given above.

$d_0 = \inf_{x \notin C} W(x)$ and $\Phi(d_0)$ are found by the resolution of equation $\Phi(d_0) > b$.

$$\lambda \in \left]0, 1 - \frac{b}{\Phi(d_0)}\right[.$$

$$\begin{aligned}\widetilde{A}_0 &= \tfrac{2}{\lambda\Phi(1)} \sup_C \left\{W(x) - 1 + \tfrac{b}{\Phi(1)} \int_0^{t^*} \Phi \circ H_\Phi^{-1}(\lambda s) ds\right\} \\ &= \tfrac{2}{\lambda\Phi(1)} \left\{ d_0 - 1 - \tfrac{b}{\lambda\Phi(1)} + \frac{b\exp\left(\left(\kappa\left(\frac{2-\alpha}{\alpha}\right)\lambda t^* + a^{\frac{2-\alpha}{\alpha}}\right)^{\frac{\alpha}{2-\alpha}}\right)}{exp(a)\lambda\Phi(1)} \right\}.\end{aligned}$$

$$\widetilde{A}_1 = \max\left\{\frac{\mathbb{E}\Big[W(X_0)\Big] + \pi(W) - 2}{H_\Phi^{-1}(\lambda\widetilde{A}_0)}, 1\right\}$$

is found from the initial distribution of our chain.

Set

$$n'_0 = \left[\frac{\left(\ln\left(\frac{1-a_\varepsilon}{1-\varepsilon}\right)\right)^{\frac{2-\alpha}{2(\alpha-1)}}}{\left(\kappa\lambda\widetilde{A}_0\right)^{\frac{\alpha}{2(\alpha-1)}}\left(\frac{2-\alpha}{\alpha}\right)} - \frac{\alpha}{(2-\alpha)\kappa\lambda\widetilde{A}_0}\right] \quad \text{where } [x] \text{ is the integer part of } x.$$

Then

$$\widetilde{A}_{\varepsilon,\Phi} = \frac{\exp\left(\left(\kappa\left(\frac{2-\alpha}{\alpha}\right)\lambda\widetilde{A}_0 n_0 + 1\right)^{\frac{\alpha}{2-\alpha}} - 1 - n_0 \ln\left(\frac{1-a_\varepsilon}{1-\varepsilon}\right)\right)}{a_\varepsilon}$$

where $n_0 = 1$ if n'_0 is non-positive, else n_0 is one of the values n'_0 or $n'_0 + 1$.

Numerical calculus

We will focus here on the one dimensional case. One easily verifies that the hypothesis of lemma 6.4.2 are satisfied.

We choose $\alpha = 0.8$, $\iota = 0.6$

Choosing $M = 1.229915$ and $a = (\kappa)^{\frac{\alpha}{2-\alpha}}$, we then have $\kappa = 1.18 \times 10^{-4}$ and $b = 0.438$. We take $d_0 = 15000$. We thus have $M' = 32.05$, $\lambda \in]0, 0.23[$ and $C = \overline{\mathcal{B}}(0, M')$. λ is choosing as desired in this interval to optimize the results.

We choose $t^* = 100$ we thus have $\varepsilon = 0.001$. We choose $a_\varepsilon = \frac{\varepsilon}{4}$. We thus get $\widetilde{A}_0 = 3 \times 10^5$, $n_0' = 2 \times 10^{10}$, $\widetilde{A}_{\varepsilon,\Phi} = \frac{\exp(2.46 \times 10^6)}{a_\varepsilon}$.

Now we choose $t^* = 1000$ and $a_\varepsilon = \frac{\varepsilon}{4}$. Then we get $\varepsilon = 0.31$; $\widetilde{A}_0 = 3 \times 10^5$, $n_0' = 338$, $\widetilde{A}_{\varepsilon,\Phi} = \frac{\exp(48.40)}{a_\varepsilon}$.

Finally, we choose $t^* = 10^6$ and $a_\varepsilon = \frac{\varepsilon}{4}$. Thus we have $\varepsilon = 0.974$; $\widetilde{A}_0 = 3.78 \times 10^5$, $n_0' = 0$ and $\widetilde{A}_{\varepsilon,\Phi} = \frac{\exp(-0.491)}{a_\varepsilon}$.

We recall that $\widetilde{A}_1$ is found from the initial distribution of our chain.

One can observe that there is a "trade-off" between the values of t^* and $\widetilde{A}_{\varepsilon,\Phi}$.

6.4.2 Comparison with functional inequalities in the reversible case

One common way to establish quantitative sub exponential convergence to equilibrium is to use functional inequalities such as weak Poincaré inequalities, originally introduced by Röckner-Wang [96], i.e. for all smooth bounded function f (say in H^1) there exists a mapping $\beta : [0, s_0] \to \mathbb{R}_+$ decreasing such that

$$\int (f - \pi(f)^2 d\pi \leq \beta(s) \int |\nabla f|^2 d\pi + s\|f - \pi(f)\|_\infty^2$$

which implies that

$$\|P^t f - \pi(f)\|_2^2 \leq \psi(t)\|f - \pi(f)\|_\infty^2$$

with $\psi(t) = 2 \inf\{s > 0;\ \beta(s)\log(1/s) \leq t\}$. One notices that the convergence obtained is not expressed in the same norm. It is however not difficult to go from a total variation (which can be seen as a L^1 convergence) to a L^2 type convergence, indeed if

$$\|P^t(x, \cdot) - \pi(\cdot)\|_{TV} \leq cr(t)V(x)$$

and $\pi(V) < \infty$ then using [8, Th. 2.1] one gets that

$$\|P^t f - \pi(f)\|_2^2 \leq 8cr(t)\pi(V)\|f - \pi(f)\|_\infty.$$

Weak Poincaré inequality are however not so easy to characterize quantitatively, given a precise probability measure π, even in dimension one where Hardy's type criterion can be used. However an interesting feature is that Cattiaux&al [25] have developed a $\Phi-$Lyapunov function technique to get quantitative estimates in the weak Poincaré inequality. Let us recall their results. Assume that a $\Phi-$Lyapunov condition holds on the generator

$$\mathcal{A}V \leq -\Phi \circ V + b1_C$$

and the measure π satisfies some local Poincaré inequality

$$\int_C f^2 d\pi \leq \kappa_C \int |\nabla f|^2 d\pi + \frac{1}{\pi(C)} \pi(f 1_C)^2$$

then a weak Poincaré inequality holds

$$\beta(s) = \frac{(1 + b\kappa_C)}{\inf\{u; \pi(\Phi(V) \leq uV) > s\}},$$

for sufficiently small s. Looking now at the way the rate of convergence is derived from $\beta(s)$, one sees that one difficulty in this approach is that every estimate has an important impact on the speed of convergence: bad estimation of b or of the local Poincaré constant has a drastic impact, for example in the sub exponential case estimated numerically in the previous subsection, whereas in our coupling approach the speed of convergence is quantified by the Lyapunov condition only but not on the local characteristics of the small set C, which appears only in the multiplicative constant. It can have an important impact on sub exponential speed of convergence, and comparable in the polynomial case.

Of course one fundamental point is that the coupling approach may work in the non reversible case without additional difficulty (at least in the strictly elliptic with bounded diffusion coefficient) even if the invariant probabilityy measure is unknown whereas very few results are known using functional inequalities (see however [8] for an attempt in this direction using very particular Lyapunov-Poincaré inequality).

6.4.3 Kinetic Fokker Planck equation

The main goal of this subsection is to emphasize the difficulty of degenerate models such as kinetic Fokker Planck equations or more degenerate models as in the chain of oscillators case. Kinetic Fokker Planck equation can be written as

$$\begin{cases} dx_t = v_t dt \\ dv_t = \sqrt{2} dB_t - v_t dt - \nabla F(x_t) dt \end{cases}$$

which describes the position and velocity of a particle submitted to friction and confinement. If for nice F, Lyapunov conditions are quite well described, see for example [116] for the exponential Lyapunov condition or [8] in the sub exponential case, very few results exist to get quantitative small set estimations in this setting, see for example [9] however hardly quantitative or [65] but which applies only for very "small" small set and thus for very particular drift function F. To this respect, the approach initiated for example by Villani [109] using the explicit form of the invariant measure linked to this SDE and Poincaré inequality furnishes explicit constants (see [110] for the sub exponential case). It should then be rather interesting to get good explicit bounds for small sets in this setting.

Appendix 6.A Integrability of hitting times under Lyapunov conditions revisited

In this section, we aim at providing a simple proof of the integrability of hitting times when starting outside a chosen small set, due to [43]. It appears first in unpublished course note done by the third author.

Let us first present a preliminary proposition linked to concave functions of semimartingales, whose proof is given for a sake of completeness

Proposition 6.A.1. *Let (Y_t) be a real valued cad lag semimartingale and let $\Phi : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ be a function C^1 in its first argument, and C^2 and concave in its second argument. Then the process*

$$\Phi(Y_t) - \int_0^t \Phi'(Y_{s-}, s)dY_s - \int_0^t \partial_t\Phi(Y_{s-}, s)ds$$

is decreasing.
Suppose moreover $\Phi' \geq 0$ and that (X_t) is a continuous time Markov process with generator $\mathcal{A}$ and let F and G be such that

$$\mathcal{A}F \leq G.$$

Then we have

$$\mathcal{A}\Phi(F) \leq \partial_t\Phi + \Phi'(F)G.$$

Proof. Since (Y_t) is a semimartingale, we can decompose it as $Y_t = A_t + M_t$ with (M_t) a martingale and (A_t) a finite variation process. Use now Itô's formula to get

$$\begin{aligned}\Phi(Y_t) &= \Phi(Y_0) + \int_0^t \Phi'(Y_{s-}, s)dY_s + \int_0^t \partial_t\Phi(Y_{s-}, s)ds \\ &\quad + \int_0^t \Phi''(Y_{s-}, s)d\langle M\rangle_t^c + \sum_{s\in[0,t]} (\Phi(Y_{s+}, s) - \Phi(Y_{s-}, s) - \Phi(Y_{s-}, s)\Delta Y_s)\end{aligned}$$

where, as usual, $\langle M\rangle_t^c$ denotes the quadratic variation of the continuous part of M and ΔY_s the size of the jump of (Y_t) at time s. The first part of the result follows as $\langle M\rangle_t^c$ is an increasing process and that Φ'' is non positive as Φ is concave.
For the second part, set $Y - t = F(X_t, t)$. It follows from the assumptions that one can write

$$dY_t = G(X_t, t)dt + dN_t + dM_t$$

where (M_t) is a cadlag martingale and N is a non increasing process. Use now the first part of the proposition to get

$$d\Phi(Y_t) \leq \Phi'(Y_{t-}, t)(G(X_t, t)dt + dN_t + dM_t) + \partial_t\Phi(Y_{t-}, t)dt.$$

The claim then follows from the fact that dN_t is a negative measure and Φ' is non negative. $\square$

The idea is then following: looking at the $\Phi-$Lyapunov condition

$$\mathcal{A}V \leq -\Phi(V) + b1_C$$

that we may simplify, if $x \in C^c$ to

$$\mathcal{A}V \leq -\Phi(V),$$

it is natural to consider the solution Φ of the ordinary differential equation

$$\partial_t \Phi = -\Phi \circ \Phi, \qquad \Phi(v, 0) = v.$$

This ODE can be solved explicitly, indeed we may choose

$$\Phi(v, t) = H_\Phi^{-1}(H_\Phi(v) - t).$$

Let us now recall that

$$\Phi^{-1}(x, t) = H_\Phi^{-1}(H_\Phi(v) + t),$$

$$\partial_x \Phi^{-1}(x, t) = \frac{\partial_t \Phi^{-1}(x, t)}{\Phi(x)} = \frac{\Phi(\Phi^{-1}(x, t))}{\Phi(x)}.$$

As Φ is concave, we easily deduce that Φ^{-1} is increasing and concave in its first argument for every fixed value of t. Using the previous proposition, we have that for $x \in C^c$

$$\begin{aligned}
\mathcal{A}\Phi^{-1}(V(x), t) &\leq \partial_t \Phi^{-1}(V(x), t) + \partial_x \Phi^{-1}(V(x), t)\mathcal{A}V \\
&\leq \partial_t \Phi^{-1}(V(x), t) - \Phi(V(x))\partial_x \Phi^{-1}(V(x), t) \\
&\leq 0.
\end{aligned}$$

Introduce now $\tau = \inf\{t \geq 0, X_t \in C\}$, and for $x \in C^c$, we immediately deduce, using Itô's formula, and the fact that for all $t < \tau$, $X_t \notin C$,

$$\mathbb{E}_x(H_\Phi^{-1}(\tau)) \leq \mathbb{E}_x \Phi^{-1}(V(X_\tau), \tau) \leq V(x)$$

which is the desired integrability of Theorem 6.1.3.

Bibliographie

[1] Aldous, D., and Thorisson, H. Shift-coupling. *Stoch. Proc. Appl. 44* (1993), 1–14.

[2] Ané, C., Blachère, S., Chafaï, D., Fougères, P., Gentil, I., Malrieu, F., Roberto, C., and Scheffer, G. *Sur les inégalités de Sobolev logarithmiques*, vol. 10 of *Panoramas et Synthèses*. Société Mathématique de France, Paris, 2000.

[3] Arcones, M. A. The large deviation principle of empirical processes. *Preprint* (2001).

[4] Athreya, K. Large deviations for branching processes-i. single type case. *Ann. Appl. Probab. 5, No. 3* (1994), 779–790.

[5] Athreya, K., and Ney, P. *Branching Process.* Springer, Berlin, 1972.

[6] Azuma, K. Weighted sums of certain dependent random variables. *Tôhoku Math. J 19, No. 3* (1967), 357–367.

[7] Bakry, D., Barthe, F., Cattiaux, P., and Guillin, A. A simple proof of the Poincaré inequality for a large class of probability measures. *Electronic Communications in Probability. 13* (2008), 60–66.

[8] Bakry, D., Cattiaux, P., and Guillin, A. Rate of convergence for ergodic continuous Markov processes : Lyapunov versus Poincaré. *J. Func. Anal. 254* (2008), 727–759.

[9] Bally, V., and Kohatsu-Higa, A. Lower bounds for densities of Asian type stochastic differential equations. *J. Funct. Anal. 258*, 9 (2010), 3134–3164.

[10] Basawa, I. V., and Huggins, R. M. Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Probab 36, 4* (1999), 1225–1233.

[11] Basawa, I. V., and Huggins, R. M. Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N. Z. J. Stat 42, 4* (2000), 423–432.

[12] Basawa, I. V., Hwang, S. Y., and Yeo, I. K. Local asymptotic normality for bifurcating autoregressive processes and related asymptotic inference. *Statistical Methodology 6* (2009), 61–69.

[13] Basawa, I. V., and Zhou, J. Non-Gaussian bifurcating models and quasi-likelihood estimation. *J. Appl. Probab 41A* (2004), 55–64.

[14] Basawa, I. V., and Zhou, J. Least-squares estimation for bifurcating autoregressive processes. *Statist. Probab.Lett. 1* (2005), 77–88.

[15] Basawa, I. V., and Zhou, J. Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors. *J. Time Ser. Anal. 26, 6* (2005), 825–842.

[16] Bennett, G. Probability inequalities for sum of independant random variables. *Journal of the American Statistical Association 57, No. 297* (1962), 33–45.

[17] Bercu, B. On large deviations in the Gaussian autoregressive process : stable, unstable and explosive cases. *Bernoulli 7, No 2* (2001), 299–316.

[18] Bercu, B., De Saporta, B., and Gégout-Petit, A. Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electronic. J. Probab. 14* (2009), 2492–2526.

[19] Bercu, B., and Proïa, F. A sharp analysis on the asymptotic behavior of the Durbin-Watson statistic for the first-order autoregressive process. *Submitted* (2011).

[20] Bercu, B., and Touati, A. Exponential inequalities for self-normalized martingales with applications. *Ann. Appl. Probab. 18, no.5* (2008), 1848–1869.

[21] Bitseki Penda, S. V., and Djellout, H. Deviation inequalities and moderate deviations for estimators of parameters in bifurcating autoregressive models. *arXiv :1204.2355v1. Soumis* (2012).

[22] Bitseki Penda, S. V., Djellout, H., and Guillin, A. Deviation inequalities, moderate deviations and some limit theorems for bifurcating Markov chains with application. *Preprint* (2011).

[23] Bitseki Penda, S. V., Djellout, H., and Proïa, F. Moderate deviations for the Durbin-Watson statistic related to the first-order autoregressive process. *arXiv 1201.3579. Soumis* (2012).

[24] Blattberg, R. Evaluation of the power of the Durbin-Watson statistic for non-first order serial correlation alternatives. *Review of Economics and Statistics 55* (1973), 508–515.

[25] Cattiaux, P., Gozlan, N., Guillin, A., and Roberto, C. Functional inequalities for heavy tailed distributions and applications to isoperimetry. *Electron. J. Probab. 13*, 15 (2010), 346–385.

[26] Chaabane, F. Version forte du théorème de la limite centrale fonctionnel pour les martingales. *C. R. Acad. Sci. Paris Série 1 323* (1996), 195–198.

[27] Chen, X. Moderate deviations for m-dependent random variables with banach space value. *Statis. and Probab. Letters. 35* (1998), 123–134.

[28] Cowan, R., and Staudte, R. G. The bifurcating autoregressive model in cell lineage studies. *Biometrics 42* (1986), 769–783.

[29] De la Peña, V. H., Lai, T. L., and Shao, Q.-M. *Self-normalized processes. Limit theory and statistical applications.* Probability and its Applications(New York). Springer-Verlag, Berlin, 2009. 275 pp.

[30] De Saporta, B., Gégout-Petit, A., and Marsalle, L. Asymmetry tests for bifurcating auto-regressive processes with missing data. *preprint*, 2011.

[31] De Saporta, B., Gégout-Petit, A., and Marsalle, L. Random coefficients bifurcating autoregressive processes. *arXiv :1205.3658v1 [math.PR]*, 2012.

[32] De Saporta, B., Gégout-Petit, A., and Marsalle, L. Parameters estimation for asymmetric bifurcating autoregressive processes with missing data. *Electronic Journal of Statistics 5* (2011), 1313–1353.

[33] Delmas, J. F., and Marsalle, L. Detection of cellular aging in Galton-Watson process. *Stochastic Processes and their Applications 120* (2010), 2495–2519.

[34] Dembo, A. Moderate deviations for martingales with bounded jumps. *Electron. Comm. Probab. 1, no. 3* (1996), 11–17.

[35] Dembo, A., and Zeitouni, O. *Large deviations techniques and applications, second edition*, vol. 38 of *Applications of Mathematics.* Springer, 1998.

[36] Deuschel, J., and Stroock, D. *Large Deviations.* American Mathematical Society, Providence, Rhode Island, 2000.

[37] Diaconis, P., and Saloff-Coste, L. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab. 6*, 3 (1996), 695–750.

[38] Djellout, H. Moderate deviations for martingale differences and applications to ϕ-mixing sequences. *Stoch. Stoch. Rep. 73, 1-2* (2002), 37–63.

[39] Djellout, H., and Guillin, A. Large and moderate deviations for moving average processes. *Ann. Fac. Sci. Toulouse Math. 10, no. 1* (2001), 23–31.

[40] Djellout, H., and Guillin, A. Moderate deviations for Markov chains with atom. *Stochastic Process. Appl. 95, no. 2* (2001), 203–217.

[41] Djellout, H., Guillin, A., and Wu, L. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Annals of Probability 32, No. 3B* (2004), 2702–2732.

[42] Djellout, H., Guillin, A., and Wu, L. Moderate deviations of empirical periodogram and non-linear functionals of moving average processes. *Ann. Inst. H. Poincaré Probab. Statist 42, no. 4* (2006), 393–416.

[43] Douc, R., Fort, G., and Guillin, A. Subgeometric rates of convergence of f-ergodic strong Markov processes. *Stochastic Process. Appl. 119*, 3 (2009), 897–923.

[44] Douc, R., Fort, G., Moulines, E., and Soulier, P. Practical drift conditions for subgeometric rates of convergence. *Ann. Appl. Probab. 14*, 3 (2004), 1353–1377.

[45] Douc, R., Guillin, A., and Najim, J. Moderate deviations for particle filtering. *Annals of Probability 15, No. 1B* (2005), 587–614.

[46] Douc, R., Moulines, E., and Rosenthal, J. S. Quantitative bounds on convergence of time-inhomogeneous Markov chains. *Ann. Appl. Probab. 14*, 4 (2004), 1643–1665.

[47] Douc, R., Moulines, E., and Sourlier, P. Computable convergence rates for sub-geometric ergodic Markov chains. *Bernouilli 13*, 3 (2004), 831–848.

[48] Down, N., Meyn, S. P., and Tweedie, R. L. Exponential and uniform ergodicity of Markov processes. *AP 23*, 4 (1995), 1671–1691.

[49] Durbin, J. Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica 38* (1970), 410–421.

[50] Durbin, J., and Watson, G. S. Testing for serial correlation in least squares regression. I. *Biometrika 37* (1950), 409–428.

[51] Durbin, J., and Watson, G. S. Testing for serial correlation in least squares regression. II. *Biometrika 38* (1951), 159–178.

[52] Durbin, J., and Watson, G. S. Testing for serial correlation in least squares regession. III. *Biometrika 58* (1971), 1–19.

[53] Eichelsbacher, P., and Löwe, M. Moderate deviations for i.i.d. random variables. *ESAIM Probab. Stat. 7* (2003), 209–218.

[54] Eichelsbacher, P., and Schmock, U. Exponential approximations in completely regular topological spaces and extensions of sanovs theorem. *Stochastic Process. Appl 77* (1998), 233–251.

[55] Fort, G., and Roberts, G. Subgeometric ergodicity of strong Markov processes. *Ann. Appl. Prob. 15*, 2 (2005), 1565–1589.

[56] Gao, F., and Zhao, X. Delta method in large deviations and moderate deviations for estimators. *The Annals of Statistics 39, No. 2* (2011), 1211–1240.

[57] Garcia1, J. An extension of the contraction principle. *Journal of Theoretical Probability 17, No. 2* (2004), 403–434.

[58] Gelfand, A., and Smith, A. Sampling based approaches to calculating marginal densities. *J. Amer. Stat. Assoc 85* (1990), 398–409.

[59] Getoor, R. Transience and recurrence of Markov processes. *Lect.Notes Math. number 784* (1980), 397–409.

[60] Gozlan, N. Integral criteria for transportation-cost inequalities. *Electron. Comm. Probab. 11* (2006), 64–77.

[61] Gozlan, N., and Léonard, C. A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields 139 ,no. 1-2* (2007), 235–283.

[62] Guillin, A. Deviations moderées pour le principe de moyennisation d'une eds avec petite diffusion. *Comptes Rendus de l'Académie des Sciences t.338, série 1* (2001), 751–754.

[63] Guillin, A. Moderate deviations of inhomogeneous functionals of Markov processes and application to averaging. *Stochastic Process. Appl. 92(2)* (2001), 287–313.

[64] Guillin, A. Averaging principle of sde with small diffusion : Moderate deviations. *Annals of Probability 31(1)* (2003), 413–443.

[65] Guillin, A., and Wang, F.-Y. Bismut formula and Harnack inequality for Fokker-planck equations. To appear in Journal of Differential Equations, 2011.

[66] Guyon, J. Limit theorems for bifurcating Markov chains. application to the detection of cellular aging. *Ann. Appl. Probab. Vol. 17, No. 5-6* (2007), 1538–1569.

[67] Guyon, J., Bize, A., Paul, G., Stewart, E., Delmas, J., and Taddéi, F. Statistical study of cellular aging. *CEMRACS 2004 Proceedings, ESAIM Proceedings 14* (2005), 100–114.

[68] Hairer, M., and Mattingly, J. C. Slow energy dissipation in anharmonic oscillator chains. To appear in Communications in Mathematical Physics, 2007.

[69] Has′minskiĭ, R. Z. *Stochastic stability of differential equations*, vol. 7 of *Monographs and Textbooks on Mechanics of Solids and Fluids : Mechanics and Analysis*. Sijthoff & Noordhoff, Alphen aan den Rijn, 1980. Translated from the Russian by D. Louvish.

[70] Hoeffding, W. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association 58, No. 301* (1963), 13–30.

[71] Inder, B. A. An approximation to the null distribution of the durbin-watson statistic in models containing lagged dependent variables. *Econometric Theory 2* (1986), 413–428.

[72] King, M. L., and Wu, P. X. Small-disturbance asymptotics and the Durbin-Watson and related tests in the dynamic regression model. *J. Econometrics 47* (1991), 145–152.

[73] Ledoux, M. Sur les déviations modérées des sommes de variables aléatoires vectorielles indépendantes de même loi. *Ann. Inst. Henri-Poincaré. 35* (1992), 123–134.

[74] Ledoux, M. *The concentration of measure phenomenon.* Mathematical Surveys and Monographs, 89. American Mathematical Society, Providence, RI, 2001.

[75] Ledoux, M., and Talagrand, M. *Probability in Banach Space.* Springer-Verlag, New York (1991).

[76] Malinvaud, E. Estimation et prévision dans les modèles économiques autorégressifs. *Review of the International Institute of Statistics 29* (1961), 1–32.

[77] Mas, A., and Menneteau, L. Large and moderate deviations for infinite-dimensional autoregessive processes. *J. Multivariate Anal. 87* (2003), 241–260.

[78] Massart, P. *Concentration inequalities and model selection.* Lecture Notes in Mathematics, 1896. Springer, Berlin, 2007.

[79] McDiarmid, C. On the method of bounded differences. *Electron. Comm. Probab. 11* (2006), 64–77.

[80] McDiarmid, C. *On the method of bounded differences.* In : Surveys in Combinatorics 1989. Cambridge University Press, Cambridge (1989) 148-188.

[81] Meyn, S. P., and Tweedie, R. L. *Markov chains and stochastic stability.* Communications and Control Engineering Series. Springer-Verlag London Ltd., London, 1993.

[82] Meyn, S. P., and Tweedie, R. L. Stability of Markovian processes II : continuous-time processes and sampled chains. *Adv. Appl. Proba. 25* (1993), 487–517.

[83] Meyn, S. P., and Tweedie, R. L. Stability of Markovian processes III : Foster-Lyapunov criteria for continuous-time processes. *Adv. Appl. Proba. 25* (1993), 518–548.

[84] Miao, Y., and Shen, S. Moderate deviation principle for autoregessive processes. *J. Multivariate Anal. 100* (2009), 1952–1961.

[85] Montenegro, R., and Tetali, P. Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci. 1*, 3 (2006), x+121.

[86] Nerlove, M., and Wallis, K. F. Use of the Durbin-Watson statistic in inappropriate situations. *Econometrica 34* (1966), 235–238.

[87] Nummelin, E. *General irreducible Markov chains and nonnegative operators.* Cambridge University Press, Cambridge, 1984.

[88] PETER, C. P., AND TASSOS, M. Limit theory for moderate deviations from a unit root. *Journal of Econometrics 136* (2007), 115–130.

[89] PUHALKSII, A. On functional principle of large deviations. *New Trends Probab.Statist 1* (1991), 198–218.

[90] PUHALKSII, A. Large deviation analysis of the single server queue. *Queueing Systems 21* (1995), 5–66.

[91] PUHALSKII, A. Large deviations of semimartingales : a maxingale problem approach. i. limits as solutions to a maxingale problem. *Stoch. Stoch. Rep. 61* (1997, no. 3-4), 141–243.

[92] ROBERTS, G., AND ROSENTHAL, J. Quantitative bounds for convergence rates of continuous time Markov processes. *Electron. J. Probab. 1*, 9 (1996), 1–21.

[93] ROBERTS, G., AND ROSENTHAL, J. Shift-coupling and convergence rates of ergodic averages. *Communications in statistics. Stochastics models 13*, 1 (1996), 147–165.

[94] ROBERTS, G., AND ROSENTHAL, J. General state space Markov chains and MCMC algorithms. *Probability Surveys 1* (2004), 20–71.

[95] ROBERTS, G. O., AND TWEEDIE, R. L. Rates of convergence of stochastically monotone and continuous time Markov models. *J. Appl. Probab. 37*, 2 (2000), 359–373.

[96] RÖCKNER, M., AND WANG, F. Y. Weak Poincaré inequalities and L^2-convergence rates of Markov semigroups. *J. Funct. Anal. 185*, 2 (2001), 564–603.

[97] ROSENTHAL, J. Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Stat. Assoc. 90* (1995), 558–566.

[98] SCHMIDT, P., AND GUILKEY. Some further evidence on the power of the Durbin-Watson and Geary tests. *Review of Economics and Statistics 57* (1975), 379–382.

[99] SMITH, A., AND ROBERTS, G. Bayesian computation via the gibbs sampler and related Markov chain Monte Carlo methods. *J. Roy. Stat. Soc. Ser B 55* (1993), 3–24.

[100] STEWART, E. J., MADDEN, R., PAUL, G., AND TADDÉI, F. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol 3(2)* (2005), e45.

[101] STOCKER, T. On the asymptotic bias of OLS in dynamic regression models with autocorrelated errors. *Statist. Papers 48* (2007), 81–93.

[102] STOUT, W. F. A martingale analogue of Kolmogorov's law of the iterated logarithm. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 15* (1970), 279–290.

[103] THORISSON, H. Coupling and shift-coupling random sequences. *Contemp. Math. 149* (1993).

[104] THORISSON, H. Shift-coupling in continuous times. *Probab. Theory Relat. Fields 99* (1994), 477–483.

[105] TILLMAN, J. A. The power of the Durbin-Watson test. *Econometrica 43* (1975), 959–974.

[106] VAN DER VAART, A. W. *Asymptotic statistics.* Cambridge University Press. New York, 1998.

[107] VARADHAN, S. R. S. Asymptotic probabilities and differential equations. *Comm.Pure Appl. Math 19* (1966), 261–286.

[108] VERETENNIKOV, A. Y. On polynomial mixing bounds for stochastic differential equations. *Stochastic Process. Appl. 70*, 1 (1997), 115–127.

[109] VILLANI, C. Hypocoercivity. Memoirs of the AMS, 2010.

[110] WANG, X.-Y. Subexponential decay in kinetic Fokker planck equation : weak hypocoercivity. Preprint, 2012.

[111] WEI, C. Z. Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. *Ann. Statist. 15, 4* (1987), 1667–1682.

[112] WORMS, J. Moderate deviations for stable Markov chains and regression models. *Electron. J. Probab. 4, no. 8* (1999), 1–28.

[113] WORMS, J. Principes de déviations modérées pour des martingales et applications statistiques. *Thèse de Doctorat à l'Université Marne-la-Vallée.* (2000).

[114] WORMS, J. Moderate deviations of some dependent variables. I. martingales. *Math. Methods Statist. 10, no. 1* (2001), 38–72.

[115] WORMS, J. Moderate deviations of some dependent variables. II. some kernel estimators. *Math. Methods Statist. 10, no. 2* (2001), 161–193.

[116] WU, L. Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stochastic Process. Appl. 91* (2001), 205–238.