Utilisation des notions de dépendance faible en statistique

Olivier Wintenberger

UNIVERSITÉ PARIS I PANTHÉON-SORBONNE
U.F.R. MATHÉMATIQUES

THÈSE

pour obtenir le grade de

DOCTEUR ÈS-SCIENCES

SPÉCIALITÉ MATHÉMATIQUES APPLIQUÉES

présentée et soutenue publiquement par

Olivier WINTENBERGER

11 juin 2007

sous le titre

UTILISATION DES NOTIONS DE DÉPENDANCE FAIBLE EN STATISTIQUE

Directeurs de Thèse

M. Paul DOUKHAN, Professeur chargé de mission à l'ENSAE,

M. Jean-Marc BARDET, Professeur à l'Université Paris 1 Panthéon-Sorbonne.

Jury

Rapporteurs : M. Emmanuel RIO, Professeur à l'Université de Versailles St-Quentin-en-Yvelines,
M. Qi-Man SHAO, Professeur à l'Université de Hong Kong et de l'Oregon.

Examinateurs : M. István BERKES, Professeur à l'Université de Graz,
M. Gérard GREGOIRE, Professeur à l'Université de Grenoble 1,
Mme. Dominique PICARD, Professeur à l'Université Paris 7,
M. Jean-Michel ZAKOIAN, Professeur à l'Université Lille 3.

A mon grand père Michel.

Remerciements

Je dois le bon accomplissement de cette thèse à plusieurs personnes : chacune, dans son rôle, m'a littéralement porté jusqu'à son terme.

En premier lieu à mes directeurs, dans un style très différent. A Paul qui m'a montré par sa motivation, son enthousiasme, son dynamisme que le métier de chercheur pouvait être passionnant. A Jean Marc qui a été pour moi un modèle d'équilibre, de calme et de patience, qui sont des valeurs aussi indispensables pour ce métier exigeant. A Marie Cottrell pour m'avoir accueilli au sein du SAMOS dans la gaieté et la bonne humeur. A toute l'équipe de ce laboratoire qui réussit à concilier rigueur intellectuelle et détente. A mes aînés dans ce domaine scientifique, Patrice Bertail, Odile Brandière, Jérôme Dedecker, Sana Louhichi, Florence Merlevède, Clémentine Prieur et Philippe Soulier qui ont toujours été de bon conseil. Merci aussi à Emmanuel Rio et Qi Man Shao qui m'ont fait profiter de leur expertise dans le domaine en rapportant sur ma thèse. A István Berkes, Gérard Grégoire, Dominique Picard et Jean Michel Zakoïan qui ont très gentiment accepté d'être les membres de mon jury. Leur jugement éclairé m'encourage à aller plus loin.

D'autres personnes m'ont beaucoup apporté, que ce soit sur le plan scientifique ou non. A mes co-auteurs Irène Gannaz et Nicolas Ragache qui, étant doctorants comme moi, ont partagé les mêmes sentiments. A d'autres doctorants, en particuliers du SAMOS comme Catherine, Hatem, Imen, Lionel, Madalina, Omar et ceux du séminaire de Chevaleret comme Christophe, Claire, Guillaume, Mathieu, Pierre, Sebastien, Stéphane, Thanh Mai, Thomas et Tu avec qui j'ai partagé de nombreux moments de détente aussi. A des collègues qui sont devenus des amis comme Cécile, Gilles, Gabriel et Jean Michel.

Enfin, il y a ceux qui ont rendu tout cela possible et qui y donnent un sens. A ma famille qui a joué un rôle très important, en particulier à mes parents qui m'ont toujours soutenu. A mes grands parents pour leur gentillesse. Et bien sûr à Nathalie qui a bien voulu partager sa vie avec moi durant ces trois années et en faire des années de bonheur.

Résumé

La dépendance faible est un outil très performant pour obtenir des résultats asymptotiques en statistique des séries chronologiques. Son atout majeur est de résumer les propriétés de dépendance de très nombreux modèles via le comportement d'une suite de coefficients. Dans un problème où le modèle n'est pas clairement identifiable, des hypothèses sur les coefficients de dépendance faible sont parfois moins contraignantes que le choix d'un modèle. De plus, pour certains modèles causaux, la dépendance faible permet d'étudier les propriétés de dépendance là où toutes les autres notions (de mélange par exemple) échouent. Les coefficients permettent d'élargir aux séries chronologiques des résultats asymptotiques classiques du cas de référence, celui d'observations indépendantes.

Le chapitre 1 est une synthèse des travaux ; il reprend les principaux résultats des chapitres suivants qui peuvent être lus séparément. Après avoir défini les notions de faible dépendance utilisées dans ce travail, nous donnons des exemples de modèles faiblement dépendants dans la partie I tels que les chaînes à mémoire infinie dans le chapitre 2 et les schémas de Bernoulli dans le chapitre 3. De nouvelles propriétés de moments et de dépendance sont données pour ces modèles. La dépendance faible permet d'atteindre des résultats asymptotiques optimaux, identiques au cas indépendant, sur une classe restreinte de modèles. Plus on étend le domaine d'étude et moins les résultats sont bons. Dans le cas de modèles causaux, les résultats obtenus sont meilleurs que pour des modèles non causaux, plus généraux.

Grâce à ses outils, nous généralisons des résultats asymptotiques classiques. Nous donnons des versions faiblement dépendantes du principe d'invariance faible dans le chapitres 3 et de convergence de l'estimateur de la densité (adaptatif ou non) dans la partie II. Des pertes dans les vitesses de convergence sont constatées dans le chapitre 4. Elles sont en partie dues à la dépendance des modèles, qui ajoute des termes perturbateurs, et en partie aux inégalités obtenues dans le cadre de la dépendance faible qui ne sont pas aussi fines que celles du cas indépendant. Nous discutons de l'origine de ces pertes en vitesse à partir de simulations fournies dans le chapitre 5.

Enfin, la normalité asymptotique de l'estimateur du Quasi Maximum de Vraisemblance dans le cas multivarié est donné pour la première fois dans la partie III. Ce problème restreint la classe des modèles à considérer à celle de modèles décorrélés. D'autres outils plus spécifiques et plus efficaces que la dépendance faible existent pour ces modèles très particuliers présentés dans le chapitre 6.

Abstract

Times series are main topics in modern statistical mathematics. They are essential for applications where randomness plays an important role. Indeed, physical constraints entail that serious modelling cannot be done using independent sequences. This represents a real problem because properties are not always known in that case. In order to generalize the main statistics of the independent case, one needs to use weakly dependent notions.

The present work, summarized in chapter 1, is devoted to provide a frame for the commonly used time series in part I. These notions are mainly divided in two different classes. The first one is the class of causal dependence. In this case the weak dependence fits the dependence structure of time series where other notions, as mixing, fail. We study in chapter 2 moment and weakly dependent properties of chains with infinite memory that are not Markovian but causally weakly dependent models. The second one is the class of non causal processes. Weak dependence in this context is very general and conditions of weak dependence are preferred to modelization. We introduce the Bernoulli shifts with dependent inputs in chapter 3. Working with these notions, we balance the optimality of the result with the size of the class of models involved. In the more restrictive causal case, the results are better than in the non causal one.

With these tools, we establish different asymptotical results. We give a Donsker principle in chapter 3 and the convergence of the density estimates (adaptive or not) in part II. Losses in the rates of convergence are observed in chapter 4. They are due to the weak dependence that puts additional terms and then perturbs the statistics. They are also due to techniques that are not as efficient as in the independent case. We discuss on simulations the origin of these losses in chapter 5.

Finally, the asymptotic normality of the Quasi Maximum Likelihood Estimator for multidimensional processes is given for the first time in part III. We use moment properties of chains with infinite memories but not their weak dependence properties. The parametric context reduces the class of models to be considered to the one of decorrelated processes. Specific tools are more efficient than weak dependence in the setting of chapter 6.

Table des matières

Chapitre 1

Synthèse des travaux

Ce chapitre reprend les principaux résultats obtenus dans les travaux [Wint1], [Wint2], [Wint3], [Wint4] et [Wint5] et détaillés dans les chapitres 2, 3, 4, 5 et 6. Après avoir défini les notions de faible dépendance utilisées dans ce travail, nous dressons une liste des principaux modèles faiblement dépendants. Nous introduisons en particulier les chaînes à mémoire infinie et les schémas de Bernoulli, avec leurs propriétés de moments et de dépendance faible. Nous généralisons grâce à ces notions des résultats asymptotiques classiques du cas de référence où les données sont indépendantes. Nous donnons des versions faiblement dépendantes du principe d'invariance faible et de convergence de l'estimateur de la densité (adaptatif ou non). Des pertes dans les vitesses de convergence sont parfois constatées. Enfin, dans le cadre de l'estimation paramétrique, la normalité asymptotique est pour la première fois donnée dans le cas multivarié. Ce résultat utilise les propriétés de moments obtenues sur les chaînes à mémoire infinie.

Voici quelques notations très utiles pour la suite de cette synthèse :
- $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probabilisé,
- $(E, \|\cdot\|)$ est un espace de Banach,
- $\Lambda^{(k)}$ est l'espace des fonctions lipschitziennes f de $E^k \to \mathbb{R}$ telles que pour tout $(x_1, \ldots, x_k)$ et $(y_1, \ldots, y_k)$ de E^k,

$$|f(x_1, \ldots, x_k) - f(y_1, \ldots, y_k)| \leqslant \mathrm{Lip}\, f \sum_{i=1}^{k} \|x_i - y_i\|, \text{ avec } \mathrm{Lip}\, f < \infty$$

- $\Lambda_\rho^{(k)}$ est l'ensemble des fonctions de $\Lambda^{(k)}$ telles que $\mathrm{Lip}\, f \leqslant \rho$,
- BV est l'espace des fonctions de $[0,1]$ dans $[0,1]$ telles que $\|h\|_{BV} < \infty$ où

$$\|h\|_{BV} = |h(0)| + \sup_{n \in \mathbb{N}} \sup_{a_0 = 0 < a_1 < \cdots < a_n = 1} \sum_{i=1}^{n} |h(a_i) - h(a_{i-1})|,$$

- BV_1 est l'ensemble des fonctions de BV telles que $\|h\|_{BV} \leqslant 1$,
- $\mathbb{L}^m(E)$ est l'espace des variables aléatoires à valeurs dans E telles que $\|X\|_m < \infty$, avec, par définition $\|X\|_m = (\mathbb{E}\|X\|^m)^{1/m}$,
- $\mathbb{L}^\infty(E)$ est l'espace des variables aléatoires à valeurs dans E telles que $\|X\|_\infty < \infty$, où, par définition $\|X\|_\infty = \inf_{A>0}\{\mathbb{P}(\|X\| < A) = 1\}$.
- Le passé au temps t d'un processus $(X_t)_{t\in\mathbb{Z}}$ est la tribu $\mathcal{P}_t = \sigma(X_s;\ s \leqslant t)$, le futur est la tribu $\mathcal{F}_t = \sigma(X_s;\ s \geqslant t)$.

1.1 Notions de dépendance faible [Wint1, Wint2, Wint3, Wint4]

Nous étudions les propriétés de dépendance de processus à temps discret $(X_t)_{t\in\mathbb{Z}}$ supposés strictement stationnaires.

Définition 1.1 *Une suite $(X_t)_{t\in\mathbb{Z}}$ est stationnaire au sens strict si $(X_1, \dots, X_n)$ et $(X_h, \dots, X_{n+h})$ ont la même loi pour tout $n > 0$ et $h \in \mathbb{Z}$.*

De plus, les séries temporelles sont supposées à "mémoire courte", c'est à dire telles que la série des auto-covariances $(\mathrm{Cov}(X_0, X_k))_{k\in\mathbb{Z}}$ soit sommable. Avec cette hypothèse, il est raisonnable d'escompter obtenir des résultats asymptotiques similaires à ceux du cas de référence des processus indépendants et identiquement distribués (i.i.d.). Dans ce but, nous introduisons des coefficients qui quantifie la dépendance de ces processus.
De nombreuses notions ont été introduites pour faciliter l'étude des processus dépendants. Chronologiquement, les notions de mélange sont les premières à être apparues. Les trois principales sont le mélange fort (α-mélange), le β-mélange et le mélange uniforme (ou ϕ-mélange) introduits respectivement par Rosenblatt [103], Rozanov et Volkonskii [104] et Ibragimov [70]. Ces notions présentent un inconvénient majeur. En pratique, le calcul de ces coefficients est délicat (voir Doukhan, [38]. De plus, certains processus ne sont pas mélangeants. Le plus célèbre est celui d'Andrews [1] :

Exemple 1.1 *Soit $(\xi_t)_{t\in\mathbb{Z}}$ une suite de variable i.i.d. telle que $\mathbb{P}(\xi_1 = 1) = \mathbb{P}(\xi_1 = 0) = 1/2$ alors la solution stationnaire $(X_t)_{t\in\mathbb{Z}}$ de l'équation*

$$X_t = \frac{1}{2}(X_{t-1} + \xi_t), \quad \forall t \in \mathbb{Z}, \tag{1.1}$$

n'est pas mélangeante au sens de Rosenblatt.

De là est venue l'idée d'introduire des notions de dépendance moins restrictives, couvrant plus de modèles, pour lesquelles nous pouvons calculer des bornes sur les coefficients et développer une théorie asymptotique intéressante. Nous présentons ci-dessous les principales notions qui seront utilisées dans la suite.

1.1.1 Cas causal

Nous appellons causaux les coefficients de dépendance faible pour lesquels le passé $\mathcal{P}_t$ et le futur $\mathcal{F}_t$ ne jouent pas un rôle symétrique. Un modèle est dit causal lorqu'on contrôle au moins un de ses coefficients causaux.

Définition 1.2 (La φ-dépendance, Dedecker et Prieur [29]) *Soit $\mathcal{M}$ une sous tribu de $\mathcal{A}$. Pour tout $X \in \mathbb{L}^\infty(E^l)$ avec $l \geqslant 1$ on définit*

$$\varphi(\mathcal{M}, X) = \sup\left\{\|\mathbb{E}(g(X)|\mathcal{M}) - \mathbb{E}(g(X))\|_\infty, g \in \Lambda_1^{(l)}\right\} .$$

La suite $\varphi_k(r)$ est définie pour tout $k \geqslant 1$ par

$$\varphi_k(r) = \max_{l \leqslant k} \frac{1}{l} \sup_{i+r \leqslant j_1 < j_2 < \cdots < j_l} \varphi(\mathcal{P}_i, (X_{j_1}, \ldots, X_{j_l})) .$$

Le processus est φ-dépendant si $\varphi(r) = \sup_{k>0} \varphi_k(r)$ tend vers 0 lorsque $r \to \infty$.

Un coefficient moins restrictif que φ et valable pour des processus non bornés est le coefficient de couplage τ.

Définition 1.3 (La τ-dépendance, Dedecker et Prieur [28]) *Soit $\mathcal{M}$ une sous tribu de $\mathcal{A}$. Pour tout $X \in \mathbb{L}^1(E^l)$ avec $l \geqslant 1$ on définit*

$$\tau(\mathcal{M}, X) = \left\|\sup\left\{|\mathbb{E}(f(X) \mid \mathcal{M}) - \mathbb{E}(f(X))|,\ f \in \Lambda_1^{(l)}\right\}\right\|_1 .$$

La suite $\tau_k(r)$ est définie pour tout $k \geqslant 1$ par

$$\tau_k(r) = \max_{l \leqslant k} \frac{1}{l} \sup_{i+r \leqslant j_1 < j_2 < \cdots < j_l} \tau(\mathcal{P}_i, (X_{j_1}, \ldots, X_{j_l})) .$$

Le processus est τ-dépendant si $\tau(r) = \sup_{k>0} \tau_k(r)$ tend vers 0 lorsque $r \to \infty$.

Dans toute la suite, nous supposons que l'espace $(E, \mathcal{A})$ est suffisamment riche pour qu'il existe une version X^* distribuée comme X et indépendante de $\mathcal{M}$ telle que $\tau(\mathcal{M}, X) = \|X - X^*\|_1$. Ce résultat de couplage définit le coefficient τ comme le minimum $\tau(\mathcal{M}, X) = \min \|X - X'\|_1$ pour toute version X' distribuée comme X et indépendante de $\mathcal{M}$ (voir Dedecker et Prieur [28] pour plus de détails).

1.1.2 Cas non causal

Le cas non causal renvoie aux coefficients pour lesquels passé et futur jouent un rôle symétrique. Les modèles correspondants, appelés modèles non causaux, sont très généraux.

Définition 1.4 (La η-dépendance, Doukhan et Louhichi [41]) *Le processus $(X_t)_{t\in\mathbb{Z}}$ est dit être (ε,ψ)-faiblement dépendant si il existe une suite $\varepsilon(r)\downarrow 0$ (lorque $r\uparrow\infty$) et une fonction ψ : $\mathbb{N}^2\times(\mathbb{R}^+)^2\to\mathbb{R}^+$ telle que*

$$|Cov(f(X_{s_1},\dots,X_{s_u}),g(X_{t_1},\dots,X_{t_v}))|\leqslant\psi(u,v,Lip\,f,Lip\,g)\varepsilon(r),$$

pour tout $r\geqslant 0$ et tout $(u+v)$-uplets tels que $s_1\leqslant\cdots\leqslant s_u\leqslant s_u+r\leqslant t_1\leqslant\cdots\leqslant t_v$ et $f,g\in\Lambda^{(u)}\times\Lambda^{(v)}$ avec $\sup_x|f(x)|\leqslant 1$ et $\sup_x|g(x)|\leqslant 1$. La η-faible dépendance correspond à $\psi(u,v,a,b)=ua+vb$, dans ce cas on écrit $\varepsilon(r)=\eta(r)$.

L'étude des propriétés de stabilité des suites faiblement dépendantes par transformation non linéaire nous a amenés à introduire le coefficient λ.

Définition 1.5 (La λ-dépendance, [Wint2]) *La λ-faible dépendance correspond à $\psi(u,v,a,b)=uvab+ua+vb$. On écrit alors $\varepsilon(r)=\lambda(r)$.*

Remarque 1.1 *D'autres coefficients (κ, κ', θ,...) découlent de choix de ψ différents, voir chapitre 3 pour plus de détails.*

1.1.3 Comparaison entre les coefficients

Remarquons la relation suivante entre les coefficients : $\lambda(r)\leqslant\eta(r)\leqslant\tau(r)\leqslant\varphi(r)$. Un processus φ-faiblement dépendant est aussi $\tau-,\eta-,\lambda-$faiblement dépendant car $\varphi(r)\downarrow 0$ avec $r\uparrow\infty$ entraîne que $\tau(r),\eta(r),\lambda(r)\downarrow 0$. Un arbitrage est possible entre l'optimalité des résultats asymptotiques et le choix d'un coefficient plus restrictif. Ainsi dans le cas causal, plus restrictif que le non causal, les résultats obtenus sont meilleurs car plus proches de ceux du cadre de référence, l'indépendance. Dans la suite, pour un modèle fixé, nous ne parlerons que de la notion de dépendance faible la plus restrictive, c'est à dire celle correspondant à la plus grande suite de coefficients tendant vers 0. Cette démarche assure d'obtenir les meilleurs résultats asymptotiques pour un modèle donné.

1.2 Les modèles causaux [Wint1, Wint3, Wint4]

1.2.1 Systèmes dynamiques

Nous avons déjà vu à l'exemple 1.1 un processus non mélangeant. Nous généralisons ici ce contre-exemple en suivant la démarche de Dedecker et Prieur [29]. La solution stationnaire de l'équation (1.1) s'écrit sous la forme

$$X_t=\sum_{j=0}^{\infty}\left(\frac{1}{2}\right)^{j+1}\xi_{t-j}.$$

On en déduit que :

- X_t est un réel dans $[0;1]$ dont le développement dyadique est $0,\xi_t\xi_{t-1}\cdots$,
- la loi marginale de X_t est la loi uniforme sur $[0;1]$,
- $X_{t-1} = T(X_t)$ avec $T(x) = 2x$ modulo 1.

$(X_t)_{t\in\mathbb{Z}}$ n'est pas mélangeant car X_{t-1} est fonction de X_t dans la dernière relation. La notion de mélange fort introduite par Rosenblatt [103] est définie par :

$$\alpha_r = \sup_{P\in\mathcal{P}_0, F\in\mathcal{F}_r} |\mathbb{P}(P\cap F) - \mathbb{P}(P)\mathbb{P}(F)|.$$

Choisissons $P = \{X_0 \in [0;1/2]\} \in \mathcal{P}_0$, un événement du passé de probabilité 1/2. Alors on a $P = \{T^r(X_r) \in [0;1/2]\} = \{X_r \in T^{r-1}([0;1/2])\}$, un événement du futur $\mathcal{F}_r$. Choisissant $F = P$ dans la définition de α, on obtient $\alpha_r \geqslant |1/2 - 1/4| = 1/2$. La suite α_r ne converge pas vers 0, $(X_t)_{t\in\mathbb{Z}}$ n'est pas mélangeant au sens de Rosenblatt.

De même, les systèmes dynamiques $X_t = T(X_{t-1})$ pour $t > 0$, $T : [0;1] \to [0;1]$ où X_0 est une variable sur $[0;1]$ ne sont pas mélangeants. Les transformations T doivent avoir un comportement chaotique pour entretenir l'aléa du processus dû à X_0. Ainsi, si T est contractante, il existe un point fixe a et le comportement asymptotique de $(X_t)_{t\geqslant 0}$ est déterministe. Les transformations T étudiées sont de type Lasota-Yorke [82].

Définition 1.6 (Transformations de type Lasota-Yorke) *T est une transformation de type Lasota-Yorke si*
- *(Régularité) Il existe une partition $0 = a_0 \leqslant a_1 \cdots \leqslant a_n = 1$ telle que $T \in \mathcal{C}_1$ et $|T'(x)| > 0$ sur $]a_{i-1}, a_i[$ pour tout $i = 1,\ldots,n$.*
- *(Expansivité) Soit I_n un intervalle tel que $(T^n)'$ y soit défini. Il existe $A > 0$ et $s > 1$ tels que $\inf_{x\in I_n} |(T^n)'| > As^n$.*
- *(Mélange topologique) Pour tous ensembles non vides U, V, il existe $n_0 \geqslant 1$ tel que $T^{-n}(U)\cap V \neq \emptyset$ pour tout $n \geqslant n_0$.*

Cette classe de transformations a de nombreuses propriétés remarquables :
- il existe une unique mesure T-invariante μ,
- μ admet une densité $f \in BV$ (voir Viana [110] pour plus de détails).

Le système dynamique (stationnaire) de type Lasota-Yorke associé à une transformation T correspond à $X_t = T(X_{t-1})$ et $X_0 \rightsquigarrow \mu$, sa mesure T-invariante. Alors
- pour toute fonction $f \in \Lambda^{(l)}$,

$$\mathbb{E}(f(X_1,\ldots,X_l) \mid X_l) = g(X_l), \tag{1.2}$$

  avec $g \in BV$ vérifiant $\|g\|_{BV} \leqslant C\, l \operatorname{Lip} f$ (Collet *et al.* [20]).

– il existe $0 \leqslant \rho < 1$ et une constante $C > 0$ tels que (Dedecker et Prieur [29])

$$\sup_{g \in BV_1} |\mathbb{E}(g(X_0) \mid X_r, X_{r+1}, \ldots) - \mathbb{E}g(X_0)| \leqslant C\rho^r.$$

– $(X_t)_{t \geqslant 0}$ est φ-faiblement dépendant avec $\varphi(r) \leqslant C\rho^r$, $0 \leqslant \rho < 1$ et $C > 0$, voir Dedecker et Prieur [29] pour plus de détails.

Remarque 1.2 *En fait il existe deux autres constantes C et C' telles que, pour toutes fonctions $f_1, \ldots, f_l \in BV_1$ on ait*

$$|\mathbb{E}(f_1(X_{i_1}) \cdots f_l(X_{i_l}) | \sigma(X_t, t \geqslant r + i_l)) - \mathbb{E}(f_1(X_{i_1}) \cdots f_l(X_{i_l}))| \leqslant C' \left(1 + C + \cdots + C^{l-1}\right) \rho^r.$$

Cette propriété de faible dépendance est plus fine que celle induite par le contrôle de φ, voir Dedecker et al. [27].

1.2.2 Chaînes de Markov

Une série stationnaire $(X_t)_{t \in \mathbb{Z}}$ vérifie la propriété de Markov (forte) à l'ordre p lorsque

$$\mathcal{L}(X_t \mid X_{t-1}, \ldots) = \mathcal{L}(X_t \mid X_{t-1}, \ldots, X_{t-p}).$$

De manière équivalente, il existe un couple $(F, (\xi_t)_{t \in \mathbb{Z}})$ tel que $X_t = F(X_{t-1}, \ldots, X_{t-p}; \xi_t)$ où $F : E^p \times E' \to E$ et $(\xi_t)_{t \in \mathbb{Z}}$ est un processus i.i.d. à valeur dans E'. On peut se restreindre au cas où ξ_t suit la loi uniforme sur $E' = [0; 1]$. Remarquons qu'une chaîne de Markov à l'ordre p à valeur dans E est en fait aussi une chaîne de Markov d'ordre 1 à valeur dans E^p.

Les premiers exemples de chaînes de Markov non mélangeants viennent naturellement dans le cas $E = [0; 1]$. Tout comme pour l'exemple 1.1 qui est une chaîne de Markov, il est toujours possible de faire correspondre à un système dynamique une chaîne de Markov associée. Barbour *et al.* [6] ont mis en évidence l'existence d'une chaîne de Markov d'ordre 1 $(X_t)_{t \in \mathbb{Z}}$ telle que $(X_0, \ldots, X_n)$ ait même loi que $(Y_n, \ldots, Y_0)$, où $(Y_t)_{t \geqslant 0}$ est un système dynamique de type Lasota-Yorke. Remarquons qu'il y a alors une inversion des indices qui s'opère, c'est à dire une inversion entre le passé et le futur.

Exemple 1.2 (Chaîne de Markov associée à un système dynamique) *Soit $(X_t)_{t \in \mathbb{Z}}$ une chaîne de Markov associée à un système dynamique de type Lasota-Yorke. Alors cette chaîne n'est pas mélangeante mais bien faiblement dépendante : il existe $0 < \rho < 1$ et $C > 0$ tels que $\varphi(r) \leqslant C\rho^r$.*

Dans la suite nous ne distinguerons pas toujours ces chaînes de Markov de leurs systèmes dynamiques associés.

Pour les modèles plus généraux suivants ($E \neq [0;1]$), des contrôles sur les coefficients de mélange existent sous des conditions d'absolue continuité sur la distribution de ξ_0 (voir Doukhan [38]). La dépendance faible permet de lever ces hypothèses (voir le chapitre 2 pour plus de détails).

Exemple 1.3 (Chaînes de Markov d'ordre p [Wint4]) *Soit l'équation*

$$X_t = F(X_{t-1}, \ldots, X_{t-p}; \xi_t), \quad \forall t \in \mathbb{Z}, \tag{1.3}$$

avec F vérifiant, pour un indice $m \geqslant 1$,

$$\begin{aligned} \|F(x_{t-1}, \ldots, x_{t-p}; \xi_0)\|_m &< \infty, \\ \|F(x_{t-1}, \ldots, x_{t-p}; \xi_0) - F(y_{t-1}, \ldots, y_{t-p}; \xi_0)\|_m &\leqslant \sum_{i=1}^{p} a_j \|x_{t-i} - y_{t-i}\|, \\ \sum_{i=1}^{p} a_j \leqslant a &< 1. \end{aligned}$$

L'existence d'une solution stationnaire $(X_t)_{t\in\mathbb{Z}}$ de l'équation (1.3) dans $\mathbb{L}^m$ est démontrée dans le chapitre 2. De plus, cette solution est τ-faiblement dépendante avec $\tau(r) \leqslant Ca^{r/p}$ avec $C > 0$.

Remarque 1.3 *Bougerol [17] prouve l'existence presque sûre de tels modèles sous des conditions plus faibles. Mais l'existence du moment d'ordre 1 n'est pas assurée par ses résultats.*

Remarque 1.4 *Duflo [51] et Dedecker et Prieur [28] ont montré l'existence de $0 < \rho < 1$ et $C > 0$ tels que $\tau(r) \leqslant C\rho^r$. Nous affinons ce résultat dans [Wint4] en établissant que $\rho \leqslant a < 1$.*

Naturellement plusieurs formes pour F sont possibles. Nous en rappelons quelques unes classiques :

- **Modèles AutoRégressifs d'ordre p ($AR(p)$)**
  Ici $E' = E$ et
  $$F(x_1, \ldots, x_p; s) = \sum_{i=1}^{p} A_i x_i + s.$$
  Dans le cas $E = \mathbb{R}^d$, les A_i sont des matrices de taille $d \times d$ et on a $\sum_{i=1}^{p} \|A_i\| = a$ avec $\|\cdot\|$ une norme matricielle.
- **Modèles autorégressifs non linéaires**
  Ici $E' = E$ et
  $$F(x_1, \ldots, x_p; s) = R(x_1, \ldots, x_p) + s.$$
  Si R vérifie $\|R(x_1, \ldots, x_p) - R(y_1, \ldots, y_p)\| \leqslant \sum_{i=1}^{p} \mathrm{Lip}\, R_i \|x_i - y_i\|$ alors $\sum_{i=1}^{p} \mathrm{Lip}\, R_i = a$.
- **Modèles $AR(1)$ à coefficients aléatoires**
  Ici on considère l'équation $X_t = A_t X_{t-1} + \zeta_t$, où $(A_t)_{t\in\mathbb{Z}}$ est une série i.i.d. de matrice aléatoire

$d \times d$ et $(\zeta_t)_{t\in\mathbb{Z}}$ un processus i.i.d. à valeur dans E. L'équation peut se mettre sous la forme (1.3) avec $F(x_1;(a,\zeta)) = ax_1 + \zeta$ où l'innovation $\xi_t = (A_t,\zeta_t)$ est double. Le modèle satisfait alors $\|A_0\|_m = a$. Ce modèle a de nombreuses propriétés remarquables et applications, voir Diaconis et Friedman [31].

1.2.3 Chaînes à mémoire infinie

Cette fois-ci nous supposons que la série $(X_t)_{t\in\mathbb{Z}}$ satisfait une équation non markovienne du type

$$X_t = F(X_{t-1},\ldots;\xi_t), \quad \forall t\in\mathbb{Z}, \quad \text{presque partout (p.p.)}, \tag{1.4}$$

avec, pour un $m \geqslant 1$,

$$\begin{aligned}
\|F(0,\ldots;\xi_0)\|_m &< \infty, &(1.5)\\
\|F(x_1,\ldots;\xi_0) - F(y_1,\ldots;\xi_0)\|_m &\leqslant \sum_{j=1}^{\infty} a_j\|x_i - y_j\|, &(1.6)\\
\sum_{j=1}^{\infty} a_j \leqslant a &< 1.
\end{aligned}$$

Demander la condition (1.6) pour n'importe quelles suites $(x_t)_{t\in\mathbb{N}}$ et $(y_t)_{t\in\mathbb{N}}$ est inadapté car on peut toujours prendre $\|x_j - y_j\| = a_j^{-1}$ et le terme de droite n'est alors pas défini. Nous nous restreignons donc à l'ensemble $\mathbb{R}^{(\mathbb{N})}$ des suites $(x_t)_{t\in\mathbb{N}}$ de E pour lesquelles il existe $N\in\mathbb{N}$ tel que $\forall t\geqslant N$ on ait $x_t = 0$.

La difficulté du problème réside dans le fait que tout le passé $(X_{t-1},\ldots)$ apparaît dans (1.4), alors que F n'est définie que sur les suites nulles à partir d'un certain rang. L'idée est d'abord de prouver l'existence du processus $(X_t)_{t\in\mathbb{Z}}$ comme étant la limite en p dans $\mathbb{L}^m(E)$ des chaînes de Markov $(X_t^{(p)})_{t\in\mathbb{Z}}$ d'ordre p vérifiant

$$X_t^{(p)} = F(X_{t-1}^{(p)},\ldots,X_{t-p}^{(p)},0,\ldots;\xi_t), \quad \forall t\in\mathbb{Z}.$$

D'après l'exemple 1.3, $(X_t^{(p)})_{t\in\mathbb{Z}}$ est une suite stationnaire de processus de $L^m(E)$. On en déduit l'existence dans $\mathbb{L}^m$ de la limite $(X_t)_{t\in\mathbb{Z}}$ en p et sa stationnarité. Puis, en utilisant (1.5) et (1.6) on obtient l'existence de la quantité $F(X_{t-1},\ldots;\xi_t)$ comme limite dans $\mathbb{L}^m(E)$ lorsque $p\to\infty$ de $F(X_{t-1}^{(p)},\ldots,X_{t-p}^{(p)},0,\ldots;\xi_t)$ pour tout $t\in\mathbb{Z}$. On a donc $\mathbb{E}\|F(X_{t-1},\ldots;\xi_t)\| < \infty$ et la quantité $F(X_{t-1},\ldots;\xi_t)$ est bien définie presque partout. Finalement l'équation (1.4) est vérifiée presque partout par $(X_t)_{t\in\mathbb{Z}}$.

Nous pouvons donc énoncer le théorème suivant, qui donne à la fois l'existence de telles chaînes à mémoires infinies et leurs propriétés de dépendance faible :

Théorème 1.1 ([Wint4]) *Si les hypothèses (1.5) et (1.6) sont vérifiées pour $m \geqslant 1$ alors il existe un processus stationnaire τ-faiblement dépendant $(X_t)_{t\in\mathbb{Z}}$ solution de l'équation 1.4 tel que :*
– $\mathbb{E}\|X_t\|^m < \infty$ *et*
– $\tau(r) \leqslant C\Big(a^{r/p} + \sum\limits_{k=p}^{\infty} a_k\Big)$ *pour tout* $p \in \mathbb{N}^*$ *où* C *ne dépend pas de* p.
Cette solution est la seule telle que $(X_j)_{j<t}$ et ξ_t soient indépendants pour tout t.

La preuve de ce théorème est donnée dans le chapitre 2.

Remarque 1.5 *Dans des cas particuliers de décroissance des coefficients $(a_j)_{j\geqslant 1}$, nous obtenons*
– *si $a_j \leqslant ce^{-\beta j}$ avec $0 < c$ alors il existe $\alpha, C > 0$ tel que $\tau(r) \leqslant Ce^{-\sqrt{\alpha\beta r}}$,*
– *si $a_j \leqslant cj^{-\beta}$ avec $\beta > 1$ et $0 < c$ alors $\tau(r) \leqslant C\,(\log r/r)^{\beta-1}$.*

Remarque 1.6 *On montre que $(X_t)_{t\in\mathbb{Z}}$ est τ-faiblement dépendante car limite en p des chaînes de Markov$(X_t^{(p)})_{t\in\mathbb{Z}}$ d'ordre p elles-mêmes τ-faiblement dépendantes (voir exemple 1.3).*

Nous donnons maintenant des exemples de fonctions F vérifiant les conditions (1.5) et (1.6). Les modèles ainsi obtenus sont classiques et utiles pour les applications en économétrie. Le premier modèle correspond au cas réel $E = E' = \mathbb{R}$.

Exemple 1.4 (Modèles affines réels [Wint4]) *Le modèle affine $(X_t)_{t\in\mathbb{Z}}$ est l'équation (1.4) où*

$$F(x_1, \ldots; u) = ug(x_1, \ldots) + f(x_1, \ldots). \tag{1.7}$$

Les conditions (1.5) et (1.6) sont vérifiées dès que g et f sont lipschitziennes.

Dans les chapitres 4 et 5 sur l'estimation de la densité, les densités marginales et jointes des couples $(X_0, X_k)_{k>0}$ doivent être bornées. Ces conditions sont satisfaites pour ce modèle d'après la proposition suivante :

Proposition 1.1 ([Wint4]) *Si il existe $\varepsilon > 0$ tel que $g > \varepsilon$ et si les innovations $(\xi_t)_{t\in\mathbb{Z}}$ admettent une densité marginale bornée f_ξ alors les densités marginales $f_{X_1,\ldots,X_n}$ de $(X_1, \ldots, X_n)$ existent pour tout $n > 0$ et satisfont pour une constante $c > 0$*

$$\|f_{X_1,\ldots,X_n}\|_\infty \leqslant c\|f_\xi\|_\infty^n.$$

Deux généralisations multivariées des modèles affines réels sont possibles. Le produit ug dans (1.7) est remplacé soit par $u \cdot g$ où u est une matrice et g un vecteur fonction du passé, soit par $M \cdot u$ où M est une matrice fonction du passé et u un vecteur. Les propriétés de ces deux généralisations sont différentes.

Exemple 1.5 (Modèles affines [Wint4]) *Le modèle affine $(X_t)_{t\in\mathbb{Z}}$ est l'équation (1.4) où*

$$F(x_1,\ldots;u) = u\cdot g(x_1,\ldots) + f(x_1,\ldots). \tag{1.8}$$

Ici $E = \mathbb{R}^d$, $u \in E' = \mathcal{M}_{d,m}$ l'ensemble des matrices de taille $d\times m$, les fonctions $g\ :\ \mathbb{R}^{d(\mathbb{N})} \to \mathbb{R}^m$ et $f\ :\ \mathbb{R}^{d(\mathbb{N})} \to \mathbb{R}^d$.

Ce cas englobe de nombreux modèles classiques.

1. **Modèles AR(∞)**, correspond au cas $d = m$, $g = Id_d$ et $f(x_1,\ldots) = a_0 + \sum_{i=1}^{\infty} a_i x_i$.
2. **Modèles bilinéaires vectoriels** avec $g(x_1,\ldots) = b_0 + \sum_{i=1}^{\infty} b_i x_i$ et $f(x_1,\ldots) = a_0 + \sum_{i=1}^{\infty} a_i x_i$.
3. **Modèles bilinéaires robustes** avec $g(x_1,\ldots) = b_0 + \sum_{i=1}^{\infty} b_i(x_i)$ et $f(x_1,\ldots) = a_0 + \sum_{i=1}^{\infty} a_i(x_i)$ où les a_i et b_i sont des fonctions lipschitziennes à valeurs dans $\mathbb{R}^d$ et $\mathbb{R}^m$ respectivement.
4. **Modèles ARCH(∞)** dans le cas réel $d = m = 1$ avec $f = 0$ et $g = \sqrt{a_0 + a_1 x_1^2 + \cdots}$.
5. **Modèle LARCH(∞) vectoriels** avec $f = 0$ et $g = a_0 + a_1 x_1 + \cdots$. Ce modèle est largement développé dans Giraitis *et al.* [58].
6. **Modèle NLARCH(∞) vectoriels** avec $f = 0$ et $g = a_0 + a_1(x_1) + \cdots$ où les a_i sont des fonctions lipschitziennes à valeurs dans $\mathbb{R}^m$.

Toutefois, l'application en statistique paramétrique (voir le chapitre 6) est délicate dans ce contexte : Les conditions d'identifiabilité n'y sont pas vérifiées. Dans le contexte paramétrique, nous préférerons l'extension suivante :

Exemple 1.6 (Processus autorégressif avec des erreurs hétéroscédastiques [Wint5]) *Ce processus $(X_t)_{t\in\mathbb{Z}}$ est la solution stationnaire de l'équation (1.4) avec*

$$F(x_1,\ldots;u) = M(x_1,\ldots)\cdot u + f(x_1,\ldots).$$

Ici $E = \mathbb{R}^d$, $E' = \mathbb{R}^m$ et les fonctons $g\ :\ \mathbb{R}^{d(\mathbb{N})} \to \mathcal{M}_{d,m}$ et $f\ :\ \mathbb{R}^{d(\mathbb{N})} \to \mathbb{R}^d$.

Les modèles GARCH multidimensionnels en sont des cas particuliers, voir le chapitre 6 pour plus de détails.

1.3 Les modèles non causaux [Wint1, Wint2, Wint3]

Nous étudions les propriétés de η et λ-faible dépendance de modèles non causaux. Nous nous plaçons dans le cas réel $E = E' = \mathbb{R}$ par commodité. Ce cas se généralise aux espaces de Banach.

1.3.1 Les modèles associés et gaussiens

Un processus $(X_t)_{t\in\mathbb{Z}}$ est associé lorsque $\mathrm{Cov}\,(f(X_1,\ldots,X_n),g(X_1,\ldots,X_n))\geqslant 0$ pour toutes fonctions croissantes $f,g:\mathbb{R}^n\to\mathbb{R}$ telles que la covariance existe. Un processus $(X_t)_{t\in\mathbb{Z}}$ est gaussien lorsque $(X_{t_1},\ldots,X_{t_n})'$ suit une loi normale vectorielle pour tout ensemble d'indices $(t_1,\ldots,t_n)$. La λ-dépendance faible est adaptée à ces deux cas puisque l'on a alors

$$\lambda(r)\leqslant\sup_{j\geqslant r}|\mathrm{Cov}(X_0,X_j)|.$$

1.3.2 Les schémas de Bernoulli avec entrées indépendantes.

Doukhan et Louhichi [41] ont donné les propriétés de dépendance faible des schémas de Bernoulli à entrées (ou innovations) indépendantes. Une entrée est un processus stationnaire $(\xi_t)_{t\in\mathbb{Z}}$ que nous supposons dans cette partie i.i.d. Si $H:\mathbb{R}^{\mathbb{Z}}\to\mathbb{R}$ est telle que

$$\sup_{t\in\mathbb{Z}}\left\|H\left(\xi_{t-j},j\in\mathbb{Z}\right)-H\left(\xi_{t-j}\mathbf{1}_{|j|\leqslant r},j\in\mathbb{Z}\right)\right\|_1\leqslant\delta_r,\tag{1.9}$$

alors le schéma de Bernoulli $(X_t)_{t\in\mathbb{Z}}$ défini par l'équation

$$X_t=H(\xi_{t-j},j\in\mathbb{Z})\qquad t\in\mathbb{Z}\tag{1.10}$$

est un processus stationnaire tel que $\|X_0\|_1<\infty$ dès que la suite $(\delta_r)_{r\in\mathbb{N}}$ converge vers 0 lorsque $r\to\infty$. De plus, ce processus est η-faiblement dépendant avec

$$\eta(r)\leqslant 2\delta_{[r/2]}.$$

Exemple 1.7 (Les moyennes mobiles infinies)
Le cas le plus simple de schéma de Bernoulli est

$$X_t=\sum_{i\in\mathbb{Z}}a_i\xi_{t-i}.\tag{1.11}$$

Si $\mathbb{E}\|\xi_0\|^2\leqslant 1$ *alors ces moyennes sont* η*-faiblement dépendantes avec*

$$\eta(r)\leqslant\sqrt{\sum_{|j|>[r/2]}\|a_j\|^2}.$$

Exemple 1.8 (LARCH(∞) Non Causaux)
On considère ici l'équation

$$X_t=\xi_t\left(a+\sum_{j\neq 0}a_jX_{t-j}\right).$$

Doukhan et al. [47] ont montré l'existence d'une solution stationnaire à cette équation de la forme

$$X_t = \xi_t \left(a + \sum_{k=1}^{\infty} \sum_{j_1,\ldots,j_k \neq 0} a_{j_1}\xi_{t-j_1} \cdots a_{j_k}\xi_{t-j_k} \right).$$

Étant donné les produits infinis d'innovations dans cette expression, les $(\xi_t)_{t\in\mathbb{Z}}$ doivent être bornés pour que X_0 admette un moment. Sous une hypothèse de contraction $\|\xi_0\|_\infty \sum_{j\neq 0} \|a_j\| = a < 1$, il existe un processus stationnaire solution de l'équation et vérifiant $\|X_0\|_\infty < \infty$. Il est η-faiblement dépendant avec

$$\eta(r) \leqslant C \left(a^{r/p} + \sum_{|j|\geqslant p} a_j \right) \quad \textit{pour tout } p \in \mathbb{N}^* \textit{et où } C \textit{ ne dépend pas de } p.$$

Remarque 1.7 *On voit distinctement ici que le coefficient non causal de faible dépendance η se comporte comme τ dans le cas causal de chaîne à mémoire infinie. De manière générale Doukhan et Truquet [48] montrent l'existence et la η faible dépendance de solutions d'équations du type*

$$X_t = F(X_{t-j},\ j \neq 0; \xi_t).$$

1.3.3 Les schémas de Bernoulli lipschitziens avec entrées faiblement dépendantes.

Soit désormais $(Y_t)_{t\in\mathbb{Z}}$ le processus des entrées du schéma (1.10) supposé $\eta-$ ou λ-faiblement dépendant. Alors le processus $(X_t)_{t\in\mathbb{Z}}$ est lui même η ou λ-faiblement dépendant dès que $H : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ est une fonction telle que pour toutes suites $(x_t)_{t\in\mathbb{Z}}, (y_t)_{t\in\mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}$ coïncidant sur tous les indices sauf un, noté $s \in \mathbb{Z}$,

$$|H(x) - H(y)| \leqslant b_s |x_s - y_s|. \tag{1.12}$$

On a alors le lemme suivant

Lemme 1.1 ([Wint2]) *Si $(Y_t)_{t\in\mathbb{Z}}$ est un processus stationnaire avec un moment $m \geqslant 1$ et si la suite $(b_s)_{s\in\mathbb{Z}}$ est telle que $L = \sum_j b_j < \infty$, alors*

- *le processus $X_t = H(Y_{t-j}, j \in \mathbb{Z}) = \lim_{I\to\infty} H\big(Y_{t-j}\mathbb{1}_{\{j\leqslant I\}}, j \in \mathbb{Z}\big)$ est un processus stationnaire tel que $\|X_0\|_m < \infty$.*
- *Si le processus d'entrée $(Y_t)_{t\in\mathbb{Z}}$ est λ-faiblement dépendant (avec des coefficients $\lambda_Y(r)$), alors $(X_t)_{t\in\mathbb{Z}}$ est λ-faiblement dépendant avec*

$$\lambda(r) \leqslant \inf_{2k\leqslant r} \left[2 \sum_{|i|\geqslant k} b_i \|Y_0\|_1 + (2k+1)^2 L^2 \lambda_Y(r-2k) \right].$$

– *Si le processus d'entrée* $(Y_t)_{t\in\mathbb{Z}}$ *est* η*-faiblement dépendant (avec des coefficients* $\eta_Y(r)$*) alors* $(X_t)_{t\in\mathbb{Z}}$ *est* η*-faiblement dépendant et*

$$\eta(r) \leqslant \inf_{2k\leqslant r}\left[2\sum_{|i|\geqslant k} b_i\|Y_0\|_1 + (2k+1)L\,\eta_Y(r-2k)\right].$$

Remarque 1.8 *Les notions de dépendance faible* η *et* λ *satisfont des propriétés d'hérédité par rapport à la classe des transformations* H *lipschitziennes vérifiant (1.12). Une telle propriété permet de décliner une infinité d'exemples de processus* η *ou* λ*-faiblement dépendants en prenant des schémas de Bernoulli d'innovations dépendantes qui peuvent elles-mêmes être des schémas de Bernoulli... Le cas d'innovations gaussiennes ou associées est particulièrement intéressant : il donne des exemples de modèles que seule la* λ*-dépendance faible permet de traiter.*

Nous allons présenter quelques modèles pour montrer la généralité des notions non causales de dépendance faible.

Exemple 1.9 (Les moyennes mobiles infinies à entrées dépendantes) *Dans ce cas* $(X_t)_{t\in\mathbb{Z}}$ *est de la forme linéaire (1.11) et les entrées sont soit* η*-dépendantes, comme par exemple* $\xi_t = H(\zeta_{t-j}; j\in\mathbb{Z})$ *avec* $(\zeta)_{t\in\mathbb{Z}}$ *i.i.d., soit* λ*-dépendantes, comme par exemple* $(\xi_t)_{t\in\mathbb{Z}}$ *gaussien ou associé.*

La structure spécifique des processus linéaires permet d'obtenir des résultats propres à ce cas (voir par exemple Peligrad et Utev [92]). Nous donnons ci-dessous des modèles non linéaires vérifiant aussi l'hypothèse (1.12) mais pour lesquels les résultats de [92]) ne s'appliquent pas.

Exemple 1.10 (Moyenne mobile absolue) *Un exemple simple de schéma non-linéaire à entrées dépendantes est*

$$X_t = \left|\sum_{j\in\mathbb{Z}} a_j\xi_{t-j}\right|.$$

Dans ce cas $b_s \leqslant |a_s|$.

Exemple 1.11 (Processus multiples) *Un autre exemple est le processus solution de l'équation*

$$X_t = \xi_t\left(a + \sum_{j\neq 0} a_j\xi_{t-j}\right),$$

où les entrées $(\xi_t)_{t\in\mathbb{Z}}$ *sont faiblement dépendantes et bornées. Dans ce cas* $b_s \leqslant 2\|\xi_0\|_\infty|a_s|$.

1.3.4 Les schémas de Bernoulli avec entrées faiblement dépendantes.

Dans cette section l'hypothèse de Lipschitziannité est affaiblie pour l'hypothèse

$$|H(x) - H(y)| \leqslant b_s(\|z\|_\infty^\ell \vee 1)|x_s - y_s|, \qquad \ell \geqslant 0, \tag{1.13}$$

où $z \in \mathbb{R}^\mathbb{Z}$ est défini par $z_s = 0$ et $z_i = x_i = y_i$ lorsque $i \neq s$. Ici $\|x\|_\infty = \sup_{i\in\mathbb{Z}} |x_i|$.

Théorème 1.2 (Schémas de Bernoulli à entrées dépendantes, [Wint2]) *Soit $(Y_t)_{t\in\mathbb{Z}}$ un processus stationnaire et $H : \mathbb{R}^\mathbb{Z} \to \mathbb{R}$ vérifiant (1.13) pour $\ell > 0$ et pour une suite de $(b_j)_{j\in\mathbb{Z}}$ telle que $\sum_j |j|b_j < \infty$. Supposons qu'il existe un couple de réels (m, m') avec $\|Y_0\|_{m'} < \infty$, $m \geqslant 1$, $m' \geqslant (\ell + 1)m$ et $m' > \ell + 1$ si $m = 1$. Alors,*

- *le processus $X_t = H(Y_{t-j}, j \in \mathbb{Z})$ est bien défini dans $\mathbb{L}^m$ et est stationnaire ;*
- *si le processus d'entrée $(Y_t)_{t\in\mathbb{Z}}$ est λ-faiblement dépendant (avec des coefficients $\lambda_Y(r)$), alors $(X_t)_{t\in\mathbb{Z}}$ est λ-faiblement dépendant et il existe une constante $c > 0$ telle que*

$$\lambda(r) = c \inf_{k\leqslant[r/2]} \left[\sum_{|j|\geqslant k} |j|b_j + (2k+1)^2 \lambda_Y(r-2k)^{\frac{m'-1-\ell}{m'-1+\ell}} \right];$$

- *si le processus d'entrée $(Y_t)_{t\in\mathbb{Z}}$ est η-faiblement dépendant (avec des coefficients $\eta_Y(r)$) alors $(X_t)_{t\in\mathbb{Z}}$ est η-faiblement dépendant et il existe une constante $c > 0$ telle que*

$$\eta(r) = c \inf_{k\leqslant[r/2]} \left[\sum_{|j|\geqslant k} |j|b_j + (2k+1)^{1+\frac{\ell}{m'-1}} \eta_Y(r-2k)^{\frac{m'-1-\ell}{m'-1}} \right].$$

Remarque 1.9 *Dans le tableau suivant, le calcul explicite des coefficients est donné d'après les bornes obtenues dans le théorème 1.2 :*

Coefficients de H	*Dépendance des innovations*	*Dépendance du schéma de Bernoulli*
$b_j \leqslant C(\lvert j\rvert+1)^{-b}$	$\lambda_Y(r) \leqslant Dr^{-\lambda}$	$\lambda(r) \leqslant cr^{-\lambda\left(1-\frac{2}{b}\right)\frac{m'-1-\ell}{m'-1+\ell}}$
$b_j \leqslant C(\lvert j\rvert+1)^{-b}$	$\eta_Y(r) \leqslant Dr^{-\eta}$	$\eta(r) \leqslant cr^{-\eta\frac{(b-2)(m'-1-\ell)}{(b-1)(m'-1)-\ell}}$
$b_j \leqslant C(\lvert j\rvert+1)^{-b}$	$\lambda_Y(r) \leqslant De^{-r\lambda}$	$\lambda(r) \leqslant cr^{2-b}$
$b_j \leqslant C(\lvert j\rvert+1)^{-b}$	$\eta_Y(r) \leqslant De^{-r\eta}$	$\eta(k) \leqslant cr^{2-b}$
$b_j \leqslant Ce^{-\lvert j\rvert b}$	$\lambda_Y(r) \leqslant Dr^{-\lambda}$	$\lambda(r) \leqslant cr^{-\lambda\frac{m'-1-\ell}{m'-1+\ell}} \log^2 r$
$b_j \leqslant Ce^{-\lvert j\rvert b}$	$\eta_Y(r) \leqslant Dr^{-\eta}$	$\eta(r) \leqslant cr^{-\eta\frac{m'-1-\ell}{m'-1}} \log^{1+\frac{\ell}{m'-1}} r$
$b_j \leqslant Ce^{-\lvert j\rvert b}$	$\lambda_Y(r) \leqslant De^{-r\lambda}$	$\lambda(r) \leqslant cr^2 e^{-\lambda r\frac{b(m'-1-\ell)}{b(m'-1+\ell)+2\eta(m'-1-\ell)}}$
$b_j \leqslant Ce^{-\lvert j\rvert b}$	$\eta_Y(r) \leqslant De^{-r\eta}$	$\eta(r) \leqslant cr^{\frac{m'-1-\ell}{m'-1}} e^{-\eta r\frac{b(m'-1-\ell)}{b(m'-1)+2\eta(m'-1-\ell)}}$

Cette condition de dépendance faible est satisfaite pour de nombreux modèles.

Exemple 1.12 (Chaos de Volterra avec entrées dépendantes) *Les schémas H de la forme*

$$H(x) = \sum_{k=0}^{K} \sum_{j_1,\ldots,j_k} a^{(k)}_{j_1,\ldots,j_k} x_{j_1} \cdots x_{j_k}$$

sont faiblement dépendants. Ces modèles sont solutions d'équations du type $X_t = F(X_{t-j}, j \neq 0; \xi_t)$. Si, par exemple, $|a^{(k)}_{j_1,\ldots,j_k}| \leqslant C\,(j_1 \vee \cdots \vee j_k)^{-\alpha}$ ou $|a^{(k)}_{j_1,\ldots,j_k}| \leqslant C \exp\left(-\alpha(j_1 \vee \cdots \vee j_k)\right)$ alors on a respectivement $b_s \leqslant C' s^{d-1-\alpha}$ ou $b_s \leqslant C' e^{-\alpha s}$ pour tout $C' > 0$.

1.3.5 Conclusion sur les modèles

Au travers de ces exemples de modèles, nous avons pu constater que la dépendance faible est plus performante que les notions de mélange dans le cas causal. Dans le cas non causal, la dépendance faible est un outil moins contraignant que le choix d'un modèle.

Dans le cas causal, la dépendance faible est un outil performant pour étudier les caractéristiques de dépendance des modèles. Notamment, elle permet d'étudier les systèmes dynamiques et les chaînes à mémoire infinie alors que les notions classiques de mélange n'en sont pas capables.

Les exemples de modèles faiblement dépendants dans le cas non causal sont infinis. Grâce aux propriétés d'hérédité, des modèles très complexes sont faiblement dépendants. Partant de ce constat, il est moins restrictif de travailler à partir d'hypothèses sur les notions de dépendance faible que sur des modèles, ce qui correspond en fait à une hypothèse plus forte. L'utilisation de la dépendance faible est moins contraignante que la modélisation.

1.4 Le principe d'invariance faible [Wint2]

Nous présentons une première extension d'un résultat asymptotique classique dans le cas indépendant, mais nouveau dans le cas faiblement dépendant. Dans toute cette section $E = \mathbb{R}$.

1.4.1 Le principe d'invariance faible

Le principe d'invariance faible est un résultat de convergence en loi. Une suite de variables aléatoires $(X_n)_{n \geqslant 0}$ converge en loi vers la variable X dès que, pour toute fonction g continue bornée de $\mathbb{R}$ dans $\mathbb{R}$, on a $\mathbb{E}(g(X_n)) \to \mathbb{E}(g(X))$ lorsque $n \to \infty$. Nous noterons dans la suite $X_n \Rightarrow X$. Pour un processus stationnaire réel centré $(X_t)_{t \in \mathbb{Z}}$, le processus des sommes partielles associé est défini par

$$W_n : t \in [0;1] \mapsto \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} X_i.$$

Ici $[nt]$ est la partie entière de nt. Pour $t = 1$, on retrouve la somme des $(X_i)_{1 \leqslant i \leqslant n}$ renormalisée par $1/\sqrt{n}$. Pour des processus à mémoire courte, la limite $\lim_{n\to\infty} \mathrm{Var}\left(\sum_{i=1}^n X_i\right)/n$ existe et on la note σ^2. Dans le cas indépendant $\sigma^2 = \mathrm{Var}\,(X_0)$ alors que dans le cas de la dépendance faible, $\sigma^2 = \sum_{k\geqslant 0} \mathrm{Cov}(X_0, X_k) \geqslant 0$ (car c'est la limite d'une suite de variances positives). On se placera dans le cas non dégénéré à savoir $\sigma^2 \neq 0$ sans le vérifier ; ce sera une hypothèse supplémentaire.

Soit $W(t)$ le mouvement brownien standard. Par définition, pour tout $\omega \in \Omega$ les trajectoires $t \in [0; 1] \mapsto W_\omega(t)$ sont des fonctions continues. De plus, $W(0) = 0$ presque sûrement et pour tous $0 \leqslant t < t' \leqslant 1$, $W(t') - W(t)$ suit une loi normale $\mathcal{N}(0, t' - t)$ indépendante de la variable aléatoire $W(t)$.

Pour tout $n \geqslant 1$, la variable aléatoire $W_n(t)$ est à valeur dans l'espace des fonctions cadlag, continue à droite et ayant une limite à gauche. Cet espace fonctionnel est noté $\mathcal{D}(0, 1)$, il est métrisable par la distance de Skorohod notée d (voir par exemple Billingsley [12] pour plus de détails). Le principe d'invariance faible est la convergence en loi du processus W_n vers le mouvement brownien W dans l'espace de Skorohod $\mathcal{D}(0, 1)$. Quelque soit la fonction g continue bornée,

$$d(\mathbb{E}(g(W_n)), \mathbb{E}(g(\sigma W))) \to_{n\to\infty} 0.$$

Par composition, on vérifie que $t \mapsto \mathbb{E}(g(W_n(t)))$ et $t \mapsto \mathbb{E}(g(W(t)))$ sont bien des fonctions de $[0; 1]$ dans $\mathbb{R}$ cadlag. Le principe d'invariance faible, appelé aussi théorème de Donsker, s'écrit plus simplement

$$W_n(t) \Rightarrow \sigma^2 W(t) \text{ dans } \mathcal{D}(0, 1).$$

Sous une condition de décroissance du coefficient de dépendance faible λ, le principe d'invariance faible est assuré (la preuve est présentée dans le chapitre 3) :

Théorème 1.3 (λ-dépendance) *Si $(X_t)_{t\in\mathbb{Z}}$ admet un moment d'ordre $m > 2$ et si $\lambda(r) = \mathcal{O}(r^{-\lambda})$ (quand $r \uparrow \infty$) avec $\lambda > 4 + \frac{2}{m-2}$ alors σ^2 est fini et*

$$W_n(t) \Rightarrow \sigma W(t), \quad \textit{dans l'espace de Skorohod } D(0, 1).$$

Remarque 1.10 *Dans le cas de η-faible dépendance le principe d'invariance faible a lieu dès que $\eta(r) = \mathcal{O}(r^{-\eta})$ avec $\eta > 3 + \frac{3}{m-2}$. Dans le cas de τ-faible dépendance (cas causal) le principe d'invariance faible a lieu avec $\tau(r) = \mathcal{O}(r^{-\tau})$ avec $\tau > 1 + \frac{1}{m-2}$. Plus la dépendance faible est restrictive et plus les conditions sur les coefficients sont faibles.*

1.4.2 Vitesses de convergence dans le TLC

Le théorème central limite (TLC) est le résultat de convergence en loi classique

$$W_n(1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \Rightarrow \sigma W(1) \rightsquigarrow \sigma \mathcal{N}(0,1) \text{ lorsque } n \to \infty.$$

Il est équivalent à la convergence des fonctions caractéristiques $\phi_{W_n(1)}(t) \to \phi_{W(1)}(t)$ pour tout $t \in \mathbb{R}$ avec $\phi_X(t) = \mathbb{E}\exp(itX)$. On obtient dans le chapitre 3 une borne sur la vitesse de convergence

$$\left|\phi_{W_n(1)}(t) - \phi_{W(1)}(t)\right| = o\left(n^{-c}\right), \text{ pour tout } t \in \mathbb{R} \text{ et pour } 0 < c < c^*.$$

La taux c^* ne dépend que des paramètres m et λ. Il vérifie $c^* < \frac{1}{4}$ et plus exactement $c^* < (m-2)/(2m-2)$ lorsque $m < 3$. Une perte est observée car dans le cas i.i.d. la vitesse $c^* = 1/2$ est atteinte.

Une autre vitesse de convergence dans le TLC est calculée. Elle correspond au choix de la norme uniforme sur les fonctions de répartitions de $W_n(1)$ et $W(1)$. Elle s'appelle l'inégalité de Berry Essen dans le cas indépendant. On montre que :

$$\sup_{x \in \mathbb{R}} \left|\mathbb{P}\left(W_n(1) \leqslant x\right) - \mathbb{P}\left(W(1) \leqslant x\right)\right| = o\left(n^{-c}\right), \qquad \text{pour } c < c'.$$

Dans le cas indépendant, cette inégalité est valable pour $c = \frac{1}{2}$, dans le cas du mélange fort $c = \frac{1}{3}$ (voir [101]). Dans notre cas, elle n'est valable que pour c' plus petit que $\frac{1}{12}$.

1.5 L'estimation de la densité non adaptative [Wint1]

Nous nous plaçons ici dans le cas $E = \mathbb{R}^d$ pour $d \geqslant 1$. Les séries étant stationnaires, elles admettent la même distribution marginale, i.e. les X_t ont même loi de probabilité P pour tout $t \in \mathbb{Z}$. On suppose dans la suite que cette loi est absolument continue par rapport à la mesure de Lebesgue et admet donc une densité f. Nous souhaitons estimer f grâce à un estimateur $\hat{f}_n$ à partir des observations $X_1, \ldots, X_n$. Ce problème est très classique dans le cas i.i.d., voir par exemple le livre de Tsybakov [108].

Nous présentons ici une large classe d'estimateurs $\hat{f}_n$ convergeants tous vers f. Nous donnons aussi des vitesses de convergences communes à tous ces estimateurs. Dans le cas de vitesses associées à la convergence uniforme, une perte apparaît par rapport au cas de référence de dépendance. De plus, cette perte dépend des propriétés de dépendance faible du modèle. Dans toute la suite $\lceil x \rceil$ est le plus petit entier majorant x.

1.5.1 L'estimateur $\hat{f}_n$

La régularité de f, notée s, est supposée connue. Nous étudions des estimateurs $\hat{f}_n$

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{m_n}(x, X_i) \ , \tag{1.14}$$

avec $(K_{m_n})_{n\geqslant 1}$ qui est une suite de fonctions numériques appelées noyaux généralisés. Les estimateurs $\hat{f}_n$ sont appelés linéaires et regroupent les estimateurs à noyau, par projection et par ondelettes. Nous travaillons sous des conditions (a), (b), (c) et (d) très générales sur la suite $(K_{m_n})_{n\geqslant 1}$

(a) Le support de K_{m_n} est un ensemble compact de diamètre $O(1/m_n^{1/d})$ lorsque $m_n \uparrow \infty$;

(b) Les fonctions $x \mapsto K_{m_n}(x, y)$ et $x \mapsto K_{m_n}(y, x)$ sont lipschitziennes avec un coefficient de Lipschitz en $O\left(m_n^{1+1/d}\right)$ lorsque $m_n \uparrow \infty$;

(c) Pour tout x dans le support de K_{m_n}, $\int K_{m_n}(x, y)\, dP(y) = 1$;

(d) Le biais de l'estimateur $\hat{f}_n$ définit par (1.14) est de l'ordre $m_n^{-\rho/d}$ lorsque $m_n \uparrow \infty$, uniformément sur toute boule de rayon $M > 0$:

$$\sup_{\|x\|\leq M} \left|\mathbb{E}[\hat{f}_n(x)] - f(x)\right| = O(m_n^{-s/d}) \ . \tag{1.15}$$

Ici m_n est un paramètre tel que $m_n \to \infty$ et $m_n/n \to 0$ lorsque $n \uparrow \infty$ (pour plus de détail voir le chapitre 4). Les conditions (a), (b), (c) et (d) sont vérifiées dans les trois grandes classes d'estimateurs (à noyau, par projection et par ondelettes) sous des hypothèses adéquates.

Estimateurs à noyau

Dans ce cas-là, m_n est appelé paramètre de fenêtre et

$$K_{m_n}(x, y) = \frac{1}{m_n} K\left(m_n^{1/d}(x - y)\right) \ .$$

Les conditions (a), (b) et (c) sont vérifiées pour $K_m(x, y) = mK\left(m^{1/d}(x - y)\right)$ dès qu'il existe $s > 0$ tel que K satisfasse pour tout $j = j_1 + \cdots + j_d$ avec $(j_1, \ldots, j_d) \in \mathbb{N}^d$:

$$\int x_1^{j_1} \cdots x_d^{j_d} K(x_1, \ldots, x_d) f(x_1) \cdots f(x_d) dx_1 \cdots dx_d = \begin{cases} 1 & \text{si } j = 0, \\ 0 & \text{pour } j \in \{1, \ldots, \lceil s-1 \rceil - 1\}, \\ \neq 0 & \text{si } j = \lceil s-1 \rceil. \end{cases}$$

L'hypothèse (d) est vérifiée dès que $f \in \mathcal{C}_s$, c'est-à-dire dès que f est telle que pour $s = \lceil s-1 \rceil + c$ avec $0 < c \leqslant 1$, f est $\lceil s-1 \rceil$-fois continûment différentiable et il existe $A > 0$ tel que $\forall (x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, $|f^{(\lceil s-1 \rceil)}(x) - f^{(\lceil s-1 \rceil)}(y)| \leqslant A|x - y|^c$.

Estimateurs par projection

On se restreint ici au cas $d = 1$. Supposons que la famille des monômes $\{1, x, x^2, \ldots\}$ appartienne bien à $L^2(I, \mu)$, où I est un intervalle borné de $\mathbb{R}$ et μ une mesure sur I. On construit une base orthonormale de $L^2(I, \mu)$ de polynômes $\{P_0, P_1, P_2, \ldots\}$ par le procédé de Gram-Schmidt. On suppose que f est nulle en dehors de I et qu'elle appartient à la classe $\mathcal{C}_s$. Alors il existe une fonction $\pi_{f,m_n} \in V_{m_n} = \text{Vect}\{P_0, \ldots, P_{m_n}\}$ telle que $\sup_{x \in I} |f(x) - \pi_{f,m_n}(x)| = O(m_n^{-s})$. Cette fonction π_{f,m_n} correspond à la projection de f :

$$\pi_{f,m_n}(x) = \int_I f(y) K_m^a(x, y) dy,$$

où $K_m^a(x, y)$ peut toujours s'écrire sous la forme d'un "noyau pondéré"

$$K_m^a(x, y) = \sum_{j=0}^{m} a_{m,j} \sum_{k=0}^{j} P_k(x) P_k(y) \ .$$

Ici $(a_{m,j})_{m \in \mathbb{N},\, 0 \leqslant j \leqslant m}$ est une séquence de poids telle que $\sum_{j=0}^{m} a_{m,j} = 1$ et pour tout j $\lim_{m \to \infty} a_{m,j} = 0$. Selon les paramètres s, μ et I, on peut trouver un tableau triangulaire $(a_{m,j})_{m \in \mathbb{N},\, 0 \leqslant j \leqslant m}$ tel que les conditions (a), (b), (c) et (d) soient satisfaites. L'exemple le plus connu est celui du noyau de Fejer lorsque $I =]0, 2\pi[$, μ est la mesure de Lebesgue renormalisée et $s = 1$.

Estimateurs par ondelettes

L'estimation par ondelettes est un cas particulier de l'estimation par projection. On se restreint aussi au cas $d = 1$.

Definition 1.1 (fonction d'échelle) *Une fonction $\phi \in \mathbb{L}^2(\mathbb{R})$ est une fonction d'échelle si la famille $\{\phi(\cdot - k) \,;\, k \in \mathbb{Z}\}$ est orthonormale.*

On choisit le paramètre de fenêtre dyadique $m_n = 2^{j(n)}$, où j est fonction de n, et on note $V_j = \text{Vect}\{\phi_{j,k}, k \in \mathbb{Z}\}$, avec $\phi_{j,k} = 2^{j/2}\phi(2^j(x - k))$. Lorsque ϕ est à support compact, on définit :

$$\hat{f}_n(x) = \frac{1}{n} \sum_{k=-\infty}^{\infty} \sum_{i=1}^{n} \phi_{j(n),k}(X_i) \phi_{j(n),k}(x) \ .$$

Cet estimateur s'écrit aussi sous la forme de (1.14) avec $K(x, y) = \sum_{k=-\infty}^{\infty} \phi(y - k)\phi(x - k)$ et $K_{m_n}(x, y) = m_n K(m_n x, m_n y)$. Remarquons que les estimateurs par ondelettes sont aussi des estimateurs à noyau. De la même manière, si ϕ est lipschitzienne telle que $\int \phi(x) x^j dx = 0$ avec $0 < j < \lceil s - 1 \rceil$ et $\int \phi(x) x^{\lceil s-1 \rceil} dx \neq 0$, alors le noyau K_{m_n} satisfait (a), (b) and (c). Si $f \in C_s$, alors l'hypothèse (d) est vérifiée.

1.5.2 La dépendance faible des modèles

Les cas de la dépendance causale φ et de la dépendance non causale η sont traités. Les résultats dépendent de la vitesse de décroissance des coefficients. Le cas géométrique correspond aux cas où les hypothèses de dépendance sont les plus fortes, la décroissance des coefficients étant la plus rapide :

[H1] : $\varphi(r) = O\left(e^{-ar^b}\right)$ avec $a > 0$ et $b > 0$,

[H1'] : $\eta(r) = O\left(e^{-ar^b}\right)$ avec $a > 0$ et $b > 0$.

Le cas riemannien a lieu lorsque l'une des deux hypothèses [H2] ou [H2'] est satisfaite :

[H2] : $\varphi(r) = \mathcal{O}(r^{-a})$ avec $a > 1$,

[H2'] : $\eta(r) = O(r^{-a})$ avec $a > 1$.

Remarquons que dans le cas géométrique les résultats sont plus proches de ceux du cas i.i.d. car les coefficients de dépendance sont plus proches de 0. L'hypothèse suivante est classique dans ce contexte :

[H3] : La densité marginale f du processus $(X_t)_{t\in\mathbb{Z}}$ existe et est bornée.

Dans le cas de processus dépendants, on demande une hypothèse supplémentaire portant sur les densités jointes des couples (X_j, X_k), $j \neq k$.

[H4] Les densités $f_{j,k}$ des couples (X_j, X_k) sont uniformément bornées par rapport à $j \neq k$.

Malheureusement l'existence de ces densités jointes n'a pas lieu dans tous les modèles présentés précédemment. Par exemple, la distribution de (X_i, X_{i+1}) n'est pas absolument continue dans l'exemple d'Andrews, solution de l'équation (1.1) et la densité $f_{j,k}$ n'existe pas. Cette remarque est valable pour les systèmes dynamiques en général. Nous traitons donc séparément les cas où :

[H4'] La série $(X_t)_{t\in\mathbb{Z}}$ est un système dynamique de type Lasota-Yorke.

Dans la suite nous nous plaçons soit dans le cas où les densité jointes $f_{j,k}$ existent et sont bornées uniformément soit dans le cas de systèmes dynamiques de type Lasota-Yorke.

1.5.3 Résultats de convergence

Résultats minimax dans le cas i.i.d.

Les résultats ci-dessous sont considérés pour des estimateurs de la forme (1.14) avec une suite de (K_{m_n}) choisie telle que (a), (b), (c) et (d) soient vérifiées. Nous rappelons les taux de convergence classiques du cas indépendant.

Lorsque $X_1, \ldots, X_n$ est un échantillon i.i.d., pour tout point fixé $x \in \mathbb{R}^d$, la suite (m_n) peut être choisie telle que pour tout $0 < q < +\infty$

$$\|\hat{f}_n(x) - f(x)\|_q = O\left(n^{-s/(2s+d)}\right) . \tag{1.16}$$

Ce taux de convergence est optimal : on peut trouver une fonction f telle que cette vitesse soit précisément atteinte (voir [108]). Dans le cas uniforme, pour tout $M > 0$ et pour une suite de paramètres $(m_n)_{n>0}$ bien choisie, on a

$$\begin{aligned}\mathbb{E}\sup_{\|x\|\leqslant M}|\hat{f}_n(x)-f(x)|^q &= \mathcal{O}\left(\frac{\log n}{n}\right)^{qs/(d+2s)}, \\ \sup_{\|x\|\leqslant M}|\hat{f}_n(x)-f(x)| &= \mathcal{O}\left(\frac{\log n}{n}\right)^{s/(d+2s)} \quad \text{presque sûrement.}\end{aligned}$$

Là encore le résultat est optimal, malgré la perte d'un terme logarithmique dans la vitesse. C'est le prix à payer pour avoir un résultat uniforme sur une boule compacte.

Résultats dans le cas dépendant

Dans le cas de la dépendance faible, nous obtenons la convergence des estimateurs linéaires $\hat{f}_n$. Toutefois une perte en vitesse est dans certains cas observée.

Théorème 1.4 (Vitesses associées au risque $\mathbb{L}^q$, [Wint1])

Cas géométrique. *Si les hypothèses [H4] ou [H4'] et [H1] ou [H1'] sont vérifiées, la suite $(m_n)_{n>0}$ peut être choisie telle que (1.16) soit valable pour tout $0 < q < +\infty$.*

Cas riemannien. *Sous les hypothèses [H4] ou [H4'] et*

- *[H2] avec $a > \max(1 + 2/d + (d+1)/s, 2 + 1/d)$ (η-dépendance),*
- *ou [H2'] avec $a > 1 + 2/d + 1/s$ (φ-dépendance),*

alors la suite $(m_n)_{n>0}$ peut être choisie telle que (1.16) soit valable pour tout $0 < q \leqslant q_0 = 2\lceil (a-1)/2\rceil$.

Remarque 1.11 *On obtient des résultats aussi bons que les vitesses optimales du cas indépendant. Le cas indépendant étant un cas particulier de dépendance faible, ces vitesses de convergences sont aussi optimales. Nos résultats généralisent ceux de Doukhan [38] obtenus dans le cas du mélange et de Ango Nze et Doukhan [2] obtenus dans le cas de la η-dépendance.*

Théorème 1.5 (Vitesses uniformes dans le cas géométrique, [Wint1]) *Pour tout $M > 0$, sous les hypothèses [H4] ou [H4'] et [H1] ou [H1'] nous avons, pour tout $0 < q < +\infty$, et pour une suite convenable de paramètres de fenêtres $(m_n)_{n>0}$,*

$$\begin{aligned}\mathbb{E}\sup_{\|x\|\leqslant M}|\hat{f}_n(x)-f(x)|^q &= O\left(\left(\frac{\log^{2(b+1)/b}(n)}{n}\right)^{qs/(d+2s)}\right), \\ \sup_{\|x\|\leqslant M}|\hat{f}_n(x)-f(x)| &=_{p.s.} O\left(\left(\frac{\log^{2(b+1)/b}(n)}{n}\right)^{s/(d+2s)}\right).\end{aligned}$$

Théorème 1.6 (Vitesses uniformes dans le cas riemannien, [Wint1]) *Pour tout $M > 0$, sous [H4] ou [H4'], [H2] ou [H2'] avec $a \geqslant 4$ et $s > 2d$, pour $q_0 = 2\lceil (a-1)/2 \rceil$ et $q \leqslant q_0$, la suite $(m_n)_{n>0}$ peut être choisie telle que*

$$\mathbb{E} \sup_{\|x\| \leqslant M} |\hat{f}_n(x) - f(x)|^q = O\left(n^{-\frac{qs}{d+2s+2d/(q_0+d)}} \right) ,$$

ou telle que

$$\sup_{\|x\| \leqslant M} |\hat{f}_n(x) - f(x)| =_{a.s.} O\left(\left(\frac{\log^{q_0+d}(n)}{n^{q_0-2}} \right)^{\frac{s}{d(q_0+2)+s(q_0+d)}} \right) .$$

Remarque 1.12 *Une perte dans les vitesses de convergence apparaît pour la convergence uniforme. Dans le cas géométrique, cette perte est logarithmique. Dans le cas riemannien, une perte d'un ordre plus élevé, en puissance de n, apparaît. Plus les coefficients de dépendance faible sont élevés, plus on s'éloigne du cas de l'indépendance et plus la perte en vitesse uniforme est grande.*

Remarque 1.13 *Dans le cas de décroissance géométrique en η les résultats sont similaires à ceux de Doukhan et Louhichi [41] pour le cas $b = 1$. Dans le cas Riemanien, ces résultats peuvent être affinés. Nous appliquons la méthode de chaînage de Liebscher [83]. Dans le cas d'estimateurs à noyau, Liebscher réussit à diminuer cette vitesse de convergence en demandant plus de régularité sur le noyau. Le taux dépend alors à la fois de la dépendance faible et de la régularité du noyau. Nous privilégions ici un résultat général qui ne dépend pas du noyau ni donc de l'estimateur $\hat{f}_n$. Nous traitons ainsi d'un coup les trois méthodes d'estimation.*

1.6 L'estimation de la densité adaptative [Wint3]

Nous donnons dans cette section des résultats théoriques sur l'estimation adaptative de la densité dans le cas faiblement dépendant (ici la régularité s de la fonction à estimer f est inconnue). Nous nous plaçons dans le cas particulier de $E = [0; 1]$. Nous présentons d'abord l'estimateur à seuillage dur par ondelette, particulièrement bien adapté au cas adaptatif : les résultats classiques du cas i.i.d. le prouvent. Nous appliquons cet estimateur dans le cas faiblement dépendant et nous obtenons des résultats quasi optimaux.

1.6.1 Estimateur par seuillage dur

Nous présentons ici les résultats de Donoho *et al.* [35]. Nous considérons les ondelettes sur $[0; 1]$ proposées par Daubechies [24].

Définition 1.7 *Une analyse multirésolution de $\mathbb{L}^2$ est une suite croissante $(V_j)_{j \in \mathbb{Z}}$ de sous ensembles fermés de $\mathbb{L}^2$ tels que*

(i) $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}$ *et* $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = \mathbb{L}^2$,

(ii) $\forall\ f \in \mathbb{L}^2$, $\forall\ j \in \mathbb{N}$, $f \in V_j$ *si et seulement si* $x \mapsto f(2^{-j}x)$ *appartient à* V_{j+1},

(iii) *Il existe une fonction* ϕ*, appelée ondelette père, telle que* $\{x \mapsto \phi(x-k)\}_{k\in\mathbb{Z}}$ *soit une base orthonormale* V_0.

Soit W_j tel que $V_{j+1} = V_j \oplus W_j$. Il existe une fonction ψ –appelée ondelette mère– telle que pour tout niveau de résolution $j \geq 0$ les familles $\{\phi_{j,k} : x \to 2^{j/2}\phi(2^j x-k)\}_{k\in\mathbb{Z}}$ et $\{\psi_{j,k} : x \mapsto 2^{j/2}\psi(2^j x-k)\}_{k\in\mathbb{Z}}$ sont des bases orthonormales de V_j et W_j. On choisit ϕ et ψ à support $[0;1]$ et qui vérifient les conditions de moments d'ordre $N \in \mathbb{N}^*$:

$$\forall k = 0\ldots N, \quad \int \phi(x)x^k dx = \delta_{0,k}, \quad \int \left|\phi(x)x^{N+1}\right| dx < \infty,$$
$$\text{et} \qquad x \mapsto \sum_k |\phi(x-k)| \in \mathbb{L}^2,$$

(ici $\delta_{0,k} = 1$ si $k = 0$ et $\delta_{0,k} = 0$ sinon). On se restreint à $N \geqslant 4$ afin d'assurer que les fonctions ϕ et ψ soient lipschitziennes.

Pour tout entier j_0 on décompose une fonction $f \in \mathbb{L}^2$

$$f = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j_0,k}\psi_{j_0,k},$$

avec $\alpha_{j,k} = \int_0^1 f(x)\phi_{j,k}(x)dx$, $\beta_{j,k} = \int_0^1 f(x)\psi_{j,k}(x)dx$. L'estimateur par projection est alors

$$\tilde{f}_n = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{j_1(n)}\sum_{k=0}^{2^j-1} \hat{\beta}_{j,k}\psi_{j,k},$$

avec $\hat{\alpha}_{j,k} = \sum_{i=1}^n \phi_{j,k}(X_i)/n$ et $\hat{\beta}_{j,k} = \sum_{i=1}^n \psi_{j,k}(X_i)/n$.
Soit (s,π,r) un triplet tel que $s > 0$, $1 \leq \pi, r \leq \infty$. L'expression

$$\|f\|_{s,\pi,r} = |\alpha_{0,0}| + \left(\sum_{j\in\mathbb{N}} 2^{j(s+1/2-1/\pi)r}\left(\sum_{k=0}^{2^j-1}|\beta_{j,k}|^\pi\right)^{r/\pi}\right)^{1/r},$$

définit une semi-norme (voir Härdle *et al.* (1998) [67] pour plus de détails). Les espaces de Besov et les boules de Besov sont alors définis par, respectivement,

$$\mathcal{B}^s_{\pi,r} := \{f \in \mathbb{L}^2 \text{ telles que } \|f\|_{s,\pi,r} < +\infty\}, \quad \mathcal{B}^s_{\pi,r}(M) := \left\{f \in \mathcal{B}^s_{\pi,r},\ \|f\|_{s,\pi,r} \leq M\right\}.$$

Remarquons que ces espaces fonctionnels ne dépendent pas de ϕ et ψ et incluent les ensembles fonctionnels classiques (par exemple $\mathcal{C}_s \subset \mathcal{B}^s_{2,2}$). On munit $\mathbb{L}^p$ de sa norme $\|f\|^p_{\mathbb{L}^p} = \int |f(x)|^p dx$ puis on définit le risque moyen de l'estimateur $\widetilde{f}_n$ par $\mathbb{E}\|\widetilde{f}_n - f\|^p_{\mathbb{L}^p}$ et sa vitesse minimax associée par

$$\inf_{f \in \mathcal{B}^s_{\pi,r}(M)} \sup_{\widetilde{f}_n \text{ un estimateur de } f} \mathbb{E}\|\widetilde{f}_n - f\|^p_{\mathbb{L}^p} = \mathcal{O}\left(n^{-\alpha}\right),$$

avec

$$\alpha = \begin{cases} s/(1+2s) & \varepsilon \geqslant 0, \\ (s - 1/\pi + 1/p)/(1+2s-2/\pi) & \varepsilon \leqslant 0, \end{cases} \qquad \text{où } \varepsilon = s\pi - (p-\pi)/2.$$

L'estimateur par projection n'atteint pas la vitesse minimax à une puissance de n près lorsque $f \in \mathcal{B}^s_{\pi,r}(M)$ avec $\pi \leqslant p$ (voir Donoho *et al.* [35]). Ce résultat de sous optimalité a motivé l'introduction d'un seuillage sur l'estimateur par projection. Plus précisément, soit $T_\lambda(\beta) = \beta \mathbb{1}_{|\beta|>\lambda}$ la fonction de seuillage dur de niveau $\lambda > 0$; on définit un nouvel estimateur $\hat{f}_n$

$$\hat{f}_n = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} T_{\lambda_j}(\hat{\beta}_{j,k})\psi_{j,k}.$$

On obtient ainsi l'estimateur seuillé $\hat{f}_n$.

1.6.2 Quasi optimalité dans le cas i.i.d.

Un estimateur quasi-optimal est un estimateur dont la vitesse est optimale à un facteur $\log n$ près. Le seuillage dur atteint cet objectif dans le cas i.i.d. avec les paramètres j_0, j_1, et λ pris de la manière suivante :

Théorème 1.7 (Donoho *et al.*, [35]) *Pour*

$$\begin{aligned} 2^{j_0} &\simeq n^{1/(1+N)}, \\ 2^{j_1} &\simeq n/\log n, \\ \lambda_j &= K\sqrt{j/n}, \text{ pour } K > 0 \text{ suffisamment grand.} \end{aligned}$$

alors $\hat{f}_n$ est un estimateur quasi-optimal dès que f appartient à $\mathcal{B}^s_{\pi,r}(M)$ avec

$$1/\pi < s \leqslant N/2,\ 1 \leqslant \pi \leqslant p,\ 1 \leqslant r \leqslant \infty.$$

Plus précisément il existe une constante $C_0(N,p,s,\pi,M)$ telle que

$$\mathbb{E}\|\hat{f}_n - f\|^p_p \leqslant C_0 \begin{cases} \left(\dfrac{\log n}{n}\right)^{p\alpha}, & si\ \varepsilon \neq 0 \\ \left(\dfrac{\log n}{n}\right)^{p\alpha} (\log n)^{(p/2-\pi/r)_+}, & si\ \varepsilon = 0. \end{cases}$$

1.6.3 Seuillage dur dans le cas dépendant

Nous obtenons une extension de ces résultats dans le cas faiblement dépendant. Le théorème 1.9 assure la quasi-optimalité de l'estimateur par seuillage dur dès que les inégalités de probabilité (1.17) et de moment (1.18) sont satisfaites.

Théorème 1.8 *Il existe des constantes $B, C > 0$ et $\gamma \geqslant 1/2$ telles que pour tout $(j,k) \in \mathbb{N}^2$ et $n \in \mathbb{N}^*$ suffisamment grand*

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\psi_{j,k}(X_i) - \mathbb{E}\psi_{j,k}(X_i))\right| \geqslant \delta\right) \leqslant B\exp\left(-C\delta^{1/\gamma}\right), \tag{1.17}$$

pour tout $\delta \geqslant 0$ tel que $\delta^{2l}\|\psi_{j,k}\|_\infty^2 \leqslant K'n$ avec $K' > 0$ et $l \geqslant 0$. Plus précisément

1. *$\gamma = 1 + 1/b$ et $l = 0$, si $\eta(r) \leqslant c\exp(-r^b)$ pour $b, c > 0$ et si pour tout $j \neq k$ les densités jointes $f_{j,k}$ de (X_j, X_k) existent et sont uniformément bornées en (j,k),*
2. *$\gamma = 0.5$ et $l = 5$, si $(X_t)_{t\in\mathbb{Z}}$ est un système dynamique de type Lasota-Yorke.*

Remarque 1.14 *Cette inégalité de probabilité est établie dans les cas de systèmes dynamiques et de processus η-faiblement dépendants. Les cas de dépendance faible causal et non causal sont donc traités simultanément.*

Remarque 1.15 *L'inégalité (1.17) est un outil pour obtenir des vitesses de convergence des estimateurs seuillés car elle contrôle le comportement des quantités $\psi_{j,k}(X_i)$ qui apparaissent dans le seuillage. Une inégalité de moment sur ces quantités est aussi nécessaire. Supposons que pour tout $n \in \mathbb{N}^*$ il existe $C > 0$ telle que pour tout $(j,k) \in \mathbb{N}^2$, et $q > 0$*

$$\mathbb{E}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\psi_{j,k}(X_i) - \mathbb{E}\psi_{j,k}(X_i))\right|^q \leqslant C. \tag{1.18}$$

Remarque 1.16 *L'inégalité de moment est atteinte dans Dedecker et Prieur [29] sous des conditions plus faibles que l'inégalité exponentielle du théorème 1.8. L'inégalité de probabilité diffère selon le cadre de dépendance. En particulier, le paramètre γ est d'autant plus grand que la dépendance est forte.*

Nous avons désormais tous les outils pour énoncer la quasi-optimalité de l'estimateur par seuillage dur dans les cas faiblement dépendant.

Théorème 1.9 *Supposons que f appartienne à une boule de Besov $\mathcal{B}^s_{\pi,r}(M)$ telle que*

$$1/\pi < s \leqslant N/2,\ 1 \leqslant \pi \leqslant p,\ 1 \leqslant r \leqslant \infty,$$

avec $N \geqslant 4$ et que les propriétés de dépendance faible du processus $(X_t)_{t\in\mathbb{Z}}$ assurent que les inégalités (1.18) et (1.17) sont satisfaites pour un couple (γ, l). Alors il existe une constante $C_0(N, p, s, \pi, M)$ telle que

$$\mathbb{E}[\|\hat{f}_n - f\|_p^p] \leqslant C_0 \begin{cases} \left(\dfrac{\log^{2\gamma} n}{n}\right)^{p\alpha} & si\ \varepsilon \neq 0 \\ \left(\dfrac{\log^{2\gamma} n}{n}\right)^{p\alpha} (\log n)^{(1-\pi/r)_+} & si\ \varepsilon = 0 \end{cases}$$

avec les paramètres fixés à

$$\begin{aligned} 2^{j_0} &\simeq n^{1/(1+N)}, \\ 2^{j_1} &\simeq n/\log^{2\gamma l} n, \\ \lambda_j &= Kj^{\gamma}/\sqrt{n},\ \textit{pour une constante } K > 0 \textit{ bien choisie.} \end{aligned}$$

Remarque 1.17 *Dans tous les cas de dépendance étudiés (η-faible dépendance et systèmes dynamiques de type Lasota Yorke), l'estimateur est quasi-optimal. Une perte logarithmique apparaît, d'autant plus grande que le paramètre γ est grand. Or ce paramètre est d'autant plus grand que la dépendance est forte.*

1.7 L'estimateur de la densité adaptif vis à vis de la dépendance [Wint3]

A ce jour, les procédures d'estimation de la densité existantes varient selon le contexte de dépendance faible. L'estimateur proposé par Tribouley et Viennet [107] ne s'applique que dans le cadre (restrictif) du β-mélange. Nous dirons que les estimateurs, adaptatifs vis à vis de la régularité de f, ne le sont pas vis à vis de la dépendance.

Dans le théorème 1.9, γ apparaît dans l'estimateur $\hat{f}_n$ via le seuillage dur, mais aussi dans le contrôle du risque. De plus, γ dépend du contexte de dépendance faible d'après le Théorème 1.9. Plutôt que de fixer ce paramètre γ selon le contexte de dépendance, nous choisissons la valeur γ qui minimise l'estimation du risque de l'estimateur $\hat{f}_n$. C'est ce qu'on appelle une méthode de validation croisée. Le coefficient γ n'étant pas fixé à priori selon la dépendance de nos observations, nous obtenons un estimateur adaptatif à la régularité de f et à la dépendance. Il n'est pas nécessaire de connaître à priori la dépendance de nos observations.

Dans un premier temps nous présentons cette procédure par validation croisée puis nous testons les résultats de l'estimateur par validation croisée sur des simulations.

1.7.1 Procédure de validation croisée

Soit $\hat{f}_n^\gamma$ l'estimateur correspondant au seuil $\lambda_j = Kj^\gamma/\sqrt{n}$. Il est quasi minimax dès qu'une inégalité de probabilité 1.17 pour un couple (γ, l) est vérifiée pour $(X_t)_{t\in\mathbb{Z}}$. Cet estimateur $\hat{f}_n^\gamma$ dépend de j_0, j_1, K et γ. On fixe $2^{j_1} = n/\log n$, $2^{j_0} = n^{1/(1+N)}$ et $K = \sqrt{2}$ comme dans Donoho et Johnstone [36]. Puis on choisit $\hat{\gamma}_n$ qui minimise un estimateur de notre risque, c'est à dire un critère de validation croisée.

Le Mean Integrated Square Error (risque moyen dans $\mathbb{L}^2$) est estimé par sous échantillonnage par bloc comme dans Hart et Vieu [68] ce qui permet d'éviter les phénomènes liés à la dépendance du processus. Soit b_n la taille d'un bloc (ici $2 * b_n \simeq n^{1/3}$, voir [68]) et

$$\mathcal{X}_{-i} = \{X_j,\ 1 \leqslant j \leqslant n,\ |i-j| \geqslant b_n\}.$$

On minimise

$$CV_b(\gamma) = \int \left(\hat{f}_n^\gamma(x)\right)^2 dx - 2n^{-1} \sum_{i=1}^n \hat{f}_{-i}^\gamma(X_i),$$

où $\hat{f}_{-i}^\gamma$ est l'estimateur $\hat{f}_{n-b_n}^\gamma$ construit à partir du sous-échantillon $\mathcal{X}_{-i}$. On choisit alors $\hat{\gamma}_n = \arg\min_\gamma CV_b(\gamma)$ et notre estimateur adaptatif correspond à $\hat{f}_n^{\hat{\gamma}_n}$.

1.7.2 Simulations

Nous simulons quatre cas de dépendance distincts sur lesquels notre procédure par validation sera testée.

1. Un processus i.i.d. dont la fonction de répartition est $F(x) = 2/\pi \arcsin(\sqrt{x})$ pour $x \in [0; 1]$.
2. Un système dynamique associé à la transformation $T(x) = 4x(1-x)$ dont la mesure invariante est F, la même que dans le cas i.i.d. (voir Prieur, [97] pour plus de propriétés sur ce système dynamique).
3. Un système dynamique non stationnaire, c'est à dire tel que $X_t = T(X_{t-1})$ (avec le même $T(x) = 4x(1-x)$) mais où X_0 ne suit plus la loi marginale F mais une loi uniforme. Ce cadre dépasse celui de la dépendance faible car alors la stationnarité n'est pas vérifiée (X_0 ne suit pas la même loi que X_t pour $t > 0$). Des propriétes d'ergodicité géométrique sur les systèmes dynamiques assurent toutefois que la loi de ce processus est proche de celle de la simulation 2 (voir Viana [110]).
4. Un analogue non causal de l'exemple d'Andrews (1.1)

   $X_t = 2(X_{t-1} + X_{t+1})/5 + 5\xi_t/21$ avec $(\xi_t)_{t\in\mathbb{Z}}$ un processus i.i.d. tel que

$$\mathbb{P}(\xi_0 = 0) = \mathbb{P}(\xi_0 = 1) = \frac{1}{2}.$$

Une solution stationnaire de cette équation est le processus non-causal $(X_t)_{t\in\mathbb{Z}}$

$$X_t = \sum_{j\in\mathbb{Z}} a_j \xi_{t-j},$$

avec $a_j = 1/3 * (1/2)^{|j|}$. Les constantes sont choisies pour que $(X_t)_{t\in\mathbb{Z}}$ prenne ses valeurs dans $[0;1]$ pour tout t. Un tel processus est η-faiblement dépendant avec $a, C > 0$ tel que $\eta_r \leqslant C\exp(-ar)$. Il n'est pas causal car dépend à la fois de son passé et de son futur. On le simule en utilisant l'algorithme de Gibbs.

1.7.3 Comportement de l'estimateur par validation croisée

La figure 1.1 illustre l'évolution du critère de validation croisée en fonction de γ et dans les différents cas de dépendance. Le risque optimal est obtenu pour un $\hat{\gamma}_n = 0,5$ dans les trois premiers cas et $\hat{\gamma}_n > 0,6$ dans le dernier. Conformément aux théorèmes 1.8 et 1.9, $\hat{\gamma}_n$ est identique dans le cas i.i.d. et des systèmes dynamiques. On constate aussi que pour un processus dont la dépendance est plus forte, $\hat{\gamma}_n$ optimal est plus grand. D'autre part, pour des valeurs γ plus petites que $\hat{\gamma}_n$, le risque est plus élevé car l'inégalité de probabilité correspondant à ces valeurs γ n'est pas vérifiée. Pour des valeurs plus grandes, le risque est plus grand car γ apparaît aussi dans le risque de l'estimateur.

Dans le cas η-faiblement dépendant le risque augmente plus lentement pour les valeurs plus grandes que $\hat{\gamma}_n$. Il est donc plus risqué de choisir $\gamma = 0,5 \leqslant \hat{\gamma}_n$ que $\gamma = 0,75 \geqslant \hat{\gamma}_n$. Dans la figure 1.2, le vrai MISE est approché par somme de Riemann et méthode de Monte-Carlo selon un nombre d'observations croissant et pour différents choix de γ. Pour le choix $\gamma = 0,5$ le risque est toujours plus fort que pour $\gamma = 0,75$, ce qui confirme qu'il vaut mieux se tromper en choisissant des valeurs γ plus grandes que $\hat{\gamma}_n$. Toutefois le choix $\gamma = 2 = 1 + 1/b$ donné par le théorème 1.8 n'est pas non plus optimal, la valeur de γ étant clairement trop élevé. Le meilleur choix possible est proche de $\hat{\gamma}_n$ et le meilleur estimateur est $\hat{f}_n^{\hat{\gamma}_n}$. L'inégalité du théorème 1.8 semble donc être vérifiée pour un couple (γ, l) avec $\gamma \leqslant 1+1/b$. Nos outils théoriques ne nous permettent pas d'obtenir cette amélioration de (1.17).

Dans la figure 1.3 nous avons représenté les BoxPlots de $\hat{\gamma}_n$ pour différents n dans les quatre cas de dépendance. Remarquons que $\hat{\gamma}_n$ dépend de n. Dans les trois premiers cas, pour un petit nombre d'observation ($n = 2^7$), un paramètre plus grand est préférable, même dans le cas i.i.d.. Puis dès $n = 2^8$, les distributions des $\hat{\gamma}_n$ sont bien centrées sur $0,5$. Dans le cas η-faiblement dépendant, la valeur moyenne de $\hat{\gamma}_n$ est plus élevée que pour les autres cas. Toutefois, cette valeur tend à diminuer avec le nombre d'observations.

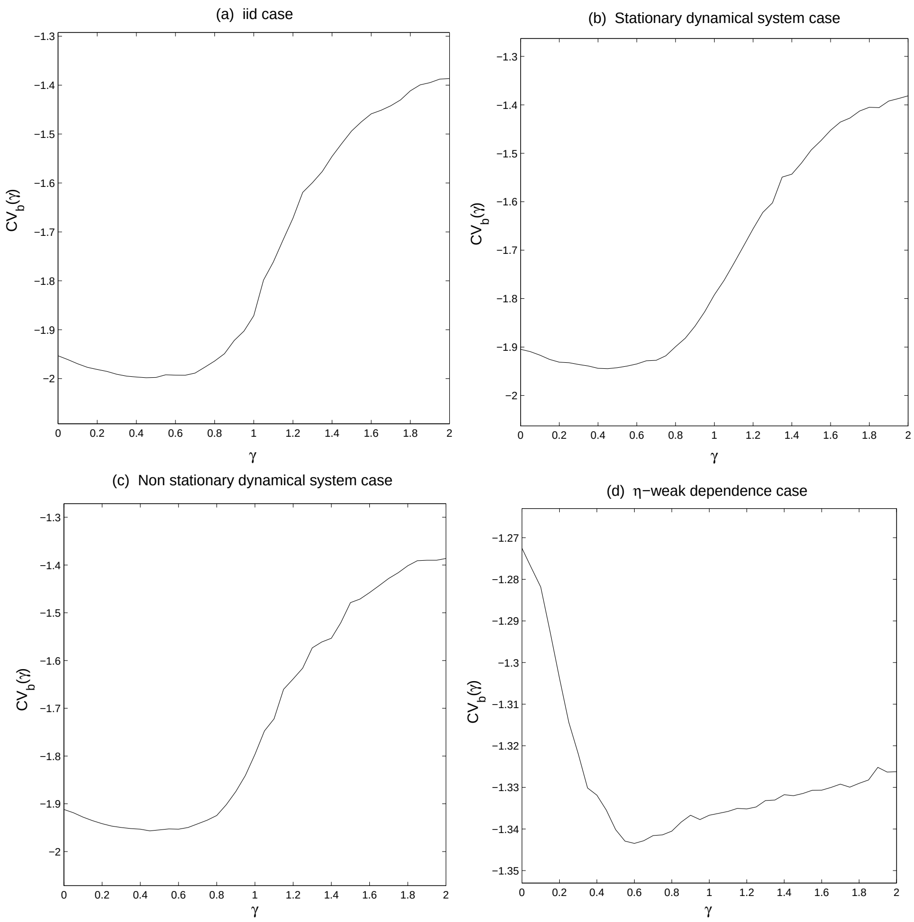

FIG. 1.1 – Cette figure présente l'évolution du critère de validation croisée en fonction de γ pour $n = 2^{10}$ observations. Le critère moyen sur 100 simulations est tracé en fonction de $\gamma = 0, 0.05 \ldots 1.95, 2$. Pour (a) les observations sont i.i.d., pour (b) un système dynamique stationnaire est simulé, (c) est le cas d'un système dynamique non-stationnaire et (d) correspond au cas η-faiblement dépendant.

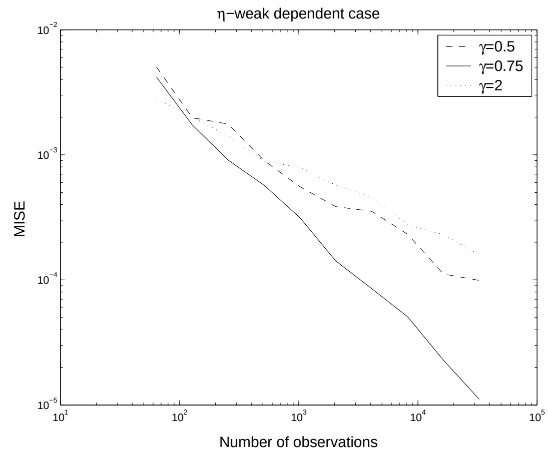

FIG. 1.2 – Évolution du MISE en fonction de la taille de l'échantillon dans le cas η-faiblement dépendant. La figure représente l'évolution du MISE sur 100 simulations en fonction de $n = 2^6, 2^7, \ldots, 2^{14}, 2^{15}$ dans une échelle log-log. Pour la courbe en tiret $\gamma = 0.5$, en plein $\gamma = 0.75$ et $\gamma = 2$ pour la courbe en pointillées.

1.7.4 L'estimateur $\hat{f}_n^{\hat{\gamma}_n}$

Pratiquement, pour un échantillon $(X_1, \ldots, X_n)$ quelconque, nous proposons d'appliquer notre procédure de validation croisée. Si la courbe du MISE estimé par validation croisée a une forme permettant clairement de distinguer un minimum, on choisit $\hat{\gamma}_n$ réalisant ce minimum. L'estimateur de la densité $\hat{f}_n^{\hat{\gamma}_n}$ a alors des chances d'être quasi-optimal.

Cette procédure ne dépend ni de la faible dépendance du modèle ni de la régularité de la fonction. Elle est adaptative à la fois à la régularité de la fonction et à la dépendance faible du modèle. Inversement, pour un modèle fixé, elle permet d'avoir une information sur l'optimalité des inégalités de probabilité obtenues théoriquement via le paramètre γ. Dans notre cas de η-faible dépendance, l'inégalité obtenue théoriquement est améliorable.

1.8 L'estimation paramétrique [Wint5]

Nous nous intéressons ici aux résultats de convergence d'un estimateur $\hat{\theta}_n$ vers le vrai paramètre θ_0. Nous restreignons les valeurs possibles de θ_0 à un sous ensemble compact de $\mathbb{R}^k$, le cadre est dit paramétrique. La méthode d'estimation considérée est le Quasi Maximum de Vraisemblance. Son champ d'application est restreint : les modèles pour lesquels s'applique cette méthode sont des cas particuliers de chaînes infinies. Ils sont bien faiblement dépendants, mais ils ont aussi beaucoup d'autres particularités qui leur sont propres. L'utilisation de ces caractéristiques spécifiques nous

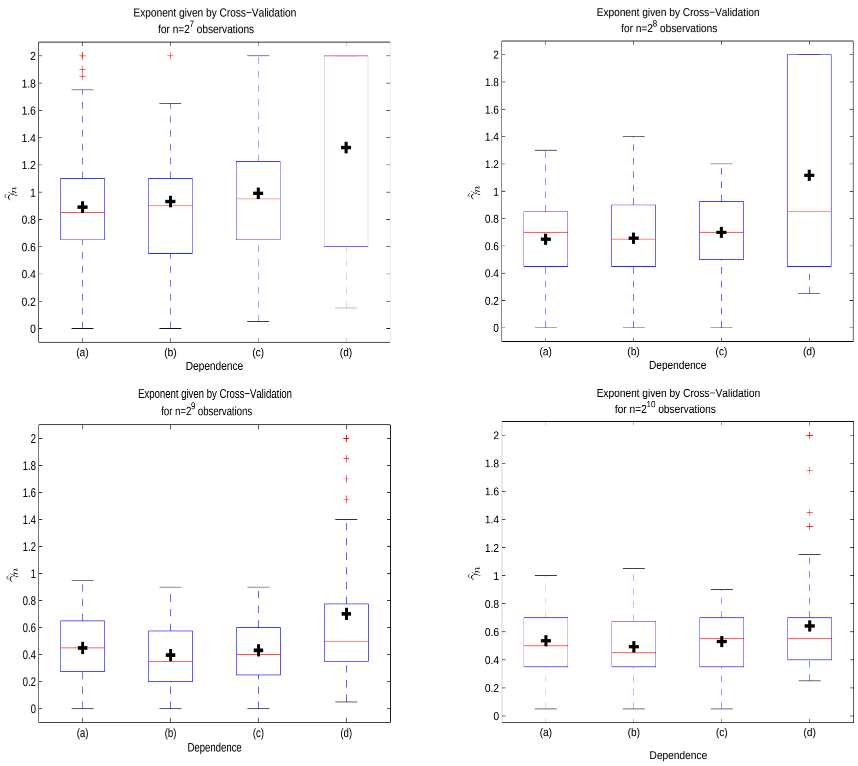

FIG. 1.3 – Boxplots (la croix figure la moyenne) pour la distribution empirique de $\hat{\gamma}_n$ obtenues pour 100 simulations pour $n = 2^7,\ 2^8,\ 2^9,\ 2^{10}$. Pour chaque valeur de n, (a) correspond au cas i.i.d., (b) au système dynamique stationnaire, (c) au système dynamique non-stationnaire et (d) au cas η-faiblement dépendant.

permet de présenter pour la première fois un résultat de normalité asymptotique dans le cas multivarié.

1.8.1 L'estimateur du QMLE

Ici $E = \mathbb{R}^d$, $E' = \mathbb{R}^m$. Nous supposons connaître le vrai modèle à un paramètre près $\theta_0 \in \mathbb{R}^k$ que nous souhaitons estimer afin d'ajuster le modèle à nos observations $X_1, \dots, X_n$. Avoir un modèle c'est avoir une équation dirigeant la dynamique du processus. En toute généralité $X_t = F_{\theta_0}(X_{t-j}; j \neq t; \xi_t)$ où $(\xi_t)_{t\in\mathbb{Z}}$ est une innovation. La forme de F sachant θ est connue et nous souhaitons estimer θ_0 qui suffit à déterminer complètement le modèle.

L'estimateur du Quasi Maximum de Vraisemblance (QMLE) est obtenu en supposant dans un premier temps que le processus $(\xi_t)_{t\in\mathbb{Z}}$ est un processus gaussien. La loi conditionnelle de X_t sachant $X_j, j \neq t$ est une gaussienne si l'innovation ξ_t au temps t est indépendante des $X_j, j \neq t$. Afin d'assurer que ce soit bien le cas, il faut qu'il y ait de la causalité dans l'équation régissant le modèle (voir Doukhan et Truquet [48]), c'est à dire que $X_t = F_{\theta_0}(X_{t-j}; j < t; \xi_t)$. Nous sommes alors dans le cas de chaînes à mémoires infinies du chapitre 2 et la solution $(X_t)_{t\in\mathbb{Z}}$ de l'équation est telle que ξ_t soit indépendante de son passé $\mathcal{P}_t$.

La loi de X_t sachant son passé $\mathcal{P}_t$ étant gaussienne, elle est complètement déterminée par sa moyenne et sa variance. Nous écrivons donc

$$X_t = M_{\theta_0}(X_{t-1}, X_{t-2}, \dots) \cdot \xi_t + f_{\theta_0}(X_{t-1}, X_{t-2}, \dots), \quad \text{pour} \quad t \in \mathbb{Z}, \tag{1.19}$$

avec
- $\theta_0 \in \Theta$ est dans un compact de $\mathbb{R}^k$;
- pour toute suite $x = (x_i)_{i\in\mathbb{Z}} \in \mathbb{R}^{d^{(\mathbb{N})}}$, $\theta \in \Theta \mapsto M_\theta(x)$ est une fonction borélienne à valeurs dans l'espace des matrices $d \times m$.
- pour toute suite $x = (x_i)_{i\in\mathbb{Z}} \in \mathbb{R}^{m(\mathbb{N})}$, $\theta \in \Theta \mapsto f_\theta(x)$ est une fonction borélienne à valeurs dans $\mathbb{R}^m$.

On suppose ξ_t normalisé, i.e. c'est un vecteur centré dont la matrice de variance covariance est I_m, la matrice identité de taille $m \times m$. Alors l'espérance et la variance de X_t sachant son passé $\mathcal{P}_t$ sont

$$\begin{aligned}
\mathbb{E}(X_t \mid \mathcal{P}_t) &= f_{\theta_0}(X_{t-1}, X_{t-2}, \dots) = f_t(\theta_0), \\
\mathrm{Var}\,(X_t \mid \mathcal{P}_t) &= M_{\theta_0}(X_{t-1}, X_{t-2}, \dots) \times M_{\theta_0}(X_{t-1}, X_{t-2}, \dots)' = M_t(\theta_0) \times M_t(\theta_0)',
\end{aligned}$$

où A' est la transposée de A. Ces quantités existent bien presque sûrement dès que $\mathbb{E}\|\xi_0\|^2 < \infty$ d'après les résultats du chapitre 2.

On dit que $(X_t)_{t\in\mathbb{Z}}$ est un processus autorégressif avec des erreurs conditionnellement hétéroscédastiques, voir Jeantheau [74]. Le modèle associé est adapté à l'estimation paramétrique par QMLE.

Les propriétés de dépendance de ce modèle sont toutefois très particulières. Par exemple, toutes les solutions de l'équation (1.19) sont telles que $\mathrm{Cov}(X_0, X_t) = 0$.

Le QMLE $\hat{\theta}_n$ est le paramètre θ maximisant la log-vraisemblance $L_{X_1,\dots,X_n}(X_1,\dots,X_n)$. On commence par approcher $L_{X_1,\dots,X_n}(X_1,\dots,X_n)$ par $L_{X_1,\dots,X_n|\mathcal{P}_0}(X_1,\dots,X_n)$. La dépendance faible nous assure que le passé $\mathcal{P}_0$ n'influence pas sur la loi de X_n pour n suffisamment grand. Puis on réécrit $L_{X_1,\dots,X_n}(X_1,\dots,X_n)$ sous la forme $\sum_{t=0}^{n-1} L_{X_t|\mathcal{P}_t}(X_1,\dots,X_n)$ afin de faire apparaître les distributions conditionnelles gaussiennes.

On a alors

$$L_{X_t|\mathcal{P}_t}(X_1,\dots,X_n) = \log\left(\frac{1}{\sqrt{2\pi \det M_t(\theta_0)M_t(\theta)'}}\right)\exp\frac{1}{2}\Big((X_t - f_t(\theta_0))'\left(M_t(\theta_0)M_t(\theta_0)\right)^{-1} \times (X_t - f_t(\theta_0))\Big).$$

Classiquement, l'estimateur du maximum de la (log-)vraisemblance (MLE) est

$$\theta_n = \arg\max_{\theta\in\Theta} -\frac{1}{2}\sum_{t=1}^{n}\big(X_t - f_\theta^t\big)'\big(M_\theta^t \cdot (M_\theta^t)'\big)^{-1}\big(X_t - f_\theta^t\big) - \frac{1}{2}\,\log\Big(\det\big(M_\theta^t(M_\theta^t)'\big)\Big).$$

Cet estimateur n'est toutefois pas constructible à partir des observations (voir Mikosch et Straumann, [105]) car il dépend de $M_\theta^t)$ et f_θ^t, donc d'une infinité de valeurs passées du processus $(X_t)_{t\in\mathbb{Z}}$. Or nous n'avons à disposition que les observations $(X_1,\dots,X_n)$. Nous appellerons QMLE l'estimateur

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} -\frac{1}{2}\sum_{t=1}^{n}\big(X_t - \widehat{f}_\theta^t\big)'\big(\widehat{M}_\theta^t \cdot (\widehat{M}_\theta^t)'\big)^{-1}\big(X_t - \widehat{f}_\theta^t\big) - \frac{1}{2}\,\log\Big(\det\big(\widehat{M}_\theta^t(\widehat{M}_\theta^t)'\big)\Big), \qquad (1.20)$$

où $\widehat{f}_\theta^t$ et $\widehat{M}_\theta^t$ sont des estimateurs de $M_\theta^t)$ et f_θ^t construit à partir des observations $(X_1,\dots,X_n)$.

1.8.2 Résultats

Les hypothèses

Nous supposons que le processus i.i.d. des innovations $(\xi_t)_{t\in\mathbb{Z}}$ est normalisé ($\mathbb{E}\xi_0 = 0$ et $\mathrm{Var}\,\xi_0 = I_m$) et tel que $\mathbb{E}\|\xi_0\|^r < \infty$. L'hypothèse de gaussiannité sur les innovations n'est plus indispensable et les résultats valables dans ce contexte le restent quelle que soit la loi de ξ_0.

– **Hypothèse ST(r) :** *Les fonctions f_θ et M_θ sont telles que :*

1. $\|f_\theta(0,0,\dots)\|_\Theta + \|M_\theta(0,0,\dots)\|_\Theta < \infty$; (1.21)
2. *il existe deux suites $(\alpha_j(f))_{j\in\mathbb{N}}$ et $(\alpha_j(M))_{j\in\mathbb{N}}$ telles que*

$$\sum_{j=1}^{\infty}\alpha_j(f) + (\mathbb{E}\|\xi_0\|^r)^{1/r}\Big(\sum_{j=1}^{\infty}\alpha_j(M)\Big) = a < 1 \text{ et } \forall (x_j)_{j\in\mathbb{N}},\ (y_j)_{j\in\mathbb{N}} \text{ dans } \mathbb{R}^{d(\mathbb{N})}, \tag{1.22}$$

$$\begin{cases} \|f_\theta((x_j)_{j\in\mathbb{N}}) - f_\theta((y_j)_{j\in\mathbb{N}})\|_\Theta & \leqslant \sum_{j=1}^{\infty}\alpha_j(f)\|x_j - y_j\|, \\ \|M_\theta((x_j)_{j\in\mathbb{N}}) - M_\theta((y_j)_{j\in\mathbb{N}})\|_\Theta & \leqslant \sum_{j=1}^{\infty}\alpha_j(M)\|x_j - y_j\|. \end{cases}$$

Ces hypothèses permettent d'assurer l'existence d'une solution stationnaire grâce aux résultats du chapitre 2. De plus, elles assurent l'existence des moments d'ordre r de la solution $(X_t)_{t\in\mathbb{Z}}$ ($\mathbb{E}\|X_0\|^r$). On en déduit l'existence des moments des quantités $f_t(\theta)$ et $M_t(\theta)$ (qui sont vues comme des limites de suites convergentes $(f_\theta(X_{t-1},\dots,X_{t-k},0,\dots,))_{k\geqslant 1}$ et $(M_\theta(X_{t-1},\dots,X_{t-k},0,\dots,))_{k\geqslant 1}$). Ainsi, ces quantités sont bien définies. De plus, on montre que les estimateurs "plug-in" $\widehat{f}_\theta^t = f_\theta(X_{t-1},\dots,X_1,0,\dots,)$ et $\widehat{M}_\theta^t = M_\theta(X_{t-1},\dots,X_1,0,\dots,)$ sont bien consistants, c'est-à-dire que $\widehat{f}_\theta^t \xrightarrow[n\to\infty]{\mathcal{P}} f_\theta^t$ et $\widehat{M}_\theta^t \xrightarrow[n\to\infty]{\mathcal{P}} M_\theta^t$.

– **Hypothèse Id :** Pour tout $\theta\in\Theta$, $M_\theta^t = M_{\theta_0}^t$ et $f_\theta^t = f_{\theta_0}^t$ p.s. si et seulement si $\theta = \theta_0$.

Cette hypothèse d'identifiabilité est incontournable. Elle assure que l'estimation du vrai paramètre θ_0 à partir des observations $(X_1,\dots,X_n)$ est effectivement possible. Si elle n'était pas vérifiée (i.e. il existe $\theta_1\neq\theta_0$ tel que $M_{\theta_1}^t = M_{\theta_0}^t$ ou $f_{\theta_1}^t = f_{\theta_0}^t$ p.s.), alors le QMLE $\hat\theta_n$ ne distingue plus le vrai paramètre θ_0 de θ_1. Pour vérifier cette hypothèse, il est préférable de se placer dans le cas de l'extension multidimensionnelle 1.19 et non 1.8. Toutefois, l'identifiabilité n'est pas assurée dans le cas général 1.19. Nous nous plaçons toujours dans des cas classiques pour lesquels l'hypothèse d'identifiabilité est vérifiable à partir de travaux existants (voir le chapitre 6).

– **Hypothèse F :** *pour toute suite $(x_j)_{j\in\mathbb{N}}\in\mathbb{R}^{d(\mathbb{N})}$ la fonction $\theta\mapsto f_\theta(x_1,\dots)$ est continue.*

– **Hypothèse M :** *pour toute suite $(x_j)_{j\in\mathbb{N}}\in\mathbb{R}^{d(\mathbb{N})}$ la fonction $\theta\mapsto M_\theta(x_1,\dots)$ est continue. De plus, il existe $\underline{M}>0$ tel que pour tout $(x_j)_{j\in\mathbb{N}}\in\mathbb{R}^{d(\mathbb{N}^*)}$*

$$\inf_{\theta\in\Theta}\det\big(M_\theta(x_1,\dots)\big(M_\theta(x_1,\dots)\big)'\big) \geqslant \underline{M} > 0. \tag{1.23}$$

Ces conditions sont classiques et assurent l'utilisation des théorèmes de convergence uniforme sur les suites de fonctions continues. La borne sur le déterminant est indispensable pour inverser la matrice et définir proprement notre estimateur.

– **Hypothèses $D^{(i)}$ ($i=1,2$) :** *les fonctions $\theta\in\Theta\mapsto f_\theta((x_j)_{j\in\mathbb{N}})$ et $\theta\in\Theta\mapsto M_\theta((x_j)_{j\in\mathbb{N}})$ sont i-fois continûment différentiables pour tout $(x_j)_{j\in\mathbb{N}}\in(\mathbb{R}^d)^{(\mathbb{N})}$. De plus,*

1. $\left\|\dfrac{\partial^i f_\theta(0,0,\ldots)}{\partial\theta_{k_1}\partial\theta_{k_i}}\right\|_\Theta + \left\|\dfrac{\partial^i M_\theta(0,0,\ldots)}{\partial\theta_{k_1}\partial\theta_{k_i}}\right\|_\Theta < \infty$ (1.24)

2. *il existe* $(\alpha_j^{(i)}(f))_{j\in\mathbb{N}^*}$ *et* $(\alpha_j^{(i)}(M))_{j\in\mathbb{N}^*}$ *vérifiant* $\displaystyle\sum_{j=1}^\infty \alpha_j^{(i)}(f) + \sum_{j=1}^\infty \alpha_j^{(i)}(M) < \infty$

   et pour tout $(x_j)_{j\in\mathbb{N}}$, $(y_j)_{j\in\mathbb{N}}$ *dans* $\mathbb{R}^{d^{(\mathbb{N})}}$ *et tout* $1 \le k \le d$, $k_1, k_i \in \{1,\ldots,d\}$

$$\left\{\begin{array}{lcl} \left\|\dfrac{\partial^i f_\theta((x_j)_{j\in\mathbb{N}})}{\partial\theta_{k_1}\partial\theta_{k_i}} - \dfrac{\partial^i f_\theta((y_j)_{j\in\mathbb{N}})}{\partial\theta_{k_1}\partial\theta_{k_i}}\right\|_\Theta & \leqslant & \sum_{j=1}^\infty \alpha_j^{(i)}(f)\|x_j - y_j\|, \\ \left\|\dfrac{\partial^i M_\theta((x_j)_{j\in\mathbb{N}})}{\partial\theta_{k_1}\partial\theta_{k_i}} - \dfrac{\partial^i M_\theta((y_j)_{j\in\mathbb{N}})}{\partial\theta_{k_1}\partial\theta_{k_i}}\right\|_\Theta & \leqslant & \sum_{j=1}^\infty \alpha_j^{(i)}(M)\|x_j - y_j\|. \end{array}\right. \qquad (1.25)$$

Les hypothèses sur les dérivées d'ordre 1 et 2 nous permettent d'obtenir la normalité asymptotique du QMLE. Les conditions de Lipschitz sur les dérivées sont indispensables dans notre cadre d'étude. Elles permettent de prouver l'existence de quantités aléatoires dépendantes de tout le passé telles que

$$\frac{\partial^i f_t(\theta)}{\partial\theta_{k_1}\partial\theta_{k_i}} \text{ et } \frac{\partial^i M_t(\theta)}{\partial\theta_{k_1}\partial\theta_{k_i}} \text{ pour } i = 1, 2.$$

Consistance forte

On obtient la consistance forte de notre estimateur dans un cadre très général.

Théorème 1.10 *Si les hypothèses ST(2), F, M et Id sont satisfaites, si* $(X_t)_{t\in\mathbb{Z}}$ *est solution de (1.19) alors dès que*

$$\alpha_j(f) + \alpha_j(M) = O\big(j^{-\ell}\big) \quad \textit{avec} \quad \ell > 3/2$$

le QMLE $\widehat{\theta}_n$ *est fortement consistant, i.e.*

$$\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta_0.$$

Remarque 1.18 *Ce théorème induit l'existence d'un moment d'ordre 1 sur le processus* $(X_t)_{t\in\mathbb{Z}}$. *Or, Jeantheau [74] a montré la consistance forte du MLE* θ_n *sous une condition de moment logarithmique* $\mathbb{E}\log\|X_0\| < \infty$ *plus faible. Toutefois, sa méthode nécessite de démontrer au cas par cas l'existence d'une solution stationnaire à l'équation (1.19). Dans notre méthode, l'existence du processus est assuré.*

Normalité asymptotique

Théorème 1.11 *Si les hypothèses ST(4),* $D^{(1)}$, $D^{(2)}$, *F, M sont satisfaites, si* $(X_t)_{t\in\mathbb{Z}}$ *est une solution de (1.19) avec* $\theta_0 \in \overset{\circ}{\Theta}$, *dès que*

$$\alpha_j^{(1)}(f) + \alpha_j^{(1)}(M) = O\big(j^{-\ell'}\big) \quad \textit{avec} \quad \ell' > 3/2,$$

alors le QMLE $\widehat{\theta}_n$ est asymptotiquement normal, i.e.

$$\sqrt{n}\big(\widehat{\theta}_n - \theta_0\big) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_d\big(0\ ,\ F(\theta_0)^{-1} \cdot G(\theta_0) \cdot F(\theta_0)^{-1}\big),$$

où $F(\theta_0)$ et $G(\theta_0)$ sont des matrices finies ($F(\theta_0)$ est supposée inversible sans vérification).

Remarque 1.19 *Les expressions de $F(\theta_0)$ et $G(\theta_0)$ sont très complexes, voir (6.51) et (6.47).*

Remarque 1.20 *D'autres versions de ces théorèmes sont fournies dans les corollaires 6.1 et 6.2. Les hypothèses ne portent plus sur les coefficients de Lipschitz de la fonction M_θ ni sur ses dérivées mais sur ceux de $M_\theta M_\theta'$. Ces hypothèses sont bien mieux adaptées dans le cas de modèles classiques comme les ARCH et les GARCH.*

Remarque 1.21 *Dans le cas réel ($m = p = 1$) pour le modèle ARCH(∞), notre résultat de consistance forte n'est pas comparable à celui de Robinson et Zaffaroni [102]. Notre hypothèse sur les coefficients de Lipschitz est plus forte alors que celle sur les moments des innovations est plus faible. Notre résultat de normalité asymptotique est quant à lui toujours meilleur que celui de Robinson et Zaffaroni [102].*

Remarque 1.22 *Dans le cas multivarié, ce résultat est unique et constitue une avancée théorique importante en économétrie. En effet, la normalité asymptotique est indispensable pour l'élaboration d'intervalles de confiance sur le paramètre θ_0. La normalité asymptotique du QMLE $\hat{\theta}_n$ est souvent présupposée sans que celle-ci soit prouvée nulle part (voir une discussion à ce sujet dans [8]). Seul le résultat de Comte et Lieberman [22] existe dans le cas très particulier du modèle multivarié BEKK (voir le chapitre 6 pour plus de détails). Notre résultat de normalité asymptotique est obtenu sous des conditions plus faibles que le leur.*

Une liste des nombreux modèles économétriques pour lesquels ce résultat s'applique est disponible au chapitre 6.

Bibliographie

[Wint1] N. Ragache, O. Wintenberger (2006). 'Convergence rates for density estimators of weakly dependent time series', Dependence in Probability and Statistics. *Lecture Notes in Statistics. P. Bertail, P. Doukhan and Ph. Soulier editors*, vol 187, Springer Verlag.

[Wint2] P. Doukhan, O. Wintenberger. 'An invariance principle for weakly dependent stationary general model' (2007), accepté dans *Probability and Mathematical Statistics.*

[Wint3] I. Gannaz, O. Wintenberger. 'Adaptative density estimation under dependence', soumis à *ESAIM P&S.*

[Wint4] P. Doukhan, O. Wintenberger. 'Weakly dependent chains with infinite memory', soumis à *Stochastic Processes and their Applications.*

[Wint5] J.M. Bardet, O. Wintenberger. 'Asymptotic normality of the Quasi Maximum Likelihood Estimator for multidimensional conditionally heteroscedastic processes', en préparation.

Première partie

The models

Chapitre 2

Weakly dependent chains with infinite memory

Abstract

The main objective of the chapter is to define strictly stationary solutions of infinite memory recurrence equations such as $X_t = F(X_{t-1}, X_{t-2}, X_{t-3}, \dots; \xi_t)$, where $(\xi_t)_{t\in\mathbb{Z}}$ denotes an independent and identically distributed sequence of random variables. To this end, we use an appropriate Lipschitz condition which also entails the integrability of $(X_t)_{t\in\mathbb{Z}}$ and weak dependence properties defined by [28]. Such models provide both continuous state space and nonlinear extensions of various examples of several well-known classes of time series.

Keywords : stationary solutions, weak dependence, coupling, Markov process, Bernoulli shifts

Note

The content of this part is based on a paper, written in collaboration with Paul Doukhan, submitted at Stochastic Processes and their Applications.

2.1 Introduction

Times series analysis is a main research field for application sake. The statistical inference deeply relies on the underlying model and model selection is very important in statistics. The choice of a model is often a compromise between antagonist criteria. For example, the principle of parsimony guides the practitioner in the choice of a model : model possessing maximum simplicity and the minimum number of parameters consonant with representational adequacy is a common objective in practice.
Regular second order stationary times series $(X_t)_{t\in\mathbb{Z}}$ are often represented as infinite autoregressive processes

$$X_t = \sum_{j=1}^{\infty} a_j X_{t-j} + \xi_t, \qquad t \in \mathbb{Z} \tag{2.1}$$

where $(\xi_t)_{t\in\mathbb{Z}}$ is the innovation process. This representation has motivated the introduction of *ARMA* processes as a tool in various research fields of the human activity such as economy, marketing, engineering, medicine and others. Many forecast methods lean on the dependence structure of the process $(X_t)_{t\in\mathbb{Z}}$. The linear representation (2.1) induces that the innovation process is a second order stationary white noise. Nonlinear models, *e.g.* Markov ones, allow the use of independent and identically distributed (iid) inputs $(\xi_t)_{t\in\mathbb{Z}}$. Such representations are preferred to (2.1) in various applications *e.g.* in finance, hydrodynamics, physics, electromagnetism, see [34, 80]. [9] obtained an existence condition for discrete state spaces when the Markov property is omitted. This condition is improved in [21, 55]. The underlying stationary models called *Random Systems with Complete Connections* are introduced in [73]. These processes are defined through their marginal conditional distributions in an extended Markov way. An alternative extension to non-Markovian context is called *Variable Length Markov Chains* (see [18]) : triangular arrays $X_{1,n}, \ldots, X_{n,n} \in \mathcal{P}_n$, are considered for some set $\mathcal{P}_n$ set of p_n-Markov chains (here ergodicity replaces the condition of stationarity). Relaxing both linearity and Markov assumptions in the representation of a random phenomenon, these models are widely used in the fields of particle systems or in DNA data analysis.
Another way to study time series is based on the conditional distributions of the process $(X_t)_{t\in\mathbb{Z}}$. Dobrushin's existence and uniqueness conditions given in [33] are widely used in this very general context ; we refer the reader to [56] for details. Dobrushin's existence condition needs regularity of the marginal distributions, closely related with sharp mixing conditions (see [38]). Mixing is a restrictive notion, as stressed in [1] and the main asymptotic results still hold under weak dependence see *e.g.* [28, 41]. Under particular sharp conditions in terms of the coupling coefficients τ, the Functional Law of the Iterated Logarithm still holds (see [28]).
Let $(E, \|\cdot\|)$ be a Banach space, called state space of the process $(X_t)_{t\in\mathbb{Z}}$. Let $(\xi_t)_{t\in\mathbb{Z}}$ be a sequence of iid random variables with values in a probability space $(E', \mathcal{B})$, called innovations or inputs of the system. The distribution of $(\xi_t)_{t\in\mathbb{N}}$ on the product probability space $(E'^{\mathbb{N}}, \mathcal{B}^{\otimes\mathbb{N}})$ is denoted by μ. In

this chapter, we introduce a new nonlinear model that is neither Markovian nor mixing. To use the notion of τ-weak dependence, integrability of the process shall be satisfied. Attractive candidates are Bernoulli shifts defined as

$$X_t = H(\xi_{t-j}, j \in \mathbb{N}), \text{ for } t \in \mathbb{Z} \text{ and } H \in \mathbb{L}^m(\mu), \text{ for } m \geqslant 1. \tag{2.2}$$

(See [41, 113] for details). Here $\mathbb{L}^m(\mu)$ denotes the set of $E-$valued random variables with finite m^{th} order moments. These quite general models are τ-weakly dependent (see [29] for examples). However, in terms of parsimony, they might not be appropriate. For example, Volterra chaotic models

$$H(x_0, x_1, \ldots) = a_0 + \sum_{k=1}^{\infty} \sum_{j_1,\ldots,j_k=0}^{\infty} a_{k;j_1,\ldots,j_k} x_{j_1} \cdots x_{j_k}$$

involve a very large family of parameters $(a_{k;j_1,\ldots,j_k})_{k,j_1,\ldots,j_k \geqslant 0}$. Moreover, even if the functional H exists, in many cases we cannot give it an explicit form.

We call *Chains with infinite memory* the stationary solutions of equations

$$X_t = F(X_{t-1}, X_{t-2}, X_{t-3}, \ldots; \xi_t), \tag{2.3}$$

where F is an E-valued function properly defined in Section 2.2.2. This dynamic behaviour corresponds to many times series encountered in practice. Such models may be seen either as linear model extensions as (2.1) or as Markov model extensions. Examples in Section 2.3 prove that parsimony is possible for such representations, reducing considerably the family of parameters. Existence of our models is proved jointly with their integrability, we quote that [17] who is interested only in *a.s.* constructions in the case of finite memory obtain optimal existence criteria (but he does not obtain the integrability of the solution).

After some notations we explicitly motivate the use of such models. Existence of moments of order m and weak dependence properties are derived as our main theorem. Approximation by suitable $p-$Markov processes is the main tool for the proof.

2.2 Main results

2.2.1 Definitions

Let us introduce some notations.

- The sequence $(\xi_t)_{t\in\mathbb{Z}}$ is iid and takes values in a probability space E'.
- $(E, \|\cdot\|)$ is a Banach space and $(X_t)_{t\in\mathbb{Z}}$ an E-valued stationary time series.
- For a random variable $Z \in E$ and a real number $m \geqslant 1$ we denote $\|Z\|_m = (\mathbb{E}\|Z\|^m)^{1/m}$.

– For $h : E \to \mathbb{R}$, we let $\|h\|_\infty = \sup_{x\in E} |h(x)|$ and

$$\mathrm{Lip}\,(h) = \sup_{x\neq y} \frac{|h(x)-h(y)|}{\|x-y\|}.$$

– $\Lambda_1\big(E\big)$ is the set of functions $h : E \to \mathbb{R}$ such that $\mathrm{Lip}\,(h) \leqslant 1$.
Now recall the notion of τ-weak dependence introduced in [28].

Definition 2.1 ([28]) *Let $(\Omega, \mathcal{C}, \mathbb{P})$ be a probability space, $\mathcal{M}$ a σ-algebra of $\mathcal{C}$ and X be a random variable with values in E. Assume that $\|X\|_1 < +\infty$ and define the coefficient τ as*

$$\tau(\mathcal{M}, X) = \left\|\sup\left\{\left|\int f(x)\mathbb{P}_{X|\mathcal{M}}(dx) - \int f(x)\mathbb{P}_X(dx)\right|,\ f \in \Lambda_1\big(E\big)\right\}\right\|_1$$

An easy way to calculate this coefficient is based on a coupling argument (see [28] for details) : $\tau(\mathcal{M}, X) \leqslant \|X - Y\|_1$ where Y, distributed as X, is independent of $\mathcal{M}$. Moreover if the probability space $(E, \mathcal{A})$ is rich enough, then there exists X^* such that $\tau(\mathcal{M}, X) = \|X - X^*\|_1$. Using the definition of τ, the dependence between the past of the sequence $(X_t)_{t\in\mathbb{Z}}$ and its future k-tuples can be assessed. More precisely, consider the norm $\|x - y\| = \|x_1 - y_1\| + \cdots + \|x_k - y_k\|$ on E^k, set $\mathcal{M}_p = \sigma(X_t, t \leqslant p)$ and

$$\begin{aligned}
\tau_k(r) &= \max_{1\leqslant l\leqslant k} \frac{1}{l} \sup\{\tau(\mathcal{M}_p, (X_{j_1}, \ldots, X_{j_l})), p + r \leqslant j_1 < \cdots < j_l\},\\
\tau_\infty(r) &= \sup_{k>0} \tau_k(r).
\end{aligned}$$

From now on, for simplicity sake and without confusion, $\tau_\infty(r)$ is denoted by $\tau(r)$. The time series $(X_t)_{t\in\mathbb{Z}}$ is τ-weakly dependent when the sequence of its coefficients $\tau(r)$ tends to 0 as r tends to infinity. Asymptotic results in this setting are given in [28].

2.2.2 The model

Bernoulli shifts given by eqn. (2.2) are not mixing in general but they are weakly dependent.

Proposition 2.1 ([28]) *Let $(\xi'_{-j})_{j\in\mathbb{N}}$ be distributed as μ and independent of $(\xi_{-j})_{j\in\mathbb{N}}$. Assume that $H \in \mathbb{L}^1(\mu)$ and that there exists a sequence $\delta_r \downarrow 0$ as $r \to \infty$ such that*

$$\left\|H\left(\xi_{r-j}, j \in \mathbb{N}\right) - H\left(\xi_r, \xi_{r-1}, \ldots, \xi_1, \xi'_0, \xi'_{-1}, \ldots\right)\right\|_1 \leqslant \delta_r, \tag{2.4}$$

then $(X_t)_{t\in\mathbb{Z}} = H(\xi_{t-j}, j \in \mathbb{N})$ is τ-weakly dependent with $\tau(r) \leqslant \delta_r$.

Such functions H need not to be defined over all $E'^{\mathbb{N}}$ but they are however *a.s.* defined on the support $\mathcal{S}$ of μ ; this set contains $S^{(\mathbb{N})}$, the set of finite sequences with elements in S where S denotes

ξ_0 distribution's support. The function H is usually unknown and we claim that the equation (2.3) is physically more significant.
In order to properly define the model we introduce projective family sets of functions $\mathcal{F}$. Let E together with its Borel $\sigma-$field $\mathcal{A}$ and E^k for $k > 1$ together with the natural $\sigma-$field $\mathcal{A}^{\otimes k}$. Then

$$\mathcal{F} = \Big\{ \left(F^{(k)}\right)_{k\in\mathbb{N}} / \ \forall k \in \mathbb{N}, F^{(k)} : E^k \times E' \to E \text{ measurable such that,} \\ \forall (x_1,\dots,x_k) \in E^k, \forall u \in E', F^{(k)}(x_1,\dots,x_k;u) = F^{(k+1)}(x_1,\dots,x_k,0;u) \Big\}.$$

When it exists, the limit in $\mathbb{L}^1$ of the random sequence $F^{(k)}(x_1,\dots,x_k;\xi_0)$ (as $k \to \infty$) is denoted by $F(x_1,\dots,x_k,\dots;\xi_0)$ for each E-valued sequence $(x_k)_{k\in\mathbb{N}}$.
The domain of F is not properly defined because it depends on F itself. For the AR(∞) models given by (2.1), it contains

$$\Big\{ (x_k)_{k\in\mathbb{N}^*} \Big/ \sum_{j\geqslant 1} |a_j x_{t-j}| < \infty \Big\}, \quad \text{if} \quad \|\xi_0\|_1 < \infty.$$

For example bounded sequences satisfy this relation but also some unbounded ones since we assumed $\sum_{j\geqslant 1}|a_j| < \infty$.
For $m \geqslant 1$, consider the subsets of $\mathcal{F}$ denoted $\mathcal{LIP}_m$ of all projective families $(F^{(k)})_k$ satisfying for all $k \in \mathbb{N}^*$

$$\left\| F^{(k)}(x_1,\dots,x_k;\xi_0) - F^{(k)}(y_1,\dots,y_k;\xi_0) \right\|_m \leqslant \sum_{j=1}^{k} a_j \|x_j - y_j\|,$$

for all $(x_1,x_2,\dots)$, $(y_1,y_2,\dots)$ and where $(a_j)_j$ is a sequence of nonnegative real numbers. In this case, it directly follows that

$$\|F(x_1,x_2,x_3,\dots;\xi_0) - F(y_1,y_2,y_3,\dots;\xi_0)\|_m \leqslant \sum_{j=1}^{\infty} a_j \|x_j - y_j\|.$$

Our main results hold if $\left(F^{(k)}\right)_k \in \mathcal{LIP}_m$ with

$$\mu_0 = \left\| F^{(0)}(\xi_0) \right\|_m < \infty, \tag{2.5}$$

$$a = \sum_{j=1}^{\infty} a_j < 1. \tag{2.6}$$

Note that from Hölder's inequality the previous conditions (2.5) and (2.6) also hold for $m = 1$. This value of $m = 1$ is the only one required to prove τ-weak dependence ; however higher order moments are needed to derive limit theorems (see [26, 40]). Assumption (2.5) leads to a moment assumption on the variable $F(0,0,0,\dots;\xi_0)$ while assumption (2.6) leads to a contraction condition on $F(\cdot;\xi_0)$ when these limits exist.

Theorem 2.1 *Assume properties (2.5) and (2.6) hold for some $m \geqslant 1$, then there exists a τ-weakly dependent stationary solution $(X_t)_{t\in\mathbb{Z}}$ of eqn. (2.3) such that $\|X_t\|_m < \infty$ with $\tau(r) \leqslant C \cdot \Big(a^{r/p} + \sum_{k=p}^{\infty} a_k\Big)$ for each $p \in \mathbb{N}^*$ and some constant C depending only on μ_0 and a. This solution is the unique one such that $(X_j)_{j<t}$ and ξ_t are independent for all t.*

Remarks

– For two special cases of decreasing of the sequence $(a_j)_{j\geqslant 1}$ the result of the theorem leads to
  - If $a_j \leqslant ce^{-\beta j}$ then $\tau(r) \leqslant Ce^{-\sqrt{\alpha\beta r}}$ where $\alpha = -\ln a$.
  - If $a_j \leqslant cj^{-\beta}$ with $\beta > 1$ then $\tau(r) \leqslant C\,(\ln r/r)^{\beta-1}$.

– As mentioned in the introduction, RSCC are defined through their marginal conditional distributions. We quote here that using such strong conditions as well as an additional irreducibility condition, the coefficient of ϕ-mixing of an RSCC reaches the bound given in Theorem 2.1.5, page 42 of [72]. This bound has exactly the same form as our bound for $\tau(r)$. For this, the authors first control the rate of convergence of such a system to its limit distribution with respect to the total variation norm, in Theorem 2.1.3, page 40 of [72]. Here we do not assume any regularity condition on the inputs (like absolute continuity), which justifies the point that the result of [72] cannot be expected here. [1] provides a non-mixing stationary solution of an equation similar to eqn. (2.3). We recall here that $(\tau(r))_{r\in\mathbb{Z}}$ ensures most of the limit theorems obtained in the $\phi-$mixing setting, [28].

– Let $(X_t)_{t\in\mathbb{Z}}$ be a solution of eqn. (2.3) such that $(X_j)_{j<t}$ and ξ_t are independent for all t. Then conditions (2.5) and (2.6) for $m = 1$ are equivalent to Dobrushin's uniqueness conditions of the stationary measure (if it exists) for $(X_t)_{t\in\mathbb{Z}}$ as given in [33]. In the case of finite state space (as in [55]), conditions also imply both existence and uniqueness of a solution of an auto-regressive equation. Following this work, we could called conditions (2.5) and (2.6) the one-sided (or causal) Dobrushin's conditions.

`Initialisation :`	$Y \leftarrow 0,$	$k \leftarrow 1.$
`While` $k \leqslant n$ `:`	$u \sim \xi_0,$	$U \leftarrow (U, u).$
	$x = F^{(k)}(Y, u),$	$X \leftarrow (X, x).$
	$Y \leftarrow (\tilde{X}_k, Y),$	$k \leftarrow k + 1.$
`End`		
	$X = (\tilde{X}_1, \ldots, \tilde{X}_n).$	

FIG. 2.1 – Algorithm

The proof involves a p-Markov approximation stated below. These p-Markov processes can be approximated by simulations. These simulations are different for each order p and then the complete method is not a recursive one. A recursive way to approximate the stationary measure is given in figure 2.1.

Corollary 2.1 *Assume that conditions (2.5) and (2.6) hold. Let* $(X_t)_{t\in\mathbb{Z}}$ *be the solution of (2.3). Conditionally to* $\{(\xi_1,\ldots,\xi_n)=U\}$,

$$\|\tilde{X}_r - X_r\|_\Phi \leqslant \frac{\|X_0\|_\Phi}{1-a}\inf_{p\geqslant 1}\left(a^{r/p} + \frac{1}{1-a}\sum_{k=p+1}^{\infty} a_k\right) \to 0 \quad as \quad r\to\infty.$$

The bound is similar than the one obtained for $\tau(r)$.

2.3 Examples

This section presents some examples for $E=\mathbb{R}^d$ with $d\geqslant 1$. Many applications may be considered with models satisfying eqn. (2.3). They could be either seen as an extension to infinite memory of solutions of Stochastic Recurrence Equations of [17] or as nonlinear extension of LARCH(∞) models of [47, 58].

Example 2.1 (Markov models) The first example is the standard and well known p-Markov system :

$$X_t = F(X_{t-1},\ldots,X_{t-p};\xi_t). \tag{2.7}$$

[77] stresses the fact that such equations describe all the p Markov processes. As soon as $\sum_{j=1}^p a_j = a < 1$, [28] prove the existence of $\alpha<1$ such that $\tau(r)\leqslant \alpha^r$ for solutions of this system for $t>0$ with X_0 following the stationary distribution. The first part of our proof is devoted to construct a solution of eqn. (2.7) on all $\mathbb{Z}$ and to improve the rate setting $\tau(r)\leqslant a^{r/p}$ (see Section 2.4 for more details).

- **Nonlinear autoregressive models**
  If we let $E'=E$ and

  $$F(x_1,\ldots,x_p;s) = R(x_1,\ldots,x_p)+s,$$

  nonlinear autoregressive models can be written as in eqn. (2.7). In this case, condition (2.5) follows from $\|\xi_0\|_m<\infty$ while condition (2.6) becomes

  $$\|R(y_1,\ldots,y_p)-R(x_1,\ldots,x_p)\| \leqslant \sum_{j=1}^p a_j\|x_j-y_j\|, \text{ with } \sum_{j=1}^p a_j < 1.$$

  The fact that the series does not depend on all its past leads to better asymptotic results than those implied by the weak dependence result of Theorem 2.1 (see [51] for detailed results).

– **Random $AR(1)$ models**
Solutions of the equation $Y_t = A_t Y_{t-1} + \zeta_t$, occur naturally as iterated random functions; [31] show that these models present fractal behaviours. Here $X_t = Y_t$ with $F(x_1; (a, \zeta)) = ax_1 + \zeta$ and $\xi_t = (A_t, \zeta_t)$. The iid input sequence can be written as $\xi_t = (A_t, \zeta_t)$ with A_t a $d \times d$-matrix and ζ_t an E-valued sequence. This model satisfies $\|F(x; \xi) - F(x'; \xi)\|_m \leqslant \|A_0\|_m \|x_1 - x_1'\|$ and condition (2.6) is satisfied as soon as $\|A_0\|_m < 1$. These models could be seen as solutions of a stochastic recurrence equation. Weaker conditions related to a Lyapunov exponent for the existence of an *a.s.* solution are obtained in [17].

Example 2.2 (Random AR(∞) models) Let us consider the iid sequence $\{\xi_t = ((A_{j,t})_{j>0}, \zeta_t)\}_{t\in\mathbb{Z}}$ with $A_{j,t}$ a $d \times d$ random matrix for all $j > 0$ and ζ_t an E-valued sequence. Here $E' = \mathcal{M}_{k,d}(\mathbb{R})^{\mathbb{N}^*} \times E$. An infinite memory extension of random $AR(1)$ models is the solution to the equation

$$X_t = \sum_{j=1}^{\infty} A_{j,t} X_{t-j} + \zeta_t.$$

If $\zeta_0 \in \mathbb{L}^m$ and $\sum_{j=1}^{\infty} \|A_{j,0}\|_m < 1$, conditions (2.5) and (2.6) hold.

Example 2.3 (Robust bilinear models) Here $d = 1$ ($E = \mathbb{R}$) and $E' = \mathbb{R}$. Solutions of the equation $X_t = A_t \xi_t + B_t$, where

$$A_t = \sum_{j=1}^{\infty} \alpha_j(X_{t-j}), \qquad B_t = \sum_{j=1}^{\infty} \beta_j(X_{t-j}).$$

extend on the bilinear models in [43] which correspond to linear functions, α_j, β_j. Classical models like ARCH or GARCH processes take this form. Assume $\sum_{j=1}^{\infty} \operatorname{Lip} \alpha_j < 1$, $\sum_{j=1}^{\infty} \operatorname{Lip} \beta_j < 1$ and $\xi_0 \in \mathbb{L}^m$, then conditions (2.5) and (2.6) are satisfied. A simple class of such models is provided with a fixed Lipschitz function h and $\alpha_j(x) = a_j h(x)$ and $\beta_j(x) = a_j h(x)$ for sequences of constants $\{a_j\}_{j>0}$, $\{b_j\}_{j>0}$. We consider bounded approximations of identity such that $h(x) = x \vee M \wedge (-M)$ to introduce more robust versions of the models in [43].

Example 2.4 (NLARCH(∞) models) A generalization of LARCH(∞) models in [58, 47] is given by equation

$$X_t \quad = \quad \xi_t \left(\alpha + \sum_{j=1}^{\infty} \alpha_j(X_{t-j}) \right).$$

where now $X_t \in E$, ξ_t is a $d \times k$ matrix (here $E' = \mathcal{M}_{k,d}(\mathbb{R})$), $\alpha \in \mathbb{R}^k$ and $\alpha_j : E \to \mathbb{R}^k$ are Lipschitz functions; in [58, 47] linear functions $\alpha_j(x) = c_j x$ are considered for $k \times d$ matrices c_j. Assumption (2.6) holds as soon as $\|\xi_0\|_m \sum_{j=1}^{\infty} \operatorname{Lip} \alpha_j < 1$.

Example 2.5 (Models with linear inputs) Let $f : \mathbb{R}^k \times E' \to E$ be measurable an satisfy $\|f(t,\xi_0) - f(s,\xi_0)\|_m \leqslant L\|t-s\|$ for some finite constant, we consider

$$X_t = f(A_t,\xi_t), \qquad A_t = \sum_{j=1}^{\infty} c_j X_{t-j},$$

here c_j denote $k \times d$ matrices. Then the previous relation (2.6) holds with $a_j = L\|c_j\|$. This is a very nonlinear case for which one only needs to produce a function of two variables and a sequence of constants. Kac used such type of mean field models in statistical mechanics in [76].

Example 2.6 (Threshold models) Setting $\xi_t = (\zeta_t, \nu_t)$, threshold models may be written as

$$X_t = \sum_{j=1}^{J} \mathbb{1}_{\nu_t \in E_j} F_j(X_{t-1}, X_{t-2}, \dots; \zeta_t)$$

for a measurable partition of $(E_j)_{1 \leqslant j \leqslant J}$ of the space of values E' of ν_t. More complicated threshold models are given by a measurable partition $E_1, \dots, E_J$ of $\mathbb{R}^k \times E'$ and

$$X_t = \sum_{j=1}^{J} \mathbb{1}_{\{(G(X_{t-1}, X_{t-2}, \dots), \nu_t) \in E_j\}} F_j(X_{t-1}, X_{t-2}, \dots; \zeta_t)$$

for a given function G of the past of the process with value in $\mathbb{R}^k$. Conditions (2.5) and (2.6) hold if the functions F_j are sufficiently contractive for $j \geqslant 1$.

Example 2.7 (Affine models) Let us consider the special case of chains with infinite memory that can be written in a bilinear form

$$X_t = M_t \xi_t + f_t, \tag{2.8}$$

where $M_t = M(X_{t-1}, \dots)$ and $f_t = f(X_{t-1}, \dots)$ are both Lipschitz functions of the past values $(X_{t-1}, X_{t-2}, X_{t-3}, \dots)$. If

$$\|\xi_0\|_\Phi \sum_{i=1}^{\infty} \operatorname{Lip} M_i + \sum_{i=1}^{\infty} \operatorname{Lip} f_i < 1,$$

the existence of a unique solution of eqn. (2.8) that has finite moment and weakly dependent properties follow from Theorem 2.1. This model enlarges many classical econometric equations (ARCH, GARCH, ARMA, ARMA-GARCH, etc. . .). Using ideas from [43], we also prove in the appendix the existence of the joint densities of the solution. Such result is necessary to obtain nonparametric estimators behavior, see [100].

Proposition 2.2 (Regularity of affine models) Here $E = E' = R^d$ for some $d \geqslant 1$. Suppose that the innovations $(\xi_t)_{t\in\mathbb{Z}}$ in the model (2.8) admit a common bounded marginal density f_ξ. Moreover, if $\inf_{(x_j)_{j>0}} \det M((x_j)_{j>0}) = \underline{M} > 0$, the marginal densities $f_{X_1,\ldots,X_n}$ of $(X_1,\ldots,X_n)$ exist for all $n > 0$ and satisfy

$$\|f_{X_1,\ldots,X_n}\|_\infty \leqslant \underline{M}^{-n}\|f_\xi\|_\infty^n.$$

Remark 2.1 *By integration we obtain the existence and the boundary of the marginal density of* X_0 *and of the joint densities of the couples* $(X_0, X_k)_{k>0}$.

Proof 1 *The solution, written as* $X_t = H(\xi_t, \xi_{t-1}, \ldots)$ *in Section 2.4.3, is independent of* $(\xi_j)_{j>t}$. *Then, if* G *is a continuous and bounded function,*

$$\begin{aligned}
\mathbb{E}\, G(X_1) &= \mathbb{E}\, G(M(X_0,\ldots)\xi_1 + f(X_0,\ldots)) \\
&= \int\!\!\int G(M(u)s_1 + f(u))\, f_\xi(s_1)ds_1\mathbb{P}_{(X_0,\ldots)}(du) \\
&\leqslant \underline{M}\int\!\!\int G(x_1)\, f_\xi(M^{-1}(u)(x_1 - f(u)))\mathbb{P}_{(X_0,\ldots)}(du)ds_1.
\end{aligned}$$

The last inequality follows from the substitution $M(u)s_1 + f(u) = x_1$, *the assumption ensuring that* $M(u)$ *is invertible. We obtain*

$$f_{X_1}(x_1) \leqslant \underline{M}^{-1}\int f_\xi(M^{-1}(u)(x_1 - f(u)))\mathbb{P}_{(X_0,\ldots)}(du) \leqslant \underline{M}^{-1}\|f_\xi\|_\infty.$$

Let prove the proposition by recursion that $\|f_{X_1,\ldots,X_n}\|_\infty \leqslant \underline{M}^{-n}\|f_\xi\|_\infty^n$.

$$\begin{aligned}
\mathbb{E}\, G(X_1,\ldots,X_n) &= \mathbb{E}\, G(X_1,\ldots,X_{n-1}, M(X_{n-1},\ldots)\xi_n + f(X_{n-1},\ldots)) \\
= \int\!\!\int\!\!\int & G(x_1,\ldots,x_{n-1}, M(x_{n-1},\ldots,x_1,u)s_n + f(x_{n-1},\ldots,x_1,u)) \\
& f_\xi(s_n)ds_n f_{(X_1,\ldots,X_{n-1})}(x_1,\ldots,x_{n-1})dx_1\cdots dx_{n-1}\mathbb{P}_{(X_0,\ldots|X_1,\ldots,X_{n-1})}(du).
\end{aligned}$$

The substitution $M(x_{n-1},\ldots,x_1,u)s_n + f(x_{n-1},\ldots,x_1,u) = x_n$ *leads to*

$$\begin{aligned}
f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) \leqslant \underline{M}^{-1}\iint & f_\xi(M^{-1}(x_{n-1},\ldots,x_1,u)(x_n - f(x_{n-1},\ldots,x_1,u))) \\
& f_{(X_1,\ldots,X_{n-1})}(x_1,\ldots,x_{n-1})dx_1\cdots dx_{n-1}\mathbb{P}_{(X_0,\ldots|X_1,\ldots,X_{n-1})}(du).
\end{aligned}$$

We obtain the recursive assumption with $\|f_{X_1,\ldots,X_{n-1}}\|_\infty \leqslant \underline{M}^{-(n-1)}\|f_\xi\|_\infty^{n-1}$.

2.4 Proofs of the main results

Firstly, we prove a useful algebraic result, see Lemma 2.1. Secondly, we construct a solution of eqn. (2.7) on all $\mathbb{Z}$ and then we use it to approximate the solution of eqn. (2.3). Then, existence of the infinite memory chain is obtained with $p \to \infty$ in Subsection 2.4.3 and weak dependence properties are derived from both a sharp control of the coefficient τ for the $p-$Markov approximations (see Lemma 2.4) and the use of coupling techniques. Finally the proof of corollary 2.1 is given.

2.4.1 An algebraic preliminary

Lemma 2.1 *Let $u_0 \geqslant 0$ and $(u_n)_{n\in\mathbb{Z}}$ be a real sequence such that $|u_n| \leqslant u_0$ if $n < 0$. Assume that*

$$u_n = \sum_{i=1}^{p} \alpha_i u_{n-i}, \qquad \forall n \geqslant 0, \tag{2.9}$$

where $(\alpha_1, \ldots, \alpha_p)$ are fixed nonnegative numbers with $\alpha = \sum_{i=1}^{p} \alpha_i$. Then,

$$u_n \leqslant \alpha^{n/p} u_0, \qquad \forall n \geqslant 0.$$

Proof 2 *Let us denote by $Q(x) = x^p - \alpha_1 x^{p-1} - \cdots - \alpha_p x^0$ and by ρ the largest modulus of the roots of the polynomial Q. We may bound this modulus ρ of the largest zero of Q, this zero is real and nonnegative as this is proved in [96] (Exercise 20, page 106) and*

$$\rho \leqslant \max_{1\leqslant i\leqslant p} \left(\frac{\alpha_i}{c_i}\right)^{1/i},$$

for all $c_1, \ldots, c_n \geqslant 0$ such that $c_1 + \cdots + c_n \leqslant 1$. With the appropriate choice $c_i \leqslant \alpha_i(1+\varepsilon)$ where $1/(1+\varepsilon) \geqslant \alpha$, we obtain $\rho \leqslant \alpha^{1/p}$.
Now, let $U_n^{(p)} = (u_n, \ldots, u_{n-p+1})'$ and let C_α be the companion matrix associated with $Q(x)$. Let $b \in]0,1]$ to be determined later. We define a norm on $\mathbb{R}^p$ by $\|(z_1, \ldots, z_p)'\|_b = \max_{1\leqslant i\leqslant p} |z_i b^{i-1}|$. Then

$$\left\|U_n^{(p)}\right\|_b = \left\|C_\alpha U_{n-1}^{(p)}\right\|_b.$$

Since $\|C_\alpha\|_b = \sup_{\|z\|_b\leqslant 1} \|C_\alpha z\|_b$ simple calculations yield

$$\|C_\alpha\|_b = \sup_{\|z\|_b\leqslant 1} \left\{ \left|\sum_{i=1}^{p} \alpha_i z_i\right| \vee \max_{1\leqslant i\leqslant p-1} b^i |z_i| \right\} \leqslant \sum_{i=1}^{p} \alpha_i b^{1-i} \vee b.$$

Now, choosing $b = \rho$, then we get

$$\|C_\alpha\|_\rho \leqslant \rho \leqslant \alpha^{1/p}.$$

To conclude we have

$$u_n \leqslant \left\|U_n^{(p)}\right\|_\rho \leqslant \|C_\alpha\|_\rho^n \left\|U_0^{(p)}\right\|_\rho \leqslant \alpha^{n/p} \left\|U_0^{(p)}\right\|_\rho \leqslant \alpha^{n/p} u_0.$$

2.4.2 Markov stationary approximation

In order to construct a solution to eqn. (2.3) we consider, for each fixed $p \geqslant 0$ and $q > 0$, the p-Markov process $(X_{p,q,t})_{t\geqslant 0}$ defined by $X_{p,q,t} = 0$ for $t \leqslant -q$ and through the recurrence equation for $t > q$

$$X_{p,q,t} = F(X_{p,q,t-1}, \ldots, X_{p,q,t-p}, 0, \ldots; \xi_t). \tag{2.10}$$

We first notice that $X_{0,q,t} = F(0, \ldots; \xi_t)$ for $t > -q$ is an iid sequence. The Lipschitz condition (2.6) implies

$$\begin{aligned}\|X_{p,q+1,0} - X_{p,q,0}\|_m &\leqslant \sum_{i=1}^{p} a_i \|X_{p,q+1,-i} - X_{p,q,-i}\|_m \\ &\leqslant \sum_{i=1}^{p} a_i \|X_{p,q+1-i,0} - X_{p,q-i,0}\|_m .\end{aligned}$$

The second inequality derives from the fact that by definition $X_{p,q,-i}$ and $X_{p,q-i,0}$ have the same law for each triplet of positive integers (p,q,i). Let us consider the sequence $v_n = \|X_{p,n+1,0} - X_{p,n,0}\|_m$ for $n \in \mathbb{Z}$. Here $v_n = 0$ if $n < 0$. For $n > 0$

$$v_n \leqslant \sum_{i=1}^{p} a_i v_{n-i}.$$

Then, $v_n \leqslant u_n$ for any real sequence $(u_n)_{n\in\mathbb{Z}}$ verifying eqn. (2.9) with $\alpha_i = a_i$, $u_n = 0$ for $n < 0$ and $u_0 = v_0$. Then, using Lemma 2.1, we reach the bound

$$v_n \leqslant a^{n/p} v_0 \leqslant a^{n/p} \|X_{p,1,0}\|_m \leqslant a^{n/p} \|F(0, \ldots; \xi_t)\|_m \leqslant a^{n/p} \mu_0.$$

Hence, for each p, $(X_{p,n,0})_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{L}^m$; it converges to some $X_{p,0} \in \mathbb{L}^m$. It is also clear that $X_{p,n,0}$ is measurable with respect to the $\sigma-$algebra generated by $\{\xi_t, t \leqslant 0\}$. The $\mathbb{L}^m$-convergence ensures that this is also the case for $X_{p,0}$. Hence $X_{p,0} = H_p(\xi_0, \xi_{-1}, \ldots)$ for some suitable measurable function H_p. The same argument for each $t \in \mathbb{Z}$ proves that $X_{p,t} = H_p(\xi_t, \xi_{t-1}, \xi_{t-2}, \ldots)$. Let $\overline{X}_{p,n,t} = (X_{p,n,t}, \ldots, X_{p,n,t-p})$ for each $t \in \mathbb{Z}$. As previously, the $(E)^{p+1}$-valued sequence $(\overline{X}_{p,n,0})_{n\in\mathbb{N}}$ converges in $\mathbb{L}^m$ (as $n \to \infty$). Its marginal distributions satisfy relation (2.10). As $n \uparrow \infty$, a continuity argument (on F) implies that $X_{p,0} = F(X_{p,-1}, \ldots, X_{p,-p}, 0, \ldots; \xi_0)$. For each $t \in \mathbb{Z}$, we apply the same argument to $\overline{X}_{p,n,t}$. Then the sequence $(X_{p,t})_{t\in\mathbb{Z}}$ is a stationary solution of the recurrence equation (2.10) for each $p \geqslant 0$.
Now, consider

$$\mu_p = \|X_{p,t}\|_m, \qquad \mu = \sup_{p\geqslant 0} \mu_p, \tag{2.11}$$

$$\Delta_{p,t} = \|X_{p+1,t} - X_{p,t}\|_m, \qquad \Delta_p = \sup_{t\in\mathbb{Z}} \Delta_{p,t}, \tag{2.12}$$

Note that the definition of μ_p given here for $p > 0$ can also be extended to $p = 0$ since $X_{0,t} = F(0, \ldots; \xi_t)$ satisfies by definition $\|X_{0,t}\|_m = \mu_0$ from eqn. (2.5). Then from eqn. (2.6), we obtain

$$\mu_p \leqslant \|X_{p,t} - X_{0,t}\|_m + \mu_0 \leqslant \sum_{j=1}^{p} a_j \|X_{p,t-j}\|_m + \mu_0 \leqslant \mu_p \sum_{j=1}^{p} a_j + \mu_0,$$

hence $\mu_p \leqslant \dfrac{1}{1-a} \cdot \mu_0$. We state this useful result as Lemma 2.2.

Lemma 2.2 *Assume properties (2.5) and (2.6) hold for some $m \geqslant 1$, then the expression defined by (2.11) satisfies the bound*

$$\mu = \sup_{p \geqslant 0} \mu_p \leqslant \frac{\mu_0}{1-a}.$$

In a similar way, we get

$$\begin{aligned}\Delta_{p,t} &= \left\| F(X_{p+1,t-1}, \ldots, X_{p+1,t-p-1}, 0, \ldots ; \xi_t) - F(X_{p,t-1}, \ldots, X_{p,t-p}, 0, \ldots ; \xi_t) \right\|_m \\ &\leqslant \sum_{j=1}^{p} a_j \|X_{p+1,t-j} - X_{p,t-j}\|_m + a_{p+1} \|X_{p+1,t-p-1}\|_m \\ &\leqslant \sum_{j=1}^{p} a_j \Delta_{p,t-j} + a_{p+1} \|X_{p+1,t-p-1}\|_m\end{aligned}$$

The previous equation implies $\Delta_p \leqslant \dfrac{a_{p+1}}{1-a} \cdot \mu$, thus with Lemma 2.2 we obtain Lemma 2.3.

Lemma 2.3 *Assume properties (2.5) and (2.6) for some $m \geqslant 1$, then the expression defined by eqn. (2.11) satisfies the bound*

$$\Delta_p \leqslant a_{p+1} \cdot \frac{\mu_0}{(1-a)^2}.$$

2.4.3 Proof of existence in Theorem 2.1

Quote first that Lemma 2.3 implies that $X_{p,t} \to X_t$ in $\mathbb{L}^m$ since this space is complete. The continuity of F ensures that X_t is a solution of eqn. (2.3). Furthermore, as a limit in $\mathbb{L}^m$ of stationary processes admitting a moment of order m, X_t is also stationary (in law) and $\|X_t\|_m < \infty$. Finally, $X_t = H(\xi_t, \xi_{t-1}, \xi_{t-2}, \ldots)$ as a limit in $\mathbb{L}^m$ of $X_{p,t} = H_p(\xi_t, \xi_{t-1}, \xi_{t-2}, \ldots)$.

2.4.4 Proof of weak dependence in Theorem 2.1

Previous assumptions with $m = 1$ are enough for this section. We need the following lemma.

Lemma 2.4 *Assume that (2.6) and (2.5) hold with $m = 1$, then the Markov chain $(X_{p,t})_t$ in eqn. (2.10) is weakly dependent with*

$$\tau_p(r) \leqslant a^{r/p}.$$

Proof 3 *Our proof follows the lines of [51]. We improve the result from [28] by using the Lemma 2.1. As quoted in Section 2.2, we use coupling to evaluate the τ-coefficient. Let $(\xi'_t)_t$ be a process distributed as $(\xi_t)_t$ and independent of the innovation. We define the process $(X^*_{p,t})_t$ as*

$$X^*_{p,t} = \begin{cases} F(X^*_{p,t-1}, \ldots, X^*_{p,t-p}, 0, \ldots ; \xi'_t), & \textit{for } t \leqslant 0; \\ F(X^*_{p,t-1}, \ldots, X^*_{p,t-p}, 0, \ldots ; \xi_t), & \textit{for } t > 0; \end{cases}$$

Using the approximation of the Section 2.4.2, we equivalently find a sequence of measurable variables with respect to the $\sigma-$algebra generated by $\xi'_t, t \leqslant 0$ denoted $(X^*_{p,n,0})_{n\in\mathbb{N}}$ such that it converges in $\mathbb{L}^m$ to $X^*_{p,0} \in \mathbb{L}^m$. The $\mathbb{L}^m$-convergence ensures this is also the case for $X_{p,0}$. Then, by definition of $\xi'_t, t \leqslant 0$, $X^*_{p,0}$ is independent of $X_{p,0}$. Due to the coupling property of the coefficient τ, we obtain $\tau_p(r) \leqslant \|X_{p,r} - X^*_{p,r}\|_1$.
It is easy to check that assumption (2.6) leads to

$$\|X_{p,r} - X^*_{p,r}\|_1 \leqslant \sum_{i=1}^{p} a_i \|X_{p,r-i} - X^*_{p,r-i}\|_1.$$

Denoting $w_r = \|X_{p,r} - X^*_{p,r}\|_1$ for $r \in \mathbb{Z}$, we use again the Lemma 2.4.2 and the relation $\|F(0,\ldots;\xi_0)\|_1 = \mu_0$ (for $m = 1$) to obtain

$$\tau_p(r) \leqslant w_r \leqslant a^{r/p} w_0 \leqslant 2\mu_p a^{r/p} \leqslant 2\frac{\|F(0,\ldots;\xi_0)\|_1}{1-a} \cdot a^{r/p}.$$

Now, let us define the process $(X^*_t)_t$ by

$$X^*_t = \begin{cases} F(X^*_{t-1}, X^*_{,t-2}, \ldots; \xi'_t), & \text{for } t \leqslant 0; \\ F(X^*_{t-1}, X^*_{t-2}, \ldots; \xi_t), & \text{for } t > 0; \end{cases}$$

We remark that $(X^*_t)_t$ is also a stationary chain with infinite memory. Lemma 2.3 gives

$$\|X_r - X_{p,r}\|_1 \leqslant \sum_{k=p}^{\infty} \Delta_k \leqslant C \sum_{k=p}^{\infty} a_{k+1}.$$

The same bound also holds for the quantity $\|X^*_r - X^*_{p,r}\|_1$. For each integer p

$$\begin{aligned} \tau(r) &\leqslant \|X_r - X^*_r\|_1 \\ &\leqslant \|X_r - X_{p,r}\|_1 + \|X_{p,r} - X^*_{p,r}\|_1 + \|X^*_r - X^*_{p,r}\|_1 \\ &\leqslant A\left(a^{r/p} + \sum_{k=p}^{\infty} a_{k+1}\right). \end{aligned}$$

If $a_j \leqslant ce^{-\beta j}$, we choose $p \approx \sqrt{\alpha r/\beta}$ where $e^{-\alpha} = a < 1$ to derive the bound $\tau(r) \leqslant Ce^{-\sqrt{\alpha\beta r}}$ for some suitable constant $C > 0$. If $a_j \leqslant cj^{-\beta}$, we choose p such that $p \ln p(1-\beta)/\ln a \approx r$ then there exists $C > 0$ such that $\tau(r) \leqslant Cp^{1-\beta}$. Finally remark that $\ln r \propto \ln p + \ln\ln p$ and then $\ln r/r \propto 1/p(1 + \ln\ln p/\ln p) \propto 1/p$ as p tends to infinity with r.

2.4.5 Proof of Corollary 2.1

2.4.6 Proof of Corollary 2.1

We define recursively the process $\left(\tilde{X}_k\right)_{k\geqslant 1}$ by the relation

$$\tilde{X}_k = F\left(\tilde{X}_{k-1}, \ldots, \tilde{X}_1, 0, \ldots; \xi_k\right).$$

We work conditionnally to the event $\{\xi_1, \dots, \xi_n\} = U$ and we fix $\tilde{X}_i = 0$ for $i \leqslant 0$. Let n be a fixed integer and $s_n \leqslant n-1$. Let $(X_t)_{t\in\mathbb{Z}}$ be the stationary solution of $X_t = F(X_{t-1}, X_{t-2}, 0, \dots; \xi_t)$. The Lipschitz assumption (2.6) leads to, for $1 \leqslant k \leqslant n$

$$\left\|\tilde{X}_k - X_k\right\|_\Phi \leqslant \sum_{i=1}^{k-1} a_i \left\|\tilde{X}_{k-i} - X_{k-i}\right\|_\Phi + \|X_0\|_\Phi \sum_{i\geqslant k} a_i.$$

The sequence $v_k = \left\|\tilde{X}_{k+1} - X_{k+1}\right\|_\Phi$ satisfies the recursion, for all $k \geqslant 1$

$$v_k \leqslant \sum_{j=1}^{k} a_j v_{k-j} + u_k$$

with $u_k := \|X_0\|_\Phi \sum_{j>k} a_j$ for $k \geqslant 1$. Remark that $u_k \downarrow_{k\to\infty} 0$. We first prove the boundness of $(v_k)_{k\in\mathbb{N}}$. Let ℓ be a fixed integer. For all k such that $\ell \geqslant k$, $v_k \leqslant a \sup_{i\leqslant \ell} v_i + u_1$. We deduce that $\sup_{i\leqslant\ell} v_i \leqslant u_1$. Finally $\|v\|_\infty \leqslant a\|X_0\|_\Phi/(1-a)$.
Let s be an integer. Now for all k such that $\ell \geqslant k+s$,

$$\begin{aligned} v_\ell &\leqslant \sum_{j=1}^{k} a_j v_{\ell-j} + \sum_{j=k+1}^{\ell} a_j v_{\ell-j} + u_\ell \\ &\leqslant a \sup_{j\geqslant s} v_j + \|v\|_\infty \sum_{j=k+1}^{\infty} a_j + u_{k+s}. \end{aligned}$$

This inequality holds for all $\ell \geqslant k+s$ then

$$\sup_{j\geqslant k+s} v_j \leqslant a \sup_{j\geqslant s} v_j + \|v\|_\infty \sum_{j=k+1}^{\infty} a_j + u_k.$$

We deduce that

$$\sup_{j\geqslant nk} v_j \leqslant a^n \|v\|_\infty + \frac{1}{1-a}\left(\|v\|_\infty \sum_{j=k+1}^{\infty} a_j + u_k\right).$$

Using the bound on $\|v\|_\infty$ leads to the result. **Final remark.** Statistical issues of the general models introduced here will be studied in more details in a further chapter ; their identification and their use for real data will be especially detailed.

Chapitre 3

An invariance principle for weakly dependent stationary general models

Abstract

The aim of this chapter is to refine a weak invariance principle for stationary sequences given by Doukhan & Louhichi (1999). Since our conditions are not causal our assumptions need to be stronger than the mixing and causal θ-weak dependence assumptions used in Dedecker & Doukhan (2003). Here, if moments of order > 2 exist, a weak invariance principle and convergence rates in the CLT are obtained; Doukhan & Louhichi (1999) assumed the existence of moments with order > 4. Besides the previously used η- and κ-weak dependence conditions, we introduce a weaker one, λ, which fits the Bernoulli shifts with dependent inputs.

Keywords : Invariance principle, weak dependence, the Bernoulli shifts

Note

The content of this part is based on a paper, written in collaboration with Paul Doukhan, accepted in Probability and Mathematical Statistics.

3.1 Introduction

Let $(X_t)_{t\in\mathbb{Z}}$ be a real-valued stationary process. A huge amount of applications make use of such times series.

Several ways of modeling weak dependence have already been proposed. One of the most popular is the notion of mixing, see [38] for bibliography ; this notion leads to a very nice asymptotic theory, in particular a weak invariance principle under very sharp conditions (see [101] for the strong mixing case). Such mixing conditions entail restrictions on the model. For example, Andrews exhibits in [1] the simple counter-example of an auto-regressive process which does not satisfy any mixing condition and innovations need much regularity in both MA(∞) and Markov models. Doukhan & Louhichi introduced in [41] new weak dependence conditions in order to solve those problems. We intend to sharpen their assumptions leading to a weak invariance principle.

A common approach to derive a weak invariance principle for stationary sequences is based on a martingale difference approximation. This approach was first explored by Gordin in [62] ; necessary and sufficient conditions were found by Heyde in [69]. Let $\mathcal{M}_t$ be a filtration. Heyde's martingale difference approximation is equivalent to the existence of moments of order 2 and

$$\sum_{t=0}^{\infty}\mathbb{E}(X_t|\mathcal{M}_0) \quad \text{and} \quad \sum_{t=0}^{\infty}(X_{-t}-\mathbb{E}(X_{-t}|\mathcal{M}_0)) \qquad \text{converge in } \mathbb{L}^2. \tag{3.1}$$

Martingale theory leads directly to invariance principles (see also [111]). In the following, the adapted case refers to the special case where X_t is $\mathcal{M}_t$-measurable. The natural filtration is written as $\mathcal{M}_t = \sigma(Y_i, i\leqslant t)$ for independent and identically distributed inputs $(Y)_{t\in\mathbb{Z}}$; thus X_t can be written as a function of the past inputs :

$$X_t = H(Y_t, Y_{t-1},\ldots). \tag{3.2}$$

Then only the first series in (3.1) needs to be considered. Using the Lindeberg technique, Dedecker & Rio relax (3.1) in [30]. Bernstein's blocks method allowed Peligrad & Utev to also improve on (3.1) in [91]. Such projective conditions are related to dependence coefficients ; Dedecker & Doukhan obtain sharp results for the causal θ-dependence in [26] and Merlevède *et al.* address the mixing cases in a nice survey paper [88].

Martingale difference approximation is not always easy, for instance in the particular case where a natural filtration does not exist. The most striking example is given by associated sequences $(X_t)_{t\in\mathbb{Z}}$. Let us recall this notion. A series is said to be associated if $\mathrm{Cov}(f_1,f_2)\geqslant 0$ for any two coordinatewise nondecreasing functions f_1 and f_2 of $(X_{t_1},\ldots,X_{t_m})$ with $\mathrm{Var}\,(f_1)+\mathrm{Var}\,(f_2)<\infty$. However, Newman & Wright obtain in [90] a weak invariance principle under the existence of second order moments and

$$\sigma^2=\sum_{t\in\mathbb{Z}}\mathrm{Cov}(X_0,X_t)<\infty. \tag{3.3}$$

Theorems 3.1 and 3.2 propose invariance principles under general assumptions : they apply to the non-causal Bernoulli shifts with weakly dependent inputs $(Y_t)_{t\in\mathbb{Z}}$,

$$X_t = H(Y_{t-j}, j \in \mathbb{Z}). \tag{3.4}$$

Heredity of weak dependence through such non-linear functionals follows from a new λ-weak dependence property ; a function of a λ-weak dependence process is λ-weakly dependent, see Section 3.3.2. Analogous models with dependent inputs are already considered by [16]. If $X_t = \sum_{j\in\mathbb{Z}} \alpha_j Y_{t-j}$, Peligrad & Utev prove in [92] that the Donsker invariance principle holds for X as soon as it holds for the innovation process Y. The non-linearity of H considered here is an important feature which has not been frequently discussed in the past. The condition of moments with order > 4 on the observations needed in [41] is reduced to a one of moments with order > 2 and the results rely on specific decays of the dependence coefficients. We do not reach the second order moment condition of [69] (or projective conditions) and [90]. We *conjecture* that some times series satisfy weak dependence conditions with fast enough decay rates in order to ensure a Donsker type theorem but they do not satisfy neither condition (3.1) nor other projective criterion (see [88]) nor association nor Gaussianity. More general models (3.4) are considered here while causal models (3.2) fit to the adapted case and to projective conditions. However, proving this conjecture is really difficult since condition (3.1) has to be checked for each σ-algebra $\mathcal{M}_0$.

The chapter is organized as follows. In Section 3.2 we introduce various weak dependent coefficients in order to state our main results. Section 3.3 is devoted to examples of weak dependent models for which we discuss our results. We shall focus on examples of λ-weakly dependent sequences. Proofs are given in the last section ; we first derive conditions ensuring the convergence of the series σ^2. A bound of the Δ-moment of a sum (with $2 < \Delta < m$) is proved in Section 3.4.2 ; this bound is of an independent interest since *eg.* it directly yields the strong laws of large number. The standard Lindeberg method with Bernstein's blocks is developed in Section 3.4.3 and yields our versions of the Donsker theorem. Convergence rates of the CLT are obtained in Section 3.4.4.

3.2 Definitions and main results

3.2.1 Weak dependence assumptions

Definition 3.1 (Doukhan & Louhichi, 1999) *The process $(X_t)_{t\in\mathbb{Z}}$ is said to be (ε,ψ)-weakly dependent if there exist a sequence $\varepsilon(r) \downarrow 0$ (as $r \uparrow \infty$) and a function $\psi : \mathbb{N}^2 \times (\mathbb{R}^+)^2 \to \mathbb{R}^+$ such that*

$$|Cov(f(X_{s_1}, \dots, X_{s_u}), g(X_{t_1}, \dots, X_{t_v}))| \leqslant \psi(u, v, Lip\, f, Lip\, g)\varepsilon(r),$$

for any $r \geqslant 0$ and any $(u+v)$-tuples such that $s_1 \leqslant \cdots \leqslant s_u \leqslant s_u + r \leqslant t_1 \leqslant \cdots \leqslant t_v$, where the real valued functions f, g are defined respectively on $\mathbb{R}^u$ and $\mathbb{R}^v$, satisfy $\|f\|_\infty \leqslant 1$, $\|g\|_\infty \leqslant 1$ and are such that $Lip\, f + Lip\, g < \infty$ where

$$Lip\, f = \sup_{(x_1,\ldots,x_u)\neq(y_1,\ldots,y_u)} \frac{|f(x_1,\ldots,x_u) - f(y_1,\ldots,y_u)|}{|x_1 - y_1| + \cdots + |x_u - y_u|}$$

Specific functions ψ yield notions of weak dependence appropriate to describe various examples of models :

- κ-weak dependence for which $\psi(u,v,a,b) = uvab$, in this case we simply denote $\varepsilon(r)$ as $\kappa(r)$;
- κ' (causal) weak dependence for which $\psi(u,v,a,b) = vab$, in this case we simply denote $\varepsilon(r)$ as $\kappa'(r)$; this is the causal counterpart of κ coefficients which is recalled only for completeness ;
- η-weak dependence, $\psi(u,v,a,b) = ua + vb$, in this case we write $\varepsilon(r) = \eta(r)$ for short ;
- θ-weak dependence is a causal dependence which refers to the function $\psi(u,v,a,b) = vb$, in this case we simply denote $\varepsilon(r) = \theta(r)$ (see [26]) ; this is the causal counterpart of η coefficients which is recalled only for completeness ;
- λ-weak dependence $\psi(u,v,a,b) = uvab + ua + vb$, in this case we write $\varepsilon(r) = \lambda(r)$.

Remarks.

- Besides the fact that it includes η and κ-weak dependences, this new notion of λ-weak dependence will be proved to be convenient, for example, for the Bernoulli shifts with associated inputs (see Lemma 3.1 below).
- If functions f and g are complex-valued, the previous inequalities remain true if we substitute $\varepsilon(r)/2$ to $\varepsilon(r)$. A useful case of such complex-valued functions is $f(x_1,\ldots,x_u) = \exp\left(it(x_1 + \cdots + x_u)\right)$ for each $t \in \mathbb{R}$, $u \in \mathbb{N}^*$ and $(x_1,\ldots,x_u) \in \mathbb{R}^u$ (see Section 3.4.3). This indeed corresponds to the characteristic function adapted to derive the convergence in distribution.

3.2.2 Main results

Let $(X_t)_{t\in\mathbb{Z}}$ be a real-valued stationary sequence of mean 0 satisfying

$$\mathbb{E}|X_0|^m < \infty, \qquad \text{for a real number } m > 2. \tag{3.5}$$

Let us assume that

$$\sigma^2 = \sum_{k\in\mathbb{Z}} \mathrm{Cov}(X_0, X_k) = \sum_{k\in\mathbb{Z}} \mathbb{E}X_0X_k \geqslant 0. \tag{3.6}$$

Denote by W the standard Brownian motion and by W_n the partial sums process

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} X_i, \quad \text{for } t \in [0,1],\ n \geqslant 1. \tag{3.7}$$

We now present our main results, which are new versions of the Donsker weak invariance principle.

Theorem 3.1 (κ-dependence) *Assume that the 0-mean κ-weakly dependent stationary process $(X_t)_{t\in\mathbb{Z}}$ satisfies eqn. (3.5) and $\kappa(r) = \mathcal{O}(r^{-\kappa})$ (as $r \uparrow \infty$) for $\kappa > 2 + \frac{1}{m-2}$ then the previous expression σ^2 is finite and*

$$W_n(t) \underset{n\to\infty}{\xrightarrow{\mathrm{D}}} \sigma W(t), \quad \textit{in the Skorohod space } D([0,1]).$$

Remark. Under the more restrictive κ' condition, Bulinski & Sashkin obtain in [19] invariance principles with the sharper assumption $\kappa' > 1 + \frac{1}{m-2}$. Our loss is explained by the fact that κ'-weakly dependent sequences satisfy $\kappa'(r) \geqslant \sum_{s\geqslant r} \kappa_s$. This simple bound directly follows from the definitions. The following result relaxes the previous dependence assumptions at the price of a faster decay of the dependence coefficients.

Theorem 3.2 (λ-dependence) *Assume that the 0-mean λ-weakly dependent stationary inputs satisfies eqn. (3.5) and $\lambda(r) = \mathcal{O}(r^{-\lambda})$ (as $r \uparrow \infty$) for $\lambda > 4 + \frac{2}{m-2}$ then σ^2 is finite and*

$$W_n(t) \underset{n\to\infty}{\xrightarrow{\mathrm{D}}} \sigma W(t), \quad \textit{in the Skorohod space } D([0,1]).$$

Remarks.

- We do not achieve better results for η or θ-weak dependence cases than the one for $\lambda-$dependence. In comparison with the result obtained by [26], our results are not as good under θ-weak dependence. We work under more restrictive moment conditions than these authors. The same remark applies for all projective measures of dependence; here, we refer to results by [69], [90], [30] and [91].
- However, the example of Section 3.3.2 stresses the fact that such results are not systematically better than those of Theorem 3.2; for such general examples, we even conjecture that theorems of [69], [90], [30] or [91] do not apply.
- The technique of the proofs is based on the Lindeberg method. In fact, we prove that $|\mathbb{E}\left(\phi(S_n/\sqrt{n}) - \phi(\sigma N)\right)| = o\left(n^{-c}\right)$ (ϕ denotes here the characteristic function) for $0 < c < c^*$ where c^* depends only on the parameters m and κ or λ respectively. If m and κ (or λ) both tend to infinity, we notice that $c^* \to \frac{1}{4}$. As κ or λ tends to infinity and $m < 3$, c^* always remains smaller than $(m-2)/(2m-2)$ (see Proposition 3.2 in Section 3.4.4 for more details).
- Using a smoothing lemma also yields an analogous bound for the uniform distance

$$\sup_{x\in\mathbb{R}} \left|\mathbb{P}\left(\frac{1}{\sqrt{n}} S_n \leqslant x\right) - \mathbb{P}\left(\sigma N \leqslant x\right)\right| = o\left(n^{-c}\right), \qquad \text{for some } c < c'.$$

A first and easy way to control c' is to let $c' = c^*/4$ but the corresponding rate is a really bad one (see *e.g.* in [45]). The Esséen inequality holds with the optimal exponent $\frac{1}{2}$ in the independent and identically distributed case (see [93]) and [101] reaches the exponent $\frac{1}{3}$ in the case of strongly mixing sequences. In Proposition 3.2 of Section 3.4.4, we achieve $c' > c^*/4$. Analogous results have been settled in Doukhan2002 for weakly dependent random fields. Previous results by [69], [90], [30] or [91] do not derive such convergence rates for the Kolmogorov distance.

Let us denote by $\mathbb{R}^{(\mathbb{Z})} = \bigcup_{I>0} \{z \in \mathbb{R}^{\mathbb{Z}} / \ z_i = 0, \ |i| > I\}$, the set of finite sequences of real numbers. We consider functions $H : \mathbb{R}^{(\mathbb{Z})} \to \mathbb{R}$ such that if $x, y \in \mathbb{R}^{(\mathbb{Z})}$ coincide for all indexes but one, let say $s \in \mathbb{Z}$, then

$$|H(x) - H(y)| \leqslant b_s(\|z\|^{\ell} \vee 1)|x_s - y_s| \tag{3.8}$$

where $z \in \mathbb{R}^{(\mathbb{Z})}$ is defined by $z_s = 0$ and $z_i = x_i = y_i$ if $i \neq s$. Here $\|x\| = \sup_{i\in\mathbb{Z}} |x_i|$. In Section 3.3.2, we prove the existence of the process $X_n = \lim_{I\to\infty} H\big((Y_{n-j}\mathbb{1}_{\{j\leqslant I\}})_{j\in\mathbb{Z}}\big)$ where $(Y_t)_{t\in\mathbb{Z}}$ is a weakly dependent real-valued input process. We denote this process by $X_n = H(Y_{n-j}, j \in \mathbb{Z})$ for simplicity and we derive its λ-weak invariance properties. Various asymptotic results follow, among which our weak invariance principle, Theorem 3.2.

Corollary 3.1 *Let $(Y_t)_{t\in\mathbb{Z}}$ be a stationary λ-weakly dependent process (with dependence coefficients $\lambda_Y(r)$) and $H : \mathbb{R}^{(\mathbb{Z})} \to \mathbb{R}$ satisfying the condition given by (3.8) for some $\ell \geqslant 0$. Let us assume that there exist real numbers m, m' with $\mathbb{E}|Y_0|^{m'} < \infty$ such that $m > 2$ and $m' \geqslant (\ell+1)m$.*
Then $X_n = H(Y_{n-i}, i \in \mathbb{Z})$ exists and satisfies the weak invariance principle in the following cases :
– ***Geometric case :*** *$b_r \leqslant Ce^{-b|r|}$ and $\lambda_Y(r) \leqslant De^{-ar}$ for $a, b, C, D > 0$.*
– ***Riemannian case :*** *If $b_r \leqslant C(1+|r|)^{-b}$ for some $b > 2$ and $\lambda_Y(r) \leqslant Dr^{-a}$ for $a, C, D > 0$ with*

$$\begin{aligned} a &> \frac{1+b}{b-1}\left(4 + \frac{2}{m-2}\right), \quad \text{if } \ell = 0, \ b > 1; \\ a &> \frac{b(m'-1+\ell)}{(b-2)(m'-1-\ell)}\left(4+\frac{2}{m-2}\right), \quad \text{if } \ell > 0, \ b > 2. \end{aligned} \tag{3.9}$$

Remark. The previous conditions are also tractable in the mixed cases. We explicitly state them for $\ell > 0$:
– $b_r \leqslant Ce^{-b|r|}$, $\lambda_Y(r) \leqslant Dr^{-a}$, if moreover $a > \dfrac{m'-1+\ell}{m'-1-\ell}\left(4+\dfrac{2}{m-2}\right)$ and $b, C, D > 0$.
– $b_r \leqslant C|r|^{-b}$ and $\lambda_Y(r) \leqslant De^{-ar}$, for $a, C, D > 0$ with $b > \dfrac{6m-10}{m-2}$.

3.3 Examples

Theorem 3.1 is useful to derive the weak invariance principle in various cases. This section is aimed at a detailed treatment of the Bernoulli shifts with dependent inputs. The important class of Lipschitz functions of dependent inputs is presented in a separate section. The importance of our results is highlighted by the models of the first subsection. More general non-linear models are considered in the second subsection. Some of those examples illustrate the conjecture we made in the introduction but we were not able to formally prove it.

3.3.1 Lipschitz processes with dependent inputs

Consider Lipschitz functions $H : \mathbb{R}^{(\mathbb{Z})} \to \mathbb{R}$, *i.e.* such that eqn. (3.8) is satisfied for $\ell = 0$. A simple example of this situation is the two-sided linear sequence

$$X_t = \sum_{i \in \mathbb{Z}} \alpha_i Y_{t-i} \tag{3.10}$$

with dependent inputs $(Y_t)_{t \in \mathbb{Z}}$. As quoted by [81] for the case of linear processes with dependent input there exists a very general solution ; essentially any Donsker type theorem for the stationary inputs implies the central limit theorem for any linear process driven by such inputs. More precisely, Theorem 5 of [92] states that this process even satisfies the Donsker invariance principle if $\sum_j |\alpha_j| < \infty$.
A simple example of Lipschitz non-linear functional of dependent inputs is

$$X_t = \left| \sum_{i \in \mathbb{Z}} \alpha_i Y_{t-i} \right| - \mathbb{E} \left| \sum_{i \in \mathbb{Z}} \alpha_i Y_{-i} \right| \tag{3.11}$$

In this case the inequality (3.8) holds with $\ell = 0$ and $b_r \leqslant |\alpha_r|$.
Another example of this situation is the following stationary process

$$X_t = Y_t \left(a + \sum_{j \neq 0} a_j Y_{t-j} \right) - \mathbb{E} Y_t \left(a + \sum_{j \neq 0} a_j Y_{t-j} \right),$$

where the inputs $(Y_t)_{t \in \mathbb{Z}}$ are bounded. In this case, the inequality (3.8) also holds with $\ell = 0$ and $b_s \leqslant 2\|Y_0\|_\infty |a_s|$.

To apply our result, we compute the weak dependence coefficients of such models.

Lemma 3.1 *Let $(Y_t)_{t \in \mathbb{Z}}$ be a strictly stationary process with a finite moment of order $m \geqslant 1$ and $H : \mathbb{R}^{(\mathbb{Z})} \to \mathbb{R}$ satisfying the condition (3.8) for $\ell = 0$ and some nonnegative sequence $(b_s)_{s \in \mathbb{Z}}$ such that $L = \sum_j b_j < \infty$. Then,*
- *the process $X_n = H(Y_{n-j}, j \in \mathbb{Z}) := \lim_{I \to \infty} H\big(Y_{n-j} 1\!\!1_{\{j \leqslant I\}}, j \in \mathbb{Z}\big)$ is a strictly stationary process with finite moments of order m.*
- *if the input process $(Y_t)_{t \in \mathbb{Z}}$ is λ-weakly dependent (the weak dependence coefficients are denoted by $\lambda_Y(r)$), then $(X_t)_{t \in \mathbb{Z}}$ is λ-weakly dependent with*

$$\lambda(k) \leqslant \inf_{2r \leqslant k} \left[2 \sum_{|i| \geqslant r} b_i \|Y_0\|_1 + (2r+1)^2 L^2 \lambda_Y(k - 2r) \right].$$

- *if the input process $(Y_t)_{t \in \mathbb{Z}}$ is η-weakly dependent (the weak dependence coefficients are denoted by $\eta_Y(r)$) then $(X_t)_{t \in \mathbb{Z}}$ is η-weakly dependent and*

$$\eta(k) \leqslant \inf_{2r \leqslant k} \left[2 \sum_{|i| \geqslant r} b_i \|Y_0\|_1 + (2r+1) L\, \eta_Y(k - 2r) \right].$$

Remark. Let $(Y_t)_{t\in\mathbb{Z}}$ be a strictly stationary process with a finite moment of order $m > 2$. If $L = \sum_j |\alpha_j| < \infty$, the process $X_n = \sum_{j\in\mathbb{Z}} \alpha_j Y_{n-j}$ is a strictly stationary process with finite moments of order m which satisfies the assumptions of Lemma 3.1 with $b_j = |\alpha_j|$. Even if the weak invariance principle is already given in [92] our result is of an independent interest, *e.g.* for functional estimation purposes. For non-linear Lipschitz functionals it yields new central limit theorems.
The result of Theorems 3.1 and 3.2 holds systematically in geometric cases. Then it is assumed Riemannian decays, *i.e.* there exists $\alpha, C > 0$ such that

$$b_r \leqslant Cr^{-\alpha}.$$

The conditions from [69] are compared below with the conditions of Theorems 3.1 and 3.2 for specific classes of inputs $(Y_t)_{t\in\mathbb{Z}}$.

LARCH(∞) inputs

A vast literature is devoted to the study of conditionally heteroskedastic models. A simple equation in terms of a vector-valued process allows a unified treatment of those models, see [47]. Let $(\xi_t)_{t\in\mathbb{Z}}$ be an independent and identically distributed centered real-valued sequence and $a, a_j, j \in \mathbb{N}^*$ be real numbers. LARCH(∞) models are solutions of the recurrence equation

$$Y_t = \xi_t \left(a + \sum_{j=1}^{\infty} a_j Y_{t-j} \right). \tag{3.12}$$

We provide below sufficient conditions for the following chaotic expansion

$$Y_t = \xi_t \left(a + \sum_{k=1}^{\infty} \sum_{j_1,\ldots,j_k \geqslant 1} a_{j_1}\xi_{t-j_1} a_{j_2} \cdots a_{j_k} \xi_{t-j_1-\cdots-j_k} a \right). \tag{3.13}$$

Assume that $\Lambda = \|\xi_0\|_m \sum_{j\geqslant 1} |a_j| < 1$ then one (essentially unique) stationary solution of eqn. (3.12) in $\mathbb{L}^m$ is given by eqn. (3.13). This solution is θ-weakly dependent with $\theta_Y(r) \leqslant Kr^{1-a}\log^{a-1} r$ for some constant $K > 0$. This implies the same bound on their coefficients $(\lambda_Y(r))_{r\geqslant 0}$. Condition (3.9) gives the weak invariant principle for $(X_t)_{t\in\mathbb{Z}}$ under the conditions $\mathbb{E}|\xi_0|^m < +\infty$ for $m > 2$, $\alpha > 1$, and

$$a > \frac{1+\alpha}{\alpha - 1}\left(4 + \frac{2}{m-2}\right) + 1.$$

The model (3.10) is also an Heyde's martingale difference approximation (3.1) as soon as

$$\sum_{k\geqslant 1} \sqrt{\sum_{i\geqslant k} \alpha_i^2} < +\infty.$$

Necessary conditions for weak invariance principle follow as $\alpha > 3/2$, $|a_j| \leqslant Cj^{-a}$ for some $a > 1$, $\mathbb{E}\xi_0^2 < +\infty$, and $\|\xi_0\|_2 \sum_{j\geqslant 1} |a_j| < 1$. These conditions are not optimal since in this case the process is adapted to the filtration $\mathcal{M}_t = \sigma(\xi_i, i \leqslant t)$. Peligrad & Utev extend in [91] the Donsker theorem to the cases where $\alpha > 1/2$. Thus, our conditions are not optimal compared to those of [92] in the linear case as in eqn. (3.10). However, for non-linear Lipschitz functional, the result seems to be new.

Non-causal LARCH(∞) inputs

The previous approach extends for the case of non-causal LARCH(∞) inputs

$$Y_t = \xi_t \left(a + \sum_{j\neq 0} a_j Y_{t-j} \right).$$

Doukhan *et al.* prove in [47] the same results of existence as for the previous causal case (just replace summation over $j > 0$ by summation over $j \neq 0$) and the dependence becomes of the η type with

$$\eta(r) = \left(\|\xi_0\|_\infty \sum_{0\leqslant 2k<r} k\Lambda^{k-1} A\left(\frac{r}{2k}\right) + \frac{\Lambda^{r/2}}{1-\Lambda} \right) \mathbb{E}|\xi_0||a|$$

where $A(x) = \sum_{|j|\geqslant x} |a_j|$, $\Lambda = \|\xi_0\|_\infty \sum_{j\geqslant 1} |a_j| < 1$. From condition (3.9) the weak invariance principle holds for $(X_t)_{t\in\mathbb{Z}}$ if $\|\xi_0\|_\infty < \infty$, $\alpha > 1$ with

$$a > \frac{1+\alpha}{\alpha-1}\left(4 + \frac{2}{m-2}\right) + 1.$$

Notice that a very restrictive new assumption is that inputs need to be uniformly bounded in this non-causal case. This result is new, a conjecture is that (3.1) does not hold.

Non-causal, non-linear inputs

The weak dependence properties of non-causal and non-linear inputs Y_t are recalled, see [41] for more details. Let $H : (\mathbb{R}^d)^{\mathbb{Z}} \to \mathbb{R}$ be a measurable function. If the sequence $(\xi_n)_{n\in\mathbb{Z}}$ is independent and identically distributed on $\mathbb{R}^d$, the Bernoulli shift with input process $(\xi_n)_{n\in\mathbb{Z}}$ is defined as

$$Y_n = H\left((\xi_{n-i})_{i\in\mathbb{Z}}\right), \qquad n \in \mathbb{Z}.$$

Such Bernoulli's shifts are η-weakly dependent (see [41]) with $\eta(r) \leqslant 2\delta_{[r/2]}$ if

$$\mathbb{E}\left|H\left(\xi_j, j\in\mathbb{Z}\right) - H\left(\xi_j \mathbb{1}_{|j|\leqslant r}, j\in\mathbb{Z}\right)\right| \leqslant \delta_r. \tag{3.14}$$

Then condition (3.9) leads to the invariance principle for $(X_t)_{t\in\mathbb{Z}}$ if $\mathbb{E}|Y_0|^m < \infty$ for $m > 2$, $\alpha > 1$ and $\delta_r \leqslant Kr^{-\delta}$ for

$$\delta > \frac{1+\alpha}{\alpha-1}\left(4 + \frac{2}{m-2}\right).$$

Conditions (3.1) of [69] do not give clear conditions on coefficients for these models. We do not know other weak invariance principle in that general context.

Associated inputs

Recall that a process is associated if $\mathrm{Cov}\left(f(Y^{(n)}), g(Y^{(n)})\right) \geqslant 0$ for any coordinatewise non-decreasing function $f, g : \mathbb{R}^n \to \mathbb{R}$ such that the previous covariance makes sense with $Y^{(n)} = (Y_1, \ldots, Y_n)$. The κ-weak dependence condition is known to hold for associated or Gaussian sequences. In both cases this condition will hold with

$$\kappa(r) = \sup_{j \geqslant r} |\mathrm{Cov}(Y_0, Y_j)|$$

Notice the absolute values are needed only in the second case since for associated processes these covariances are nonnegative. Independent sequences as well are associated and Pitt proves in [95] that a Gaussian process with nonnegative covariances is also associated. Finally, we recall that non-decreasing functions of associated sequences remain associated. Associated models are classically built this way from independent and identically distributed sequences, see [85].

Suppose that the inputs $(Y_t)_{t \in \mathbb{Z}}$ are such that $\kappa(r) \leqslant Cr^{-a}$ (for some $a, C > 0$). For the associated cases and model (3.10), the invariance principle of [90] follows from remark of [86] as soon as $\mathbb{E}Y^2 < +\infty$, $a > 1$ and $\alpha > 1$. These conditions are optimal, they correspond to $\sum_j \mathrm{Cov}(X_0, X_j) < \infty$. Such strong conditions are due to the fact that zero correlation implies independence for associated processes. Our conditions for invariance principle are much stronger : $\mathbb{E}|Y|^m < +\infty$ with $m > 2$, $\alpha > 1$ and

$$a > \frac{1+\alpha}{\alpha - 1}\left(4 + \frac{2}{m-2}\right).$$

For non-linear Lipschitz cases as in eqn. (3.11) the result seems to be new. In the special case of κ-weak dependent inputs that are not associated, the optimal weak invariance principle of [90] does not apply, see [41] for examples.

3.3.2 The Bernoulli shifts with dependent inputs

Let $H : \mathbb{R}^{(\mathbb{Z})} \to \mathbb{R}$ be a measurable (non necessarily Lipschitz) function and $X_n = H(Y_{n-i}, i \in \mathbb{Z})$. Such models are proved to exhibit either λ or η-weak dependence properties. Because the Bernoulli shifts of κ-weak dependent inputs are neither κ nor η-weakly dependent, the κ case is here included in the λ one.

Consider the non-Lipschitz function H defined by

$$H(x) = \sum_{k=0}^{K} \sum_{j_1, \ldots, j_k} a^{(k)}_{j_1, \ldots, j_k} x_{j_1} \cdots x_{j_k}.$$

In this case, Lemma 3.1 does not apply. To derive weak dependence properties of such processes, we assume that H satisfies the condition (3.8) with $\ell \neq 0$, which remains a stronger assumption than for the case of independent inputs, see eqn. (3.14). Relaxing Lipschitz assumption on H is possible if we assume the existence of higher moments for the inputs. The following lemma gives both the existence and the weak dependence properties of such models

Lemma 3.2 *Let $(Y_i)_{i\in\mathbb{Z}}$ be a stationary process and $H : \mathbb{R}^{(\mathbb{Z})} \to \mathbb{R}$ satisfies condition (3.8) for some $\ell > 0$ and some sequence $b_j \geqslant 0$ such that $\sum_j |j|b_j < \infty$. Let us assume that there exist a pair of real numbers (m, m') with $\mathbb{E}|Y_0|^{m'} < \infty$ such that $m \geqslant 1$, $m' \geqslant (\ell+1)m$ and $m' > \ell + 1$ if $m = 1$. Then,*

- *the process $X_n = H(Y_{n-i}, i \in \mathbb{Z})$ is well defined in $\mathbb{L}^m$: it is a strictly stationary process ;*
- *if the input process $(Y_i)_{i\in\mathbb{Z}}$ is λ-weakly dependent (the weak dependence coefficients are denoted by $\lambda_Y(r)$), then X_n is λ-weakly dependent and there exists a constant $c > 0$ such that*

$$\lambda(k) \leqslant c \inf_{r\leqslant[k/2]} \left[\sum_{|j|\geqslant r} |j|b_j + (2r+1)^2 \lambda_Y(k-2r)^{\frac{m'-1-\ell}{m'-1+\ell}} \right];$$

- *if the input process $(Y_i)_{i\in\mathbb{Z}}$ is η-weakly dependent (the weak dependence coefficients are denoted by $\eta_Y(r)$) then X_n is η-weakly dependent and there exists a constant $c > 0$ such that*

$$\eta(k) \leqslant c \inf_{r\leqslant[k/2]} \left[\sum_{|j|\geqslant r} |j|b_j + (2r+1)^{1+\frac{\ell}{m'-1}} \eta_Y(k-2r)^{\frac{m'-1-\ell}{m'-1}} \right].$$

Such models where already mentioned in the mixing case by [12] and [16]. The proofs are deferred to Section 3.4.6.

Volterra models with dependent inputs

Consider the function H defined by

$$H(x) = \sum_{k=0}^{K} \sum_{j_1,\dots,j_k} a^{(k)}_{j_1,\dots,j_k} x_{j_1} \cdots x_{j_k},$$

then if x, y are as in eqn. (3.8)

$$H(x) - H(y) = \sum_{k=1}^{K} \sum_{u=1}^{k} \sum_{\substack{j_1,\dots,j_{u-1}\\ j_{u+1},\dots,j_k}} a^{(k)}_{j_1,\dots,j_{u-1},s,j_{u+1},\dots,j_k} \times x_{j_1} \cdots x_{j_{u-1}} (x_s - y_s) x_{j_{u+1}} \cdots x_{j_k}.$$

From the triangular inequality we thus derive that the previous lemma 3.2 may be written with $\ell = K - 1$,

$$b_s = \sum_{k=1}^{K} \sum\nolimits^{(k,s)} |a^{(k)}_{j_1,\dots,j_k}|$$

where $\sum^{(k,s)}$ stands for the sums over all indices in $\mathbb{Z}^k$ where one of the indices $j_1, \dots, j_k$ takes on the value s and

$$L \equiv \sum_{k=0}^{K} \sum_{j_1,\dots,j_k} \left| a^{(k)}_{j_1,\dots,j_k} \right| .$$

For example, $|a^{(k)}_{j_1,\dots,j_k}| \leqslant C\,(j_1 \vee \cdots \vee j_k)^{-\alpha}$ or $\leqslant C \exp\left(-\alpha(j_1 \vee \cdots \vee j_k)\right)$ respectively yield $b_s \leqslant C' s^{d-1-\alpha}$ or $b_s \leqslant C' e^{-\alpha s}$ for some constant $C' > 0$.

Markov stationary inputs

Markov stationary sequences satisfy a recurrence equation

$$Z_n = F(Z_{n-1}, \dots, Z_{n-d}, \xi_n)$$

where (ξ_n) is a sequence of independent and identically distributed random variables. In this case $Y_n = (Z_n, \dots, Z_{n-d+1})$ is a Markov chain such that $Y_n = M(Y_{n-1}, \xi_n)$ with

$$M(x_1, \dots, x_d, \xi) = (F(x_1, \dots, x_d, \xi), x_1, \dots, x_{d-1}). \tag{3.15}$$

Theorem 1.IV.24 of [50] proves that eqn. (3.15) has a stationary solution $(Z_n)_{n\in\mathbb{Z}}$ in $\mathbb{L}^m$ for $m \geqslant 1$ as soon as $\|F(0,\xi)\|_m < \infty$ and there exist a norm $\|\cdot\|$ on $\mathbb{R}^d$ and a real number $a \in [0, 1[$ such that $\|F(x,\xi) - F(y,\xi)\|_m \leqslant a\|x - y\|$. In this setting θ-dependence holds with $\theta_Z(r) = \mathcal{O}(a^{r/d})$ (as $r \uparrow \infty$). We shall not give more details about the significative examples provided in Doukhan2005. Indeed, we already mentioned that our results are sub-optimal in such causal cases ; such dependent sequences may however also be used as inputs for the Bernoulli shifts.

Explicit dependence rates

We now specify the decay rates from Lemma 3.2. For standard decays of the previous sequences, it is easy to get the following explicit bounds. Here $b, c, C, D, \lambda, \eta > 0$ are constants which may differ from one case to the other.

- If $b_j \leqslant C(|j|+1)^{-b}$ and $\lambda_Y(j) \leqslant Dj^{-\lambda}$, resp. $\eta_Y(j) \leqslant Dj^{-\eta}$ then from a simple calculation, we optimize both terms in order to prove that $\lambda(k) \leqslant ck^{-\lambda\left(1-\frac{2}{b}\right)\frac{m'-1-\ell}{m'-1+\ell}}$, resp. $\eta(k) \leqslant ck^{-\eta\frac{(b-2)(m'-2)}{(b-1)(m'-1)-\ell}}$. Note that in the case where $m' = \infty$ this exponent may be arbitrarily close to λ for large values of $b > 0$. This exponent may thus take all possible values between 0 and λ.
- If $b_j \leqslant Ce^{-|j|b}$ and $\lambda_Y(j) \leqslant De^{-j\lambda}$, and, respectively, if $\eta_Y(j) \leqslant De^{-j\eta}$, we obtain the bounds $\lambda(k) \leqslant ck^2 e^{-\lambda k\frac{b(m'-1-\ell)}{b(m'-1+\ell)+2\eta(m'-1-\ell)}}$, resp. $\eta(k) \leqslant ck^{\frac{m'-1-\ell}{m'-1}} e^{-\eta k\frac{b(m'-2)}{b(m'-1)+2\eta(m'-2)}}$.
  The geometric decay of both $(b_j)_{j\in\mathbb{Z}}$ and the weak dependence coefficients of the inputs ensure the geometric decay of the weak dependence coefficients of the Bernoulli shift.

– If we assume that the coefficients $(b_j)_{j\in\mathbb{Z}}$ associated with the Bernoulli shift have a geometric decay, say $b_j \leqslant Ce^{-|j|b}$ and that $\lambda_Y(j) \leqslant Dj^{-\lambda}$ (resp. $\eta_Y(j) \leqslant Dj^{-\eta}$) we find $\lambda(k) = ck^{-\lambda\frac{m'-1-l}{m'-1+\ell}}\log^2 k$, resp. $\eta(k) = ck^{-\eta\frac{m'-2}{m'-1}}\log^{1+\frac{\ell}{m'-1}} k$.
  If $m' = \infty$ tightness is reduced by a factor $\log^2 k$ with respect to the dependence coefficients of the input dependent series $(Y_t)_{t\in\mathbb{Z}}$.
– If we assume that the coefficients $(b_j)_{j\in\mathbb{Z}}$ associated with the Bernoulli shift have a Riemannian decay, say $b_j \leqslant C(|j|+1)^{-b}$ and that $\lambda_Y(j) \leqslant De^{-j\lambda}$ (resp. $\eta_Y(j) \leqslant De^{-j\eta}$) we find $\lambda(k) \leqslant ck^{2-b}$, resp. $\eta(k) \leqslant ck^{2-b}$.

All models or functions of models we present here are λ-weakly dependent. We treat some basic examples in detail when a discussion with other results is possible. We believe that for some models, λ-weak invariance properties follow from easy computations, and then, statistical results like our weak invariance principle.

3.4 Proofs of the main results

Our proof for central limit theorems is based on a truncation method. For a truncation level $T \geqslant 1$ we shall denote $\overline{X}_k = f_T(X_k) - \mathbb{E}f_T(X_k)$ with $f_T(X) = X \vee (-T) \wedge T$. From now on, we shall use the convenient notation $a_n \preceq b_n$ for two real sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ when there exists some constant $C > 0$ such that $|a_n| \leqslant Cb_n$ for each integer n. We also remark that $\overline{X}_k$ has moments of all orders because it is bounded. In the entire sequel, we denote $\mu = \mathbb{E}|X_0|^m$. For any $a \leqslant m$, we control the moment $\mathbb{E}|f_T(X_0) - X_0|^a$ with Markov inequality

$$\mathbb{E}|f_T(X_0) - X_0|^a \leqslant \mathbb{E}|X_0|^a \mathbb{1}_{\{|X_0|\geqslant T\}} \leqslant \mu T^{a-m},$$

thus using Jensen inequality yields

$$\|\overline{X}_0 - X_0\|_a \leqslant 2\mu^{\frac{1}{a}} T^{1-\frac{m}{a}}. \tag{3.16}$$

Starting from this truncation, we are now able to control the limiting variance as well as the higher order moments.

In this section we prove that the central limit theorems corresponding to the convergence $W_n(1) \to W(1)$ in both Theorems 3.1 and 3.2 hold and we shall provide convergence rates corresponding to these central limit theorems. The weak invariance principle is obtained in a standard way from such central limit theorems and tightness, which follows from Lemma 3.2, by using the classical Kolmogorov-Centsov tightness criterion, see [12]. In the last subsection, we prove Lemma 3.4 that states the properties of our (new) Bernoulli's shifts with dependent inputs.

3.4.1 Variances

Lemma 3.3 (Variances) *If one of the following conditions holds*

$$\sum_{k=0}^{\infty}\kappa(k) < \infty \tag{3.17}$$

$$\sum_{k=0}^{\infty}\lambda(k)^{\frac{m-2}{m-1}} < \infty \tag{3.18}$$

then the series σ^2 is convergent.

Proof. Using the fact that $\overline{X}_0 = g_T(X_0)$ is a function of X_0 with $\mathrm{Lip}\, g_T = 1$ snd $\|g_T\|_\infty \leqslant 2T$, we derive

$$|\mathrm{Cov}(\overline{X}_0, \overline{X}_k)| \leqslant \kappa(k) \text{ or } (4T+1)\lambda(k), \text{ respectively.} \tag{3.19}$$

In the κ dependent case, the truncation may thus be omitted and

$$|\mathrm{Cov}(X_0, X_k)| \leqslant \kappa(k). \tag{3.20}$$

In the following, we shall only consider λ dependence. We develop

$$\mathrm{Cov}(X_0, X_k) = \mathrm{Cov}(\overline{X}_0, \overline{X}_k) + \mathrm{Cov}(X_0 - \overline{X}_0, X_k) + \mathrm{Cov}(\overline{X}_0, X_k - \overline{X}_k).$$

We use a truncation T (to be determined) and the two previous bounds eqn. (3.16) and eqn. (3.19) ; then the Hölder inequality with the exponents $1/a + 1/m = 1$ yields

$$\begin{aligned} |\mathrm{Cov}(X_0, X_k)| &\leqslant (4T+1)\lambda(k) + 2\|X_0\|_m \|\overline{X}_0 - X_0\|_a \\ &\leqslant (4T+1)\lambda(k) + 4\mu^{1/a+1/m} T^{1-m/a} \\ &\leqslant (4T+1)\lambda(k) + 4\mu T^{2-m}. \end{aligned}$$

Chosing $T^{m-1} = \mu/\lambda(k)$ we obtain

$$|\mathrm{Cov}(X_0, X_k)| \leqslant 9\mu^{\frac{1}{m-1}}\lambda(k)^{\frac{m-2}{m-1}}. \tag{3.21}$$

3.4.2 A Δ-order moment bound

Lemma 3.4 *Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary and centered process. Let us assume that $\mathbb{E}|X_0|^m < \infty$, and that this process is either κ-weakly dependent with $\kappa(r) = \mathcal{O}\left(r^{-\kappa}\right)$ or λ-weakly dependent with $\lambda(r) = \mathcal{O}\left(r^{-\lambda}\right)$. If $\kappa > 2 + \frac{1}{m-2}$, or $\lambda > 4 + \frac{2}{m-2}$, then for all $\Delta > 2$ small enough there exists a constant $C > 0$ such that*

$$\|S_n\|_\Delta \leqslant C\sqrt{n}.$$

Remarks.

- $\Delta \in]2, 2 + A \wedge B \wedge 1[$ where A and B are constants smaller than $m - 2$ and depend on m and respectively κ or λ. Equations eqn. (3.25) and eqn. (3.26) precise the previous involved constants A and B.
- The constant C satisfies $C > \left(\frac{5}{2^{(\Delta-2)/2} - 1}\right)^{1/\Delta} \sum_{k\in\mathbb{Z}} |\mathrm{Cov}(X_0, X_k)|$. Under the conditions of this lemma, Lemma 3.3 entails
$$c \equiv \sum_{k\in\mathbb{Z}} |\mathrm{Cov}(X_0, X_k)| < \infty.$$
- The result is sketched from [19]. However, their dependence condition is of a causal nature while our is not. It explains a loss with respect to the exponents λ and κ. In their κ'-weak dependence setting the best possible value of the exponent is 1 while it is 2 for our non-causal dependence.

Proof of Lemma 3.4. For convenience, let denote in the sequel $\Delta = 2 + \delta$ and $m = 2 + \zeta$. Like in [71] or [19], we proceed by induction on k for $n \leqslant 2^k$ to prove that

$$\|1 + |S_n|\|_\Delta \leqslant C\sqrt{n}. \tag{3.22}$$

We assume that eqn. (3.22) is satisfied for all $n \leqslant 2^{K-1}$. Setting $N = 2^K$ we have to find a bound for $\|1 + |S_N|\|_\Delta$. We can divide the sum S_N into three blocks : the first two blocks have the same size $n \leqslant 2^{K-1}$ and are denoted by Q and R ; the third block V, located between Q and R, has cardinality $q < n$. We then have $\|1 + |S_N|\|_\Delta \leqslant \|1 + |Q| + |R|\|_\Delta + \|V\|_\Delta$. The term $\|V\|_\Delta$ is directly bounded by $\|1+|V|\|_\Delta \leqslant C\sqrt{q}$ from the recurrence assumption. Writing $q = \xi N^b$ with $0 < \xi, b < 1$, then this term is of order strictly smaller than $\sqrt{\xi N}$. For $\|1 + |Q| + |R|\|_\Delta$, we have

$$\begin{aligned}
\mathbb{E}(1 + |Q| + |R|)^\Delta &\leqslant \mathbb{E}(1 + |Q| + |R|)^2(1 + |Q| + |R|)^\delta,\\
&\leqslant \mathbb{E}(1 + 2|Q| + 2|R| + (|Q| + |R|)^2)(1 + |Q| + |R|)^\delta.
\end{aligned}$$

We expand the right-hand side of this expression ; the following terms appear

- $\mathbb{E}(1 + |Q| + |R|)^\delta \leqslant 1 + |Q|_2^\delta + |R|_2^\delta \leqslant 1 + 2c^\delta(\sqrt{n})^\delta$,
- $\mathbb{E}|Q|(1 + |Q| + |R|)^\delta \leqslant \mathbb{E}|Q|((1 + |R|)^\delta + |Q|^\delta) \leqslant \mathbb{E}|Q|(1 + |R|)^\delta + \mathbb{E}|Q|^{1+\delta}$. The term $\mathbb{E}|Q|^{1+\delta}$ is bounded by $\|Q\|_2^{1+\delta}$ and then by $c^{1+\delta}(\sqrt{n})^{1+\delta}$. The term $\mathbb{E}|Q|(1+|R|)^\delta$ is bounded by $\|Q\|_{1+\delta/2}\|1+|R|\|_\Delta^\delta$ using Hölder inequality. It is at least of order $cC^\delta(\sqrt{n})^{1+\delta}$, analogous to the latter one, where we exchange the roles of Q and R.
- $\mathbb{E}(|Q| + |R|)^2(1 + |Q| + |R|)^\delta$. For this term, we use an inequality from [19]
$$\mathbb{E}(|Q| + |R|)^2(1 + |Q| + |R|)^\delta \leqslant \mathbb{E}|Q|^\Delta + \mathbb{E}|R|^\Delta + 5(\mathbb{E}Q^2(1 + |R|)^\delta + \mathbb{E}R^2(1 + |Q|)^\delta).$$
Now $\mathbb{E}|Q|^\Delta \leqslant C^\Delta(\sqrt{n})^\Delta$ is bounded by using eqn. (3.22). The second term is its analogous with R substituted to Q. The third term has to be handled with a particular care, as follows.

We use the weak dependence notion to control $\mathbb{E}Q^2(1+|R|)^\delta$ and $\mathbb{E}R^2(1+|Q|)^\delta$. Denote by $\overline{X}$ the variable $X \vee T \wedge (-T)$ for a real $T>0$ to be determined later. By extension $\overline{Q}$ and $\overline{R}$ denote the truncated sums of the variables X_i. We have

$$\mathbb{E}|Q|^2(1+|R|)^\delta \leqslant \mathbb{E}Q^2||R|-|\overline{R}||^\delta + \mathbb{E}(Q^2-\overline{Q}^2)(1+|\overline{R}|)^\delta + \mathbb{E}\overline{Q}^2(1+|\overline{R}|)^\delta.$$

We begin with a control of $\mathbb{E}Q^2||R|-|\overline{R}||^\delta$. Using the Hölder inequality with $2/m+1/m'=1$ yields

$$\mathbb{E}Q^2||R|-|\overline{R}||^\delta \leqslant \|Q\|_m^2 \| ||R|-|\overline{R}||^\delta \|_{m'}$$

$\|Q\|_\Delta$ is bounded using eqn. (3.22) and

$$||R|-|\overline{R}||^{\delta m'} \leqslant |R|^{\delta m'} 1\!\!1_{\{|R|>T\}} \leqslant |R|^{\delta m'} 1\!\!1_{|R|>T}.$$

We then bound $1\!\!1_{|R|>T} \leqslant (|R|/T)^\alpha$ with $\alpha = m-\delta m'$, hence

$$\mathbb{E}||R|-|\overline{R}||^{\delta m'} \leqslant E|R|^m T^{\delta m'-m}.$$

By convexity and stationarity, we have $\mathbb{E}|R|^m \leqslant n^m \mathbb{E}|X_0|^m$, so that

$$\mathbb{E}Q^2(|R|-|\overline{R}|)^\delta \preceq n^{2+m/m'} T^{\delta - m/m'}.$$

Finally, remarking that $m/m' = m-2$, we obtain

$$\mathbb{E}Q^2(|R|-|\overline{R}|)^\delta \preceq n^m T^{\Delta-m}.$$

We get the same bound for the second term

$$\mathbb{E}(Q^2-\overline{Q}^2)(1+|\overline{R}|)^\delta \preceq n^m T^{\Delta-m}.$$

For the third one, we introduce a covariance term

$$\mathbb{E}\overline{Q}^2(1+|\overline{R}|)^\delta \leqslant \mathrm{Cov}(\overline{Q}^2,(1+|\overline{R}|)^\delta) + \mathbb{E}\overline{Q}^2\mathbb{E}(1+|\overline{R}|)^\delta.$$

The latter is bounded with $|Q|_2^2|R|_2^\delta \leqslant c^\Delta \sqrt{n}^\Delta$. The covariance is controlled as follows by using weak-dependence

– in the κ-dependent case : $n^2 T \kappa(q)$,

– in the λ-dependent case : $n^3 T^2 \lambda(q)$.

We then choose either the truncation $T^{m-\delta-1} = n^{m-2}/\kappa(q)$ or $T^{m-\delta} = n^{m-3}/\lambda(q)$. At this point, the three terms of the decomposition are of the same order

$$\begin{aligned}
\mathbb{E}|Q|^2(1+|R|)^\delta &\preceq \left(n^{3m-2\Delta}\kappa(q)^{m-\Delta}\right)^{1/(m-\delta-1)}, \text{ under } \kappa\text{-dependence},\\
\mathbb{E}|Q|^2(1+|R|)^\delta &\preceq \left(n^{5m-3\Delta}\lambda(q)^{m-\Delta}\right)^{1/(m-\delta)}, \text{ under } \lambda\text{-dependence}.
\end{aligned}$$

Let $q = N^b$, we note that $n \leqslant N/2$ and this term is of order $N^{\frac{3m-2\Delta+b\kappa(\Delta-m)}{m-\delta-1}}$ under κ-weak dependence and the order $N^{\frac{5m-3\Delta+b\lambda(\Delta-m)}{m-\delta}}$ under λ-weak dependence. Those terms are thus negligible with respect to $N^{\Delta/2}$ if

$$\kappa > \frac{3m - 2\Delta - \Delta/2(m-\delta-1)}{b(m-\Delta)}, \text{ under } \kappa\text{-dependence,} \tag{3.23}$$

$$\lambda > \frac{5m - 3\Delta - \Delta/2(m-\delta)}{b(m-\Delta)}, \text{ under } \lambda\text{-dependence.} \tag{3.24}$$

Finally, using this assumption, $b < 1$ and $n \leqslant N/2$, we derive the bound for some suitable constants $a_1, a_2 > 0$

$$\mathbb{E}(1+|S_N|)^\Delta \leqslant \left((2^{-\delta/2} + \xi^\Delta)C^\Delta + 5\cdot 2^{-\delta/2}c^\Delta + a_1 N^{-a_2}\right)\left(\sqrt{N}\right)^\Delta.$$

Using the relation between C and c, we conclude that eqn. (3.22) is also true at the step N if the constant C satisfies $(2^{-\delta/2} + \xi^\Delta)C^\Delta + 5\cdot 2^{-\delta/2}c^\Delta + a_1 N^{-a_2} \leqslant C^\Delta$. Choose $C > \left(\frac{5c^\Delta + a_1 2^{\delta/2}}{2^{\delta/2}-1}\right)^{1/\Delta}$ with $c = \sum_{k\in\mathbb{Z}} |\mathrm{Cov}(X_0, X_k)|$, then the previous relation holds for some $0 < \xi < 1$. Finally, we use eqn. (3.23) and eqn. (3.24) to find a condition on δ.
In the case of κ-weak dependence, we rewrite inequality eqn. (3.23) as

$$0 > \delta^2 + \delta(2\kappa - 3 - \zeta) - \kappa\zeta + 2\zeta + 1.$$

It leads to the following condition on δ

$$\delta < \frac{\sqrt{(2\kappa-3-\zeta)^2 + 4(\kappa\zeta - 2\zeta - 1)} + \zeta + 3 - 2\kappa}{2} \wedge 1 = A. \tag{3.25}$$

We do the same in the case of the λ-weak dependence

$$\delta < \frac{\sqrt{(2\lambda-6-\zeta)^2 + 4(\lambda\zeta - 4\zeta - 2)} + \zeta + 6 - 2\lambda}{2} \wedge 1 = B. \quad \square \tag{3.26}$$

Remark. The bounds A and B are always smaller than ζ.

3.4.3 Proofs of Theorems 3.1 and 3.2

Let $S = \frac{1}{\sqrt{n}} S_n$ and consider $p = p(n)$ and $q = q(n)$ in such a way that

$$\lim_{n\to\infty} \frac{1}{q(n)} = \lim_{n\to\infty} \frac{q(n)}{p(n)} = \lim_{n\to\infty} \frac{p(n)}{n} = 0$$

and $k = k(n) = n/[p(n) + q(n)]$

$$Z = \frac{1}{\sqrt{n}}\left(U_1 + \cdots + U_k\right), \qquad \text{with } U_j = \sum_{i\in B_j} X_i$$

where $B_j =](p+q)(j-1),\ (p+q)(j-1)+p] \cap \mathbb{N}$ is a subset of p successive integers from $\{1,\dots,n\}$ such that, for $j \neq j'$, B_j and $B_{j'}$ are at least distant of $q = q(n)$ from each other. We denote by B'_j the block between B_j and B_{j+1} and $V_j = \sum_{i \in B'_j} X_i$. V_k is the last block of X_i between the end of B_k and n. Furthermore, let $\sigma_p^2 = \mathrm{Var}\,(U_1)/p = \sum_{|i|<p}(1-|i|/p)\mathbb{E}X_0X_i$, and let

$$Y = \frac{U'_1 + \cdots + U'_k}{\sqrt{n}}, \qquad U'_j \sim \mathcal{N}(0, p\,\sigma_p^2)$$

where the Gaussian variables V_j are mutually independent and also independent of the sequence $(X_n)_{n\in\mathbb{Z}}$. We also consider a sequence $U_1^*, \dots, U_k^*$ of mutually independent random variables with the same distribution as U_1 and we let $Z^* = (U_1^* + \cdots + U_k^*)/\sqrt{n}$. In the entire section, we fix $t \in \mathbb{R}$ and we define $f : \mathbb{R} \to \mathbb{C}$ by $f(x) = \exp\{itx\}$. Then $\mathbb{E}f(S) - f(\sigma N) = \mathbb{E}f(S) - f(Z) + \mathbb{E}f(Z) - f(Z^*) + \mathbb{E}f(Z^*) - f(Y) + \mathbb{E}f(Y) - f(\sigma N)$. Lindeberg method is devoted to prove that this expression converges to 0 as $n \to \infty$. The first and the last terms in this inequality are referred to as the auxiliary terms in this Bernstein-Lindeberg method. They come from the replacement of the individual initial – non-Gaussian and Gaussian respectively – random variables by their block counterparts. The second term is analogous to that obtained with decoupling and turns the proof of the central limit theorem to the independent case. The third term is referred to as the main term and following the proof under independence it will be bounded above by using a Taylor expansion. Because of the dependence structure, in the corresponding bounds, some additional covariance terms will appear.
The following subsections are organized as follows : we first consider the auxiliary terms and the main terms are then decomposed by the usual Lindeberg method and the corresponding terms coming from the dependence or the usual remainder terms (standard for the independent case) are considered in separate subsections. In the last one, we collect these calculations to obtain the central limit theorem.

Auxiliary terms

Using Taylor expansions up to the second order, we obtain

$$|\mathbb{E}f(S) - f(Z)| \leqslant \|f'\|_\infty \mathbb{E}|S - Z|$$
$$\text{and} \quad |\mathbb{E}f(Y) - f(\sigma N)| \leqslant \frac{\|f''\|_\infty^2}{2}|\mathbb{E}Y^2 - \sigma^2|.$$

We note that $Z - S = (V_1 + \cdots + V_k)/\sqrt{n}$ is a sum of X_i's for which the number of terms is $\leqslant (k+1)q + p$. Then eqn. (3.21) and eqn. (3.20), under conditions (3.18) or (3.17) respectively, entail :

$$(\mathbb{E}|Z - S|)^2 \leqslant \mathbb{E}|Z - S|^2 \preceq \big((k+1)q + p\big)/n.$$

Now $Y \sim \sqrt{\dfrac{kp}{n}}\sigma_p N$, thus

$$|\mathbb{E}Y^2 - \sigma^2| \leqslant \left|\frac{kp}{n} - 1\right| \sigma_p^2 + |\sigma_p^2 - \sigma^2|.$$

Remarking that $|kp/n-1|^2 \preceq \big((k+1)q+p\big)/n$, it remains to bound the quantity

$$|\sigma_p^2-\sigma^2| \;\leqslant\; \sum_{|i|<p}\frac{|i|}{p}|\mathbb{E}X_0X_i|+\sum_{|i|>p}|\mathbb{E}X_0X_i|.$$

Let $a_i=|\mathbb{E}X_0X_i|$, under conditions (3.18) or (3.17) (respectively), the series $\sum_{i=0}^{\infty}a_i$ converge thus $s_j=\sum_{i=j}^{\infty}a_i \underset{j\to\infty}{\to} 0$ and

$$|\sigma_p^2-\sigma^2|\leqslant 2\sum_{i=0}^{p-1}\frac{i}{p}\cdot a_i+2s_p\leqslant \frac{2}{p}\sum_{i=0}^{p-1}s_i+2s_p.$$

Cesaro lemma entails that term $|\sigma_p^2-\sigma^2|$ converges to 0.
Hence $|\mathbb{E}f(S)-f(Z)|+|\mathbb{E}f(Y)-f(\sigma N)|$ tends to 0 as $n\uparrow\infty$.
To determine the convergence rate, we assume that $a_i=\mathcal{O}(i^{-\alpha})$ for some $\alpha>1$; then

$$|\sigma_p^2-\sigma^2|\preceq p^{(1-\alpha)\vee -1}.$$

Remarking that $a_i=\mathbb{E}X_0X_i=\mathrm{Cov}(X_0,X_i)$, we then use equations (3.20) and (3.21) and we find $\alpha=\kappa$ or $\alpha=\lambda(m-2)/(m-1)$ depending of the weak-dependence setting. With $p=n^a$, $q=n^b$ for 2 constants a and b and from the relation $\|f^{(j)}\|_\infty\leqslant|t|^j$, those bounds become, up to a constant

$|t|\left(n^{(b-a)/2}+n^{(a-1)/2}\right)+t^2\left(n^{b-a}+n^{a\{(1-\kappa)\vee-1\}}\right)$, in the κ-weak dependence setting,

$|t|\left(n^{(b-a)/2}+n^{(a-1)/2}\right)+t^2\left(n^{b-a}+n^{a\{(1-\lambda(m-2)/(m-1))\vee-1\}}\right)$, under λ-weak dependence.

Main terms

It remains to control the second and the third terms of the sum. They are bounded as usual by

$$|\mathbb{E}f(Z)-f(Z^*)|\leqslant\sum_{j=1}^{k}|\mathbb{E}\Delta_j|,\quad |\mathbb{E}f(Z^*)-f(Y)|\leqslant\sum_{j=1}^{k}|\mathbb{E}\Delta'_j|,$$

where $\Delta_j=f(W_j+x_j)-f(W_j+x_j^*)$, for $j=1,\ldots,k$ with $x_j=\frac{1}{\sqrt{n}}U_j$, $x_j^*=\frac{1}{\sqrt{n}}U_j^*$, $W_j=w_j+\sum_{i>j}x_i^*$, $w_j=\sum_{i<j}x_i$ and $\Delta'_j=f(W'_j+x_j^*)-f(W'_j+x'_j)$, for $j=1,\ldots,k$ with $x'_j=\frac{1}{\sqrt{n}}U'_j$, $W'_j=\sum_{i<j}x_i^*+\sum_{i>j}x'_i$.
Exploiting the special form of f and the independence properties of the variables U_i^* and U'_i, we can write

$$\begin{aligned}\mathbb{E}\Delta_j &= \big(\mathbb{E}f(w_j)f(x_j)-\mathbb{E}f(w_j)\mathbb{E}f(x_j^*)\big)\,\mathbb{E}f\left(\sum_{i>j}x_i^*\right),\\ \mathbb{E}\Delta'_j &= \big(\mathbb{E}f(x_j^*)-\mathbb{E}f(x'_j)\big)\,\mathbb{E}f\left(W'_j\right).\end{aligned}$$

We then control the two terms $\mathbb{E}f\left(\sum_{i>j} x_i^*\right)$ and $\mathbb{E}f\left(W_j'\right)$ by the fact that $\|f\|_\infty \leqslant 1$ and we use the coupling to introduce a covariance term

$$\begin{aligned} |\mathbb{E}\Delta_j| &\leqslant \left|\mathrm{Cov}\left(f\left(\sum_{i<j} x_i\right), f(x_j)\right)\right|, \\ |\mathbb{E}\Delta_j'| &\leqslant \left|\mathbb{E}f(x_j^*) - \mathbb{E}f(x_j')\right|. \end{aligned}$$

– For Δ_j, we use weak dependence.
To do so, write $|\mathbb{E}\Delta_j| = |\mathrm{Cov}[F(X_m, m \in B_i, i < j), G(X_m, m \in B_j)]|$, with $F(z_1, \ldots, z_{kp}) = f\Big(\sum_{i<j} u_i/\sqrt{n}\Big)$ where $u_i = \sum_{\ell \in B_i} z_\ell$. We verify that $\|F\|_\infty \leqslant 1$ and we control $\mathrm{Lip}\, F$:

$$\begin{aligned} &\left| f\left(\frac{1}{\sqrt{n}} \sum_{i<j} \sum_{\ell \in B_i} z_\ell\right) - f\left(\frac{1}{\sqrt{n}} \sum_{i<j} \sum_{\ell \in B_i} z_\ell'\right)\right| \\ &\qquad\qquad \leqslant \left|1 - \exp it\left(\frac{1}{\sqrt{n}} \sum_{i<j} \sum_{\ell \in B_i} (z_\ell - z_\ell')\right)\right| \leqslant \frac{|t|}{\sqrt{n}} \sum_{\ell=1}^{kp} |z_\ell - z_\ell'|. \end{aligned}$$

Similarly, for $G(z_1, \ldots, z_p) = f\left(\sum_{i=1}^{p} z_i/\sqrt{n}\right)$, we have $\|G\|_\infty = 1$ and $\mathrm{Lip}\, G \preceq |t|/\sqrt{n}$. We then distinguish the two cases of weak dependence, remarking the gap between the left and the right terms in the covariance is at least q.
– In the κ-weak dependent setting : $|\mathbb{E}\Delta_j| \preceq kp \cdot p \cdot \frac{|t|}{\sqrt{n}} \cdot \frac{|t|}{\sqrt{n}} \cdot \kappa(q)$.
– Under the λ dependence condition :
$|\mathbb{E}\Delta_j| \preceq \left(kp \cdot p \cdot \frac{|t|}{\sqrt{n}} \cdot \frac{|t|}{\sqrt{n}} + kp \cdot \frac{|t|}{\sqrt{n}} + p \cdot \frac{|t|}{\sqrt{n}}\right) \cdot \lambda(q)$.
Note that these bounds do not depend on j :

$$\begin{aligned} |\mathbb{E}f(Z) - f(Z^*)| &\preceq kp \cdot t^2 \cdot \kappa(q), && \text{under } \kappa, \\ &\preceq kp \cdot (t^2 + |t|\sqrt{k/p}) \cdot \lambda(q), && \text{under } \lambda. \end{aligned}$$

Knowing that $p = n^a$, $q = n^b$, $\kappa(r) = \mathcal{O}\left(r^{-\kappa}\right)$ or $\lambda(r) = \mathcal{O}\left(r^{-\lambda}\right)$, these convergence rates respectively become $n^{1-\kappa b}$ or $n^{1+(1/2-a)_+ - \lambda b}$ in the κ or the λ dependence context.
– For Δ_j', Taylor expansions up to order 2 or 3 respectively give :

$$\begin{aligned} |f(x_j^*) - f(x_j')| &\leqslant |x_j^* - x_j'|\|f'\|_\infty + \frac{1}{2}(x_j^* - x_j')^2\|f''\|_\infty + r_j \\ r_j &\leqslant \frac{1}{2}\|f''\|_\infty (x_j^* - x_j')^2, \text{ or} \\ &\leqslant \frac{1}{6}\|f'''\|_\infty |x_j^* - x_j'|^3, \end{aligned}$$

For an arbitrary $\delta \in [0,1]$, we have :

$$
\begin{aligned}
\mathbb{E}r_j &\preceq \mathbb{E}(t^2(|x_j^*|^2 + |x_j'|^2) \wedge |t|^3(|x_j^*|^3 + |x_j'|^3)) \\
&\preceq \mathbb{E}(t^2|x_j^*|^2 \wedge |t|^3|x_j^*|^3) + \mathbb{E}(t^2|x_j'|^2 \wedge |t|^3|x_j'|^3) \\
&\preceq |t|^{2+\delta}\left(\mathbb{E}|x_j^*|^{2+\delta} + \mathbb{E}|x_j'|^{2+\delta}\right).
\end{aligned}
$$

By the stationarity of the sequence $(X_i)_{i\in\mathbb{Z}}$, we obtain

$$
|\mathbb{E}\Delta_j'| \preceq |t|^{2+\delta} n^{-1-\frac{\delta}{2}}\left(\mathbb{E}|S_p|^{2+\delta} \vee p^{1+\frac{\delta}{2}}\right).
$$

Lemma 3.4 allows us to fin a bound for $\mathbb{E}|S_p|^{2+\delta}$. If $\kappa > 2 + \frac{1}{\zeta}$, or $\lambda > 4 + \frac{2}{\zeta}$, where $\kappa(r) = \mathcal{O}\left(r^{-\kappa}\right)$ or $\lambda(r) = \mathcal{O}\left(r^{-\lambda}\right)$ then there exist $\delta \in]0, \zeta \wedge 1[$ and $C > 0$ such that

$$
\mathbb{E}|S_p|^{2+\delta} \leqslant C p^{1+\delta/2}.
$$

We then obtain

$$
|\mathbb{E}f(Z^*) - f(Y)| \preceq |t|^{2+\delta} k(p/n)^{1+\delta/2}.
$$

Because $p = n^a$, this bound is of order $n^{(a-1)\delta/2}$ in both κ and λ-weak dependence settings. We now collect the previous bounds to conclude that a multidimensional CLT holds under assumptions of both Theorems 3.1 and 3.2. Tightness follows from the Kolmogorov-Chentsov criterion (see [12]) and Lemma 3.4 ; thus both Theorems 3.1 and 3.2 follow from repeated application of the previous CLT. □

3.4.4 Rates of convergence

Rates of convergence are now presented in two propositions of independent interest. We compute explicit bounds for both the difference of characteristic functions and the Berry-Esséen inequalities.

Proposition 3.1 *Let $(X_t)_{t\in\mathbb{Z}}$ be a weakly dependent stationary process satisfying eqn. (3.5) with $m = 2 + \zeta$ then the difference between the characteristic functions is bounded by*

$$
\left|\mathbb{E}\left(e^{itS_n/\sqrt{n}} - e^{it\sigma N}\right)\right| = o(n^{-c}),
$$

for some $c < c^*$ and all $t \in \mathbb{R}$ where c^* depends of the weak dependent coefficients

– *under κ-weak dependence, if $\kappa(r) = \mathcal{O}(r^{-\kappa})$ for $\kappa > 2 + \frac{1}{\zeta}$, then $c^* = \dfrac{(\kappa-1)A}{A + 2\kappa(1+A)}$ where*

$$
A = \frac{\sqrt{(2\kappa - 3 - \zeta)^2 + 4(\kappa\zeta - 2\zeta - 1)} + \zeta + 3 - 2\kappa}{2} \wedge 1.
$$

– *under λ-weak dependence, if $\lambda(r)=\mathcal{O}(r^{-\lambda})$ for $\lambda>4+\frac{2}{\zeta}$, then* $c^*=\dfrac{(\lambda+1)B}{2+B+2\lambda(1+B)}$ *where*

$$B=\frac{\sqrt{(2\lambda-6-\zeta)^2+4(\lambda\zeta-4\zeta-2)}+\zeta+6-2\lambda}{2}\wedge 1,$$

We use the following Esséen inequality in Proposition 3.2

Theorem 3.3 (Theorem 5.1 p.142 of [93]) *Let X and Y be 2 random variables and assume that Y is Gaussian. Let F and G be their distribution functions with corresponding characteristic functions f and g. Then, for every $T>0$, we have for suitable constants b and c*

$$\sup_{x\in\mathbb{R}}|F(x)-G(x)|\leqslant b\int_{-T}^{T}\left|\frac{f(t)-g(t)}{t}\right|dt+\frac{c}{T}. \tag{3.27}$$

Proposition 3.2 (A rate in the Berry Essen bounds) *Let $(X_t)_{t\in\mathbb{Z}}$ be a real stationary process satisfying Proposition 3.1 assumptions. We obtain*

$$\sup_x|F_n(x)-\Phi(x)|=o\left(n^{-c}\right)$$

with $c<c'$ where $c'=c^*/(3+A)$ or $c'=c^*/(3+B)$, respectively, in κ or λ-weak dependence contexts (A, B and c^* are defined in Proposition 3.1).

Proof of proposition 3.1. In the previous section, the different terms have already been bounded as follows :

– In the κ-weak dependence case, the exponents of n in the bounds obtained in section 3.4.3 are
  – for the auxiliary terms : $(b-a)/2$, $(a-1)/2$ and $a(1-\kappa)$,
  – for the main terms : $1-\kappa b$ and $(a-1)\delta/2$.

  Because $\delta<1$ and $b<a<1$, we remark that $(a-1)\delta/2>(a-1)/2$ and $1-\kappa b>a(1-\kappa)$. The only rate of the auxiliary term it remains to consider is $(b-a)/2$ and we obtain

$$a^*=\frac{2+\delta+2\kappa\delta}{\delta+2\kappa(1+\delta)}\in\left]b^*,\frac{\delta}{1+\delta}\right[,\quad b^*=\frac{2+a^*}{1+2\kappa}\in\left]0,a^*\right[.$$

  We conclude with standard calculations and with the help of the inequality $\delta<A$ (see eqn. (3.25)).

– We have the equivalent in the λ-weak dependence case
  – for the auxiliary terms : $(b-a)/2$, $(a-1)/2$ and $a(1-\lambda)$,
  – for the main terms : $1+(1/2-a)_+-\lambda b$ and $(a-1)\delta/2$.

  Only three rates give the asymptotic : $(a-1)\delta/2$, $1+(1/2-a)_+-\lambda b$ and $(b-a)/2$. In the previous case, the optimal choice of a^* was smaller than $1/2$. Then we have to consider here the rate $3/2-a-\lambda b$ and not $1-\lambda b$. Thus

$$a^*=\frac{3+\delta+2\lambda\delta}{2+\delta+2\lambda(1+\delta)}\in\left]b^*,\frac{\delta}{1+\delta}\right[,$$
$$b^*=\frac{3+2\delta}{2+\delta+2\lambda(1+\delta)}\in\left]0,a^*\right[$$

Finally, we obtain a rate of n^{-c^*} using the inequality eqn. (3.26). $\square$

Proof of proposition 3.2. Let choose a^* and b^* as in the proof of proposition 3.1. We now need to make precise the impact of t on the different term of the bound of the $\mathbb{L}^1$ distance between the characteristic functions of S and σN. Up to a constant independent of t, the Kolmogorov distance is bounded by $\left(|t| + t^2 + |t|^{2+C}\right) n^{-c^*}$. Here $C = A$ or B in the two contexts of dependence. Using Theorem 3.3 for a well chosen value of T, we obtain the result of proposition 3.2. $\square$

3.4.5 Proof of Lemma 3.1

The case of Lipschitz functions of dependent inputs is divided in two sections devoted respectively to the definition of such models and to their weak dependence properties.

Existence

Let $Y^{(s)} = (Y_{-i}\mathbb{1}_{|i|<s})_{i\in\mathbb{Z}}$, $Y_+^{(s)} = (Y_{-i}\mathbb{1}_{-s<i\leqslant s})_{i\in\mathbb{Z}}$ for $s \in \mathbb{Z}$ and $H\left(Y^{(\infty)}\right) = \lim_{s\to\infty} H\left(Y^{(s)}\right)$. In order to prove the existence of the Bernoulli shift with dependent inputs, we show that X_0 is the sum of a normally convergent series in $\mathbb{L}^m$; formally

$$X_0 = H\big(Y^{(\infty)}\big) = H(0) + \Big(H\big(Y^{(1)}\big) - H(0)\Big) + \sum_{s=1}^{\infty} H\big(Y^{(s+1)}\big) - H\big(Y_+^{(s)}\big) + \Big(H\big(Y_+^{(s)}\big) - H\big(Y^{(s)}\big)\Big).$$

From eqn. (3.8) we obtain

$$\begin{aligned}
\left\|H\left(Y^{(1)}\right) - H(0)\right\|_m &\leqslant b_0\|Y_0\|_m,\\
\left\|H\left(Y^{(s+1)}\right) - H\left(Y_+^{(s)}\right)\right\|_m &\leqslant b_{-s}\|Y_{-s}\|_m,\\
\left\|H\left(Y_+^{(s)}\right) - H\left(Y^{(s)}\right)\right\|_m &\leqslant b_s\|Y_s\|_m.
\end{aligned}$$

By $(Y_t)_{t\in\mathbb{Z}}$'s stationarity we get

$$\begin{aligned}
\|X_0\|_m \leqslant \left\|H\left(Y^{(1)}\right) - H(0)\right\|_m + \sum_{s=1}^{\infty} \left\|H\left(Y^{(s+1)}\right) - H\left(Y_+^{(s)}\right)\right\|_m &\\
+ \left\|H\left(Y_+^{(s)}\right) - H\left(Y^{(s)}\right)\right\|_m \leqslant \sum_{i\in\mathbb{Z}} b_i\|Y_0\|_m & \quad (3.28)
\end{aligned}$$

Analogously, the process $X_t = H(Y_{t-i}, i \in \mathbb{Z})$ is well defined as the sum of a normally convergent series in $\mathbb{L}^m$. The stationarity of $(X_t)_{t\in\mathbb{Z}}$ holds from that of the input process $(Y_t)_{t\in\mathbb{Z}}$.

Weak dependence properties

Let $X_n^{(r)} = H(Y^{(r)})$ and $X_{\mathbf{s}} = (X_{s_1}, \ldots, X_{s_u})$, $X_{\mathbf{t}} = (X_{t_1}, \ldots, X_{t_v})$ for any $k \geqslant 0$ and any $(u+v)$-tuple such that $s_1 < \cdots < s_u \leqslant s_u + k \leqslant t_1 < \cdots < t_v$. Then we have for all f, g satisfying $\|f\|_\infty, \|g\|_\infty \leqslant 1$

and $\mathrm{Lip}\, f + \mathrm{Lip}\, g < \infty$

$$|\mathrm{Cov}(f(X_{\mathbf{s}}), g(X_{\mathbf{t}}))| \leqslant |\mathrm{Cov}(f(X_{\mathbf{s}}) - f(X_{\mathbf{s}}^{(r)}), g(X_{\mathbf{t}}))| \tag{3.29}$$
$$+ \ |\mathrm{Cov}(f(X_{\mathbf{s}}^{(r)}), g(X_{\mathbf{t}}) - g(X_{\mathbf{t}}^{(r)}))| \tag{3.30}$$
$$+ \ |\mathrm{Cov}(f(X_{\mathbf{s}}^{(r)}), g(X_{\mathbf{t}}^{(r)}))|. \tag{3.31}$$

Using the fact that $\|g\|_\infty \leqslant 1$, we bound the term eqn. (3.29)

$$2\mathrm{Lip}\, f \cdot \mathbb{E}\left|\sum_{i=1}^{u}(X_{s_i} - X_{s_i}^{(r)})\right| \leqslant 2u\mathrm{Lip}\, f \max_{1\leqslant i\leqslant u} \mathbb{E}\left|X_{s_i} - X_{s_i}^{(r)}\right|.$$

Applying inequality (3.28) in the case where $m = 1$, we bound this term by $\mathbb{E}\left|X_{s_i} - X_{s_i}^{(r)}\right| \leqslant \sum_{i\geqslant r} b_i \|Y_0\|_1$. The second term (3.30) is bounded in a similar way.
The last term (3.31) can be written as

$$\left|\mathrm{Cov}(F^{(r)}(Y_{s_i+j}, 1 \leqslant i \leqslant u, |j| \leqslant r), G^{(r)}(Y_{t_i+j}, 1 \leqslant i \leqslant v, |j| \leqslant r)\right|,$$

where $F^{(r)} : \mathbb{R}^{u(2r+1)} \to \mathbb{R}$ and $G^{(r)} : \mathbb{R}^{v(2r+1)} \to \mathbb{R}$. Under the assumption $r \leqslant [k/2]$, we use the $\varepsilon = \eta$ or λ-weak dependence of Y in order to bound this covariance term by $\psi(\mathrm{Lip}\, F^{(r)}, \mathrm{Lip}\, G^{(r)}, u(2r+1), v(2r+1))\varepsilon_{k-2r}$, with respectively $\psi(u, v, a, b) = ua + vb$ or $\psi(u, v, a, b) = uvab + ua + vb$. We compute

$$\mathrm{Lip}\, F^{(r)} = \sup \frac{|f(H(x_{s_i+l}, 1 \leqslant i \leqslant u, |l| \leqslant r) - f(H(y_{s_i+l}, 1 \leqslant i \leqslant u, |l| \leqslant r)|}{\sum_{i=1}^{u} \sum_{-r\leqslant l\leqslant r} |x_{s_i+l} - y_{s_i+l}|},$$

where the sup extends to $x \neq y$ where $x, y \in \mathbb{R}^{u(2r+1)}$. Notice now that if x, y are sequences with $x_i = y_i = 0$ if $|i| \geqslant r$ then repeated applications of the condition (3.8) yields

$$|H(x) - H(y)| \leqslant \sum_{|i|\leqslant r} b_i |x_i - y_i| \leqslant L \sum_{|i|\leqslant r} |x_i - y_i| \tag{3.32}$$

where $L = \sum_{i\in\mathbb{Z}} b_i$. Repeating inequality eqn. (3.32), we obtain

$$|F^{(r)}(x) - F^{(r)}(y)| \leqslant \mathrm{Lip}\, f\, L \sum_{i=1}^{u} \sum_{-r\leqslant l\leqslant r} |x_{s_i+l} - y_{s_i+l}|$$

and we get $\mathrm{Lip}\, F^{(r)} \leqslant \mathrm{Lip}\, f \cdot L$. Similarly $\mathrm{Lip}\, G^{(r)} \leqslant \mathrm{Lip}\, g\, L$.
Under η-weak dependent inputs, we bound the covariance

$$|\mathrm{Cov}(f(X_{\mathbf{s}}), g(X_{\mathbf{t}}))| \leqslant (u\mathrm{Lip}\, f + v\mathrm{Lip}\, g) \times \left[2 \sum_{|i|\geqslant r} b_i \|Y_0\|_1 + (2r+1) L \eta_Y(k - 2r)\right].$$

Under λ-weak dependent inputs

$$|\mathrm{Cov}(f(X_{\mathbf{s}}), g(X_{\mathbf{t}}))| \leqslant (u\mathrm{Lip}\, f + v\mathrm{Lip}\, g + uv\mathrm{Lip}\, f\mathrm{Lip}\, g)\times$$
$$\times\left(\left\{2\sum_{|i|\geqslant r} b_i\|Y_0\|_1 + (2r+1)L\,\lambda_Y(k-2r)\right\} \vee (2r+1)^2L^2\lambda_Y(k-2r)\right). \square$$

3.4.6 Proof of Lemma 3.2

Existence

We decompose X_0 as above in the case $\ell = 0$. Here, we bound each terms by

$$\begin{array}{rcl} |H(Y^{(1)}) - H(0)| & \leqslant & b_0|Y_0| \\ |H(Y^{(s+1)}) - H(Y_+^{(s)})| & \leqslant & b_{-s}(\|Y_+^{(s)}\|_\infty^l \vee 1)|Y_{-s}| \\ |H(Y_+^{(s)}) - H(Y^{(s)}| & \leqslant & b_s(\|Y^{(s)}\|_\infty^l \vee 1)|Y_s| \end{array}$$

Using Hölder inequality yields

$$\mathbb{E}\left|H(Y^{(1)}) - H(0)\right| + \sum_{s=1}^{\infty}\mathbb{E}\left|H(Y^{(s+1)}) - H(Y_+^{(s)})\right|$$
$$+\,\mathbb{E}\left|H(Y_+^{(s)}) - H(Y^{(s)})\right| \leqslant \sum_{i\in\mathbb{Z}} 2|i|b_i(\|Y_0\|_1 + \|Y_0\|_{l+1}^{l+1})$$

Hence assumptions $\ell + 1 \leqslant m'$ and $\sum_{i\in\mathbb{Z}}|i|b_i < \infty$ together imply that the variable $H(Y)$ is well defined in $\mathbb{L}^1$. In the same manner, the process $X_n = H(Y_{n-i}, i \in \mathbb{Z})$ is well defined. The proof extends in $\mathbb{L}^m$ if $m \geqslant 1$ is such that $(\ell+1)m \leqslant m'$.

Weak dependence properties

Here, we exhibit some Lipschitz functions and we then truncate inputs. We write $\overline{Y} = Y \vee (-T) \wedge T$ for a truncation T set below. Denote $X_n^{(r)} = H(Y^{(r)})$ and $\overline{X}_n^{(r)} = H(\overline{Y}^{(r)})$. Furthermore, for any $k \geqslant 0$ and any $(u+v)$-tuple such that $s_1 < \cdots < s_u \leqslant s_u + k \leqslant t_1 < \cdots < t_v$, we set $X_{\mathbf{s}} = (X_{s_1}, \ldots, X_{s_u})$, $X_{\mathbf{t}} = (X_{t_1}, \ldots, X_{t_v})$ and $\overline{X}_{\mathbf{s}}^{(r)} = (\overline{X}_{s_1}^{(r)}, \ldots, \overline{X}_{s_u}^{(r)})$, $\overline{X}_{\mathbf{t}}^{(r)} = (\overline{X}_{t_1}^{(r)}, \ldots, \overline{X}_{t_v}^{(r)})$. Then we have for all f, g satisfying $\|f\|_\infty, \|g\|_\infty \leqslant 1$ and $\mathrm{Lip}\, f + \mathrm{Lip}\, g < \infty$

$$|\mathrm{Cov}(f(X_{\mathbf{s}}), g(X_{\mathbf{t}}))| \leqslant |\mathrm{Cov}(f(X_{\mathbf{s}}) - f(\overline{X}_{\mathbf{s}}^{(r)}), g(X_{\mathbf{t}}))| \tag{3.33}$$
$$+\ |\mathrm{Cov}(f(\overline{X}_{\mathbf{s}}^{(r)}), g(X_{\mathbf{t}}) - g(\overline{X}_{\mathbf{t}}^{(r)}))| \tag{3.34}$$
$$+\ |\mathrm{Cov}(f(\overline{X}_{\mathbf{s}}^{(r)}), g(\overline{X}_{\mathbf{t}}^{(r)}))|. \tag{3.35}$$

Using the fact that $\|g\|_\infty \leqslant 1$, the term (3.33) is bounded by

$$2u\mathrm{Lip}\, f\left(\max_{1\leqslant i\leqslant u}\mathbb{E}\left|X_{s_i}-X_{s_i}^{(r)}\right|+\max_{1\leqslant i\leqslant u}\mathbb{E}\left|X_{s_i}^{(r)}-\overline{X}_{s_i}^{(r)}\right|\right).$$

With the same arguments used in the proof of the existence of $H(Y^{(\infty)})$, the first term in the right-hand side of the inequality is bounded by

$$\sum_{i\geqslant s}2|i|b_i(\|Y_0\|_1+\|Y_0\|_{l+1}^{l+1}).$$

Notice now that if x, y are sequences with $x_i = y_i = 0$ if $|i| \geqslant r$ then an infinitely repeated application of the previous inequality (3.8) yields

$$|H(x)-H(y)|\leqslant L(\|x\|_\infty^l\vee\|y\|_\infty^l\vee 1)\|x-y\| \tag{3.36}$$

where $L=\sum_{i\in\mathbb{Z}}b_i<\infty$ because $\sum_{i\in\mathbb{Z}}|i|b_i<\infty$. The second term is bounded by using eqn. (3.36)

$$\begin{aligned}\mathbb{E}\left|X_{s_i}^{(r)}-\overline{X}_{s_i}^{(r)}\right| &= \mathbb{E}\left|H\left(Y^{(r)}\right)-H\left(\overline{Y}^{(r)}\right)\right|\\ &\leqslant L\mathbb{E}\left(\left(\max_{-r\leqslant i\leqslant r}|Y_i|\right)^l\sum_{-r\leqslant j\leqslant r}|Y_j|1\!\!1_{Y_j\geqslant T}\right)\\ &\leqslant L(2r+1)^2\mathbb{E}\left(\max_{-r\leqslant i,j\leqslant r}|Y_i|^l|Y_j|1\!\!1_{|Y_j|\geqslant T}\right)\\ &\leqslant L(2r+1)^2\|Y_0\|_{m'}^{m'}T^{\ell+1-m'}\end{aligned}$$

The second term (3.34) of the sum is analogously bounded. The last term (3.35) can be written as

$$\left|\mathrm{Cov}\left(\overline{F}^{(r)}(Y_{s_i+j},1\leqslant i\leqslant u,|j|\leqslant r),\overline{G}^{(r)}(Y_{t_i+j},1\leqslant i\leqslant v,|j|\leqslant r)\right)\right|,$$

where $\overline{F}^{(r)}:\mathbb{R}^{u(2r+1)}\to\mathbb{R}$ and $\overline{G}^{(r)}:\mathbb{R}^{u(2r+1)}\to\mathbb{R}$. Under the assumption $r\leqslant[k/2]$, we use the $\varepsilon=\eta$ or $\varepsilon=\lambda$-weak dependence of Y to bound this term by $\psi\left(\mathrm{Lip}\,\overline{F}^{(r)},\mathrm{Lip}\,\overline{G}^{(r)},u(2r+1),v(2r+1)\right)\varepsilon_{k-2r}$, with respectively $\psi(u,v,a,b)=uvab$ or $\psi(u,v,a,b)=uvab+ua+vb$.

$$\mathrm{Lip}\,\overline{F}^{(r)}=\sup\frac{|f(H(\overline{x}_{s_i+l},1\leqslant i\leqslant u,|l|\leqslant r)-f(H(\overline{y}_{s_i+l},1\leqslant i\leqslant u,|l|\leqslant r)|}{\sum_{j=1}^u\|x_j-y_j\|},$$

where the sup extends to $(x_1,\ldots,x_u)\neq(y_1,\ldots,y_u)$ where $x_i,y_i\in\mathbb{R}^{2r+1}$. Using eqn. (3.36)

$$\begin{aligned}|\overline{F}^{(r)}(x)-\overline{F}^{(r)}(y)| &\leqslant \mathrm{Lip}\,fL\sum_{i=1}^u\left(\|\overline{x}_{s_i}\|_\infty\vee\|\overline{y}_{s_i}\|_\infty\vee 1\right)^l\|\overline{x}_{s_i}-\overline{y}_{s_i}\|\\ &\leqslant \mathrm{Lip}\,fLT^l\sum_{i=1}^u\sum_{-r\leqslant l\leqslant r}|x_{s_i+l}-y_{s_i+l}|.\end{aligned}$$

We thus obtain $\operatorname{Lip} F^{(r)} \leqslant \operatorname{Lip} f \cdot L \cdot T^l$. Similarly $\operatorname{Lip} G^{(r)} \leqslant \operatorname{Lip} g \cdot L \cdot T^l$.
Under η-weak dependent inputs, we bound the covariance

$$|\mathrm{Cov}(f(X_{\mathbf{s}}), g(X_{\mathbf{t}}))| \leqslant (u\mathrm{Lip}\, f + v\mathrm{Lip}\, g)\Big\{4 \sum_{|i|\geqslant r} |i| b_i(\|Y_0\|_1 + \|Y_0\|_{l+1}^{l+1}) \\ + (2r+1)L\left((2r+1)2\|Y_0\|_{m'}^{m'} T^{l+1-m'} + T^l \eta_Y(k-2r)\right)\Big\}$$

We then fix the truncation $T^{m'-1} = \dfrac{2(2r+1)\|Y_0\|_{m'}^{m'}}{\eta_Y(k-2r)}$ to obtain the result of Lemma 3.2 in the η-weak dependent case.
Under λ-weak dependent inputs

$$|\mathrm{Cov}(f(X_{\mathbf{s}}), g(X_{\mathbf{t}}))| \leqslant (u\mathrm{Lip}\, f + v\mathrm{Lip}\, g + uv\mathrm{Lip}\, f\mathrm{Lip}\, g) \times \Bigg(\Big\{4 \sum_{|i|\geqslant r} |i| b_i(\|Y_0\|_1 + \|Y_0\|_{l+1}^{l+1}) \\ + (2r+1)L\left(2(2r+1)T^{l+1-m'}\|Y_0\|_{m'}^{m'} + T^l \lambda_Y(k-2r)\right)\Big\} \vee \Big\{(2r+1)^2 L^2 T^{2l} \lambda_Y(k-2r)\Big\}\Bigg)$$

We then set a truncation such that $T^{l+m'-1} = \dfrac{2\|Y_0\|_{m'}^{m'}}{L\lambda_Y(k-2r)}$ to obtain the result of Lemma 3.2 in the η-weak dependent case. $\square$

Deuxième partie

The Density Estimation

Chapitre 4

Convergence rates for density estimators of weakly dependent time series

Abstract

Assuming that $(X_t)_{t\in\mathbb{Z}}$ is a vector valued time series with a common marginal distribution admitting a density f, our aim is to provide a wide range of consistent estimators of f. We consider different methods of estimation of the density as kernel, projection or wavelets ones. Various cases of weakly dependent series are investigated including the Doukhan & Louhichi (1999)'s η-weak dependence condition, and the φ-dependence of Dedecker & Prieur (2005). We thus obtain results for Markov chains, dynamical systems, bilinear models, non causal Moving Average... From a moment inequality of Doukhan & Louhichi (1999), we provide convergence rates of the term of error for the estimation with the $\mathbb{L}^q$ loss or almost surely, uniformly on compact subsets.

Keywords : Weak Dependence, Non-parametric density estimation.

Note

The content of this part is based on a paper, written in collaboration with Nicolas Ragache, published in Dependence in Probability and Statistics. Springer (Ed.) (2006) 349-372.

4.1 Introduction

Assume that $(X_n)_{n\in\mathbb{Z}}$ is a sequence of $\mathbb{R}^d$ valued random variables with common distribution which is absolutely continuous with respect to Lebesgue's measure, with density f. Stationarity is not assumed so that the case of a sampled process $\{X_{i,n} = x_{h_n(i)}\}_{1\leqslant i\leqslant n}$ for any sequence of monotonic functions $(h_n(.))_{n\in\mathbb{Z}}$ and any stationary process $(x_n)_{n\in\mathbb{Z}}$ that admits a marginal density is included. This chapter investigates convergence rates for density estimation in different cases. First, we consider two concepts of weak dependence :

– Non-causal η-dependence introduced in [41] by Doukhan & Louhichi,
– Dedecker & Prieur's φ-dependence (see [29]).

These two notions of dependence cover a large number of examples of time series (see section § 3). Next, following Doukhan (see [37]) we propose a unified study of linear density estimators $\hat{f}_n$ of the form

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{m_n}(x, X_i)\ , \tag{4.1}$$

where $\{K_{m_n}\}$ is a sequence of kernels. Under classical assumptions on $\{K_{m_n}\}$ (see section § 2.2), the results in the case of independent and identically distributed (i.i.d. in short) observations X_i are well known (see for instance [108]). At a fixed point $x \in \mathbb{R}^d$, the sequence m_n can be chosen such that

$$\|\hat{f}_n(x) - f(x)\|_q = O\left(n^{-\rho/(2\rho+d)}\right)\ , \tag{4.2}$$

where $\|X\|_q^q = \mathbb{E}|X|^q$. The coefficient $\rho > 0$ measures the regularity of f (see Section 4.2.2 for the definition of the notion of regularity). The same rate of convergence also holds for the Mean Integrated Square Error (MISE), defined as $\int \|\hat{f}_n(x) - f(x)\|_2^2 p(x)\, dP(x)$ for some nonnegative and integrable function p. The rate of uniform convergence on a compact set incurs a logarithmic loss appears. For all $M > 0$ and for a suitable choice of the sequence m_n,

$$\mathbb{E} \sup_{\|x\|\leqslant M} |\hat{f}_n(x) - f(x)|^q = \mathcal{O}\left(\frac{\log n}{n}\right)^{q\rho/(d+2\rho)}\ , \tag{4.3}$$

and

$$\sup_{\|x\|\leqslant M} |\hat{f}_n(x) - f(x)| =_{a.s.} \mathcal{O}\left(\frac{\log n}{n}\right)^{\rho/(d+2\rho)}\ . \tag{4.4}$$

These rates are optimal in the minimax sense. We thus have no hope to improve on them in the dependent setting. A wide literature deals with density estimation for absolutely regular or β-mixing processes (for a definition of mixing coefficients, see [38]). For instance, under the assumption $\beta_r = o\left(r^{-3-2d/\rho}\right)$, Ango Nze & Doukhan prove in [3] that (4.2), (4.3) and (4.4) still hold. The sharper condition $\sum_r |\beta_r| < \infty$ entails the optimal rate of convergence for the MISE (see [107]). Results for the MISE have been extended to the more general φ- and η-dependence contexts by Dedecker &

Prieur ([29]) and Doukhan & Louhichi in [42]. In this chapter, our aim is to extend the bounds (4.2), (4.3) and (4.4) in the η- and φ-weak dependence contexts.

We use the same method as in [41] based on the following moment inequality for weakly dependent and centered sequences $(Z_n)_{n\in\mathbb{Z}}$. For each even integer q and for each integer $n \geqslant 2$:

$$\left\| \sum_{i=1}^{n} Z_i \right\|_q^q \leqslant \frac{(2q-2)!}{(q-1)!} \left\{ V_{2,n}^{q/2} \vee V_{q,n} \right\} , \tag{4.5}$$

where $\|X\|_q^q = \mathbb{E}|X|^q$ and for $k = 2, \dots, q$,

$$V_{k,n} = n \sum_{r=0}^{n-1} (r+1)^{k-2} C_k(r) ,$$

with

$$C_k(r) := \sup\{|\mathrm{cov}(Z_{t_1} \cdots Z_{t_p}, Z_{t_{p+1}} \cdots Z_{t_k})|\} , \tag{4.6}$$

where the supremum is over all the ordered k-tuples $t_1 \leqslant \cdots \leqslant t_k$ such that $\sup_{1\leqslant i\leqslant k-1} t_{i+1} - t_i = r$. We will apply this bound when the Z_is are defined in such a way that $\sum_{i=1}^{n} Z_i$ is proportional to the fluctuation term $\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)$. The inequality (4.5) gives a bound for this part of the deviation of the estimator which depends on the covariance bounds $C_k(r)$. The other part of the deviation is the bias, which is treated by deterministic methods. In order to obtain suitable controls of the fluctuation term, we need two different type of bounds for $C_k(r)$. Conditions on the decay of the weak dependence coefficients give a first bound. Another type of condition is also required to bound $C_k(r)$ for the smaller values of r ; this is classically achieved with a regularity condition on the joint law of the pairs (X_j, X_k) for all $j \neq k$. In Doukhan & Louhichi (see [42]), rates of convergence are obtained when the coefficient η decays geometrically fast and the joint densities are bounded. We relax these conditions to cover the case when the joint distributions are not absolutely continuous and when the η- and φ-dependence coefficients decrease slowly (sub-geometric and Riemannian decays are considered).

Under our assumptions, we prove that (4.2) still holds (see Theorem 4.1). Unfortunately, additional losses appear for the uniform bounds. When $\eta(r)$ or $\varphi(r) = O(e^{-ar^b})$ with $a > 0$ and $b > 0$, we prove in Theorem 4.2 that (4.3) and (4.4) hold with $\log(n)$ replaced by $\log^{2(b+1)/b}(n)$. If $\eta(r)$ or $\varphi(r) = O(r^{-a})$ with $a > 1$, Theorem 4.3 gives bounds similar to (4.3) and (4.4) with the right hand side replaced by $O(n^{-q\rho/\{d+2\rho+2d/(q_0+d)\}}$ and $O(\{\log^{q_0+d}(n)/n^{q_0-2}\}^{\rho/\{2\rho q_0+d(q_0+2)\}})$, respectively, and with $q_0 = 2\lceil(a-1)/2\rceil$ (by definition $\lceil x\rceil$ is the smallest integer larger than or equal to the real number x). As already noticed in [42], the loss w.r.t the i.i.d. case highly depends on the decay of the dependence coefficients. In the case of geometric decay, the loss is logarithmic while it is polynomial in the case of polynomial decays.

The chapter is organized as follows. In Section 4.2.1, we introduce the notions of η and φ dependence. We give the notation and hypothesis in Section 4.2.2. The main results are presented in Section 4.2.3. We then apply these results to particular cases of weak dependence processes, and we provide examples of kernel K_m in Section 4.3. Section 4.4 contains the proof of the Theorems and three important lemmas.

4.2 Main results

We first describe the notions of dependence considered in this chapter, then we introduce assumptions and formulate the main results of the chapter (convergence rates).

4.2.1 Weak dependence

We consider a sequence $(X_i)_{i\in\mathbb{Z}}$ of $\mathbb{R}^d$ valued random variables, and we fix a norm $\|\cdot\|$ on $\mathbb{R}^d$. Moreover, if $h:\mathbb{R}^{du}\to\mathbb{R}$ for some $u\geqslant 1$, we define

$$\text{Lip }(h) \quad = \sup_{(a_1,\ldots,a_u)\neq(b_1,\ldots,b_u)} \frac{|h(a_1,\ldots,a_u)-h(b_1,\ldots,b_u)|}{\|a_1-b_1\|+\cdots+\|a_u-b_u\|} .$$

Definition 4.1 (η-dependence, Doukhan & Louhichi (1999)) *The process $(X_i)_{i\in\mathbb{Z}}$ is η-weakly dependent if there exists a sequence of non-negative real numbers $(\eta(r))_{r\geqslant 0}$ satisfying $\eta(r)\to 0$ when $r\to\infty$ and*

$$\left|\text{cov}\left(h\left(X_{i_1},\ldots X_{i_u}\right),k\left(X_{i_{u+1}},\ldots,X_{i_{u+v}}\right)\right)\right| \quad \leq \quad (u\text{Lip}(h)+v\text{Lip}(k))\eta(r) ,$$

for all $(u+v)$-tuples, $(i_1,\ldots,i_{u+v})$ with $i_1\leq\cdots\leq i_u\leqslant i_u+r\leq i_{u+1}\leqslant\cdots\leqslant i_{u+v}$, and $h,k\in\Lambda^{(1)}$ where

$$\Lambda^{(1)} \quad = \quad \left\{h:\exists u\geqslant 0, h:\mathbb{R}^{du}\to\ \mathbb{R}, \text{Lip }(h)<\infty, \|h\|_\infty=\sup_{x\in\mathbb{R}^{du}}|h(x)|\leqslant 1\right\} .$$

Remark The η-dependence condition can be applied to non-causal sequences because information "from the future" (i.e. on the right of the covariance) contributes to the dependence coefficient in the same way as information "from the past" (i.e. on the left). It is the non-causal alternative to the θ condition in [26] and [41].

Definition 4.2 (φ-dependence, Dedecker & Prieur (2004)) *Let $(\Omega,\mathcal{A},\mathbb{P})$ be a probability space and $\mathcal{M}$ a σ-algebra of $\mathcal{A}$. For any $l\in\mathbb{N}^*$, any random variable $X\in\mathbb{R}^{dl}$ we define :*

$$\varphi(\mathcal{M},X)=\sup\{\|\mathbb{E}(g(X)|\mathcal{M})-\mathbb{E}(g(X))\|_\infty, g\in\Lambda_{1,l}\} ,$$

where $\Lambda_{1,l} = \{h : \mathbb{R}^{dl} \mapsto \mathbb{R}/\mathrm{Lip}\,(h) < 1\}$. *The sequence of coefficients* $\varphi_k(r)$ *is then defined by*

$$\varphi_k(r) = \max_{l \leqslant k} \frac{1}{l} \sup_{i+r \leqslant j_1 < j_2 < \cdots < j_l} \varphi(\sigma(\{X_j ; j \leqslant i\}), (X_{j_1}, \ldots, X_{j_l})) \,.$$

The process is φ*-dependent if* $\tilde{\phi}(r) = \sup_{k>0} \varphi_k(r)$ *tends to* 0 *with* r.

Remark The φ dependence coefficients provide covariance bounds. For a Lipschitz function k and a bounded function h,

$$\left|\mathrm{cov}\left(h\left(X_{i_1}, \ldots, X_{i_u}\right), k\left(X_{i_{u+1}}, \ldots, X_{i_{u+v}}\right)\right)\right| \leq v\mathbb{E}\left|h\left(X_{i_1}, \ldots, X_{i_u}\right)\right| \mathrm{Lip}\,(k)\varphi(r) \,. \tag{4.7}$$

4.2.2 Notations and definitions

Assume that $(X_n)_{n\in\mathbb{Z}}$ is an η or $\tilde{\phi}$ dependent sequence of $\mathbb{R}^d$ valued random variables. We consider two types of decays for the coefficients. The geometric case is the case when Assumption [H1] or [H1'] holds.

[H1] : $\eta(r) = O\left(e^{-ar^b}\right)$ with $a > 0$ and $b > 0$,

[H1'] : $\varphi(r) = O\left(e^{-ar^b}\right)$ with $a > 0$ and $b > 0$.

The Riemannian case is the case when Assumption [H2] or [H2'] holds.

[H2] : $\eta(r) = O(r^{-a})$ with $a > 1$,

[H2'] : $\varphi(r) = \mathcal{O}(r^{-a})$ with $a > 1$.

As usual in density estimation, we shall assume :

[H3] : The common marginal distribution of the random variables X_n, $n \in \mathbb{Z}$ is absolutely continuous with respect to Lebesgue's measure, with common bounded density f.

The next assumption is on the density with respect to Lebesgue's measure (if it exists) of the joint distribution of the pairs (X_j, X_k), $j \neq k$.

[H4] The density $f_{j,k}$ of the joint distribution of the pair (X_j, X_k) is uniformly bounded with respect to $j \neq k$.

Unfortunately, for some processes, these densities may not even exist. For example, the joint distributions of Markov chains $X_n = G(X_{n-1}, \varepsilon_n)$ may not be absolutely continuous. One of the simplest example is

$$X_k = \frac{1}{2}\left(X_{k-1} + \varepsilon_k\right) \;,$$

where $\{\varepsilon_k\}$ is an i.i.d. sequence of Bernoulli random variables and X_0 is uniformly distributed on $[0, 1]$. The process $\{X_n\}$ is strictly stationary but the joint distributions of the pairs (X_0, X_k) are degenerated for any k. This Markov chain can also be represented (through an inversion of the time) as a dynamical system $(T_{-n}, \ldots, T_{-1}, T_0)$ which has the same law as $(X_0, X_1, \ldots, X_n)$ (T_0 and X_0 are random variables distributed according to the invariant measure, see [6] for more details). Let us recall the definition of a dynamical system.

Definition 4.3 (dynamical system) *A one-dimensional dynamical system is defined by*

$$\forall k \in \mathbb{N}\,,\, T_k := F^k(T_0)\ ,$$

where $F : I \to I$, I is a compact subset of $\mathbb{R}$ and in this context, F^k denotes the k-th iterate of the appplication F : $F^1 = F$, $F^{k+1} = F \circ F^k$, $k \geq 1$. We assume that there exists an invariant probability measure μ_0, i.e. $F(\mu_0) = \mu_0$, absolutely continuous with respect to Lebesgue's measure, and that T_0 is a random variable with distribution μ_0.

We restrict our study to one-dimensional dynamical systems T in the class $\mathcal{F}$ of dynamical systems defined by a transformation F that satisfies the following assumptions (see [98]).
- $\forall k \in \mathbb{N}$, $\forall x \in \mathrm{int}(I)$, $\lim_{t\to 0^+} F^k(x+t) = F^k(x^+)$ and $\lim_{t\to 0^-} F^k(x+t) = F^k(x^-)$ exist ;
- $\forall k \in \mathbb{N}^*$, denoting $D_+^k = \{x \in \mathrm{int}(I), F^k(x^+) = x\}$ and $D_-^k = \{x \in \mathrm{int}(I), F^k(x^-) = x\}$, we assume $\lambda\left(\bigcup_{k\in\mathbb{N}^*}\left(D_+^k \bigcup D_-^k\right)\right) = 0$, where λ is the Lebesgue measure.

When the joint distributions of the pairs (X_j, X_k) are not assumed absolutely continuous (and then [H4] is not satisfied), we shall instead assume :

[H5] The dynamical system $(X_n)_{n\in\mathbb{Z}}$ belongs to $\mathcal{F}$.

We consider in this chapter linear estimators as in (4.1). The sequence of kernels K_m is assumed to satisfy the following assumptions.

(a) The support of K_m is a compact set with diameter $O(1/m^{1/d})$;

(b) The functions $x \mapsto K_m(x,y)$ and $x \mapsto K_m(y,x)$ are Lipschitz functions with Lipschitz constant $O\left(m^{1+1/d}\right)$;

(c) For all x in the support of K_m, $\int K_m(x,y)\,dy = 1$;

(d) The bias of the estimator $\hat{f}_n$ defined in (4.1) is of order $m_n^{-\rho/d}$, uniformly on compact sets.

$$\sup_{\|x\|\leq M}\left|\mathbb{E}[\hat{f}_n(x)] - f(x)\right| = O(m_n^{-\rho/d})\ . \tag{4.8}$$

4.2.3 Results

In all our results we consider kernels K_m and a density estimator of the form (4.1) such that assumptions (a), (b), (c) and (d) hold.

Theorem 4.1 ($\mathbb{L}^q$-convergence) Geometric case. *Under Assumptions [H4] or [H5] and [H1] or [H1'], the sequence m_n can be chosen such that inequality (4.2) holds for all $0 < q < +\infty$.*

Riemannian case. *Under the assumptions [H4] or [H5], if additionally*
- *[H2] holds with $a > \max(1 + 2/d + (d+1)/\rho, 2 + 1/d)$ (η-dependence),*
- *or [H2'] holds with $a > 1 + 2/d + 1/\rho$ ($\tilde{\phi}$-dependence),*

then the sequence m_n can be chosen such that inequality (4.2) holds for all $0 < q \leqslant q_0 = 2\lceil (a-1)/2 \rceil$.

Theorem 4.2 (Uniform rates, geometric decays) *For any $M > 0$, under Assumptions [H4] or [H5] and [H1] or [H1'] we have, for all $0 < q < +\infty$, and for a suitable choice of the sequence m_n,*

$$\begin{aligned}
\mathbb{E} \sup_{\|x\| \leqslant M} |\hat{f}_n(x) - f(x)|^q &= O\left(\left(\frac{\log^{2(b+1)/b}(n)}{n} \right)^{q\rho/(d+2\rho)} \right), \\
\sup_{\|x\| \leqslant M} |\hat{f}_n(x) - f(x)| &=_{a.s.} O\left(\left(\frac{\log^{2(b+1)/b}(n)}{n} \right)^{\rho/(d+2\rho)} \right).
\end{aligned}$$

Theorem 4.3 (Uniform rates, Riemannian decays) *For any $M > 0$, under Assumptions [H4] or [H5], [H2] or [H2'] with $a \geqslant 4$ and $\rho > 2d$, for $q_0 = 2\lceil (a-1)/2 \rceil$ and $q \leqslant q_0$, the sequence m_n can be chosen such that*

$$\mathbb{E} \sup_{\|x\| \leqslant M} |\hat{f}_n(x) - f(x)|^q = O\left(n^{-\frac{q\rho}{d+2\rho+2d/(q_0+d)}} \right),$$

or such that

$$\sup_{\|x\| \leqslant M} |\hat{f}_n(x) - f(x)| =_{a.s.} O\left(\left(\frac{\log^{q_0+d}(n)}{n^{q_0-2}} \right)^{\frac{\rho}{d(q_0+2)+\rho(q_0+d)}} \right).$$

Remarks.

- Theorem 4.1 shows that the optimal convergence rate of (4.2) still holds in the weak dependence context. In the Riemannian case, when $a \geqslant 4$, the conditions are satisfied if the density function f is sufficient regular, namely, if $\rho > d+1$.
- The loss with respect to the i.i.d. case in the uniform convergence rates (Theorems 4.2 and 4.3) is due to the fact that the probability inequalities for dependent observations are not as good as Bernstein's inequality for i.i.d. random variables (Bernstein inequalities in weak dependence context are proved in [44]). The convergence rates depend on the decay of the weak dependence coefficients. This is in contrast to the case of independent observations.
- In Theorem 4.2 the loss is a power of the logarithm of the number of observations. Let us remark that this loss is reduced when b tends to infinity. In the case of η-dependence and geometric decreasing, the same result is in [41] for the special case $b = 1$. In the framework of φ-dependence, Theorem 4.2 seems to provide the first result on uniform rates of convergence for density estimators.
- In Theorem 4.3, the rate of convergence in the mean is better than the almost sure rate for technical reasons. Contrary to the geometric case, the loss is no longer logarithmic but is a power of n. The rate gets closer to the optimal rate as $q_0 \to \infty$, or equivalently $a \to \infty$.

– These results are new under the assumption of Riemannian decay of the weak dependence coefficients. The condition on a is similar to the condition on β in [4]. Even if the rates are better than in [42], there is a huge loss with respect to the mixing case. It would be interesting to know the minimax rates of convergence in this framework.

4.3 Models, applications and extensions

The class of weak dependent processes is very large. We apply our results to three examples : **two-sided moving averages**, **bilinear models** and **expanding maps**. The first two will be handled with the help of the coefficients η, the third one with the coefficients φ.

4.3.1 Examples of η-dependent time series.

It is of course possible to define η-dependent random fields (see [26] for further details) ; for simplicity, we only consider processes indexed by $\mathbb{Z}$.

Definition 4.4 (Bernoulli shifts) *Let $H : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ be a measurable function. A Bernoulli shift is defined as $X_n = H(\xi_{n-i}, i \in \mathbb{Z})$ where $(\xi_i)_{i\in\mathbb{Z}}$ is a sequence of i.i.d random variables called the innovation process.*

In order to obtain a bound for the coefficients $\{\eta(r)\}$, we introduce the following regularity condition on H. There exists a sequence $\{\delta_r\}$ such that

$$\sup_{i\in\mathbb{Z}} \mathbb{E}\left|H\left(\xi_{i-j}, j \in \mathbb{Z}\right) - H\left(\xi_{i-j}\mathbb{1}_{|j|<r}, j \in \mathbb{Z}\right)\right| \leqslant \delta_r \ ,$$

Bernoulli shifts are η-dependent with $\eta(r) = 2\delta_{r/2}$ (see [41]). In the following, we consider two special cases of **Bernoulli shifts**.

1. **Non causal linear processes.** A real valued sequence $(a_i)_{i\in\mathbb{Z}}$ such that $\sum_{j\in\mathbb{Z}} a_j^2 < \infty$ and the innovation process $\{\xi_n\}$ define a **non-causal linear process** $X_n = \sum_{-\infty}^{+\infty} a_i \xi_{n-i}$. If we control a moment of the innovations, the **linear process** (X_n) is η-dependent. The sequence $\{\eta(r)\}_{r\in\mathbb{N}}$ is directly linked to the coefficients $\{a_i\}_{i\in\mathbb{Z}}$ and various types of decay may occur. We consider only Riemannian decays $a_i = \mathcal{O}\left(i^{-A}\right)$ with $A \geqslant 5$ since results for geometric decays are already known. Here $\eta(r) = \mathcal{O}\left(\sum_{|i|>r/2} a_i\right) = O(r^{1-A})$ and [H2] holds. Furthermore, we assume that the sequence $(\xi_i)_{i\in\mathbb{Z}}$ is i.i.d. and satisfies the condition $|\mathbb{E}e^{iu\xi_0}| \leqslant C(1+|u|)^{-\delta}$, for all $u \in \mathbb{R}$ and for some $\delta > 0$ and $C < \infty$. Then, the densities f and $f_{j,k}$ exist for all $j \neq k$ and they are uniformly bounded (see the proof in the causal case in Lemma 1 and Lemma 2 in [57]) ; hence [H4] holds. If the density f of X_0 is ρ-regular with $\rho > 2$, our estimators converge to the density with the rates :

– $n^{-\rho/(2\rho+1)}$ in $\mathbb{L}^q$-norm ($q \leqslant 4$) at each point x,
– $n^{-\rho/(2\rho+3/2)}$ in $\mathbb{L}^q$-norm ($q \leqslant 4$) uniformly on an interval,
– $\left(\log^4(n)/n\right)^{\rho/(4\rho+3)}$ almost surely on an interval.

In the first case, the rate we obtain is the same as in the i.i.d. case. For such linear models, the density estimator also satisfies the Central Limit Theorem (see [65] and [25]).

2. **Bilinear model.** The process $\{X_t\}$ is a **bilinear model** if there exist two sequences $(a_i)_{i\in\mathbb{N}^*}$ and $(b_i)_{i\in\mathbb{N}^*}$ of real numbers and real numbers a and b such that :

$$X_t = \xi_t \left(a + \sum_{j=1}^{\infty} a_j X_{t-j} \right) + b + \sum_{j=1}^{\infty} b_j X_{t-j} \ . \tag{4.9}$$

Squared **ARCH(∞)** or **GARCH(p, q)** processes satisfy such an equation, with $b = b_j = 0$ for all $j \geq 1$. Define

$$\lambda = \|\xi_0\|_p \sum_{j=1}^{\infty} a_j + \sum_{j=1}^{\infty} b_j \ .$$

If $\lambda < 1$, then the equation (4.9) has a strictly stationary solution in L^p (see [43]). This solution is a **Bernoulli shift** for which we have the behavior of the coefficient η :
– $\eta(r) = O\left(\mathbb{E}^{-\lambda r}\right)$ for some $\lambda > 0$ if there exists an integer N such that $a_i = b_i = 0$ for $i \geqslant N$.
– $\eta(r) = O(\mathbb{E}^{-\lambda\sqrt{r}})$ for some $\lambda > 0$ if $a_i = O(\mathbb{E}^{-Ai})$ and $b_i = O(\mathbb{E}^{-Bi})$ with $A > 0$ and $B > 0$.
– $\eta(r) = O(\{r/\log(r)\}^{-\lambda})$ for some $\lambda > 0$ if $a_i = O(i^{-A})$ and $b_i = O(i^{-B})$ with $A > 1$ and $B > 1$.

Let us assume that the i.i.d. sequence $\{\xi_t\}$ has a marginal density $f_\xi \in C_\rho$, for some $\rho > 2$. The density of X_t conditionally to the past can be written as a function of f_ξ. We then check recursively that the common density of X_t for all t, say f, also belongs to C_ρ. Furthermore, the regularity of f_ξ ensures that f and the joint densities $f_{j,k}$ for all $j \neq k$ are bounded (see [43]) and [H4] holds. The assumptions of Theorem 4.1 are satisfied, and the estimator $\hat{f}_n$ achieves the minimax bound (4.2) if either :
– There exists an integer N such that $a_i = b_i = 0$ for $i \geqslant N$;
– There exist $A > 0$ and $B > 0$ such that $a_i = O(e^{-Ai})$ and $b_i = O(e^{-Bi})$;
– There exist $A \geqslant 4$ and $B \geqslant 5$ such that $a_i = O(i^{-A})$ and $b_i = O(i^{-B})$. Then, this optimal bound holds only for $2 \leqslant q < q(A, B)$ where $q(A, B) = 2[((B - 1) \wedge A)/2]$.

Note finally that the rates of uniform convergence provided by Theorems 4.2 and 4.3 are sub-optimal.

4.3.2 Examples of φ-dependent time series.

Let us introduce an important class of **dynamical systems** :

Example 4.1 *$(T_i = F^i(T_0))_{i\in\mathbb{N}}$ is an* **expanding map** *or equivalently F is a Lasota-Yorke function if it satisfies the three following criteria.*

- *(Regularity) There exists a grid $0 = a_0 \leqslant a_1 \cdots \leqslant a_n = 1$ such as $F \in \mathcal{C}_1$ and $|F'(x)| > 0$ on $]a_{i-1}, a_i[$ for each $i = 1, \dots, n$.*
- *(Expansivity) Let I_n be the set on which $(F^n)'$ is defined. There exists $A > 0$ and $s > 1$ such that $\inf_{x\in I_n} |(F^n)'| > As^n$.*
- *(Topological mixing) For any nonempty open sets U, V, there exists $n_0 \geqslant 1$ such as $F^{-n}(U)\cap V \neq \varnothing$ for all $n \geqslant n_0$.*

Examples of **Markov chains** $X_n = G(X_{n+1}, \varepsilon_n)$ associated to an **expanding map** $\{T_n\}$ belonging to $\mathcal{F}$ are given in [6] and [29]. The simplest one is $X_k = (X_{k-1} + \varepsilon_k)/2$ where the ε_k follows a binomial law and X_0 is uniformly distributed on $[0, 1]$. We easily check that $F(x) = 2x \bmod 1$, the transformation of the associated **dynamical system** T_n, satisfies all the assumptions such as T_n is an **expanding map** belonging to $\mathcal{F}$.

The coefficients of φ-dependence of such a **Markov chain** satisfy $\varphi(r) = O(e^{-ar})$ for some $a > 0$ (see [29]). Theorems 4.1 and 4.2 give the $\mathbb{L}^q$ rate $n^{-\rho/(2\rho+1)}$, the uniform $\mathbb{L}^q$ rate and the almost sure rate $\left(\log^4(n)/n\right)^{\rho/(2\rho+1)}$ of the estimators of the density of μ_0.

4.3.3 Sampled process

Since we do not assume stationarity of the observed process, the following observation scheme is covered by our results. Let $(x_n)_{n\in\mathbb{Z}}$ be a stationary process whose marginal distribution is absolutely continuous, let $(h_n)_{n\in\mathbb{Z}}$ be a sequence of monotone functions and consider the sampled process $\{X_{i,n}\}_{1\leqslant i\leqslant n}$ defined by $X_{i,n} = x_{h_n(i)}$. The dependence coefficients of the sampled process may decay to zero faster than the underlying unobserved process. For instance, if the dependence coefficients of the process $(x_n)_{n\in\mathbb{Z}}$ have a Riemannian decay, those of the sampled process $\{x_{h_n(i)}\}$ with $h_n(i) = i2^n$ decay geometrically fast. The observation scheme is thus a crucial factor that determines the rate of convergence of density estimators.

4.3.4 Density estimators and bias

In this section, we provide examples of kernels K_m and smoothness assumptions on the density f such that assumptions (a), (b), (c) and (d) of subsection 4.2.2 are satisfied.

Kernel estimators The kernel estimator associated to the bandwidth parameter m_n is defined by :

$$\hat{f}_n(x) = \frac{m_n}{n} \sum_{i=1}^{n} K\left(m_n^{1/d}(x - X_i)\right) .$$

We briefly recall the classical analysis for the deterministic part R_n in this case (see [108]). Since the sequence $\{X_n\}$ has a constant marginal distribution, we have $\mathbb{E}[\hat{f}_n(x)] = f_n(x)$ with $f_n(x) = \int_D K(s) f\left(x - s/m_n^{1/d}\right) ds$. Let us assume that K is a Lipschitz function compactly supported in $D \subset \mathbb{R}^d$. For $\rho > 0$, let K satisfy, for all $j = j_1 + \cdots + j_d$ with $(j_1, \ldots, j_d) \in \mathbb{N}^d$:

$$\int x_1^{j_1} \cdots x_d^{j_d} K(x_1, \ldots, x_d) dx_1 \cdots dx_d = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{for } j \in \{1, \ldots, \lceil \rho - 1 \rceil - 1\}, \\ \neq 0 & \text{if } j = \lceil \rho - 1 \rceil. \end{cases}$$

Then the kernels $K_m(x,y) = mK\left(m^{1/d}(x-y)\right)$ satisfy (a), (b) and (c). Assumption (d) holds and if $f \in \mathcal{C}_\rho$, where C_ρ is the class of function f such that for $\rho = \lceil \rho - 1 \rceil + c$ with $0 < c \leqslant 1$, f is $\lceil \rho - 1 \rceil$-times continuously differentiable and there exists $A > 0$ such that $\forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d$, $|f^{(\lceil \rho - 1 \rceil)}(x) - f^{(\lceil \rho - 1 \rceil)}(y)| \leqslant A|x-y|^c$.

Projection estimators We only consider in this section the case $d = 1$. Under the assumption that the family $\{1, x, x^2, \ldots\}$ belongs to $L^2(I, \mu)$, where I is a bounded interval of $\mathbb{R}$ and μ is a measure on I, an orthonormal basis of $L^2(I, \mu)$ can be defined which consists of polynomials $\{P_0, P_1, P_2, \ldots\}$. We assume that f belongs to a class $\mathcal{C}'_\rho$ which is slightly more restrictive than the class $\mathcal{C}_\rho$ (see Theorem 6.23 p.218 in [46] for details). Then for any $f \in L^2(I, \mu) \cap \mathcal{C}'_\rho$, there exists a function $\pi_{f,m_n} \in V_{m_n}$ such that $\sup_{x \in I} |f(x) - \pi_{f,m_n}(x)| = O(m_n^{-\rho})$. Consider then the projection $\pi_{m_n} f$ of f on the subspace V_{m_n} =Vect$\{P_0, P_1, \ldots, P_{m_n}\}$. It can be expressed as

$$\pi_{m_n} f(x) = \sum_{j=0}^{m_n} \left\{ \int_I P_j(s) f(s) d\mu(s) \right\} P_j(x).$$

The projection estimator of the density f of the real valued random variables $\{X_i\}_{1 \leqslant i \leqslant n}$ is naturally defined as

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{m_n}(x, X_i) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=0}^{m_n} P_j(X_i) P_j(x) .$$

Then $\mathbb{E}\hat{f}_n(x) = \pi_{m_n} f(x)$ is an approximation of $f(x)$ in V_{m_n}. The fact that I is compact and the Christoffel-Darboux formula and its corollary (see [106]) ensure properties (a) and (b) for the kernels K_m. We easily check that properties (c) also holds. Unfortunately, the optimal rate $(m_n^{-\rho})$ does not necessarily hold. We then have to consider the weighted kernels $K_m^a(x,y)$ defined by :

$$K_m^a(x,y) = \sum_{j=0}^{m} a_{m,j} \sum_{k=0}^{j} P_k(x) P_k(y) ,$$

where $\{a_{m,j};\, m \in \mathbb{N}, 0 \leqslant j \leqslant m\}$ is a weight sequence satisfying $\sum_{j=0}^{m} a_{m,j} = 1$ and for all j : $\lim_{m \to \infty} a_{m,j} = 0$. If the sequence $\{a_{m,j}\}$ is such that K_m^a is a nonnegative kernel then $\|K_m^a\|_1 =$

$\int_I K_m^a(x,s)d\mu(s) = 1$ and the kernel K_m^a satisfies (a), (b) and (c). Moreover, the uniform norm of the operator $f \mapsto K_m^a * f(x)$ is $\sup_{\|f\|_\infty=1} \|K_m^a * f\|_\infty = \|K_m^a\|_1 = 1$. The linear estimator built with this kernel is

$$\hat{f}_n^a(x) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{m_n} a_{m_n,j}\sum_{k=0}^{j} P_k(X_i)P_k(x) \ ,$$

and its bias has the optimal rate :

$$\begin{aligned}
|\mathbb{E}\hat{f}_n^a(x) - f(x)| &= |K_{m_n}^a * f(x) - \pi_{f,m_n}f(x) + \pi_{f,m_n}f(x) - f(x)| \ , \\
&\leqslant |K_{m_n}^a * (f(x) - \pi_{f,m_n}f(x)) + \pi_{f,m_n}f(x) - f(x)| \ , \\
&\leqslant (\|K_{m_n}^a\|_1 + 1)m_n^{-\rho} = \mathcal{O}(m_n^{-\rho}) \ .
\end{aligned}$$

Such an array $\{a_{m,j}\}$ cannot always be defined. We give an example where it is possible.

Example 4.2 (Fejer kernel) *For the trigonometric basis $\{\cos(nx), \sin(nx)\}_{n\in\mathbb{N}}$, we can find a 2π-periodic function $f \in \mathcal{C}_1'$ such that $\sup_{x\in[-\pi;\pi]}|f(x) - \pi_m f(x)| = O(m^{-1}\log m)$. The associated estimator reads :*

$$\hat{f}_n(x) = \frac{1}{2\pi} + \frac{1}{n\pi}\sum_{i=1}^{n}\sum_{k=1}^{m_n}\cos(kX_i)\cos(kx) + \sin(kX_i)\sin(kx) \ .$$

We remark that $\mathbb{E}\hat{f}_n$ is the Fourier series of f truncated at order m_n :

$$D_{m_n}f(x) = \frac{1}{2\pi}\int_0^{2\pi} f(t)D_{m_n}(x-t)\mathbb{D}t \ .$$

where

$$D_m(x) = \sum_{k=-m}^{m} e^{ikx} = \frac{\sin(\{2m+1\}x/2)}{\sin(x/2)}$$

is (the symmetric) Dirichlet's kernel. Recall that Fejer's kernel is defined as

$$F_m(x) = \frac{1}{m}\sum_{k=0}^{m-1} D_k(x) = \sum_{k=-(m-1)}^{m-1}\left(1 - \frac{|k|}{m}\right)e^{ikx} = \frac{\sin^2(mx/2)}{m\,\sin^2(x/2)} \ .$$

The kernel F_m is a nonnegative weighted kernel corresponding to Dirichlet's kernel and the sequence of weights $a_{m,j} = 1/m$ and satisfies (a), (b) and (c). The estimator associated to the Fejer's kernels is defined by

$$\tilde{f}_n(x) = \frac{1}{2\pi} + \frac{1}{n\pi}\sum_{i=1}^{n}\sum_{j=1}^{m_n}\frac{1}{m_n}\sum_{k=1}^{j}\cos kX_i\cos kx + \sin kX_i \sin kx \ ,$$

If the common density f is 2π-periodic and belongs to $\mathcal{C}_1'$, then assumption (d) holds.

Using general Jackson's kernels (see [89]), we can find an estimator such that $R_n = O(m_n^{-\rho/d})$ for other values of ρ, but the weight sequence $a_{m,j}$ highly depends of the value of ρ.

Wavelet estimation Wavelet estimation is a particular case of projection estimation. For the sake of simplicity, we restrict hte study to $d = 1$.

Definition 4.5 (Scaling function [38]) *A function $\phi \in L^2(\mathbb{R})$ is called a scaling function if the family $\{\phi(\cdot - k)\,;\, k \in \mathbb{Z}\}$ is orthonormal.*

We choose the bandwidth parameter $m_n = 2^{j(n)}$ and define $V_j = \mathrm{Vect}\{\phi_{j,k}, k \in \mathbb{Z}\}$, where $\phi_{j,k} = 2^{j/2}\phi(2^j(x-k))$. Under the assumption that ϕ is compactly supported, we define (the sum over the index k is in fact finite) :

$$\hat{f}_n(x) = \frac{1}{n}\sum_{k=-\infty}^{\infty}\sum_{i=1}^{n}\phi_{j(n),k}(X_i)\phi_{j(n),k}(x)\;.$$

The wavelets estimator is of the form (4.1) with $K(x,y) = \sum_{k=-\infty}^{\infty}\phi(y-k)\phi(x-k)$ and $K_m(x,y) = mK(mx, my)$. Under the additionnal assumption that $\sum_{k\in\mathbb{Z}}\phi(x-k) = 1$ for almost all x, we can write :

$$\begin{aligned}
\left|\mathbb{E}(\hat{f}_n(x) - f(x))\right| &\leqslant \left|\int K_{m_n}(y,x)f(y)dy - f(x)\right|\;, \\
&= \left|\int m_nK(m_ny, m_nx)(f(y) - f(x))dy\right|\;, \\
&= \left|\int m_nK(m_nx + t, m_nx)(f(x + t/m_n) - f(x))dt\right|\;.
\end{aligned}$$

If ϕ is a Lipschitz function such that $\int \phi(x)x^j dx = 0$ if $0 < j < \lceil \rho - 1\rceil$ and $\int \phi(x)x^{\lceil \rho-1\rceil}dx \neq 0$, then the kernel K_m satisfy properties (a), (b) and (c). If $f \in C_\rho$, then Assumption (d) holds.

4.4 Proof of the Theorems

The proof of our results is based on the decomposition :

$$\hat{f}_n(x) - f(x) = \underbrace{\hat{f}_n(x) - \mathbb{E}\left(\hat{f}_n(x)\right)}_{FL_n(x)=\text{fluctuation}} + \underbrace{\mathbb{E}\left(\hat{f}_n(x)\right) - f(x)}_{\text{bias}}\;. \tag{4.10}$$

The bias term is of order $m_n^{-\rho/d}$ by Assumption (d). We now present three lemmas useful to derive the rate of the fluctuation term.

Lemma 4.1 (Moment inequalities) *For each even integer q, under the assumption [H4] or [H5] and if moreover one of the following assumption holds :*

– *[H1] or [H1'] holds (geometric case) ;*
– *[H2] holds, $m_n = n^\delta \log(n)^\gamma$ with $\delta > 0$, $\gamma \in \mathbb{R}$ and*

$$a > \max\left(q-1, \frac{(q-1)\delta(4+2/d)}{q-2+\delta(4-q)}, 2+\frac{1}{d}\right) ,$$

– *[H2'] holds, $m_n = n^\delta \log(n)^\gamma$ with $\delta > 0$ and $\gamma \in \mathbb{R}$ and*

$$a > \max\left(q-1, \frac{(q-1)\delta(2+2/d)}{q-2+\delta(4-q)}, 1+\frac{1}{d}\right) .$$

Then, for each $x \in \mathbb{R}^d$,

$$\limsup_{n\to\infty} (n/m_n)^{q/2} \|FL_n(x)\|_q^q < +\infty .$$

Lemma 4.2 (Probability inequalities)
– *Geometric case. Under Assumptions [H4] or [H5] and [H1] or [H1'] there exist positive constants C_1, C_2 such that*

$$\mathbb{P}\left(|FL_n(x)| \geqslant \varepsilon\sqrt{m_n/n}\right) C_1 \leq \exp\{-C_2 \varepsilon^{b/(b+1)}\} .$$

– *Riemannian case. Under Assumptions [H4] or [H5], if $m_n = n^\delta \log(n)^\gamma$ and if one of the following assumtions holds :*
  – *[H2] with $a > \max\{1 + 2(\delta + 1/d)/(1-\delta), 2 + 1/d\}$,*
  – *[H2'] with $a > \max(1 + 2\{1/d(1-\delta)\}, 1 + 1/d)$,*

  then,

$$\mathbb{P}\left(|FL_n(x)| \geqslant \varepsilon\sqrt{m_n/n}\right) \leq C\varepsilon^{-q_0} ,$$

  with $q_0 = 2\lceil (a-1)/2 \rceil$.

Lemma 4.3 (Fluctuation rates) *Under the assumptions of Lemma 4.2, we have for any $M > 0$,*
– *Geometric case.*

$$\sup_{\|x\|\leqslant M} |FL_n(x)| =_{a.s.} O\left(\sqrt{\frac{m_n}{n}} \log^{(b+1)/b}(n)\right) ;$$

– *Riemannian case.*

$$\sup_{\|x\|\leqslant M} |FL_n(x)| =_{a.s.} O\left(\left(\frac{m_n^{1+2/q_0}}{n^{1-2/q_0}}\right)^{\frac{1}{1+d/q_0}} \log n\right) ,$$

with $q_0 = 2\lceil (a-1)/2 \rceil$.

Remarks.
– In Lemma 4.1, we improve the moment inequality of [42], where the condition in the case of coefficient η is $a > 3(q-1)$, which is always stronger than our condition.

– In the i.i.d. case a Bernstein type inequality is available :

$$\mathbb{P}\left(|FL_n(x)| \geq \varepsilon\sqrt{\frac{m}{n}}\right) \leq C_1 \exp\left(-C_2\varepsilon^2\right) ,$$

Lemma 4.2 provides a weaker inequality for dependent sequences. Other probability inequalities for dependent sequences are presented in [29] and [44].

– Lemma 4.3 gives the almost sure bounds for the fluctuation. It is derived directly from the two previous lemmas.

Proof of the lemmas

Proof. [Proof of Lemma 4.1] Let x be a fixed point in $\mathbb{R}^d$. Denote $Z_i = u_n(X_i) - \mathbb{E}u_n(X_i)$ where $u_n(.) = K_{m_n}(., x)/\sqrt{m_n}$. Then

$$\sum_{i=1}^{n} Z_i = \sum_{i=1}^{n} u_n(X_i) - \mathbb{E}u_n(X_i) = \frac{n}{\sqrt{m_n}}(\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)) = \frac{n}{\sqrt{m_n}} FL_n(x) . \tag{4.11}$$

The order of magnitude of the fluctuation $FL_n(x)$ is obtained by applying the inequality (4.5) to the centered sequence $\{Z_i\}_{1\leqslant i\leqslant n}$ defined above. We then control the normalized fluctuation of (4.11) with the covariance terms $C_k(r)$ defined in equation (4.6). Firstly, we bound the covariance terms :

– **Case** $r = 0$**.** Here $t_1 = \cdots = t_k = i$. Then we get :

$$C_k(r) = \left|\mathrm{cov}\left(Z_{t_1}\cdots Z_{t_p}, Z_{t_{p+1}}\cdots Z_{t_k}\right)\right| \leqslant 2\mathbb{E}|Z_i|^k .$$

By definition of Z_i :

$$\mathbb{E}|Z_i|^k \leqslant 2^k\mathbb{E}|u_n(X_i)|^k \leqslant 2^k\|u_n\|_\infty^{k-1}\mathbb{E}|u_n(X_0)| . \tag{4.12}$$

– **Case** $r > 0$**.** $C_k(r) = \left|\mathrm{cov}\left(Z_{t_1}\cdots Z_{t_p}, Z_{t_{p+1}}\cdots Z_{t_k}\right)\right|$ is bounded in different ways, either using weak-dependence property or by direct bound.
  – **Weak-dependence bounds :**
    – *η-dependence :* Consider the following application :

$$\phi_p : (x_1, \ldots, x_p) \mapsto (u_n(x_1)\cdots u_n(x_p)) .$$

Then $\|\phi_p\|_\infty \leqslant 2^p\|u_n\|_\infty^p$ and $\mathrm{Lip}\,\phi_p \leqslant 2^p\|u_n\|_\infty^{p-1}\mathrm{Lip}\,u_n$. Thus by η-dependence, for all $k \geqslant 2$ we have :

$$\begin{aligned} C_k(r) &\leqslant \left(p2^p\|u_n\|_\infty^{p-1} + (k-p)2^{p-k}\|u_n\|_\infty^{p-k-1}\right)\mathrm{Lip}\,u_n\eta(r) ,\\ &\leqslant k2^k\|u_n\|_\infty^{k-1}\mathrm{Lip}\,u_n\eta(r) . \end{aligned} \tag{4.13}$$

– *φ-dependence :* We use the inequality (4.7). Using the bound

$$\mathbb{E}|\phi_p(X_1,\dots,X_p)| \leqslant \|u_n\|_\infty^{p-1}\mathbb{E}|u_n(X_0)| \ ,$$

we derive a bound for the covariance terms :

$$C_k(r) \quad \leqslant \quad k2^k\|u_n\|_\infty^{k-2}\mathbb{E}|u_n(X_0)|\mathrm{Lip}\, u_n\tilde{\phi}(r) \ . \tag{4.14}$$

– **Direct bound :** Triangular inequality implies for $C_k(r)$:

$$\left|\mathrm{cov}\left(Z_{t_1}\cdots Z_{t_p}, Z_{t_{p+1}}\cdots Z_{t_k}\right)\right| \leqslant \underbrace{\left|\mathbb{E}\prod_{i=1}^k Z_{t_i}\right|}_{A} + \underbrace{\left|\mathbb{E}\prod_{i=1}^p Z_{t_i}\right|}_{B_p}\underbrace{\left|\mathbb{E}\prod_{i=p+1}^k Z_{t_i}\right|}_{B_{k-p}} ,$$

$$\begin{aligned} A &= |\mathbb{E}\left(u_n(X_{t_1}) - \mathbb{E}u_n(X_{t_1})\right)\cdots\left(u_n(X_{t_k}) - \mathbb{E}u_n(X_{t_k})\right)| \ , \\ &= |\mathbb{E}u_n(X_0)|^k + |\mathbb{E}\left(u_n(X_{t_1})\cdots u_n(X_{t_k})\right)| \\ &\quad + \sum_{s=1}^{k-1}|\mathbb{E}u_n(X_0)|^{k-s}\sum_{t_{i_1}\leqslant\cdots\leqslant t_{i_s}}\left|\mathbb{E}\left(u_n(X_{t_{i_1}})\cdots u_n(X_{t_{i_s}})\right)\right| \ . \end{aligned}$$

Firstly, with $k \geqslant 2$:

$$|\mathbb{E}u_n(X_0)|^k \leqslant \|u_n\|_\infty^{k-2}(\mathbb{E}|u_n(X_0)|)^2 \ .$$

Secondly, if $1 \leqslant s \leqslant k-1$:

$$\begin{aligned} \left|\mathbb{E}\left(u_n(X_{t_{i_1}})\cdots u_n(X_{t_{i_s}})\right)\right| &\leqslant \mathbb{E}|u_n(X_{t_{i_1}})\cdots u_n(X_{t_{i_s}})| \ , \\ &\leqslant \|u_n\|_\infty^{s-1}\mathbb{E}|u_n(X_0)| \ , \\ |\mathbb{E}u_n(X_0)|^{k-s} &\leqslant \|u_n\|_\infty^{k-s-1}\mathbb{E}|u_n(X_0)| \ . \end{aligned}$$

Thirdly there is at least two different observations with a gap of $r > 0$ among $X_{t_1},\dots,X_{t_k}$ so for any integer $k \geqslant 2$:

$$|\mathbb{E}\left(u_n(X_{t_1})\cdots u_n(X_{t_k})\right)| \leqslant \|u_n\|_\infty^{k-2}\mathbb{E}|u_n(X_0)u_n(X_r)| \ .$$

Then, collecting the last four inequations yields :

$$\begin{aligned} A \quad &\leqslant \quad \|u_n\|_\infty^{k-2}(\mathbb{E}|u_n(X_0)|)^2 \\ &\quad + (\mathbb{E}|u_n(X_0)|)^2\sum_{s=1}^{k-1}C_s^k\|u_n(X_0)\|_\infty^{k-2} + \|u_n\|_\infty^{k-2}\mathbb{E}|u_n(X_0)u_n(X_r)| \ . \end{aligned}$$

So :
$$A \leqslant \|u_n\|_\infty^{k-2}\left((2^k-1)(\mathbb{E}|u_n(X_0)|)^2+\mathbb{E}|u_n(X_0)u_n(X_r)|\right) . \tag{4.15}$$
Now, we bound B_i with $i<k$. As before :
$$\begin{aligned} B_i &= \left|\mathbb{E}\left(u_n(X_{t_1})-\mathbb{E}u_n(X_{t_1})\right)\cdots\left(u_n(X_{t_i})-\mathbb{E}u_n(X_{t_i})\right)\right| ,\\ &= \sum_{s=0}^{i}|\mathbb{E}(u_n(X_0)|^{i-s}\sum_{t_{j_1}\leqslant\ldots\leqslant t_{j_s}}\left|\mathbb{E}\left(u_n(X_{t_{j_1}})\cdots u_n(X_{t_{j_s}})\right)\right| ,\\ &\leqslant 2^i\|u_n\|_\infty^{i-2}(\mathbb{E}|u_n(X_0)|)^2 . \end{aligned}$$

Then :
$$B_p\times B_{k-p}\leqslant 2^k\|u_n\|_\infty^{k-4}(\mathbb{E}|u_n(X_0)|)^4\leqslant 2^k\|u_n\|_\infty^{k-2}(\mathbb{E}|u_n(X_0)|)^2 . \tag{4.16}$$
Another interesting bound for $r>0$ follows, because according to inequalities (4.15) and (4.16) we have :
$$C_k(r)\leqslant \|u_n\|_\infty^{k-2}\left((2^{k+1}-1)(\mathbb{E}|u_n(X_0)|)^2+\mathbb{E}|u_n(X_0)u_n(X_r)|\right) .$$
Noting $\gamma_n(r)=\mathbb{E}|u_n(X_0)u_n(X_r)|\vee(\mathbb{E}|u_n(X_0)|)^2$, we have :
$$C_k(r)\leqslant 2^{k+1}\|u_n\|_\infty^{k-2}\gamma_n(r) . \tag{4.17}$$

We now use the different values of the bounds in inequalities (4.12), (4.13), (4.14) and (4.17). If we define the sequence $(w_r)_{0\leqslant r\leqslant n-1}$ as :
- $w_0=1$,
- $w_r=\gamma_n(r)\wedge\|u_n\|_\infty\mathrm{Lip}\,u_n\eta(r)\wedge\mathbb{E}|u_n(X_0)|\mathrm{Lip}\,u_n\varphi(r)$,

then, for all r such that $0\leqslant r\leqslant n-1$ and for all $k\geqslant 2$:
$$C_k(r)\leqslant k2^k\|u_n\|_\infty^{k-2}w_r .$$

We derive from this inequality and from (4.5) :
$$\begin{aligned} \left\|\sum_{i=1}^n Z_i\right\|_q^q &\leqslant \frac{(2q-2)!}{(q-1)!}\left\{\left(n\sum_{r=0}^{n-1}C_2(r)\right)^{q/2}\vee n\sum_{r=0}^{n-1}(r+1)^{q-2}C_q(r)\right\} ,\\ &\leqslant (4q\sqrt{n})^q\left\{\left(\sum_{r=0}^{n-1}w_r\right)^{q/2}\vee\left(\frac{\|u_n\|_\infty}{\sqrt{n}}\right)^{q-2}\sum_{r=0}^{n-1}(r+1)^{q-2}w_r\right\} . \end{aligned}$$

In order to control w_r, we give bounds for the terms $\gamma_n(r)=\mathbb{E}|u_n(X_0)u_n(X_r)|\vee(\mathbb{E}|u_n(X_0)|)^2$:
- In the case of [H4], we have :
$$\mathbb{E}|u_n(X_0)u_n(X_r)|\leqslant \sup_{j,k}\|f_{j,k}\|_\infty\|u_n\|_1^2 ,$$
$$(\mathbb{E}|u_n(X_0)|)^2\leqslant\|f\|_\infty^2\|u_n\|_1^2 .$$

– In the case of [H5], Lemma 2.3 of [98] proves that $\mathbb{E}|u_n(X_0)u_n(X_r)| \leqslant (\mathbb{E}|u_n(X_0)|)^2$ for n sufficiently large and the same bound as above remains true for the last term.

In both cases, we conclude that $\gamma_n(r) \preceq \|u_n\|_1^2$. The properties (a), (b) and (c) of section 4.2.2 and the expression $u_n = K_{m_n}/\sqrt{m_n}$ ensure that $\|u_n\|_1 \preceq \dfrac{1}{\sqrt{m_n}}$ and $\operatorname{Lip} u_n \preceq m_n^{1/2+1/d}$. We deduce using the compact support of K that $\|u_n\|_\infty \preceq \sqrt{m_n}$. We then have for $r \geqslant 1$:

$$w_r \preceq \frac{1}{m_n} \wedge m_n^{1+1/d}\eta(r) \wedge m_n^{1/d}\varphi(r) \; . \tag{4.18}$$

In order to prove Lemma 4.1, it remains to control the sums

$$\left(\frac{\|u_n\|_\infty}{\sqrt{n}}\right)^{k-2} \sum_{r=0}^{n-1} (r+1)^{k-2} \frac{1}{m_n} \wedge m_n^{1+1/d}\eta(r) \wedge m_n^{1/d}\varphi(r) \; , \tag{4.19}$$

for $k=2$ and $k=q$ in both Riemannian and geometric cases.

– **Geometric case.**

Under [H1] or [H1'] : We remark that $a \wedge b \leqslant a^\alpha b^{1-\alpha}$ for all $\alpha \in [0;1]$. Using (4.18), we obtain first that $w_r \preceq (\eta(r) \wedge \varphi(r))^\alpha m_n^{\alpha(1+1/d)-(1-\alpha)}$ for n sufficiently large. Then for $0 < \alpha \leqslant \frac{d}{2d+1}$ we bound w_r independently of m_n : $w_r \preceq (\eta(r) \wedge \varphi(r))^\alpha$. For all even integer $k \geqslant 2$ we derive from the form of $\eta(r) \wedge \tilde{\phi}_r$ that (in the third inequality $u = ar^b$) :

$$\begin{aligned}
\sum_{r=1}^{n-1} (r+1)^{k-2} w_r &\preceq \sum_{r=0}^{n-1} (r+1)^{k-2} \exp(-\alpha a r^b) \; , \\
&\preceq \int_0^\infty r^{k-2} \exp(-\alpha a r^b) dr \; , \\
&\preceq \frac{1}{b a^{\frac{k-1}{b}}} \int_1^\infty u^{\frac{k-1}{b}-1} \exp(-u) du \; , \\
&\preceq \frac{1}{b a^{\frac{k-1}{b}}} \Gamma\left(\frac{k-1}{b}\right) \; .
\end{aligned}$$

Using the Stirling formula, we can find a constant B such that, for the special cases $k=2$ and $k=q$:

$$\sum_{r=1}^{n-1} (r+1)^{k-2} w_r \preceq \frac{1}{b a^{\frac{k-1}{b}}} \Gamma\left(\frac{k-1}{b}\right) \preceq (Bk)^{\frac{k}{b}} \; .$$

– **Riemannian case.**

Under [H6] and [H2] : Let us recall that [H6] implies that $m_n \leqslant n^\delta$ for n sufficiently large and $0 < \delta < 1$ and that the assumption of Lemma 4.1 implies that :

$$a > \max\left(q-1, \frac{\delta(q-1)(4+2/d)}{q-2+\delta(4-q)}, 2+\frac{1}{d}\right) \; .$$

Then, we have $a > \max\left(k-1, \dfrac{\delta(k-1)(4+2/d)}{k-2+\delta(4-k)}\right)$ for both cases $k=q$ or $k=2$. This assumption on a implies that :

$$\frac{(k+2/d)\delta+2-k}{2(a-k+1)} < \frac{(4-k)\delta+k-2}{2(k-1)} .$$

Furthermore, reminding that $0<\delta<1$:

$$0 < \frac{(4-k)\delta+k-2}{2(k-1)} = 1 - \frac{k(1+\delta)-4\delta}{2(k-1)} \leqslant 1 .$$

We derive from the two previous inequalities that there exists $\zeta_k \in]0,1[$ verifying

$$\frac{(k+2/d)\delta+2-k}{2(a-k+1)} < \zeta_k < \frac{(4-k)\delta+k-2}{2(k-1)}.$$

For $k=q$ or $k=2$, we now use Tran's technique as in [2]. We divide the sum (4.19) in two parts in order to bound it by sequences tending to 0, due to the choice of ζ_k :

$$\begin{aligned}
\left(\sqrt{\frac{m_n}{n}}\right)^{k-2} \sum_{r=0}^{[n^{\zeta_k}]-1} (r+1)^{k-2} w_r &\preceq \left(\sqrt{\frac{m_n}{n}}\right)^{k-2} \frac{[n^{\zeta_k}]^{k-1}}{m_n} , \\
&\preceq n^{(2\zeta_k(k-1)-((4-k)\delta+k-2))/2} , \\
&= O(1) , \\
\left(\sqrt{\frac{m_n}{n}}\right)^{k-2} \sum_{r=[n^{\zeta_k}]}^{n-1} (r+1)^{k-2} w_r &\leqslant \left(\sqrt{\frac{m_n}{n}}\right)^{k-2} m_n^{1+1/d} [n^{\zeta_k}]^{k-1-a} , \\
&\leqslant n^{(-2\zeta_k(a-k-1)+((k+2/d)\delta+2-k))/2} , \\
&= O(1) .
\end{aligned}$$

Under [H6] and [H2'] : Under the assumption of Lemma 4.1 :

$$a > \max\left(q-1, \frac{\delta(q-1)(2+2/d)}{q-2+\delta(4-q)}, 1+\frac{1}{d}\right) ,$$

we derive exactly as in the previous case that there exists $\zeta_k \in]0;1[$ for $k=q$ or $k=2$ such that

$$\frac{(k-2+2/d)\delta+2-k}{2(a-k+1)} < \zeta_k < \frac{(4-k)\delta+k-2}{2(k-1)} .$$

We then apply again the Tran's technique that bound the sum (4.19) in that case.

Lemma 4.1 directly follow from (4.11). □

Remarks. We have in fact proved the following sharper result.In the geometric case, there exists a constant $C>0$ such that for every even integer q and all x in the support of K_{m_n} :

$$\left\| \frac{1}{\sqrt{nm_n}} \sum_{i=1}^n K_{m_n}(X_i,x) - \mathbb{E}K_{m_n}(X_i,x) \right\|_q^q \leq \left(Cq^{1+1/b}\right)^q. \tag{4.20}$$

Proof. [Proof of Lemma 4.2] The cases of Riemannian or geometric decay of the dependence coefficients are considered separately.

- **Geometric decay** We present a technical lemma useful to deduce exponential probabilities from moment inequalities at any even order.

  Lemma 4.4 *If the variables $\{V_n\}_{n\in\mathbb{Z}}$ satisfies, for all $k \in \mathbb{N}^*$*
  $$\|V_n\|_{2k} \leqslant \phi(2k)\ , \tag{4.21}$$
  where ϕ is an increasing function with $\phi(0) = 0$. Then :
  $$\mathbb{P}(|V_n| \geqslant \varepsilon) \leqslant e^2 \exp\left(-\phi^{-1}(\varepsilon/e)\right)\ .$$

  Proof. By Markov's inequality and Assumption (4.21), we obtain
  $$\mathbb{P}\left(|V_n| \geqslant \varepsilon\right) \leqslant \left(\frac{\phi(2k)}{\varepsilon}\right)^{2k}\ .$$
  With the convention $0^0 = 1$, the inequality is true for all $k \in \mathbb{N}$. Reminding that $\phi(0) = 0$, there exists an integer k_0 such that $\phi(2k_0) \leqslant \varepsilon/e < \phi(2(k_0+1))$. Noting ϕ^{-1} the generalized inverse of ϕ, we have :
  $$\begin{aligned}\mathbb{P}\left(|V_n| \geqslant \varepsilon\right) &\leqslant \left(\frac{\phi(2k_0)}{\varepsilon}\right)^{2k_0} \leqslant e^{-2k_0} = e^2 e^{-2(k_0+1)}\ ,\\ &\leqslant e^2 \exp\left(-\phi^{-1}(\varepsilon/e)\right)\ .\end{aligned}$$
  □

  We rewrite the inequality (4.20) : $\left\|\sqrt{\frac{n}{m_n}} FL_n\right\|_{2k} \leqslant \phi(2k)$ with $\phi(x) = Cx^{\frac{b+1}{b}}$ for a convenient constant C. Applying Lemma 4.4 to $V_n = \sqrt{\frac{n}{m_n}} FL_n$ we obtain :
  $$\mathbb{P}\left(|FL_n| \geqslant \varepsilon\sqrt{\frac{m_n}{n}}\right) \leqslant e^2 \exp\left(-\phi^{-1}(\varepsilon/e)\right)\ ,$$
  and we obtain the result of the Lemma 4.2.

- **Riemannian decay** In this case, the result of Lemma 4.1 is obtained only for some values of q depending of the value of the parameter a :
  - In the case of η-dependence :
  $$a > \max\left(q-1, \frac{1+\delta+2/d}{1-\delta}, 2+\frac{1}{d}\right)\ .$$
  - In the case of φ-dependence :
  $$a > \max\left(q-1, 1+\frac{2}{d(1-\delta)}, 1+\frac{1}{d}\right)\ .$$

We consider that the assumptions of the Lemma 4.2 on a are satisfied in both cases of dependence. Then $q_0 = 2\left\lceil \frac{a-1}{2} \right\rceil$ is the even integer such that $a-1 \leqslant q_0 < a+1$. It is the largest order such that the assumptions of Lemma 4.1 (recalled above) are satisfied and then the Lemma 4.1 gives us directly the rate of the moment : $\lim_{n\to\infty} \sup \left(\frac{n}{m_n}\right)^{q_0/2} \|FL_n(x)\|_{q_0}^{q_0} < +\infty$. We apply Markov to obtain the result of Lemma 4.2 :

$$\mathbb{P}\left(|FL_n(x)| \geqslant \varepsilon\sqrt{\frac{m_n}{n}}\right) \leqslant \frac{\left(\sqrt{\frac{n}{m_n}}\|FL_n(x)\|_{q_0}\right)^{q_0}}{\varepsilon^{q_0}} .$$

□

Proof. [Proof of Lemma 4.3] We follow here Liebscher's strategy as in [4]. We recover $B := B(0, M)$, the ball of center 0 and radius M, by at least $(4M\mu + 1)^d$ balls $B_j = B(x_j, 1/\mu)$. Then, under the assumption that $K_m(., y)$ is supported on a compact of diameter proportional smaller than $1/m^{1/d}$, we have, for all j :

$$\sup_{x\in B_j} |FL_n(x)| \leqslant |\hat{f}_n(x_j) - \mathbb{E}\hat{f}_n(x_j)| + C\frac{{m_n}^{1/d}}{\mu}(|\tilde{f}_n(x_j) - \mathbb{E}\tilde{f}_n(x_j)| + 2|\mathbb{E}\tilde{f}_n(x_j)|) , \tag{4.22}$$

with C a constant and $\tilde{f}_n(x) = \frac{1}{n}\sum_{i=1}^n \tilde{K}_{m_n}(x, X_i)$ where $\tilde{K}_{m_n}$ is a kernel of type $\tilde{K}_m(x, y) = K_0 m k(x, y, x_j, 1/m^{1/d})$. The $1/b$-Lipschitz function $k(x, y, a, b)$ is equal to 1 on $B(a, b)$ and null outside $B(a, b+1/b)$. The constant K_0 is fixed in order that $\tilde{K}_{m_n}$ satisfies properties (a), (b) and (c) of section 4.2.2. Then using (4.22) and with obvious short notation :

$$\begin{aligned}
\mathbb{P}\left(\sup_{\|x\|\leqslant M} |FL_n(x)| > \varepsilon\sqrt{\frac{m_n}{n}}\right) &\leqslant \sum_{j=1}^{(4M\mu+1)^d} \mathbb{P}\left(\sup_{x\in B_j} |FL_n(x)| > \varepsilon\sqrt{\frac{m_n}{n}}\right) , \\
&\leqslant (4M\mu+1)^d \left[\sup_{x\in B_j} \mathbb{P}\left(|FL_n(x_j)| > \varepsilon\sqrt{\frac{m_n}{n}}\right)\right. \\
&\quad + \mathbb{P}\left(C\frac{m_n^{1/d}}{\mu}|\tilde{F}L_n(x)| > \varepsilon\sqrt{\frac{m_n}{n}}\right) \\
&\quad \left. + \mathbb{P}\left(2C\frac{m_n^{1/d}}{\mu}|\mathbb{E}\tilde{f}_n(x_j)| > \varepsilon\sqrt{\frac{m_n}{n}}\right)\right] .
\end{aligned}$$

Using the fact that f is bounded, $\mathbb{E}\tilde{f}_n = \int \tilde{K}_{m_n}(x_j, s)f(s)ds$ is bounded independently of n. Then taking $\mu = m_n^{1/d-1/2} n^{1/2} L(n)/\varepsilon$ ensures that $\mathbb{P}\left(2C\frac{m_n^{1/d}}{\mu}|\mathbb{E}\tilde{f}_n(x_j)| > \varepsilon\sqrt{\frac{m_n}{n}}\right)$ is null for n sufficiently large. Applying Lemma 4.2 on f and $\tilde{f}$, uniform probability inequality in both cases of geometric and Riemannian decays become :

$$\mathbb{P}\left(\sup_{\|x\|\leqslant M}|FL_n(x)|\geqslant \varepsilon_n\sqrt{\frac{m_n}{n}}\right) \preceq \mu^d\exp\left(-C\varepsilon_n^{\frac{b}{b+1}}\right), \tag{4.23}$$

$$\mathbb{P}\left(\sup_{\|x\|\leqslant M}|FL_n(x)|\geqslant \varepsilon_n\sqrt{\frac{m_n}{n}}\right) \preceq \mu^d\varepsilon_n^{-q_0}. \tag{4.24}$$

In the geometric case, fix ε_n as $G(\log n)^{(b+1)/b}$ such that the bound becomes $\mu^d n^{-GC}$. Reminding that $\mu\leqslant n$, the sequence $\mu^d n^{-GC}$, bounded by n^{d-GC}, is summable for a conveniently chosen constant G. Borel-Cantelli's Lemma then concludes the proof in this case.
In the Riemannian case, take $\varepsilon_n=(m_n^{1-d/2}n^{1+d/2})^{\frac{1}{q_0+d}}\log n$ in order that the bound becomes $n^{-1}\log^{-q_0}nL(n)$. Reminding that $q_0\geqslant 2$, this sequence is summable and here again we conclude by applying Borel-Cantelli's Lemma. □

Proof of the theorems

The order of magnitude of the bias is given by Assumption (d) and the Lemmas provide bounds for fluctuation term. There only remain to determine the optimal bandwidth m_n in each case.
Proof. [Proof of Theorem 4.1] Applying Lemma 4.1 yields Theorem 4.1 when q is an even integer. For any real q, Lemma 4.1 with $2(\lceil q/2\rceil+1)\geqslant 2$ and Jensen's inequalities yields :

$$\begin{aligned}\left(\frac{n}{m_n}\right)^{q/2}\mathbb{E}|FL_n(x)|^q &= \left(\frac{n}{m_n}\right)^{q/2}\mathbb{E}\left(FL_n(x)^{2(\lceil q/2\rceil+1)}\right)^{q/\{2(\lceil q/2\rceil+1)\}},\\ &\leqslant \left(\left(\frac{n}{m_n}\right)^{\lceil q/2\rceil+1}\mathbb{E}FL_n(x)^{2(\lceil q/2\rceil+1)}\right)^{q/\{2(\lceil q/2\rceil+1)\}}.\end{aligned}$$

Plugging this bound and the bound for the bias in (4.10), we obtain a bound for the $\mathbb{L}^q$-error of estimation :

$$\|\hat{f}_n(x)-f(x)\|_q\leqslant\|FL_n(x)\|_q+|R_n(x)|=O\left(\sqrt{\frac{m_n}{n}}+m_n^{-\rho/d}\right).$$

The optimal bandwidth $m_n^*=n^{\frac{d}{2\rho+d}}$ is the same as in the i.i.d. case. Thus [H6] holds with $\delta=\frac{d}{2\rho+d}$. For this valued of δ, the conditions on the parameter a of Lemma 4.2 are equivalent to those of Theorem 4.1. □
Proof. [Proof of Theorem 4.2] Applying the probability inequality (4.23) in the proof of Lemma 4.3 and the identity $\mathbb{E}|Y|^q=\int_0^{+\infty}\mathbb{P}\left(|Y|\geqslant t^{1/q}\right)dt$, we obtain

$$\mathbb{E}\sup_{\|x\|\leqslant M}|\hat{f}_n(x)-f(x)|^q=O\left(\left\{\sqrt{\frac{m_n}{n}}\log^{(b+1)/b}(n)\right\}^q+m_n^{-q\rho/d}\right).$$

Lemma 4.3 gives the rate of almost sure convergence :

$$\sup_{\|x\|\leqslant M} |\hat{f}_n(x) - f(x)| =_{a.s.} O\left(\sqrt{\frac{m_n}{n}} \log^{\frac{b+1}{b}} n + m_n^{-\rho/d}\right) .$$

In both cases, the optimal bandwidth is $m_n^* = (n/\log^{2(b+1)/b}(n))^{d/(2\rho+d)}$, which yields the rates claimed in Theorem 4.2. □

Proof. [Proof of Theorem 4.3] Applying the probability inequality (4.24) and the same line of reasoning as in the previous proof, we obtain

$$\mathbb{E} \sup_{\|x\|\leqslant M} |\hat{f}_n(x) - f(x)|^q = O\left(\left(\sqrt{\frac{m_n}{n}} m_n^{\frac{2}{q_0+d}}\right)^q + m_n^{-q\rho/d}\right) ,$$

where $q_0 = 2\lceil (a-1)/2 \rceil$. The optimal bandwidth $m_n^* = n^{d/(d+2\rho+2d/(q_0+d))}$ implies [H6] with $\delta = d/(d+2\rho+2d/(q_0+d))$. For this value of δ, the conditions on a of Lemma 4.2 are satisfied as soon as $a \geqslant 4$ and $\rho > 2d$.

Lemma 4.3 gives the rate for the fluctuation in the almost sure case. This leads the optimal bandwidth

$$m_n^* = \left(n^{q_0-2}/\log^{q_0+d}(n)\right)^{\frac{d}{d(q_0+2)+\rho(q_0+d)}} .$$

We then deduce the two different rates of Theorem 4.3, either in the almost sure or in the $\mathbb{L}^q$ framework. □

Chapitre 5

Adaptive density estimation under dependence

Abstract

Assume that $(X_t)_{t\in\mathbb{Z}}$ is a real valued time series admitting a common marginal density f with respect to Lebesgue measure. Donoho *et al.* (1996) propose a near-minimax method based on thresholding wavelets to estimate f on a compact set in an independent and identically distributed setting. The aim of the present work is to extend this methodology to different weakly dependent cases. Bernstein's type inequalities are proved to be sufficient to extend near-minimax results. Assumptions are detailed for dynamical systems and under the η-weak dependence condition from Doukhan & Louhichi (1999). The threshold levels in our estimator integrates the dependence structure of the sequence $(X_t)_{t\in\mathbb{Z}}$ through one parameter γ. The near minimaxity is obtained for $\mathbb{L}^p$-convergence rates ($p \geqslant 1$). An estimator of γ is obtained by a cross-validation procedure. The procedure is illustrated via a simulation study of some dynamical systems and non Markovian η-weakly dependent sequences.

Keywords : Adaptive estimation, Asymptotic minimax, Hard thresholding, Nonparametric density estimation, Wavelets, Weak dependence.

Note

The content of this part is based on a paper, written in collaboration with Irène Gannaz, submitted to ESAIM P& S.

Introduction

Let $(X_t)_{t\in\mathbb{Z}}$ be a real valued time series admitting a common marginal density f that is compactly supported. The general purpose of this chapter is to estimate f by a wavelet estimator $\hat{f}_n$ constructed from n observations $(X_1,\dots,X_n)$. We refer to Vannucci (1998) [109] for a survey of the use of wavelet bases in density estimation. Wavelets are interesting because they are localised both in time and frequency. Donoho *et al.* (1996) [35] showed that projection-like linear estimators are not optimal in a minimax approach, i.e. that their rates of convergence are slower than the minimax one. Introducing nonlinearity by thresholding wavelet coefficients allow them to obtain quasi-optimal results. The corresponding estimators are called near-minimax because their rates differ from the minimax one only up to a logarithm term. The present work extends wavelet density estimation from the independent and identically distributed (iid for short) framework to cases where dependence between variables occurs. Many kinds of thresholding exist, such as for example Hall and Patil's (1995) [64] or those presented in Antoniadis and Fan (2001) [5], but we restrict ourselves to hard-thresholding.

Several ways of quantifying dependence have already been worked out. One of the most popular is the notion of mixing. For β-mixing, Tribouley and Viennet (1998) [107] have proved the optimality of an estimator $\hat{f}_n$ for which the Mean Integrated Square Error (MISE for short) is the minimax one. Comte and Merlevède (2002) [23] obtain a near-minimax result for more general α-mixing models, paying the gain of generality by a logarithm loss in the convergence rate. However, the class of α and β-mixing models is quite restrictive. Andrews (1984) [1] exhibits the simple non-mixing process

$$X_t = \frac{1}{2}\left(X_{t-1}+\xi_t\right), \text{ where } (\xi_t)_{t\in\mathbb{Z}} \text{ iid with law } Bern(p). \tag{5.1}$$

New coefficients have been recently introduced to study such non mixing models. Via a time reversion, the Markov chain (5.1) is distributed as the dynamical system $(Y_t = T^t Y_0)_{t\in\mathbb{Z}}$, where Y_0 follow the uniform distribution on $[0,1]$ and $T(x) = 2x \bmod 1$ (see Barbour *et al.* (2000) [6] for more details). As noted by Dedecker and Prieur (2005) [29], such dynamical systems (and associated Markov chains) are dependent but not mixing. The above cited authors introduce the notion of φ-weak dependence to quantify dependence of such processes. Another (eventually) non-mixing class of models is the one of non-causal Bernoulli shifts –$X_t = H(\xi_{t-j}, j\in\mathbb{Z}))_{t\in\mathbb{Z}}$ where $(\xi_t)_{t\in\mathbb{Z}}$ is an iid process–. These models belong to the class of η-weak dependent processes (see Doukhan and Louhichi, 1999, [41]).

The estimation scheme is based on Donoho *et al.* (1996)'s procedure, in [35], and it is adaptative with respect to the regularity of f. The hard-threshold levels of the estimator $\hat{f}_n$ integrate the dependence of the observations via a parameter γ. If the weak dependence context is known, the estimator is near minimax : in the case of dynamical systems the same rate as in iid setting is achieved thanks to inequalities obtained by Dedecker *et al.* (2007) [27] and by Doukhan & Neumann [44] and for non-causal η-weakly dependent observations another logarithmic loss appears. If the weak dependence context is unknown, a cross validation procedure gives an estimator $\hat{\gamma}_n$. The resulting estimator $\hat{f}_n^{\hat{\gamma}_n}$ is adaptative with respect to the dependence and it is available for a large weak dependence range in an unified approach. We believe that this is a real improvement on existing results because no restrictive mixing assumptions on the observations are required.

The chapter is structured as follows. The context of estimation is presented in the next section. Examples of different weakly dependent sequences of interest are given in Section 5.2. Section 5.3 is devoted to the main results where relevent probability inequalities and near minimax rates are obtained. In Section 5.4 some numerical applications are discussed and near minimax density estimators are given by a new cross validation procedure. The proofs are relegated in the last Section.

5.1 Estimation framework

Without loss of generality f is considered supported by $[0, 1]$. Let us recall some useful facts about wavelet estimation (see Hardle, Kerkyacharian, Picard and Tsybakov (1998) [67] for more details on wavelet estimation). For all $p \geqslant 1$, $\mathbb{L}^p$ denotes in the sequel the space of all functions f supported by $[0, 1]$ such that $\|f\|_p^p = \int |f(x)|^p dx < \infty$.

Definition 5.1 *An orthogonal multiresolution analysis of $\mathbb{L}^2$ is an increasing sequence $(V_j)_{j\in\mathbb{N}}$ of closed subsets of $\mathbb{L}^2$ satisfying*

(i) $\bigcap_{j\in\mathbb{N}} V_j = \{0\}$ *and* $\overline{\bigcup_{j\in\mathbb{N}} V_j} = \mathbb{L}^2$,

(ii) $\forall\ f \in \mathbf{L}^2$, $\forall\ j \in \mathbb{N}$, $f \in V_j$ *if and only if* $x \mapsto f(2^{-j}x)$ *belongs to* V_{j+1},

(iii) *There exists a function ϕ, called father wavelet, such that $\{x \mapsto \phi(x-k)\}_{k\in\mathbb{Z}}$ forms an orthonormal base of V_0.*

If W_j denotes the orthogonal supplement of V_j in V_{j+1}, i.e $V_{j+1} = V_j \oplus W_j$, then there exists a function ψ –called mother wavelet– such that $\{x \mapsto \psi(x-k)\}_{k\in\mathbb{Z}}$ forms an orthonormal basis of W_0. At every resolution level $j \geq 0$ the families $\{\phi_{j,k} : x \rightarrow 2^{j/2}\phi(2^j x - k)\}_{k\in\mathbb{Z}}$ and $\{\psi_{j,k} : x \mapsto 2^{j/2}\psi(2^j x - k)\}_{k\in\mathbb{Z}}$ are orthonormal bases of respectively V_j and W_j. Assume that ϕ is supported by $[0, 1]$ and has a zero-moments property of order $N \in \mathbb{N}^*$, i.e. :

$$\forall k = 0 \ldots N, \quad \int \phi(x)x^k dx = \delta_{0,k}, \quad \int \left|\phi(x)x^{N+1}\right| dx < \infty, \tag{5.2}$$

$$\text{and} \qquad x \mapsto \sum_k |\phi(x-k)| \in \mathbb{L}^2, \tag{5.3}$$

(here $\delta_{0,k} = 1$ if $k = 0$ and else $\delta_{0,k} = 0$). The above assumptions imply that the associated mother wavelet ψ also satisfies

$$\forall k = 0 \dots N, \quad \int \psi(x) x^k dx = 0, \quad \int \left|\psi(x)x^{N+1}\right| dx < \infty.$$

Wavelet bases on $[0,1]$ proposed by Daubechies (1992) [24] are considered with a sufficient number ($N \geqslant 4$) of vanishing moments to be Lipschitz functions. Recall that a Lipschitz function $h : \mathbb{R}^u \to \mathbb{R}$ for some $u \in \mathbb{N}^*$ is a function such that $\mathrm{Lip}\,(h) < \infty$ with

$$\mathrm{Lip}\,(h) \quad = \quad \sup_{(a_1,\dots,a_u)\neq(b_1,\dots,b_u)} \frac{|h(a_1,\dots,a_u) - h(b_1,\dots,b_u)|}{|a_1 - b_1| + \cdots + |a_u - b_u|}.$$

Note that wavelets ϕ and ψ are bounded as Lipschitz functions supported by $[0,1]$.

For any fixed integer j_0, an arbitrary function $f \in \mathbb{L}^2$ can be decomposed as

$$f = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j_0,k}\psi_{j_0,k},$$

where $\alpha_{j,k} = \int_0^1 f(x)\phi_{j,k}(x)dx$, $\beta_{j,k} = \int_0^1 f(x)\psi_{j,k}(x)dx$. The projection-type estimator is

$$\tilde{f}_n = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{j_1(n)}\sum_{k=0}^{2^j-1} \hat{\beta}_{j,k}\psi_{j,k},$$

where $\hat{\alpha}_{j,k} = 1/n\sum_{i=1}^n \phi_{j,k}(X_i)$ and $\hat{\beta}_{j,k} = 1/n\sum_{i=1}^n \psi_{j,k}(X_i)$ are the empirical estimators of the coefficients $\alpha_{j,k}$ and $\beta_{j,k}$. Such density estimators can be write on the form : $\tilde{f}_n(x) = \sum_{i=1}^n K_n(x, X_i)$, with $(K_n)_{n\geqslant 1}$ sequence of kernels defined by $K_n(x,y) = \sum_{k=0}^{2^{j_0}-1} \phi_{j_0,k}(x)\phi_{j_0,k}(y) + \sum_{j=j_0}^{j_1(n)}\sum_{k=0}^{2^j-1} \psi_{j,k}(x)\psi_{j,k}(y)$. The kernels K_n are linear (they can be written as an integral with respect to the empirical distribution) and thus the estimator $\tilde{f}_n$ is linear.

Let (s,π,r) be a triplet such that $s > 0$, $1 \leq \pi, r \leq \infty$. The expression

$$\|f\|_{s,\pi,r} = |\alpha_{0,0}| + \left(\sum_{j\in\mathbb{N}} 2^{j(s+1/2-1/\pi)r}\left(\sum_{k=0}^{2^j-1}|\beta_{j,k}|^\pi\right)^{r/\pi}\right)^{1/r},$$

defines a semi-norm on $\mathbb{L}^2$ (see Härdle *et al.* (1998) [67] for more details). Besov spaces and Besov balls are respectively denoted

$$\mathcal{B}^s_{\pi,r} := \{f \in \mathbb{L}^2 \text{ such that } \|f\|_{s,\pi,r} < +\infty\}, \quad \mathcal{B}^s_{\pi,r}(M) := \left\{f \in \mathcal{B}^s_{\pi,r},\ \|f\|_{s,\pi,r} \leq M\right\}.$$

Note that Besov spaces do not depend on ϕ and ψ. For $f \in \mathcal{B}^s_{\pi,r}(M)$, the $\mathbb{L}^p$-mean error of an estimator f_n is defined by $\mathbb{E}\|f_n - f\|_p^p$. The associated minimax rate verify

$$\inf_{f \in \mathcal{B}^s_{\pi,r}(C)} \sup_{f_n \text{ estimator of } f} \mathbb{E}\|f_n - f\|_p^p = \mathcal{O}\left(n^{-\alpha}\right),$$

with

$$\alpha = \begin{cases} s/(1+2s) & \varepsilon \geqslant 0, \\ (s - 1/\pi + 1/p)/(1 + 2s - 2/\pi) & \varepsilon \leqslant 0, \end{cases} \qquad \text{where } \varepsilon = s\pi - (p - \pi)/2. \tag{5.4}$$

Linear estimators, including $\tilde{f}_n$, do not achieve such rates for $f \in \mathcal{B}^s_{\pi,r}(M)$ with $\pi \leqslant p$. In order to bypass this drawback, Donoho *et al.* (1996) introduce in [35] non-linear estimators $\hat{f}_n$ via thresholding. Let $T_\lambda(\beta) = \beta \mathbb{1}_{|\beta|>\lambda}$ be the hard-threshold function of level $\lambda > 0$; given n they consider integers j_0, j_1 and parameters $(\lambda_j)_{j=j_0 \dots j_1}$ and define

$$\hat{f}_n = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0,k}\phi_{j_0,k} + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} T_{\lambda_j}(\hat{\beta}_{j,k})\psi_{j,k},$$

In order to achieve a minimax rate –up to a logarithmic term– the paramters defining $\hat{f}_n$ have to be chosen appropriately. Donoho *et al.* (1996) obtain in the iid framework the following result :

Theorem 5.1 (Donoho *et al.*, 1996) *Suppose that f belongs to a Besov ball $\mathcal{B}^s_{\pi,r}(M)$ with*

$$1/\pi < s \leqslant N/2,\ 1 \leqslant \pi \leqslant p,\ 1 \leqslant r \leqslant \infty,$$

where N is the regularity of the wavelet. If $(X_t)_{t\in\mathbb{Z}}$ is an iid sequence, there exists a constant $C_0(N, p, s, \pi, M)$ such that

$$\mathbb{E}\|\hat{f}_n - f\|_p^p \leqslant C_0 \begin{cases} \left(\dfrac{\log n}{n}\right)^{p\alpha}, & \text{if } \varepsilon \neq 0 \\ \left(\dfrac{\log n}{n}\right)^{p\alpha} (\log n)^{(p/2-\pi/r)_+}, & \text{if } \varepsilon = 0, \end{cases}$$

where the minimax rate α and ε are given in (5.4) and

$$2^{j_0} \simeq n^{1/(1+N)}, \tag{5.5}$$

$$2^{j_1} \simeq n/\log n, \tag{5.6}$$

$$\lambda_j = K\sqrt{j/n}, \text{ for a suitable constant } K > 0. \tag{5.7}$$

In a regression framework, Donoho and Johnstone (1995) [36] fix the constant $K = \sqrt{2}$ for practical implementation. Hereafter, we will adopt the same choice.

5.2 Dependent models

Models of this section satisfy the conditions of the theorems stated in Section 5.3.

5.2.1 Expanding maps

Dynamical systems $(X_t)_{t\geqslant 0}$ are defined through a function $T:[0,1]\to[0,1]$ by

$$X_i = T^i(X_0), \qquad \forall i\in\mathbb{N}$$

where X_0 is distributed as a Lebesgue dominated measure μ on $[0,1]$ and T^i denotes $\underbrace{T\circ T\circ\cdots\circ T}_{i \text{ terms}}$.
Expanding maps $(X_t = T^t(X_0))_{t\in\mathbb{N}}$ (or equivalently Lasota-Yorke functions T) are dynamical systems such that

- *(Regularity)* The function T is differentiable, with a continuous derivate T' and there exists a grid $0=a_0\leqslant a_1\cdots\leqslant a_k=1$ such that $|T'(x)|>0$ on $]a_{i-1},a_i[$ for each $i=1,\dots,k$.
- *(Expansivity)* For any integer i, let I_i be the set on which the first derivate of T^i, $(T^i)'$, is defined. There exists $a>0$ and $s>1$ such that $\inf_{x\in I_i}\{|(T^i)'(x)|\}>as^i$.
- *(Topological mixing)* For any nonempty open sets U,V, there exists $i_0\geqslant 1$ such that $T^{-i}(U)\cap V\neq\varnothing$ for all $i\geqslant i_0$.

The class of dynamical systems has remarkable properties (see Viana, 1997, [110]). Its members admit an invariant measure μ_0 with a density $f\in BV$ where BV is the set of bounded variation functions h defined on $[0,1]$ such that $\|h\|_{BV}<+\infty$ where

$$\|h\|_{BV} = |h(0)| + \sup_{n\in\mathbb{N}}\ \sup_{a_0=0<a_1<\cdots<a_n=1}\sum_{i=1}^{n}|h(a_i)-h(a_{i-1})|.$$

Note that $\mathcal{B}^1_{1,1}\subset BV\subset\mathcal{B}^1_{1,\infty}$ (see e.g. Donoho *et al.* (1996), [35]). Expanding maps are also geometrically ergodic in mean, i.e. there exists constants $\alpha, C>0$ with $\alpha<1$ such that $\mathbb{E}|X_t-\tilde{X}_t|\leqslant C\alpha^t$ for all $t\geqslant 0$, where $(\tilde{X}_t)_{t\geqslant 0}$ is the stationary expanding map $(\tilde{X}_t = T^t(\tilde{X}_0))_{t\in\mathbb{N}}$ obtained with $\tilde{X}_0$ following a distribution μ_0. These models are not mixing (see Dedecker and Prieur, 2005, [29]).

5.2.2 η-weakly dependence

Doukhan and Louhchi have introduced this notion in 1999 in [41].

Definition 5.2 (Doukhan and Louhichi, 1999) *The stationary process $(X_t)_{t\in\mathbb{Z}}$ is η-weakly dependent if there exists a sequence of non-negative real numbers $(\eta(r))_{r\in\mathbb{N}}$ satisfying $\eta(r)\to 0$ when $r\to\infty$ and such that :*

$$\left|Cov\left(h\left(X_{i_1},\dots X_{i_u}\right), k\left(X_{i_{u+1}},\dots,X_{i_{u+v}}\right)\right)\right| \leq (u\,\mathrm{Lip}(h) + v\,\mathrm{Lip}(k))\,\eta(r)$$

for all $(u+v)$*-tuples,* $(i_1, \ldots, i_{u+v})$ *with* $i_1 \leq \cdots \leq i_u \leqslant i_u + r \leq i_{u+1} \leqslant \cdots \leqslant i_{u+v}$*, and for all* $h, k \in \Lambda^{(1)}$ *where*

$$\Lambda^{(1)} = \left\{ h : \exists u \geqslant 0,\ h : \mathbb{R}^u \to \mathbb{R},\ Lip\,(h) < \infty,\ \|h\|_\infty = \sup_{x \in \mathbb{R}^u} |h(x)| \leqslant 1 \right\}.$$

The η-dependence refers to non-causal situations because information "from the future" (i.e. on the right of the covariance) contributes as much as information "from the past" (i.e. on the left) in the dependence scheme. This notion of dependence includes general models which may be non-mixing. We will consider a subgeometric decay, meaning that :

$$\text{There exist } a, b, C > 0 \text{ such that } \eta(r) \leqslant C \exp(-ar^b), \tag{5.8}$$

and we will also assume that the joint densities $f_{j,k}$ of (X_j, X_k) exist and are uniformly bounded for $j \neq k$.

5.2.3 Bernoulli shifts

Let $H : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ be a measurable function. If the sequence $(\xi_t)_{t\in\mathbb{Z}}$ is iid on $\mathbb{R}$, a Bernoulli shift with input process $(\xi_t)_{t\in\mathbb{Z}}$ is defined as

$$X_t = H\left((\xi_{t-i})_{i\in\mathbb{Z}}\right), \qquad t \in \mathbb{Z}.$$

Such Bernoulli shifts are $\eta-$weakly dependent (see Doukhan and Louhichi, 1999, [41]) with $\eta(r) \leqslant 2\delta_{[r/2]}$ if

$$\mathbb{E}\left| H\left(\xi_j, j \in \mathbb{Z}\right) - H\left(\xi_j \mathbb{1}_{|j| \leqslant r}, j \in \mathbb{Z}\right)\right| \leqslant \delta_r.$$

Different values of b in equation (5.8) arise naturally for specific functions H.

Infinite moving average

The most simple case of infinitely dependent Bernoulli shift is the infinite moving average process

$$X_t = \sum_{i \in \mathbb{Z}} \alpha_i \xi_{t-i}.$$

Doukhan and Lang (2002) [39] prove they are η-weakly dependent with

$$\eta(r) = \sqrt{\sum_{|j| > [r/2]} a_j^2}. \tag{5.9}$$

If $(a_j)_{j \neq 0}$ satisfies $a_j \leqslant K\alpha^{|j|}$ for $j \neq 0$, $K > 0$ and $0 < \alpha < 1$ then equation (5.8) holds with $b = 1$.

LARCH(∞) inputs

A vast literature is devoted to the study of conditionally heteroscedastic models. A simple equation in terms of a vector valued process allows a unified treatment of those models, see [47]. Let $(\xi_t)_{t\in\mathbb{Z}}$ be an iid centered real valued sequence and $a, a_j, j \in \mathbb{N}^*$ be real numbers. LARCH(∞) models are solutions of the recurrence equation

$$X_t = \xi_t \left(a + \sum_{j=1}^{\infty} a_j X_{t-j} \right). \tag{5.10}$$

Assume that $\Lambda = \mathbb{E}|\xi_0| \sum_{j\geqslant 1} |a_j| < 1$ then one (essentially unique) stationary solution of eqn. (5.10) is given by

$$X_t = \xi_t \left(a + \sum_{k=1}^{\infty} \sum_{j_1,\ldots,j_k \geqslant 1} a_{j_1} \xi_{t-j_1} a_{j_2} \cdots a_{j_k} \xi_{t-j_1-\cdots-j_k} a \right). \tag{5.11}$$

The solution (5.11) of equation (5.10) is geometrically $\eta-$weakly dependent with $b = 1/2$ if there exists $0 < \alpha < 1$ and $K > 0$ such that $a_j \leqslant K\alpha^{j-1}$ for all $j > 0$ and $\mathbb{E}|\xi_0| < 1 - \alpha$.

Non-Causal LARCH(∞) inputs

The previous approach extends for the case of Non-Causal LARCH(∞) inputs

$$X_t = \xi_t \left(a + \sum_{j \neq 0} a_j X_{t-j} \right).$$

Doukhan, Teyssière and Winant (2005) prove in [47] the same results of existence of a stationary solution as for the previous causal case (only replace summation for $j > 0$ by summation for $j \neq 0$). This solution satisfies equation (5.8) with $b = 1/2$ if there exists $K, \alpha > 0$ and $\alpha < 1$ such that $a_j \leqslant K\alpha^{|j|}$ for all $j \neq 0$.

5.3 Main results

In all this section, let $(X_t)_{t\in\mathbb{Z}}$ be a stationary real valued sequence, ϕ a father wavelet in $\mathbb{L}^2$ satisfying (5.2) for $N \geqslant 4$ and the condition (5.3), and $\{\psi_{j,k}, j \in \mathbb{N}, k \in \mathbb{Z}\}$ the wavelet functions associated with. A probability inequality is given in an unified way for the different cases of dependence introduced in Section 5.2. Theorem 5.3 gives near-minimax estimators for weakly dependent observations.

5.3.1 Probability inequalities

Remark that $\psi_{j,k}$ is a Lipschitz function with Lipschitz constant $O\left(2^{3j/2}\right)$, its support is a compact set with diameter $O(2^{-j})$ and $\int \psi_{j,k}(x)\,dx = 2^{-j/2}$. We then rewrite it as $\psi_{j,k}(x) = K_m(0, x-k)/\sqrt{m}$, where $m = 2^j$ and $K_m(x, y)$ is a general kernel (see [100] for the definition of this notion). For η-weakly dependent sequences, Ragache and Wintenberger [100] prove, under condition (5.8) that there exists a constant $C > 0$ such that for all $(j, k) \in \mathbb{N}^2$, $n \in \mathbb{N}^*$ and every even integer q

$$\mathbb{E}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\psi_{j,k}(X_i) - \mathbb{E}\psi_{j,k}(X_i))\right|^q \leqslant \left(Cq^{1+1/b}\right)^q. \tag{5.12}$$

These bounds of q-th moments are linked to probability inequalities by the following useful lemma :

Lemma 5.1 (Ragache and Wintenberger, 2006) *If the variables $\{V_n\}_{n\in\mathbb{Z}}$ satisfy, for all $k \in \mathbb{N}^*$*

$$\|V_n\|_{2k} \leqslant \Phi(2k)$$

where Φ is an increasing function with $\Phi(0) = 0$ and $\lim_{x\to+\infty} \Phi(x) = +\infty$, then :

$$\mathbb{P}(|V_n| > \delta) \leq e^2 \exp\left(-\Phi^{-1}(\delta/e)\right).$$

Taking $V_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_{j,k}(X_i) - \mathbb{E}\psi_{j,k}(X_i)$ and $\Phi(x) = Kx^{1+1/b}$ leads directly to the first assertion of the following theorem

Theorem 5.2 *There exists constants $B, C > 0$ and $\gamma \geqslant 1/2$ such that for all $(j, k) \in \mathbb{N}^2$ and $n \in \mathbb{N}^*$*

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\psi_{j,k}(X_i) - \mathbb{E}\psi_{j,k}(X_i))\right| \geqslant \delta\right) \leqslant B\exp\left(-C\delta^{1/\gamma}\right), \tag{5.13}$$

for all $\delta \geqslant 0$ such that there exists $K' > 0$ with $\delta^{2l}\|\psi_{j,k}\|_\infty^2 \leqslant K'n$ for some constant $l \geqslant 0$. More precisely

1. *$\gamma = 1 + 1/b$ and $l = 0$, if coefficients $\eta(r)$ of $(X_t)_{t\in\mathbb{Z}}$ satisfy (5.8) and if for all $j \neq k$ the joints densities $f_{j,k}$ of (X_j, X_k) exist and are uniformly bounded.*
2. *$\gamma = 0.5$ and $l = 1$, if $(X_t)_{t\in\mathbb{Z}}$ is an iid process.*
3. *$\gamma = 0.5$ and $l = 5$, if $(X_t)_{t\in\mathbb{Z}}$ is a stationary expanding map.*

Assertions 2 and 3 of Theorem 5.2 follow directly from the following Bernstein's type inequality : for all $\lambda > 0$

$$P\left(\left|\sum_{i=1}^{n}\psi_{j,k}(X_i) - \mathbb{E}\psi_{j,k}(X_i)\right| \geqslant \lambda\right) \leqslant B\exp\left(-C\frac{\lambda^2}{\sigma_n^2 + \lambda^{\frac{2l}{l+1}}\|\psi_{j,k}\|_\infty^{\frac{2}{l+1}}}\right), \tag{5.14}$$

where $\sigma_n^2 = \operatorname{Var}\sum_{i=1}^n \psi_{j,k}(X_i)$ and $B, C > 0$ do not depend on j and k. From equation (5.15) in the case $q = 2$, we know that σ_n^2 has the rate n. Fixing $\delta = \lambda/\sqrt{n}$ and under $\delta^{2l}\|\psi_{j,k}\|_\infty^2 \leqslant K'n$, $\lambda^{2l/(l+1)}\|\psi_{j,k}\|_\infty^{2/(l+1)} \leqslant K''n$ with $K'' > 0$, the result of Theorem 5.2 with $\gamma = 0.5$ follows. Equation

(5.14) with $B = 2$ and $C = 1$ is the classical Bernstein's one in the iid case, see for instance Petrov (1995), [93]. The result is new for expanding maps and its proof is given in the last section.

The bounds of q-th moments of inequality (5.12) could be weakened as, for every $q > 0$, there exists a constant $C(q) > 0$ such that for all $(j, k) \in \mathbb{N}^2$, $n \in \mathbb{N}^*$

$$\mathbb{E}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\psi_{j,k}(X_i) - \mathbb{E}\psi_{j,k}(X_i))\right|^q \leqslant C(q). \tag{5.15}$$

This property also holds for dynamical systems as Lasota-Yorke ones (see Ragache & Wintenberger, [100]).

5.3.2 Near-minimaxity of the estimation scheme

This result extends the one of Donoho *et al.* (1996), [35], to dependent settings.

Theorem 5.3 *Suppose that f belongs to a Besov ball $\mathcal{B}^s_{\pi,r}(M)$ with*

$$1/\pi < s \leqslant N/2,\ 1 \leqslant \pi \leqslant p,\ 1 \leqslant r \leqslant \infty,$$

where $N \geqslant 4$ is the regularity of the wavelet. Suppose that there exists a couple (γ, l) and that conditions of dependence on $(X_t)_{t\in\mathbb{Z}}$ hold such that the inequalities (5.15) and (5.13) are satisfied. Then there exists a constant $C_0(N, p, s, \pi, M)$ such that

$$\mathbb{E}[\|\hat{f}_n - f\|_p^p] \leqslant C_0 \begin{cases} \left(\dfrac{\log^{2\gamma} n}{n}\right)^{p\alpha} & \textit{if } \varepsilon \neq 0 \\ \left(\dfrac{\log^{2\gamma} n}{n}\right)^{p\alpha} (\log n)^{(1-\pi/r)_+} & \textit{if } \varepsilon = 0 \end{cases}$$

where the minimax rate α and ε are given in (5.4) and

$$2^{j_0} \simeq n^{1/(1+N)}, \tag{5.16}$$

$$2^{j_1} \simeq n/\log^{2\gamma l} n, \tag{5.17}$$

$$\lambda_j = Kj^\gamma/\sqrt{n},\ \textit{for a well chosen constant } K > 0. \tag{5.18}$$

The above result takes into account the dependence of the observations through the threshold. The dependence of the observations presently changes the behavior of the moments of wavelet coefficients, determined by Theorem 5.2. The idea of adapting the threshold level to the behavior of the moments of the wavelet coefficients estimators is not new but it is usually developped in ill-posed inverse problems. As it is put in evidence in the Theorem 5.1 of Kerkyachrian and Picard (2000), [78], the quality of the estimator relies on an adapted threshold level, that is given by moments and probability inequalities on the estimators of the wavelet coefficients. For example, Johnstone *et al.* (2004) apply this principle in an ill-posed inverse problems in [75]. Here the idea is exactly the same, but the moment inequalities obtained relie on the dependence of the setting rather than on the inversion of the problem.

The result for $\gamma = 0.5$ is the same as in Donoho *et al.* (1996), [35]. As noted previously, such values of γ arise in iid context and for dynamical systems. Consequently, Donoho *et al.* (1996) result is extended to dynamical systems. It is also extended to expanding maps, where the assertion 1. of Theorem 2 holds, but the rate obtained differs from a logarithmic term.

The parameter γ determines the convergence rates through Theorem 5.3. It also calibrates the degree of dependence of the observations through Theorem 5.2. The optimal γ_0 corresponds to the smallest error of estimation. This value could change with the dependence of the observations. Theorem 5.2 gives us a theoretical value γ in inequality (5.13) (see section 5.3.1) that leads to near minimax estimators. Such probability inequality exhibits a bound $\gamma_0 \leqslant \gamma$ but not necessarily the optimal γ_0.

5.4 Numerical results

In the next section, we propose to estimate numerically the optimal parameter γ_0 by an estimator $\hat{\gamma}_n$ obtained by a cross-validation procedure.

5.4.1 Cross-validation procedure

According to Theorem 5.3, near-minimax estimators $\hat{f}_n^\gamma$ with hard-threshold of the form $\lambda_j = Kj^\gamma/\sqrt{n}$ (with $K, \gamma > 0$) are considered in the sequel. Let us fix $K = \sqrt{2}$, $2^{j_0} = n^{1/(1+N)}$ and $2^{j_1} = n/\log n$ as in Donoho and Johnstone (1995), [36], in order that $\hat{f}_n^\gamma$ depends only on γ. The MISE, corresponding to the $\mathbb{L}^p$-mean error for $p = 2$, is :

$$MISE(\hat{f}_n^\gamma) = \mathbb{E}\int \left(\hat{f}_n^\gamma(x) - f(x)\right)^2 dx = \mathbb{E}L(\hat{f}_n^\gamma).$$

Note that

$$L(\hat{f}_n^\gamma) = \underbrace{\int \left(\hat{f}_n^\gamma(x)\right)^2 dx - 2\int \hat{f}_n^\gamma(x) f(x) dx}_{J(\hat{f}_n^\gamma)} + \underbrace{\int f^2 dx}_{constant}.$$

Minimizing the MISE is then equivalent to minimize $\mathbb{E}J(\hat{f}_n^\gamma)$. Following Hart and Vieu (1990), [68], we define a leave-out procedure. Let b_n be positive integers and $\mathcal{X}_{-i}$, $i = 1 \dots n$, be the associated sub-sampling :

$$\mathcal{X}_{-i} = \{X_j,\ 1 \leqslant j \leqslant n,\ |i-j| \geqslant b_n\}.$$

We define the "leave-out b cross validation function" by :

$$CV_b(\gamma) = \int \left(\hat{f}_n^\gamma(x)\right)^2 dx - 2n^{-1} \sum_{i=1}^n \hat{f}_{-i}^\gamma(X_i),$$

where $\hat{f}_{-i}^\gamma$ is the hard threshold estimator based on the sub-sampling $\mathcal{X}_{-i}$.
The idea is to break the dependence by considering blocks of size $2 * b_n + 1 \to_{n\to\infty} \infty$ around the observation X_i in order to obtain $n^{-1}\sum_{i=1}^n \hat{f}_{-i}^\gamma(X_i) \sim_{n\to\infty} \mathbb{E}\hat{f}_{-i}^\gamma(X) \sim_{n\to\infty} \int \hat{f}_n^\gamma(x)f(x)dx$. Lahiri (2003) shows in [79] the efficiency of such sub-sampling methods in a weakly dependent context. Hall *et al.* (1995) prove in [64] that for kernel estimators the bandwith chosen by minimizing the corresponding "leave-out b cross validation function" is a good estimator of the optimal bandwith. Theoretical results are not developed here but we conjecture that such result remains true for two reasons :

- all cases considered here are (sub-)geometrically weakly dependent and restrictions on decays of coefficient ensure asymptotic results as law of large numbers,
- as shown by Theorem 5.3, our parameter γ plays a fundamental role on the convergence rates as much as the classical bandwith does.

According to remark 2.1 of Hart and Vieu (1990), [68], we fix $2 * b_n \sim n^{1/3}$. Note that the orthonormality of wavelets basis gives

$$CV_b(\gamma) = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0,k}^2 + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} T^2_{\sqrt{2j^\gamma/\sqrt{n}}}(\hat{\beta}_{j,k}) - 2n^{-1}\sum_{i=1}^n \hat{f}_{-i}^\gamma(X_i).$$

Our proposed procedure consists in choosing $\hat{\gamma}_n = \arg\min_{\gamma=0,0.05,\dots,1.95,2} CV_b(\gamma)$.

The implementation has been done in Matlab using the package Wavelab, heely available on the net on [112]. In order to implement numerically the procedure, we have used some approximations ; indeed, computation is based on an equispaced grid, while the data considered is not equispaced. In order to reduce the resulting approximation error, we consider a finre grid than the number n of observations. Here $I = 4n$ in order to increase the precision without araising too much calculus time. Actually, we will consider in implementation values of the form $\psi_{j,k}(l_i/I)$, wether than $\psi_{j,k}(X_i)$, with l_i the integer part of X_iI. The bias caused by implementation is not studied here.

5.4.2 The estimator $\hat{\gamma}_n$

This section illustrates and evaluates our estimation procedure on simulated examples.

Consider first the case of an iid sample $(X_1, \dots, X_n)$ generated according to the cumulative distribution function $F(x) = 2/\pi \arcsin(\sqrt{x})$ for $x \in [0,1]$. The criteria is minimized for $\gamma = 0.5$ in Figure

5.1. Thus, the cross validation procedure selects $\hat{\gamma}_n$ near 0.5 for $n = 2^{10}$ (see Figure 5.2). For a sufficiently large number of observations this estimator $\hat{\gamma}_n$ obtained numerically has value concording with $\gamma = 0.5$ chosen by Donoho *et al.* (1996) in [35].

Let us now consider the case of an expanding map $T(x) = 4x(1-x)$. The invariant measure has the same distribution F as in the previous iid case (see Prieur, 2001, [97]). We then first simulate $\tilde{X}_1, \ldots, \tilde{X}_n$ such that $\tilde{X}_1 \rightsquigarrow F$ and then $\tilde{X}_i = T^{i-1}(\tilde{X}_1)$ for $2 \leqslant i \leqslant n$. As for the iid case, the cross validation criteria in Figure 5.1(a) chooses $\hat{\gamma}_n = 0.5$. Figure 5.2 corroborates this value : for high values of n, the distribution of $\hat{\gamma}_n$ obtained by croos-validation is centered on the value 0.5.

The assumption of stationarity could be irrelevant in many cases. Simulating the first observation $\tilde{X}_1 \rightsquigarrow F$ to estimate the (unknown) distribution F of the invariant measure of a given transformation T is impossible. Thus we simulate $X_i = T^i(X_0)$ for $2 \leqslant i \leqslant 2n$ with $X_0 \rightsquigarrow \mathcal{U}([0,1])$ and retain only the n last terms. The process $(X_t)_{t\geqslant 1}$ is geometrically ergodic in mean and the error with respect to the stationary case is negligeable regarding with the error of the density estimation. Results remain valid and the procedure acts as in previous cases of stationary and iid cases, see Figures 5.1(b) and 5.2.

Study now a non-Markovian extension of Andrews' example (5.1)

$$X_t = 2(X_{t-1} + X_{t+1})/5 + 5\xi_t/21 \text{ where } (\xi_t)_{t\in\mathbb{Z}} \text{ are iid with the common law } Bern(1/2). \tag{5.19}$$

A stationary solution of this AR(2) equation is the non-causal process $(X_t)_{t\in\mathbb{Z}}$

$$X_t = \sum_{j\in\mathbb{Z}} a_j \xi_{t-j},$$

where $a_j = 1/3 * (1/2)^{|j|}$. This solution belongs to $[0,1]$ and its density is the one of $(U + U' + \xi_0)/3$ where U and U' are independent variables following $\mathcal{U}([0,1])$. As in Andrews' example such processes are not mixing. However, using equation (5.9) there exists $a, C > 0$ such that $\eta(r) \leqslant C\exp(-ar)$. Assumption (5.8) is satisfied with $b = 1$.

Perfect simulation is not available in this framework and Gibbs algorithm is appropriate for approching the true distribution of $(X_1, \ldots, X_n)$. A vectorial Markov Chain $(X_0^{(t)}, \ldots, X_{n+1}^{(t)})_{t\in\mathbb{N}^*}$ is simulated by :

$$\begin{aligned} \text{forall } j \in \{0 \ldots n+1\}, &\ \text{ the distribution of } X_j^{(0)} \text{ is uniform on } [0,1], \\ X_j^{(t)} &= 2(X_{j-1}^{(t)} + X_{j+1}^{(t-1)})/5 + 5 * \xi_t/21. \end{aligned}$$

This chain is uniformly geometrically ergodic and its invariant distribution is the true one of $(X_0, \ldots, X_{n+1}$
The dependence on the initial values is then geometrically decreasing with the number of iteration

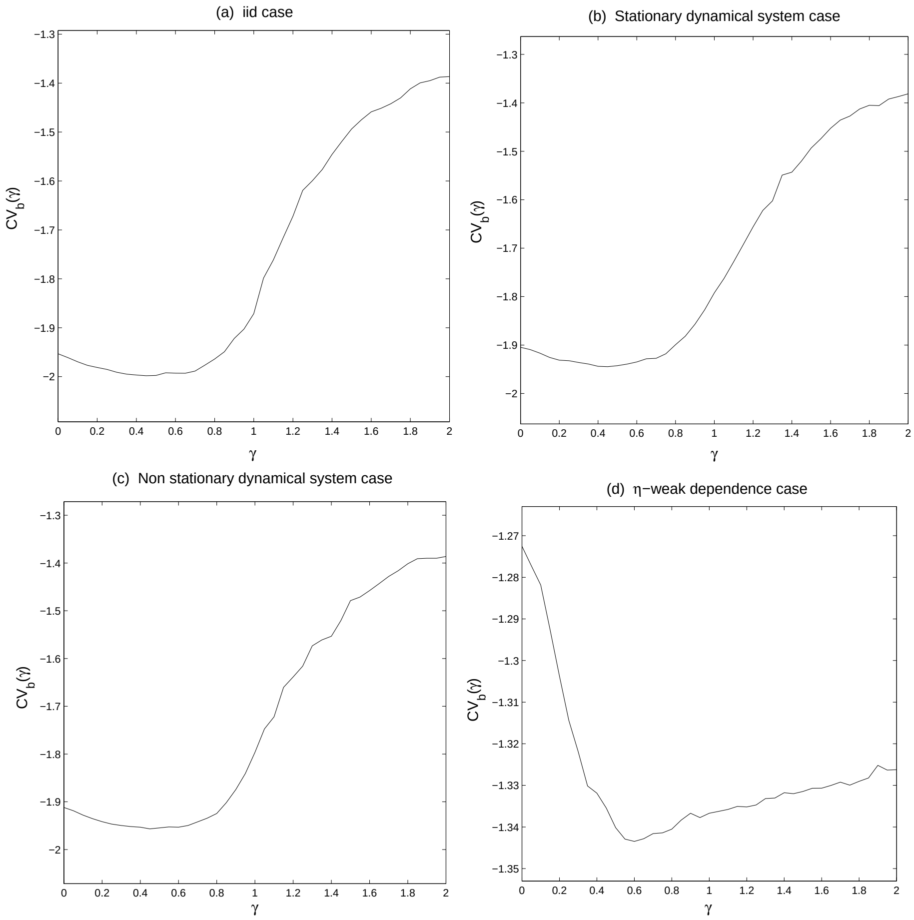

FIG. 5.1 – Cross validation criterion's evolution with respect to γ for $n = 2^{10}$ observations. The curves represent the evolution of the mean of the criteria calculated on 100 simulations with respect to $\gamma = 0, 0.05 \ldots 1.95, 2$. Four cases were considered : in (a) the observations are iid, in (b) we simulate stationary dynamical system, in (c) a non stationary dynamical system and in (d) we consider a η-weak dependent case.

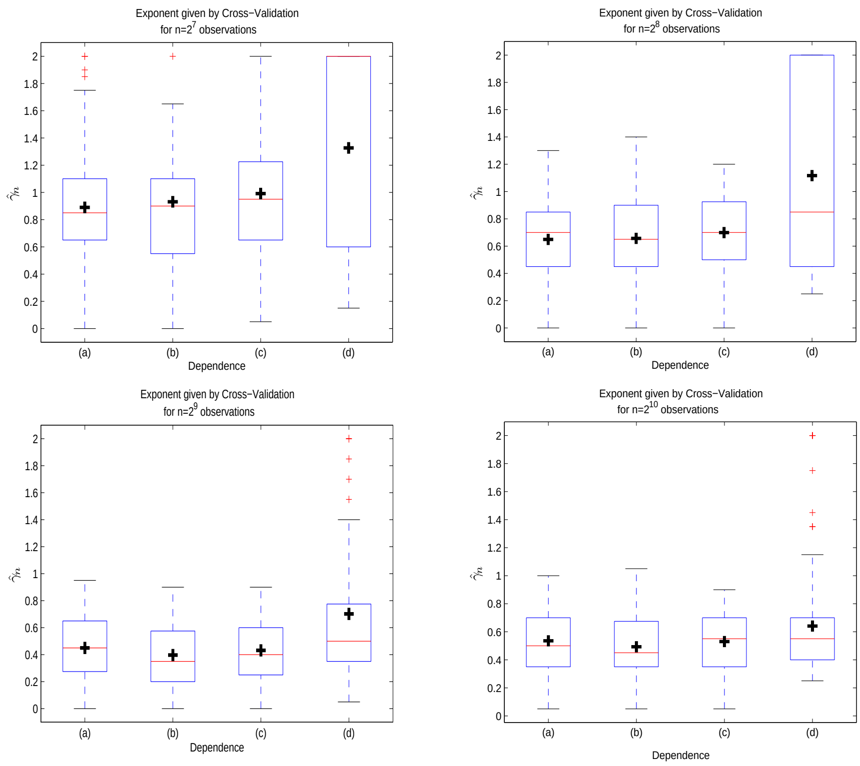

FIG. 5.2 – Empirical distribution of $\hat{\gamma}_n$. The figures represent the boxplots (the black cross is the mean) of the empirical ditribution of $\hat{\gamma}_n$ obtained on 100 simulations for different initial sample sizes : $n = 2^7,\ 2^8,\ 2^9,\ 2^{10}$. For each value of n, we abut the boxplots of differents cases ; From left to right : (a) iid case, (b) stationary dynamical system case, (c) non-stationary dynamical system case and (d) the η-weak dependent case.

of the Gibbs sampler (see Guyon, 1995, [63]). One chooses to take as observations $(X_1^{(n)}, \dots, X_n^{(n)})$ in order to reduce edge effects. The error coming from the simulation compared with the error of estimation is negligable. The shape of the criteria in Figure 5.1 is completely different in this non Markovian context. The cross-validation procedure leads to larger choices of $\hat{\gamma}_n$ than before (see Figure 5.2).

This numerical procedure has a practical interest especially when dealing with density estimation for time series. In practice, when the independence of the observations is not acquired, the cross-validation estimator $\hat{f}_n^{\hat{\gamma}_n}$ seems more adapted than the classical one with $\gamma = 0.5$.

5.4.3 Numerical study of the convergence rates

This section does not deal directly with our cross validation procedure. We study numerically the rates of convergence in different contexts by Monte Carlo simulations. The true densities of the simulated sequences are known. The $\mathbb{L}^2$-error rates is approximated by the Riemannian sum :

$$2^{-5} \sum_{i=0}^{2^5-1} \hat{f}_n^{\gamma}(2^{-6} + i2^{-5}) - f(2^{-6} + i2^{-5}).$$

With the parameter γ fixed at 0.5, we estimate the MISE on 100 iterations using this formula for the different cases of dependence. Figure 5.3 represents the evolution of the values of MISE obtained with respect to the sample-size n in the cases of iid series, stationary and non stationary expanding maps. We can observe that the evolution is the same for the differents cases considered in this figure. Added to the fact that the value of γ given by Theorem 2 is the same and that the evolutions of the criterion in Figure 2.(a) to (c) are very similar, compared with the η-weak dependent case, this means that in our example, these kinds of causal dependence do not modify the behaviour of the classical estimator from the iid case. In particular, we can note that the framework of non stationnary expanding maps is very similar to the iid case for our estimation procedure.

As already mentioned by Tribouley and Viennet (1998) in [107], a safe strategy to avoid large errors dealing with dependent data is to increase the threshold. A way to enlarge the threshold is to choose a larger γ. The special shape of the criteria shows that $\gamma = 0.5$ is a critical value and larger ones seems much more stable in the non causal context. In figure 5.4 are ploted estimations on 100 iterations of the MISE for three choices of $\gamma = 0.5, 0.75, 2$ and for various numbers of observations $n = 2^6, 2^7, \dots, 2^{14}, 2^{15}$. The MISE is clearly the smallest for $\gamma = 0.75$.

Theorically Theorem 5.2 tolds us inequality (5.13) holds with $\gamma = 2$ (see Section 5.3.1 for more details). Actually, this Theorem only gives an upper bound of γ veryfing inequality (5.13) and smaller

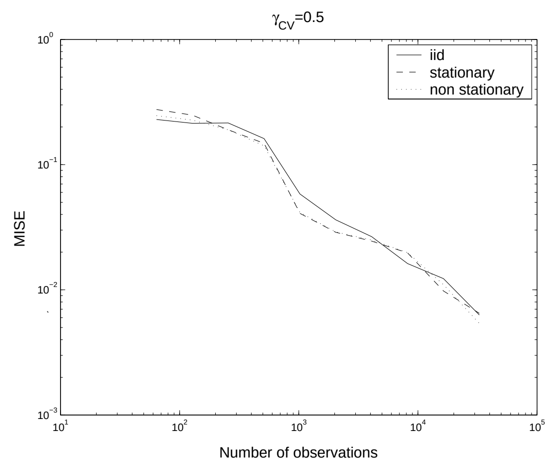

FIG. 5.3 – Evolution of the MISE with respect to the sample size. The figure represents the estimation of the MISE obtained by 100 simulations for $n = 2^6, 2^7, \dots, 2^{14}, 2^{15}$ in a log-log scale. Three cases were considered : iid observations (solid), stationary expanding maps (dash) and non stationary expanding maps (dots)

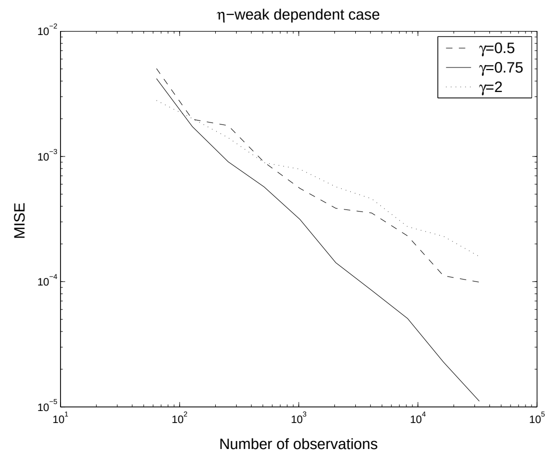

FIG. 5.4 – Evolution of the MISE with respect to the sample size for η-weak dependent data. The figure represents the estimation of the MISE obtained by 100 simulations for $n = 2^6, 2^7, \dots, 2^{14}, 2^{15}$ in a log-log scale. Three values of γ were considered : $\gamma = 0.5$ (dash), $\gamma = 0.75$ (solid) and $\gamma = 2$ (dots)

choices of γ can be possible. Choosing a lower γ such that inequality (5.13) holds allows to reduce the convergence rate via Theorem 5.3. But as we do not have a lower bound, we do not have access to the optimal choice of such γ. Figure 5.4 confirms that $\gamma = 2$ is probably not the optimal choice of γ in the studied example because we observe that the MISE of $\hat{f}_n^2$ is always greater than the one obtained with $\hat{f}_n^{0.75}$ even for large values of n. The fact $\hat{f}_n^{0.5}$ also have less satisfying results than $\hat{f}_n^{0.75}$ means a priori that inequality (5.13) is not available with $\gamma = 0.5$ in this context.

This study put in evidence the limitation of our theorical study : Theorem 5.2 does not gives the optimal choice of γ but only a possible choice. Using a cross-validation procedure allows us to bypass this drawback, giving the value of γ the most adapted to the observations and leading to a better estimation than an arbitrary choice of the threshold.

Conclusion

Probability inequalities give a γ that controls the logarithmic loss in the convergence rate. It leads to near minimax estimators $\hat{f}_n^{\gamma}$. This value of γ is not necessarly the optimal one, i.e. the one that minimizes the error of the estimator. The proposed estimator $\hat{f}_n^{\hat{\gamma}_n}$, where $\hat{\gamma}_n$ is obtained by a cross-validation procedure, seems a better density estimator when dealing with time series. For iid sequences or expanding maps the estimator $\hat{\gamma}_n$ converges to 0.5 according to Theorem 5.3. Even for non stationary expanding maps the previous result remains valid. In the η-weak dependence case, larger values $\gamma > 0.5$ are preferable, like the ones given by the cross validation procedure. Both theorical results and implementation on simulated examples tolds us that the behavior of the hard-threshold density estimator is similar for expanding maps (stationary or not) than for the iid case, while non causal η-weak dependence needs to calibrate differently the threshold. In our opinion, the cross-validation procedure presented gives a satisfactory unified approach of these different cases.

5.5 Proofs

In this section, proofs of Theorem 5.3 and inequality (5.14) are collected.

5.5.1 Proof of the Theorem 5.3

We restrict ourselves to the case of compactly supported distributions. We consider that f is defined on $[0,1]$ without loss of generality. We assume furthermore that the function f we wish to estimate belongs to a Besov Ball $\mathcal{B}^s_{\pi,r}(M)$ and that the assumptions on the indexes given in Theorem 5.3 hold. The density f can be written as follows :

$$f = \underbrace{\sum_{k=0}^{2^j_0-1} \alpha_{j_0,k}\phi_{j_0,k}}_{E_{j_0}f} + \underbrace{\sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}}_{D_{j_0,j_1}f} + \underbrace{\sum_{j=j_1+1}^{\infty}\sum_{k=0}^{2^j-1} \beta_{j,k}\psi_{j,k}}_{D_{j_1,\infty}f}.$$

This decomposition is linked with the properties of the Besov Balls : the first term $E_{j_0} f$ represents the projection of f over the space generated by scale functions of order j_0 whereas the part $D_{j_0,j_1} f + D_{j_1,\infty} f$ are the "details", it is to say the projection over the wavelet spaces. (See Härdle *et al.* (1998), [67]).

Let us recall below the form of the estimator :

$$\hat{f}_n = \underbrace{\sum_{k=0}^{2^j_0-1} \hat{\alpha}_{j_0,k}\phi_{j_0,k}}_{\hat{E}_{j_0} f} + \underbrace{\sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \gamma_{\lambda_j}(\hat{\beta}_{j,k})\psi_{j,k}}_{\hat{D}_{j_0,j_1} f},$$

where γ_{λ_j} is the hard-thresholding function with the level dependent threshold $\lambda_j = K\frac{j^\gamma}{\sqrt{n}}$. The functions j_0 and j_1 are given by $j_0 = \log_2(n^{1/(1+N)})$ and $j_1 = \log_2(n/\log^{2\gamma l} n)$.

Thanks to Minkowski's inequality, we can decompose the risk of $\hat{f}_n$ in three parts :

$$\mathbb{E}[\|\hat{f}_n - f\|_p^p] \leq 3^{p-1}\left(\underbrace{\mathbb{E}[\|\hat{E}_{j0} f - E_{j0} f\|_p^p]}_{T_1} + \underbrace{\mathbb{E}[\|\hat{D}_{j_0,j_1} f - D_{j_0,j_1} f\|_p^p]}_{T_2} + \underbrace{\|D_{j_1,\infty} f\|_p^p}_{T_3}\right).$$

We may now study the convergence rate of each of these terms. We will not consider them in their order of appearance but of difficulty.

5.5.2 Some technical tools

We suppose the series (X_i) satisfy the conditions (5.15) and (5.13) with γ a positive constant. Those assumptions provide us moments inequalities for the estimation of the scale and wavelet coefficients. We have, for all $j \in \mathbb{N}$ and $k \in \mathbb{Z}$:

$$\begin{cases} \hat{\alpha}_{j,k} - \alpha_{j,k} &= \dfrac{1}{n}\displaystyle\sum_{i=1}^n (\phi_{j,k}(X_i) - \mathbb{E}\phi_{j,k}(X_i)), \\ \hat{\beta}_{j,k} - \beta_{j,k} &= \dfrac{1}{n}\displaystyle\sum_{i=1}^n (\psi_{j,k}(X_i) - \mathbb{E}\psi_{j,k}(X_i)). \end{cases}$$

When we restrict ourselves to the cases where $j \leqslant j_1$, we can also control more generally the terms : $\mathbb{E}|\hat{\alpha}_{j,k} - \alpha_{j,k}|^p$ and $\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^p$ for a fixed real $p > 0$ and uniformly for all $j \leqslant j_1$ and $k \in \mathbb{Z}$. Then $\|\phi_{j,k}\|_\infty$ and $\|\psi_{j,k}\|_\infty$ are bounded, up to a constant, by $2^{j_1/2}$, which is always smaller than $\sqrt{n}$. Through the inequality (5.15) it leads there exists $C > 0$ such that

$$\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^p \leqslant Cn^{-p/2}. \tag{5.20}$$

The functions ϕ satisfy also the condition (5.2) and the same inequalities hold for each $\alpha_{j,k}$.

5.5.3 Approximation error T_3

This term is the bias introduced by the fact that in reality we do not estimate f but only its projection over a space of scale functions of order j_1. The fact that we observe dependent data does not affect this term because it is deterministic, so we can apply the usual bounds. As in the proof of Theorem 5 of Donoho et al. (1996), [35], or by applying Theorem 9.3 of Härdle *et al.* (1998), [67], there exists a constant $C > 0$ such that

$$T_3 \leqslant C \left(2^{-j_1 s'}\right)^p \text{ where } s' = s - 1/\pi + 1/p.$$

The index s' is coming from the Sobolev inclusion $\mathcal{B}^s_{\pi,r} \subset \mathcal{B}^{s'}_{p,r}$. Recall that we have taken $2^{-j_1} = \log^{2\gamma l} n/n$. Moreover, $p \geqslant \pi$ and $s > 1/\pi$ implies $s' > \alpha$:

- If $\varepsilon \leq 0$, $\alpha = s'/(1 + 2(s - 1/\pi))$. Then $\alpha < s'$ because $s > 1/\pi$ by hypothesis.
- If $\varepsilon \geq 0$, $\alpha - s' = s/(1 + 2s) - s' = -(2s^2 + (1 + 2s)(1/\pi - 1/p))/(1 + 2s)$. As $s > 0$, we conclude noting that $s' > \alpha$.

As a consequence T_3 has always a smaller rate than the convergence one, i.e. $T_3 \leqslant C \left(\log^{2\gamma l} n/n\right)^{s'p} \leqslant C' \left(\log^{2\gamma} n/n\right)^{\alpha p}$ for a well chosen constant $C' > 0$.

5.5.4 Bias of scale estimation T_1

Paralleling the proof of Theorem 5 of Donoho *et al.* (1996)„ [35], we can apply the result of Meyer (1992), [89], for the scaling function ϕ satisfying a concentration assumption. It gives, for a suitable $C > 0$

$$T_1 = \mathbb{E} \left\| \sum_{k=0}^{2^j_0-1} (\hat{\alpha}_{j_0,k} - \alpha_{j_0,k}) \phi_{j_0,k} \right\|_p^p \leqslant C\, 2^{j_0(p/2-1)} \sum_{k=0}^{2^{j_0}-1} \mathbb{E}|\hat{\alpha}_{j_0,k} - \alpha_{j_0,k}|^p.$$

Thanks to the inequality (5.20), the rate of convergence of T_1 is bounded by

$$T_1 \leqslant C\, 2^{j_0 p/2} n^{-p/2}.$$

Note that the choice (5.16) of j_0 implies that the order of the bound is $(2^{j_0}/n)^{p/2} = n^{-pN/(2+2N)}$. We conclude that T_1 is negligible thanks to the hypothesis $N \geq 2s$.

5.5.5 Details term T_2

We apply the same result from Meyer (1992), [89], on the wavelet basis $\{\psi_{j,k}\}$:

$$T_2 = \mathbb{E} \left\| \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} (\gamma_\lambda(\hat{\beta}_{j,k}) - \beta_{j,k}) \psi_{j,k} \right\|_p^p \sim \sum_{j=j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} \mathbb{E}|\gamma_\lambda(\hat{\beta}_{j,k}) - \beta_{j,k}|^p.$$

In order to prove that this term achieves the desired rate of convergence, we need to decompose it again. We will need in particular to distinguish whether the estimation of the coefficients are thresholded or not. We therefore introduce the following sets :

$$
\begin{aligned}
\hat{B}_j &:= \left\{k = 0 \ldots 2^j - 1, \quad |\hat{\beta}_{j,k}| > \lambda_j\right\}, \\
B_j^+ &:= \left\{k = 0 \ldots 2^j - 1, \quad |\beta_{j,k}| > 2\lambda_j\right\}, \\
B_j^- &:= \left\{k = 0 \ldots 2^j - 1, \quad |\beta_{j,k}| < \lambda_j/2\right\}.
\end{aligned}
$$

Then $T_2 \leq T_{21} + T_{22} + T_{23} + T_{24}$ with :

$$
\begin{aligned}
T_{21} &= \sum_{j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} \mathbb{E}\left[|\hat{\beta}_{j,k} - \beta_{j,k}|^p \mathbb{1}_{\hat{B}_j \cap B_j^-}\right], \\
T_{22} &= \sum_{j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} \mathbb{E}\left[|\hat{\beta}_{j,k} - \beta_{j,k}|^p \mathbb{1}_{\hat{B}_j \cap B_j^{-c}}\right], \\
T_{23} &= \sum_{j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{P}(\hat{B}_j^c \cap B_j^+), \\
T_{24} &= \sum_{j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{P}(\hat{B}_j^c \cap B_j^{+c}).
\end{aligned}
$$

where the exponent c denotes the complementary.

Term T_{23}

Within the set $\hat{B}_j \cap B_j^-$ or the set $\hat{B}_j^C \cap B_j^+$, we can notice that $|\hat{\beta}_{j,k} - \beta_{j,k}|$ is lower bounded by $\lambda_j/2$. We then have for a constant $C > 0$:

$$
T_{23} \leqslant C \sum_{j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{P}(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j/2).
$$

We can have an upper bound of $\mathbb{P}(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j/2)$ thanks to Theorem 5.2, choosing $\delta = \sqrt{n}\lambda_j/2 = Kj^\gamma/2$. With $2^{j_1} = n(\log n)^{-2\gamma}$, for a sufficiently large n :

$$
\delta^{2l} \|\psi_{j,k}\|_\infty^2 \leqslant K' j^{\gamma 2l} 2^{j_1} \leqslant K' \log^{\gamma 2l} n * n / \log^{2\gamma l} n \leqslant K' n.
$$

We thus obtain that $\delta^{2l} \|\psi_{j,k}\|_\infty^2 \leqslant K' n$ and consequently Theorem 5.2 can be applied with this choice of δ. It leads to the following bound

$$
\mathbb{P}(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j/2) \leqslant e^{-CK^{1/\gamma} j}. \tag{5.21}
$$

Thus, there exists an increasing function of K denoted $\nu(K)$ such that

$$\mathbb{P}(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j/2) \leqslant 2^{-j\nu}.$$

Added to the fact that we control $\sum_k |\beta_{j,k}|^p 2^{jp(s'+1/2-1/p)}$ since the assumptions in Theorem 5.3 imply that f belongs to a Besov space $\mathcal{B}^{s'}_{p,\infty}$ with $s' = s - 1/\pi + 1/p$, we obtain for $C > 0$,

$$T_{23} \leqslant C \sum_{j_0}^{j_1} 2^{-j(s'p+\nu)}.$$

It follows that T_{23} has the rate $2^{-j_0(s'p+\nu)} \leqslant n^{-(s'p+\nu)/(1+N)}$.
Taking K sufficiently large, we can control as accurately as we want the parameter ν and then T_{23} becomes asymptotically negligible compared to the other terms.

Term T_{21}

We first introduce $\mathbb{P}(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j/2)$ as we did it for the term T_{23}. By the Cauchy-Shwarz inequality there exists $C > 0$

$$T_{21} \leqslant C \sum_{j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} \left[\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^{2p}\right]^{1/2} \mathbb{P}(|\hat{\beta}_{j,k} - \beta_{j,k}| > \lambda_j/2)^{1/2}.$$

We achieve the following rate for T_{21} using (5.21)

$$\sum_{j_0}^{j_1} 2^{j(p/2-1-\nu/2)} \sum_{k=0}^{2^j-1} \left[\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^{2p}\right]^{1/2}.$$

Inequality (5.20) leads to the rate $n^{-p/2} \sum_{j_0}^{j_1} 2^{jp/2-\nu/2}$. As for T_{23}, we may choose the constant K large enough to get $\nu \geq p$. Then, we can write

$$T_{21} \leqslant C\, n^{-p/2} 2^{j_1(p-\nu)/2}.$$

Here again, we can choose the constant K sufficiently large, in order to achieve a sufficiently large ν such that T_{21} becomes negligible compared to the other terms.

Term T_{24}

This term corresponds to the leading one, meaning it is the one which determines the convergence rate. As

$$\mathbb{1}_{\hat{B}_j^C \cap B_j^{+C}} \leqslant \mathbb{1}_{|\beta_{j,k}| \leq 2\lambda_j},$$

there exists $C > 0$ such that

$$T_{24} \leqslant C \sum_{j=j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{1}_{\{|\beta_{j,k}| \leqslant 2\lambda_j\}}. \tag{5.22}$$

We now must distinguish on the values of ε.

– If $\varepsilon > 0$, we introduce j_{0+} as the largest integer such that

$$2^{j_{0+}} \leqslant \left(\frac{n}{\log^{2\gamma} n}\right)^{\frac{1}{1+2s}}. \tag{5.23}$$

The hypothesis $s \leq N/2$ implies $j_{0+} \geq j_0$. We decompose the inequality (5.22) as follows

$$T_{24} \leq \underbrace{\sum_{j=j_0}^{j_{0+}} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{1}_{\{|\beta_{j,k}| \leqslant 2\lambda_j\}}}_{T_{241+}} + \underbrace{\sum_{j=j_{0+}}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{1}_{\{|\beta_{j,k}| \leqslant 2\lambda_j\}}}_{T_{242+}}.$$

We study separately the behaviours of T_{241+} and T_{242+}.

Term T_{241+}**.** This term satisfy the following inequality $T_{241+} \leq \sum_{j=j_0}^{j_{0+}} 2^{j(p/2-1)} 2^j (2\lambda_j)^p$. With λ_j given by (5.18), we have $T_{241+} \leq n^{-p/2} 2^{j_{0+}p/2} (j_{0+})^{\gamma p}$. The choice of j_{0+} in (5.23) gives $T_{241+} \leq C(\log^{2\gamma} n/n)^{\alpha p}$, with C positive constant.

Term T_{242+}**.** As $\pi - p < 0$, we have the inequality

$$\mathbb{1}_{|\beta_{j,k}| \leq 2\lambda_j} \leqslant \left|\frac{\beta_{j,k}}{2\lambda_j}\right|^{(\pi-p)}. \tag{5.24}$$

We thus can easily bound T_{242+} by

$$\sum_{j=j_{0+}}^{j_1} 2^{j(p/2-1)} \lambda_j^{p-\pi} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi. \tag{5.25}$$

Replacing λ_j by its value,

$$T_{242+} \leq C \left(\frac{j_1^{2\gamma}}{n}\right)^{(p-\pi)/2} \sum_{j=j_{0+}}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi. \tag{5.26}$$

Using the control of $\sum_k |\beta_{j,k}|^\pi 2^{j\pi(s+1/2-1/\pi)}$ given by the inclusion $\mathcal{B}^s_{\pi,r} \subset \mathcal{B}^s_{\pi\infty}$

$$T_{242+} \leqslant C \sum_{j=j_{0+}}^{j_1} 2^{jp/2} \lambda_j^{p-\pi} 2^{-j(s+1/2)\pi} \|f\|^\pi_{s,\pi,\infty} \leqslant C \left(\frac{j_1^{2\gamma}}{n}\right)^{(p-\pi)/2} \sum_{j=j_0}^{j_1} 2^{-j\varepsilon}. \tag{5.27}$$

Let us define $v_n^+ := n^{-(p-\pi)/2}\sum_{j=j_{0+}}^{j_1} 2^{-j\varepsilon}$.
As $\varepsilon > 0$, v_n^+ is bounded by $n^{-(p-\pi)/2}2^{-j_{0+}\varepsilon}$. So, it gives with j_{0+} as in (5.23) : $v_n^+ = (\log n)^{2\gamma\varepsilon/(1+2s)}n^{-\varepsilon/(1+2s)-(p-\pi)/2}$. The equality $\varepsilon/(1+2s)+(p-\pi)/2 = \alpha p$ if $\varepsilon \geqslant 0$ implies obtain that v_n^+ is equal to $(\log^{-2\gamma}(n)n)^{-\alpha p}(\log^{\gamma(p-\pi)}(n)$.
Finally, we have

$$T_{242+} \leq C\,(\log^{2\gamma} n/n)^{\alpha p},$$

which is the rate we are looking for.

To conclude, when $\varepsilon > 0$, we obtain

$$T_{24} \leq C\,(\log^{2\gamma} n/n)^{\alpha p}.$$

– If $\varepsilon < 0$, we introduce j_{1-} as the largest integer satisfying

$$2^{j_{1-}} \leqslant \left(\frac{n}{\log^{2\gamma} n}\right)^{\frac{\alpha}{s'}}. \tag{5.28}$$

When $\varepsilon < 0$, $\alpha < s'$ and so $j_{1-} \leq j_1$ for a sufficient n. We decompose the inequality (5.22) as follows

$$T_{24} \leq \underbrace{\sum_{j=j_0}^{j_{1-}} 2^{j(p/2-1)}\sum_{k=0}^{2^j-1}|\beta_{j,k}|^p\mathbb{1}_{\{|\beta_{j,k}|\leqslant 2\lambda_j\}}}_{T_{241-}} + \underbrace{\sum_{j=j_{1-}}^{j_1} 2^{j(p/2-1)}\sum_{k=0}^{2^j-1}|\beta_{j,k}|^p\mathbb{1}_{\{|\beta_{j,k}|\leqslant 2\lambda_j\}}}_{T_{242-}}.$$

We consider separately T_{241-} and T_{242-}.

Term T_{241-}. Following the scheme used for T_{242+}, the inequality (5.27) becomes

$$T_{241-} \leqslant C\sum_{j=j_0}^{j_{1-}} 2^{jp/2}\lambda_j^{p-\pi}2^{-j(s+1/2)\pi}\|f\|_{s,\pi,\infty}^{\pi} \leqslant C\left(\frac{j_1^{2\gamma}}{n}\right)^{(p-\pi)/2}\sum_{j=j_0}^{j_{1-}} 2^{-j\varepsilon}. \tag{5.29}$$

Let us define $v_n^- = n^{-(p-\pi)/2}\sum_{j=j_0}^{j_{1-}} 2^{-j\varepsilon}$.
As $\varepsilon < 0$, v_n^- is bounded by $n^{-(p-\pi)/2}2^{-j_{1-}\varepsilon}$. Thanks to the choice of j_{1-} this bound is equal to $(\log^{2\gamma l} n)^{\varepsilon}n^{-\varepsilon\alpha/s'-(p-\pi)/2}$. Like above, we can notice that if $\varepsilon \leqslant 0$, we have the equality $\varepsilon\alpha/s'+(p-\pi)/2 = \alpha p$. Then $v_n^- = (\log(n)^{2\gamma})^{\varepsilon\alpha/s'}n^{-\alpha p} = (\log(n)^{2\gamma})^{\alpha p-(p-\pi)/2}n^{-\alpha p}$. Together with (5.29), we obtain

$$T_{241-} \leq C\,(\log^{2\gamma} n/n)^{\alpha p}.$$

Term T_{242-}. If h is defined as

$$h = \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1}\left(\beta_{j,k}\mathbb{1}_{|\beta_{jk}|\leq 2\lambda_j}\right)\psi_{j,k},$$

then h is a fonction belonging to the Besov ball $\mathcal{B}^s_{\pi,r}$. Then we can write T_{242-} as $T_{242-} = \mathbb{E}\left[\left\|h-\hat{h}\right\|_p^p\right]$ with

$$\hat{h} = \sum_{j=j_0}^{j_{1-}} \sum_{k=0}^{2^j-1} \left(\beta_{j,k}\mathbb{1}_{|\beta_{jk}|\leq 2\lambda_j}\right)\psi_{j,k}.$$

This remark allows us to apply the same bound as for the term T_3 :

$$T_{242-} \leq C\,(2^{-j_{1-}s'})^p.$$

The choice of j_{1-} in (5.28) gives the exact bound $(\log^{2\gamma} n/n)^{\alpha p}$.

Combining the bounds of T_{241-} and T_{242-}, we have :

$$T_{24} \leq (\log^{2\gamma} n/n)^{\alpha p},$$

when $\varepsilon < 0$.

– We finally consider the case $\varepsilon = 0$. First, if $\varepsilon = 0$, inequality (5.26) becomes

$$T_{24} \leqslant C\left(\frac{j_1^{2\gamma}}{n}\right)^{(p-\pi)/2} \sum_{j=j_0}^{j_1} 2^{j(s+1/2-1/\pi)\pi} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi.$$

- – Noting that $\mathcal{B}^s_{\pi,r}$ is included in $\mathcal{B}^s_{\pi,r'}$ for all $r' \geq r$, we first can control the term T_{24} by $\|f\|^\pi_{s,\pi,\pi}$ if $\pi \geq r$.
- – When $\pi < r$, then we can apply Hölder inequality to obtain $\|f\|_{s,\pi,r}$. Let for every integer j

$$t_j = 2^{j(s+1/2-1/\pi)\pi} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi.$$

With this notation and using Hölder inequality :

$$T_{24} \leqslant C\left(\frac{j_1^{2\gamma}}{n}\right)^{(p-\pi)/2} \sum_{j=j_0}^{j_1} t_j^\pi \leqslant C\left(\frac{j_1^{2\gamma}}{n}\right)^{(p-\pi)/2} \left(\sum_{j=j_0}^{j_1} t_j^r\right)^{\pi/r} \left(\sum_{j=j_0}^{j_1} t_j^{r'}\right)^{\pi/r'},$$

with $1/r + 1/r' = 1/\pi$. As f belongs to $\mathcal{B}^s_{\pi,r}$ which is included in $\mathcal{B}^s_{\pi,\infty}$, we have :

$$T_{24} \leqslant C\left(\frac{j_1^{2\gamma}}{n}\right)^{(p-\pi)/2} \|f\|^\pi_{s,\pi,r}\left((j_1-j_0)\|f\|^{r'}_{s,\pi,\infty}\right)^{\pi/r'} \leqslant C\left(\frac{j_1^{2\gamma}}{n}\right)^{(p-\pi)/2} j_1^{(1-\pi/r)}.$$

To conclude, for a suitable constant $C > 0$

$$T_{24} \leqslant C\left(\frac{\log^{-2\gamma} n}{n}\right)^{-\alpha p} \begin{cases} 1 & \text{if } \varepsilon \neq 0, \\ (\log n)^{(1-\pi/r)_+} & \text{if } \varepsilon = 0. \end{cases} \tag{5.30}$$

Actually, this rate is the one given in Theorem 5.3.

Term T_{22}

The scheme of the proof of the convergence of this term is very similar to the term T_{24}.
– If $\varepsilon > 0$, we introduce j_{0+} like in (5.23) and decompose T_{22} as follows :

$$T_{22} = \underbrace{\sum_{j=j_0}^{j_{0+}} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{1}_{\{|\beta_{j,k}|\geqslant 2\lambda_j\}}}_{T_{221+}} + \underbrace{\sum_{j=j_{0+}}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{1}_{\{|\beta_{j,k}|\geqslant 2\lambda_j\}}}_{T_{222+}}.$$

Term T_{221+}. As told in Subsection 5.5.2, we bound $\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^p$ by $n^{-p/2}$ uniformly in j and k. We can deduce

$$T_{221+} \leq Cn^{-p/2} \sum_{j_0}^{j_{0+}} 2^{jp/2} \leq Cn^{-p/2} 2^{j_{0+}p/2}.$$

Replacing j_{0+} by its value, we see that this rate is smaller than $(\log^{2\gamma} n/n)^{\alpha p}$.

Term T_{222+}. Using the same method as in (5.24), we note that

$$\mathbb{1}_{\{|\beta_{j,k}|\geqslant \lambda_j/2\}} \leqslant \left|\frac{2\beta_{j,k}}{\lambda_j}\right|^\pi. \tag{5.31}$$

Recall the inequality : $\mathbb{E}|\hat{\beta}_{j,k} - \beta_{j,k}|^p \leq n^{-p/2}$. There exists $C > 0$ such that

$$T_{222+} \leqslant C\, n^{-p/2} \sum_{j=j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} \left|\frac{\beta_{j,k}}{\lambda_j}\right|^\pi.$$

Replacing λ_j by its value the rate becomes

$$\frac{j_0^{-\gamma\pi}}{n^{(p-\pi)/2}} \sum_{j=j_{0+}}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi. \tag{5.32}$$

At this step, we recognize a bound of the same form than the one obtained in inequality (5.27) for T_{242+}. Actually, for $C > 0$

$$T_{222+} \leqslant C\, j_{0+}^{-\gamma\pi} \left(n^{-(p-\pi)/2} \sum_{j=j_{0+}}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi \right).$$

Considering the proof for the term T_{242+} this leads to

$$T_{222+} \leqslant C\, {j_{0+}}^{-\gamma\pi} (\log^{2\gamma} n/n)^{\alpha p},$$

which converges faster than $(\log^{2\gamma} n/n)^{\alpha p}$.

– When $\varepsilon < 0$, we decompose as follows :

$$T_{24} \leq \underbrace{\sum_{j=j_0}^{j_{1-}} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{1}_{\{|\beta_{j,k}|\geqslant 2\lambda_j\}}}_{T_{241-}} + \underbrace{\sum_{j=j_{1-}}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \mathbb{1}_{\{|\beta_{j,k}|\geqslant 2\lambda_j\}}}_{T_{242-}},$$

with j_{1-} defined in (5.28). We then consider the terms separately.

Term T_{221-}. Exactly like for T_{222+}, we can prove the inequality

$$T_{222+} \leqslant C\, j_0^{-\gamma\pi} \left(n^{-(p-\pi)/2} \sum_{j=j_0}^{j_{1-}} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi \right),$$

Applying the same developments than for T_{241-} we then obtain $T_{221-} \leq j_0^{-\gamma\pi}(\log^{2\gamma} n/n)^{\alpha p}$ which proves that T_{221-} converges to 0 with a better rate than the one wanted.

Term T_{222-}. Exactly like for the term T_{242-}, this term can be seen as the L^p-risk of an estimator and it can be bounded like T_3. We obtain the same rate than for T_{242-}.

– If $\varepsilon = 0$, inequality (5.31) leads to an inequality of the form (5.32)

$$T_{22} \leq j_0^{-\gamma\pi} n^{-(p-\pi)/2} \sum_{j_0}^{j_1} 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\beta_{j,k}|^p.$$

When $\varepsilon = 0$, we have $p/2-1 = s+1/2-1/\pi$. Consequently, using the Sobolev inclusion $\mathcal{B}^s_{\pi,r} \subset \mathcal{B}^s_{\pi,\infty}$, we have :

$$T_{22} \leq j_1 j_0^{-\gamma\pi} n^{-(p-\pi)/2} \|f\|_{s,\pi,\infty}.$$

We can notice $(p-\pi)/2 = \alpha p$ because of the condition $\varepsilon = 0$ and thus

$$T_{22} \leq C\, \log(n)^{2\gamma\alpha p - (\gamma(\pi+2\alpha p)-1)} n^{-\alpha p}.$$

It is then sufficient to prove $(\gamma(\pi-2\alpha p)-1) \geq 0$. Replacing αp by $(p-\pi)/2$, we have $\gamma(\pi-2\alpha p)-1 = \gamma p - 1$. As $p/2-1 = s+1/2-1/\pi > 0$, we have $p > 2$, and γ is always greater or equal to $1/2$. Consequently, $\gamma p - 1$ is always positive, which achieves the proof. □

5.5.6 Proof of inequality (5.14)

We only consider the case of expanding maps and $l = 5$. The $\tilde{\phi}$-dependence was introduced by Dedecker and Prieur (2005) in [29]. It seems well-adapted to study Lasota-Yorke functions. Let denote $BV_1 = \{f \in BV;\ \|f\|_{BV} \leqslant 1\}$.

Definition 5.3 (Dedecker and Prieur (2005)) *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space and* $\mathcal{M}$ *a* σ*-algebra of* $\mathcal{A}$*. For any random variable* $X \in \mathbb{R}^d$ *we define :*

$$\tilde{\phi}(\mathcal{M}, X) = \sup\left\{\|\mathbb{E}(g(X)|\mathcal{M}) - \mathbb{E}(g(X))\|_\infty,\ g \in BV_1\right\},$$

where, in our case, the coefficients $(\tilde{\phi}(r))_{r\in\mathbb{N}}$ *are defined by*

$$\tilde{\phi}(r) = \sup_{i+r \leqslant s}\left\{\tilde{\phi}(\sigma(\{X_j/j \geqslant i\}), X_s)\right\}.$$

The process is $\tilde{\phi}$*-dependent if* $\tilde{\phi}(r)$ *tends to* 0 *as* r *tends to infinity.*

Dedecker and Prieur (2005) prove in [29] that expanding maps are geometrically $\tilde{\phi}$-weakly dependent, with $\tilde{\phi}(r) = 2\exp(-ar))$ with $a > 0$. Moreover, Dedecker *et al.* give an extended version of that inequality in [27] for expanding maps; there exist two constants $C, C' > 0$ such that for all $f_{s_1}, \dots, f_{s_u} \in BV_1$

$$|\mathbb{E}(f_{s_1}(X_{s_1})\cdots f_{s_u}(X_{s_u})|\sigma(X_t, t \geqslant r+s_u)) - \mathbb{E}(f_{s_1}(X_{s_1})\cdots f_{s_u}(X_{s_u}))| \leqslant C'\left(1 + C + \cdots + C^{u-1}\right)\tilde{\phi}(r). \tag{5.33}$$

The probability inequality of theorem 4.21 is a version of a Bernstein's inequality in Doukhan and Neumann, [44], for weakly dependent random variables.

Theorem 5.4 (Doukhan and Neumann) *Suppose that* $Y_1, \dots, Y_n$ *are real-valued random variables with* $E[Y_i] = 0$ *and* $P(|Y_i| \le M) = 1$ *, for all* $i = 1, \dots, n$ *and some* $M < \infty$*. We assume that there exist constants* $K < \infty$*,* $\mu \ge 0$ *and a non increasing sequence of real coefficients* $(\rho_n)_{n\in\mathbb{N}_0}$ *such that, for all u-tuples* $(s_1, \dots, s_u)$ *and all v-tuples* $(t_1, \dots, t_v)$ *with* $1 \le s_1 \le \cdots \le s_u \le t_1 \le \cdots \le t_v \le n$ *the following inequality is fulfilled :*

$$|Cov(Y_{s_1}\cdots Y_{s_u}, Y_{t_1}\cdots Y_{t_v})| \le u\, K^2\, M^{u+v-2}\, \rho_{t_1 - s_u}, \tag{5.34}$$

and

$$\sum_{s=0}^{\infty}(s+1)^{k-2}\rho_s \le (k!)^\mu. \tag{5.35}$$

Then

$$P\left(\sum_{i=1}^n Y_i \ge \lambda\right) \le \exp\left(-\frac{\lambda^2/2}{\sigma_n^2 + B_n^{1/(\mu+2)}\lambda^{(2\mu+3)/(\mu+2)}}\right),$$

where $B_n = 2\,(K \vee M)\left(\frac{2^{4+\mu}\, n\, K^2}{\sigma_n^2} \vee 1\right)$ *and* $\sigma_n^2 = Var(\sum_{i=1}^n Y_i)$*.*

We apply this inequality to $Y_i = \psi_{j,k}(X_i) - \mathbb{E}\psi_{j,k}(X_i)$ for all $i \in \mathbb{Z}$, with $j \leqslant j_1$ and $k \in \mathbb{Z}$. Remark that $\psi_{j,k}$ are Lipschitzian functions with Lipschitz constant smaller than $2^{j3/2}\mathrm{Lip}\,\psi$. Reminding that $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ is null outside $[2^{-j}k; 2^{-j}(k+1)]$ for all $0 \leqslant k \leqslant 2^j - 1$, we verify that

$Y_i = f_i(X_i)$ with $\|f_i\|_{BV} \leqslant |\mathbb{E}\psi_{j,k}(X_0)| + 2^{j3/2}\mathrm{Lip}\,\psi\, 2^{-j}$ for all $i \in \mathbb{Z}$. It remains to control the order of $|\mathbb{E}\psi_{j,k}(X_0)| \leqslant 2^{j/2}\int |\psi(2^j x - k)|f(x)dx$. Classically, posing $u = 2^j x - k$, using the fact that f is uniformly bounded and that ψ is integrable, we obtain $\mathbb{E}\,|\psi_{j,k}(X_0)| \leqslant C'' 2^{-j/2}$ with C'' depending on $\int |\psi|$ and $\|f\|_\infty$. Finally, for all $i \in \mathbb{Z}$, we have $\|f_i\|_{BV} \leqslant c2^{j/2}\mathrm{Lip}\,\psi$ for some $c > 0$. Then, following Dedecker and Prieur (2005), [29], we can rewrite the covariance term 5.34 in order to use the notion of the $\tilde{\phi}$-dependence :

$$\begin{aligned} |\mathrm{Cov}\,(Y_{s_1}\cdots Y_{s_u}, Y_{t_1}\cdots Y_{t_v})| &\leqslant \mathbb{E}|(\mathbb{E}(Y_{s_1}\cdots Y_{s_u}|Y_{t_1}\cdots Y_{t_v}) - \mathbb{E}(Y_{s_1}\cdots Y_{s_u}))Y_{t_1}\cdots Y_{t_v}|, \\ &\leqslant \|\mathbb{E}(Y_{s_1}\cdots Y_{s_u}|Y_{t_1}) - \mathbb{E}(Y_{s_1}\cdots Y_{s_u})\|_\infty \mathbb{E}|Y_{t_1}\cdots Y_{t_v}|, \end{aligned}$$

Using inequality 5.33 and the $\tilde{\phi}$-dependence properties of expanding maps,

$$\begin{aligned} &\|\mathbb{E}(Y_{s_1}\cdots Y_{s_u}|Y_{t_1}) - \mathbb{E}(Y_{s_1}\cdots Y_{s_u})\|_\infty \\ &\qquad \leqslant \|f_0\|_{BV}^u\, C' \left(1 + C + \cdots + C^{u-1}\right) \tilde{\phi}(t_1 - s_u), \\ &\qquad\qquad \leqslant c2^{ju/2}(\mathrm{Lip}\,\psi)^u C' u\,(C \vee 1)^{u-1}\, \tilde{\phi}(t_1 - s_u), \end{aligned}$$

with some $c > 0$ depending on $\int |\psi|$ and $\|f\|_\infty$. To control the covariance term 5.34, let bound

$$\mathbb{E}|Y_{t_1}\cdots Y_{t_v}| \leqslant \mathbb{E}|Y_0| \cdot \|Y_0\|_\infty^{v-1} \leqslant C'' 2^{-j/2} 2^{j(v-1)/2} \|\psi\|_\infty^{v-1}$$

Collecting all the bounds, we finally achieve to control the covariance term by

$$|\mathrm{Cov}\,(Y_{s_1}\cdots Y_{s_u}, Y_{t_1}\cdots Y_{t_v})| \leqslant cu2^{j(u+v-2)/2}(\mathrm{Lip}\,\psi)^u \|\psi\|_\infty^{v-1}\,(C \vee 1)^{u-1}\, \tilde{\phi}(t_1 - s_u),$$

for some $c > 0$ depending on $\int |\psi|$ and $\|f\|_\infty$.

We apply Theorem 5.4 with $\rho = \tilde{\phi}$, $M = 2^{j/2}\mathrm{Lip}\,\psi\|\psi\|_\infty^u\,(C \vee 1)$, K^2 as a well chosen constant depending on $\mathrm{Lip}\,(\psi)$, $\int |\psi|$ and $\|f\|_\infty$. It remains to study (5.35) in order to determine μ. Previous results on expanding maps imply that $\tilde{\phi}_r \leqslant \exp(-ar)$ with $a > 0$. Quote that

$$\sum_{r=0}^{n-1}(r+1)^{q-2}\exp(-ar) \;\leqslant\; \int_0^\infty r^{q-2}\exp(-ar)dr.$$

Then, the change of variable $u = ar$ gives

$$\sum_{r=0}^{n-1}(r+1)^{q-2}\exp(-ar) \leqslant \frac{1}{a^{q-1}}\int_0^\infty u^{q-2}\exp(-u)du = \frac{1}{a^{q-1}}\Gamma\,(q-1) \leqslant \frac{C}{a^{q-1}}(q-1)!$$

The last inequality follows from Stirling's formula which entails that for any constant $A > 0$, and any $\varepsilon > 0$ there exists $B_\varepsilon > 0$ such that $A^k \leqslant B_\varepsilon k!^\varepsilon$. Thus, under this rate of dependence, assumptions of Theorem 5.4 hold with $\mu = 1$. It is easy to check that the order of B_n is the same as $2^{j/2} \propto \|\psi_{j,k}\|_\infty$ and Theorem 4.21 is proved. $\square$

Troisième partie

The Parametric Estimation

Chapitre 6

Asymptotic normality of the Quasi Maximum Likelihood Estimator for multidimensional causal processes

Abstract

The consistence and asymptotic normality of the Quasi Maximum Likelihood Estimator (QMLE) are given for a general class of multidimensional causal processes including for instance multivariate AR(∞) or ARCH(∞) processes. For particular cases of processes already studied in the literature (for instance univariate or multivariate GARCH, ARCH, ARMA-GARCH processes) the assumptions required for establishing this results are often weaker than existing conditions. New results concerning the QMLE asymptotic behavior applied to numerous other examples of univariate or multivariate processes (for instance univariate TARCH, bilinear, or multivariate NLARCH processes) are also provided.

Keywords : Multidimensional Causal Processes, QMLE, Strong Consistency, Asymptotic Normality.

Note

The content of this part is based on a paper, written in collaboration with Jean Marc Bardet.

6.1 Introduction

In this paper we study the quasi maximum likelihood estimator (QMLE) for general $\mathbb{R}^m$-valued causal stationary processes $X = (X_t,\ t \in \mathbb{Z})$ (with $m \geq 1$), defined by the following relations :

$$X_t = M_{\theta_0}(X_{t-1}, X_{t-2}, \ldots) \cdot \xi_t + f_{\theta_0}(X_{t-1}, X_{t-2}, \ldots), \quad \text{for} \quad t \in \mathbb{Z}, \tag{6.1}$$

where :
– $\theta_0 \in \Theta$ a compact subset of $\mathbb{R}^d$;
– for all sequences $(x_i)_{i\in\mathbb{N}} \in (\mathbb{R}^m)^{\mathbb{N}}$ such that it exists $N \in \mathbb{N}$ satisfying $x_i = 0$ for $i \geq N$, $\theta \in \Theta \mapsto M_\theta((x_i)_{i\in\mathbb{N}^*})$ is a Borelian function and a $(m \times p)$-matrix (with $1 \leq m \leq p$), such that $\text{Rank}(M_\theta((x_i)_{i\in\mathbb{N}^*})) = m$.
– for all sequences $(x_i)_{i\in\mathbb{N}} \in (\mathbb{R}^m)^{\mathbb{N}}$ such that it exists $N \in \mathbb{N}$ satisfying $x_i = 0$ for $i \geq N$, $\theta \in \Theta \mapsto f_\theta((x_i)_{i\in\mathbb{N}^*})$ is a Borelian function with values in $\mathbb{R}^m$.
– the sequence of $\mathbb{R}^p$-random vectors $(\xi_t)_{t\in\mathbb{Z}} = (\xi_t^{(k)})_{t\in\mathbb{Z},\ 1\leq k\leq p}$ is independent and identically distributed. The random variable ξ_0 is such that $\kappa_2 = \mathbb{E}\big[|\xi_0^{(k)}|^2\big] = \text{Var}\,(\xi_0^{(k)}) = 1$ and $\mathbb{E}\xi_0^{(k)}\xi_0^{(k')} = 0$ for $k \neq k'$.

Additional assumptions on M_θ and f_θ ensuring the existence and the stationarity of X will be provided above. Note that the class of such processes X satisfying relation (6.1) is a specific case of the class of infinite memory processes considered in Doukhan and Wintenberger [49]. If $f_\theta \equiv 0$ and with :

$$H_\theta(X_{t-1}, X_{t-2}, \ldots) = C_0 + \sum_{i=1}^{q'} \Big(\sum_{j=1}^{k} C_{ij} X_{t-i} X'_{t-i} C'_{ij} \Big) + \sum_{i=1}^{q} \Big(\sum_{j=1}^{k} D_{ij} H_\theta(X_{t-i-1}, X_{t-i-2}, \ldots) D'_{ij} \Big),$$
$$\text{where} \quad H_\theta = M_\theta \cdot M'_\theta, \tag{6.2}$$

and $q, q' \in \mathbb{N}$, $k \leq d(d+1)/2$, C_0 is positive definite and C_{ij} and D_{ij} are real $p \times p$ matrix of constants, the process X is a multivariate GARCH(q, q') using the BEKK representation (defined by Engle and Kroner [54] ; another representation was also introduced by Bollerslev [14]). Thus, the class of such multivariate GARCH(q, q') is a particular case of processes X satisfying relation (6.1) as well as their natural generalization, the class of multivariate ARCH(∞) processes defined by :

$$H_\theta(X_{t-1}, X_{t-2}, \ldots) := B_0 + \sum_{i=1}^{\infty} B_i X_{t-i} X'_{t-i} B'_i,$$

with B_0 and $B_1, B_2, \ldots$ positive definite $p \times p$ matrix of constants.

If $M_\theta \equiv I_d$, the process X satisfying relation (6.1) is a multivariate AR(∞) process. Section 6.4 below is devoted to these particular cases of multivariate GARCH(q, q'), ARCH(∞) or AR(∞) processes as well as to the other examples.

The aim of this paper is to provide a method for estimating the parameter θ from a sample $(X_1, \ldots, X_n)$ of X. Various methods can be envisaged in such a parametric frame, by beginning with the Maximum Likelihood Estimation (MLE in the sequel). Different authors studied asymptotic behaviour of MLE for particular cases of multivariate processes satisfying (6.1) (see for instance, Bollerslev and Wooldridge [15], Jeantheau [74] for multivariate GARCH(q, q') processes or Dunsmuir and Hannan [52], Brockwell and DaviesBrockwell1991, or Mauricio [87] for multivariate ARMA processes). Even if the convergence rate of MLE is certainly optimal (a proof of efficiency was obtained in Berkes and Horváth, [10], in the case of 1-dimensional GARCH(q, q')), this method presents numerous drawbacks. Indeed, in one hand, the theoretical expression of the likelihood depends on the distribution of the sequence of innovations (ξ_t). In another hand, even if (ξ_t) is a Gaussian sequence, this expression is rarely explicit.
Thus, two extensions of exact MLE which are in fact approximations of MLE are preferably used. The most popular is the Whittle approximation of MLE which does not present both those drawbacks : only the spectral density of the process is required. For instance, it was successfully applied in the case of 1-dimensional linear processes (see Hannan [66]). However, the case of 1-dimensional conditionally heteroscedastic processes leads to another problem : the spectral density of X^2 has to be considered instead of that of X (which is a constant). Thus, the order of innovation r has to be greater than 8 for obtaining an asymptotical normality of the Whittle MLE (see for instance Giraitis and Robinson [60] for ARCH(∞) processes, or Bardet *et al.* [7] for more general models). Under usual regularity conditions on the spectral density, the convergence rate is $\sqrt{n}$ but the asymptotic covariance matrix is larger than the Cramer-Rao bound (see Giraitis and Robinson [60]).
The second approximation of exact MLE is so-called the Quasi-Maximum Likelihood Estimation (QMLE in the sequel). For respectively 1-dimensional GARCH(q, q') and ARCH(∞) processes, Berkes and Horváth [10] and Robinson and Zaffaroni (2006) proved the asymptotic normality of this estimator under usual regularity conditions. Straumann and Mikosch [105] has also obtained the asymptotic normality for a more general class of 1-dimensional processes. For all those processes, the QMLE is more interesting than Whittle MLE because it requires only $r = 4$ and its asymptotic covariance is smaller (but not always equal to Cramer-Rao bound). For multivariate conditionally heteroscedastic processes, Jeantheau [74] for GARCH$(1, 1)$ and especially Comte and Lieberman (2003) for GARCH(q, q') with $r = 8$, and Ling and McAleer [22] for ARMA-GARCH processes with $r = 6$ obtained the same kind of results. In the sequel, we extend their results to the very general class of processes satisfying (6.1) using only a order moment condition $r = 4$.

The principle of QMLE is based on the expression of the likelihood when the sequence $(\xi_t)_t$ is a sequence of standardized Gaussian vectors. In such a case, the exact (up to an additional constant)

Gaussian log-likelihood is,

$$L_n(\theta) = -\frac{1}{2} \sum_{t=1}^{n} \left(X_t - f_\theta^t\right)' \left(M_\theta^t \cdot (M_\theta^t)'\right)^{-1} \left(X_t - f_\theta^t\right) - \frac{1}{2} \log\Big(\det\left(M_\theta^t (M_\theta^t)'\right)\Big) = -\frac{1}{2}\sum_{t=1}^{n} q_t(\theta), \quad (6.3)$$

where A' denotes the translated matrix of A and for all $t \in \mathbb{Z}$ and $\theta \in \Theta$,

$$f_\theta^t = f_\theta(X_{t-1}, X_{t-2}, \ldots) \quad \text{and} \quad M_\theta^t = M_\theta(X_{t-1}, X_{t-2}, \ldots). \quad (6.4)$$

Remark that $q_t(\theta)$ are real valued variables. Except for very particular cases, this expression is not computable from the sample $(X_1, \ldots, X_n)$. As a consequence and it is the main idea of the QMLE, approximations $\widehat{f}_\theta^t$ and $\widehat{M}_\theta^t$ can be plugged in (6.3) for defining :

$$\widehat{L}_n(\theta) = -\frac{1}{2} \sum_{t=1}^{n} \left(X_t - \widehat{f}_\theta^t\right)' \left(\widehat{M}_\theta^t \cdot (\widehat{M}_\theta^t)'\right)^{-1} \left(X_t - \widehat{f}_\theta^t\right) - \frac{1}{2} \log\Big(\det\left(\widehat{M}_\theta^t (\widehat{M}_\theta^t)'\right)\Big) = -\frac{1}{2}\sum_{t=1}^{n} \widehat{q}_t(\theta), \quad (6.5)$$

with for $t \geq 2$,

$$\widehat{f}_\theta^t := f_\theta(X_{t-1}, X_{t-2}, \ldots, X_1, 0, \ldots) \quad \text{and} \quad \widehat{M}_\theta^t := M_\theta(X_{t-1}, X_{t-2}, \ldots, X_1, 0, \ldots). \quad (6.6)$$

Then, the QMLE $\widehat{\theta}_n$ is defined as the M-estimator of the form

$$\widehat{\theta}_n := \arg\max_{\theta \in \Theta} \widehat{L}_n(\theta). \quad (6.7)$$

In Section 6.3 below the asymptotic behavior of $\widehat{\theta}_n$ is provided. Strong consistency is proved as well as a central limit theorem under regularity conditions on functions M_θ and f_θ (the proofs are stated in Section 6.5). In Section 6.4, these results are compared with Straumann and Mikosch [105] or Robinson and Zaffaroni [102] results in the case of one dimensional processes and with Comte and Lieberman [22] or Ling and McAleer [84] results in the case of multidimensional processes. Other examples are also studied providing new results of asymptotic normality of the QMLE. But to begin with, the following Section 6.2 deals with the problem of existence and stationarity of processes satisfying (6.1).

6.2 Properties of the model

In the sequel, several notations will be used :

- $\|.\|$ denotes indifferently the usual Euclidean norm of a vector or a matrix (for A a $(n \times p)$-matrix, $\|A\| = \sup_{\|X\| \leq 1} \left\{\|AX\|,\ X \in \mathbb{R}^p\right\})$;
- For $\theta \in \Theta \mapsto g(\theta)$ a measurable function which can be a vector or a matrix, $\|g\|_\Theta = \sup_{\theta \in \Theta} \|g(\theta)\|$;
- If V denote a vector space then $V^{(\mathbb{N})} \subset V^{\mathbb{N}}$ denotes the set of $v = (v_j)_{j \in \mathbb{N}}$ such that there exists some integer N satisfying $v_j = 0$ for each $j \geqslant N$.

Now, the following assumption specifies the choice of functions M_θ and f_θ for insuring the stationarity of a solution of (6.1), depending on the order-moment of the innovations $(\xi_t)_t$:

Assumption ST(r) **:** *The family* $(\xi_t)_{t\in\mathbb{Z}} = (\xi_t^{(k)})_{t\in\mathbb{Z},\ 1\le k\le d}$ *of innovation vectors is such that*

$$\kappa_r := \mathbb{E}\big[|\xi_0^{(k)}|^r\big] < +\infty, \textit{with } r \geq 2,$$

and the functions f_θ and M_θ are such that :

1. $\|f_\theta(0,0,\ldots)\|_\Theta + \|M_\theta(0,0,\ldots)\|_\Theta < \infty;$ (6.8)
2. *There exist two sequences $(\alpha_j(f))_{j\in\mathbb{N}}$ and $(\alpha_j(M))_{j\in\mathbb{N}}$ of non-negative real numbers such that*
$$\sum_{j=1}^{\infty}\alpha_j(f) + (\mathbb{E}\|\xi_0\|^r)^{1/r}\Big(\sum_{j=1}^{\infty}\alpha_j(M)\Big) := a < 1 \tag{6.9}$$
with for all $(x_j)_{j\in\mathbb{N}}$, $(y_j)_{j\in\mathbb{N}}$ in $(\mathbb{R}^m)^{(\mathbb{N})}$,
$$\begin{cases} \|f_\theta((x_j)_{j\in\mathbb{N}}) - f_\theta((y_j)_{j\in\mathbb{N}})\|_\Theta & \leqslant \sum_{j=1}^{\infty}\alpha_j(f)\|x_j-y_j\|, \\ \|M_\theta((x_j)_{j\in\mathbb{N}}) - M_\theta((y_j)_{j\in\mathbb{N}})\|_\Theta & \leqslant \sum_{j=1}^{\infty}\alpha_j(M)\|x_j-y_j\|. \end{cases}$$

In Doukhan and Wintenberger [49], under conditions on F, the existence of a r-order stationary solution $X = (X_t)_{t\in\mathbb{Z}}$ of the equation

$$X_t = F\big(X_{t-1}, X_{t-2}, \ldots; \xi_t\big) \quad \text{for} \quad t \in \mathbb{Z}, \tag{6.10}$$

have been established. Using this result, one deduces,

Proposition 6.1 *Let $(\xi_t)_{t\in\mathbb{Z}}$, f_θ and M_θ satisfy Assumption ST(r) with $r \geq 2$. Then there exists a unique stationary solution of the equation (6.1) such that $\mathbb{E}\|X_t\|^r < \infty$. Moreover, for all $\theta \in \Theta$, this solution can be written as a causal Bernoulli shift of the innovation, i.e. there exists a measurable function H_θ such that $X_t = H_\theta(\xi_t, \xi_{t-1}, \ldots)$ for all $t \in \mathbb{Z}$.*

Proof : With $\mathbb{E}\|\xi_t\|^r < \infty$, if F is such that for all $(x_i)_{i\in\mathbb{N}}$, $(y_i)_{i\in\mathbb{N}}$ in $(\mathbb{R}^m)^{(\mathbb{N})}$,
– $\mathbb{E}\big[\|F(0,0,\ldots;\xi_0)\|^r\big] < \infty\,;$
– $\Big(\mathbb{E}\big[\|F\big((x_i)_{i\in\mathbb{N}}\,;\xi_0\big) - F\big((y_i)_{i\in\mathbb{N}}\,;\xi_0\big)\|^r\big]\Big)^{1/r} \leq \sum_{j\in\mathbb{N}} a_j\|x_j-y_j\|$, with $\sum_{j\in\mathbb{N}} a_j < 1$,

then Doukhan and Wintenberger [49] have shown that it exists a r-order stationary solution X of (6.10). Therefore, if $F\big(X_{t-1}, X_{t-2}, \ldots; \xi_t\big) = M_\theta(X_{t-1}, X_{t-2}, \ldots)\cdot\xi_t + f_\theta(X_{t-1}, X_{t-2}, \ldots)$, it is obvious that condition (6.8) of Assumption ST(r) implies $\mathbb{E}\big[\|F(0,0,\ldots,\xi_0)\|^r\big] < \infty$. Moreover, Minkowsky inequality implies

$$\begin{aligned}\Big(\mathbb{E}\Big[\big\|F\big((x_i)_{i\in\mathbb{N}}\,;\xi_0\big)-F\big((y_i)_{i\in\mathbb{N}}\,;\xi_0\big)\big\|_\Theta^r\Big]\Big)^{1/r} &\leq \Big(\mathbb{E}\Big[\big\|(M_\theta((x_i)_{i\in\mathbb{N}})-M_\theta((y_i)_{i\in\mathbb{N}}))\cdot\xi_0\big\|_\Theta^r\Big]\Big)^{1/r}\\ &\qquad +\big\|f_\theta((x_i)_{i\in\mathbb{N}})-f_\theta((y_i)_{i\in\mathbb{N}})\big\|_\Theta\\ &\leq \big(\mathbb{E}\|\xi_0\|^r\big)^{1/r}\big\|M_\theta((x_i)_{i\in\mathbb{N}})-M_\theta((y_i)_{i\in\mathbb{N}})\big\|_\Theta\\ &\qquad +\big\|f_\theta((x_i)_{i\in\mathbb{N}})-f_\theta((y_i)_{i\in\mathbb{N}})\big\|_\Theta.\end{aligned}$$

Therefore, from condition (6.9) of Assumption ST(r), one deduces that the second condition of Doukhan and Wintenberger holds. □

Therefore,

$$\mathbb{E}\Big[\|f_\theta^t\|_\Theta^r\Big]=\mathbb{E}\Big[\|f_\theta(X_{t-1},X_{t-2},\ldots)\|_\Theta^r\Big]<\infty \ \text{ and } \ \mathbb{E}\Big[\|M_\theta^t\|_\Theta^r\Big]=\mathbb{E}\Big[\|M_\theta(X_{t-1},X_{t-2},\ldots)\|_\Theta^r\Big]<\infty.$$

Thus the limits $f_\theta^t=f_\theta(X_{t-1},X_{t-2},\ldots)$ and $M_\theta^t=M_\theta(X_{t-1},X_{t-2},\ldots)$ are almost surely finite. With the $0-1$ law of Kolmogorov, it implies the ergodicity of this solution X,

Proposition 6.2 *Under assumptions of Proposition 6.1, X is an ergodic stationary process.*

Proof : It is involved by Proposition 4.3 in Krengel (1985). Indeed, if $(E,\mathcal{E})$ and $(\tilde{E},\tilde{\mathcal{E}})$ are measurable spaces, $(v_t)_{t\in\mathbb{Z}}$ is a stationary ergodic sequence of E-valued random elements and $H:(E^{\mathbb{N}},\mathcal{E}^{\mathbb{N}})\mapsto(\tilde{E},\tilde{\mathcal{E}})$ is a measurable function then the sequence $(\tilde{v}_t)_{t\in\mathbb{Z}}$ defined by $\tilde{v}_t=H(v_t,v_{t-1},\ldots)$ is stationary and ergodic. □

For X satisfying (6.1), $\widehat{f}_\theta^t$ and $\widehat{M}_\theta^t(\widehat{M}_\theta^t)'$ are respectively estimators of the conditional expectation $f_\theta^t=\mathbb{E}(X_t\mid X_{t-1},X_{t-2},\ldots)$ and the conditional covariance matrix $M_\theta^t\big(M_\theta^t\big)'=\mathbb{E}\big((X_t-f_\theta^t)(X_t-f_\theta^t)'\mid X_{t-1},X_{t-2},\ldots\big)$. In order to prove the consistence of those estimators, two conditions on f_θ and M_θ have to be added :

- **Assumption F :** *For all sequence $(x_j)_{j\in\mathbb{N}}\in(\mathbb{R}^m)^{(\mathbb{N})}$ the function $\theta\mapsto f_\theta(x_1,\ldots)$ is continuous.*
- **Assumption M :** *For all sequence $(x_j)_{j\in\mathbb{N}}\in(\mathbb{R}^m)^{(\mathbb{N})}$ the function $\theta\mapsto M_\theta(x_1,\ldots)$ is continuous. Moreover it exists $\underline{M}>0$ satisfying such that for all $(x_j)_{j\in\mathbb{N}}\in(\mathbb{R}^m)^{(\mathbb{N})}$*

$$\inf_{\theta\in\Theta}\det\big(M_\theta(x_1,\ldots)\big(M_\theta(x_1,\ldots)\big)'\big)\geqslant\underline{M}>0. \tag{6.11}$$

Then, using essentially condition (6.9) of Assumption ST(r), one obtains :

Lemma 6.1 *Assume that Assumptions ST(r), F and M are satisfied with $r\geq 2$. Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary solution of equation (6.1). Then, for all $t\in\mathbb{N}^*$,*

1. *the function $\theta \to f_\theta^t$ is continuous a.e.,* $\mathbb{E}\Big[\|f_\theta^t\|_\Theta^r\Big] < \infty$ *and for all* $t \in \mathbb{N}^*$,

$$\mathbb{E}\Big[\|\widehat{f}_\theta^t - f_\theta^t\|_\Theta^r\Big] \leqslant \mathbb{E}\|X_0\|^r \Big(\sum_{j \geqslant t} \alpha_j(f)\Big)^r. \tag{6.12}$$

2. *the function $\theta \to M_\theta^t(M_\theta^t)'$ is continuous a.e., satisfies for all $t \in \mathbb{Z}$,* $\mathbb{E}\big[\|M_\theta^t(M_\theta^t)'\|_\Theta^{r/2}\big] < \infty$ *and* $\big(\underline{M}\big)^{1/m} \leq \inf_{\theta \in \Theta} \|M_\theta^t(M_\theta^t)'\|$ *a.s. Moreover, there exists $C > 0$ not depending on t and such that*

$$\mathbb{E}\big[\|\widehat{M}_\theta^t(\widehat{M}_\theta^t)' - M_\theta^t(M_\theta^t)'\|_\Theta^{r/2}\big] \leqslant C\Big(\sum_{j \geqslant t} \alpha_j(M)\Big)^{r/2} \quad \textit{for all } t \in \mathbb{N}^*. \tag{6.13}$$

<u>*Proof :*</u> 1. First, condition (6.9) of Assumption ST(r) implies,

$$\begin{aligned}\mathbb{E}\Big[\|\widehat{f}_\theta^t - f_\theta^t\|_\Theta^r\Big] &\leq \mathbb{E}\Big[\Big\|\sum_{j \geqslant t} \alpha_j(f) X_{t-j}\Big\|^r\Big] \\ &\leq \mathbb{E}\|X_0\|^r \Big(\sum_{j \geqslant t} \alpha_j(f)\Big)^r,\end{aligned}$$

because $\sum_{j \geqslant 1} \alpha_j(f) < 1 < \infty$. Thus, (6.12) is established. Moreover, the space of continuous functions on Θ, denoted $\mathcal{C}(\Theta)$ is a Banach space equipped with the uniform norm $\|\cdot\|_\Theta$. It implies that $\mathbb{L}^r(\mathcal{C}(\Theta))$ is also a Banach space (because $r \geq 2 \geq 1$). Because f_θ^∞ is a limit in this Banach space of a.s. continuous functions $(\widehat{f}_\theta^t)_t$ and because for all $t \in \mathbb{Z}$, f_θ^∞ and f_θ^t have the same distributions, one deduces that $\theta \to f_\theta^t$ is also a.s uniformly continuous on Θ. Finally, once again from the Minkowsky inequality, for all $\theta \in \Theta$ and $t \in \mathbb{Z}$,

$$\begin{aligned}\mathbb{E}\Big[\|f_\theta^t\|^r\Big] &\leq \mathbb{E}\Big[\|\widehat{f}_\theta^0 - f_\theta^t\|^r\Big] + \mathbb{E}\Big[\|\widehat{f}_\theta^0\|^r\Big] \\ &\leq \mathbb{E}\|X_0\|^r \Big(\sum_{j \geqslant 1} \alpha_j(f)\Big)^r + \|f_\theta(0, 0, \ldots)\|_\Theta^r,\end{aligned}$$

and therefore, from condition (6.8) of Assumption ST(r), $\mathbb{E}\Big[\|f_\theta^t\|_\Theta^r\Big] \leq A < \infty$, with $A > 0$ not depending on θ and t.

2. Here $(X_t)_{t \in \mathbb{Z}}$ is a stationary solution of equation (6.1) with θ_0 a fixed vector of Θ. Let $\theta \in \Theta$ (remark that X does not depend on the vector θ). Let $\mathcal{C}(\Theta, \mathcal{M}_m)$ be the Banach space of continuous functions from Θ to $\mathcal{M}_m$, the space of $m \times m$ real squared matrix, equipped with the uniform norm $\sup_{\theta \in \Theta} \|\cdot\|$. Let $\mathbb{L}^{r/2}(\mathcal{C}(\Theta, \mathcal{M}_m))$ be the Banach spaces of $r/2$-order random functions of $\mathcal{C}(\Theta, \mathcal{M}_m)$ equipped with the uniform norm. From Assumptions A and M, for all $p \in \mathbb{N}$ and $t \in \mathbb{N}$, the function

$$\theta \in \Theta \mapsto h^{t,p}(\theta) := M_\theta(X_{t-1}, \ldots, X_{t-p}, 0, \ldots)\big(M_\theta(X_{t-1}, \ldots, X_{t-p}, 0, \ldots)\big)'$$

belongs to $\mathbb{L}^{r/2}(\mathcal{C}(\Theta,\mathcal{M}_m))$ and is a continuous function in θ. Moreover, using condition (6.9) of Assumption ST(r) and usual inequalities satisfied by matrix norms, one obtains

$$\begin{aligned}\|h^{t,p}(\theta)\|_\Theta^{r/2} &= \sup_{\theta\in\Theta}\|M_\theta(X_{t-1},\ldots,X_{t-p},0,\ldots)\big(M_\theta(X_{t-1},\ldots,X_{t-p},0,\ldots)\big)'\|^{r/2}\\ &\le \sup_{\theta\in\Theta}\|M_\theta(X_{t-1},\ldots,X_{t-p},0,\ldots)\|^r\\ &\le \Big(\|M_\theta(0,0,\ldots)\|_\Theta+\sum_{j=1}^\infty\|X_{t-j}\|\alpha_j(M)\Big)^r.\end{aligned}$$

Using Doukhan and Wintenberger (2006), one obtains $\mathbb{E}\|X_0\|^r<\infty$ for $r\ge 2$ such that $\mathbb{E}\|\xi_0\|^r<\infty$. Thus, with $\sum_{j=1}^\infty\alpha_j(M)<\infty$,

$$\mathbb{E}\big[\|h^{t,p}(\theta)\|_\Theta^{r/2}\big]\leqslant 2^{r-1}\;\|M_\theta(0,0,\ldots)\|_\Theta^r+2^{r-1}\;\mathbb{E}\big[\|X_0\|^r\big]\Big(\sum_{j=1}^\infty\alpha_j(M)\Big)^r=B^2<\infty, \qquad (6.14)$$

with a bound B which does not depend on p and t.

Now the sequence $(h^{t,p}(\theta))_{p\in\mathbb{N}^*}$ satisfies the Cauchy criteria (in $\mathbb{L}^{r/2}(\mathcal{C}(\Theta,\mathcal{M}_m))$). Indeed, let p and q be two integers such that $p<q$. Remark that $\|h^{t,p}(\theta)\|^{r/2}=\|M_\theta(X_{t-1},\ldots,X_{t-p},0,\ldots)\|^r$ and $\|h^{t,q}(\theta)\|^{r/2}=\|M_\theta(X_{t-1},\ldots,X_{t-q},0,\ldots)\|^r$. Then,

$$\begin{aligned}\|h^{t,p}-h^{t,q}\|_\Theta^{r/2}\leqslant \Big(&\|M_\theta(X_{t-1},\ldots,X_{t-p},0,\ldots)\|_\Theta^{r/2}+\|M_\theta(X_{t-1},\ldots,X_{t-q},0,\ldots)\|_\Theta^{r/2}\Big)\\ &\times\big\|M_\theta(X_{t-1},\ldots,X_{t-p},0,\ldots)-M_\theta(X_{t-1},\ldots,X_{t-q},0,\ldots)\big\|_\Theta^{r/2}.\end{aligned}$$

Relation (6.14) and the Cauchy-Schwartz inequality imply

$$\begin{aligned}\mathbb{E}\big[\|h^{t,p}-h^{t,q}\|_\Theta^{r/2}\big] &\le \cdot\Big(2\mathbb{E}\big[\|M_\theta(X_{t-1},\ldots,X_{t-p},0,\ldots)\|_\Theta^r\big]+2\mathbb{E}\big[\|M_\theta(X_{t-1},\ldots,X_{t-q},0,\ldots)\|_\Theta^r\big]\Big)^{1/2}\\ &\quad\times\Big(\mathbb{E}\big[\|(M_\theta(X_{t-1},\ldots,X_{t-p},0,\ldots)-M_\theta(X_{t-1},\ldots,X_{t-q},0,\ldots)\|_\Theta^r\big]\Big)^{1/2}\\ &\le 2B\cdot\Big(\mathbb{E}\Big[\Big(\sum_{j=p+1}^q\alpha_j(M)\|X_{t-j}\|\Big)^r\Big]\Big)^{1/2},\\ &\le 2B\cdot(\mathbb{E}\|X_0\|^r)^{1/2}\Big(\sum_{j=p+1}^q\alpha_j(M)\Big)^{r/2},\end{aligned}$$

from condition (6.8) of Assumption ST(r). Thus, there exists a constant $C>0$ not depending on t, p and q such that

$$\mathbb{E}\|h^{t,p}-h^{t,q}\|_\Theta^{r/2}\leqslant C\Big(\sum_{j=p+1}^q\alpha_j(M)\Big)^{r/2}.$$

Thanks to $\sum_{j=1}^{\infty} \alpha_j(M) < \infty$, one deduces that $(h^{t,p}(\theta))_{p\in\mathbb{N}^*}$ is a Cauchy sequence in $\mathbb{L}^{r/2}(\mathcal{C}(\Theta, \mathcal{M}_m))$ of a.e. continuous functions and therefore it converges to $h^t(\theta) := M_\theta^t\big(M_\theta^t\big)' \in \mathbb{L}^{r/2}(\mathcal{C}(\Theta, \mathcal{M}_m))$. As a consequence, $\theta \to h^t(\theta)$ is a.e. continuous function such that $\mathbb{E}\big[\|h^t(\theta)\|_\Theta^{r/2}\big] < \infty$.
Thus it exists a subsequence $(p_k)_{k\in\mathbb{N}}$ such that $\|h^{t,p_k} - h^t\|_\Theta \xrightarrow[k\to\infty]{a.s.} 0$. Therefore, Assumption M implies that $\det\big(M_\theta^t\big(M_\theta^t\big)'\big) \geq \underline{M}$ for all $\theta \in \Theta$ with probability 1. From usual relations between $\|\cdot\|$ and det for a matrix, one deduces that $\big(\underline{M}\big)^{1/m} \leq \inf_{\theta\in\Theta} \|M_\theta^t\big(M_\theta^t\big)'\|$ a.s.
Finally, using exactly the same arguments as previously, it follows that for all $t \in \mathbb{N}^*$,

$$\mathbb{E}\big[\|h^{t,t} - h^t\|_\Theta^{r/2}\big] \leqslant 2B \cdot \Big(\mathbb{E}\|X_0\|^{r/2}\Big)^{1/2} \Big(\sum_{j\geqslant t} \alpha_j(M)\Big)^{r/2}.$$

The relation $h^{t,t}(\theta)) = \widehat{M}_\theta^t\big(\widehat{M}_\theta^t\big)'$ implies (6.13). $\square$

From this Lemma, one deduces that $\widehat{f}^t$ and $\widehat{M}_\theta^t\big(\widehat{M}_\theta^t\big)'$ are respectively consistent estimators (uniformly in θ) of conditional expectation and covariance matrix of X. It follows that the stationary solution X of equation (6.1) for any $\theta = \theta_0$ is invertible in the sense of Straumann [105].

6.3 Asymptotic behavior of the QMLE

Let $\widehat{\theta}_n$ define as in (6.7). An identifiability condition, has to be added for the consistence of $\widehat{\theta}_n$:

Assumption Id : For all $\theta \in \Theta$, $M_\theta^t = M_{\theta_0}^t$ and $f_\theta^t = f_{\theta_0}^t$ a.s. if and only if $\theta = \theta_0$.

Then, using similar arguments than in Straumann [105], one obtains :

Theorem 6.1 *Let Assumptions ST(2), F, M and Id hold. Assume also that*

$$\alpha_j(f) + \alpha_j(M) = O\big(j^{-\ell}\big) \quad \textit{with} \quad \ell > 3/2. \tag{6.15}$$

Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary solution of equation (6.1). Then the QMLE $\widehat{\theta}_n$ defined by (6.7) is strongly consistent, i.e.

$$\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta_0.$$

From the proof of this Theorem, an obvious but interesting (notably for ARCH processes) corollary can also be added. Indeed, additional conditions on matrix H_θ (defined in (6.2)) can also be provided for establishing the consistence of $\widehat{\theta}_n$. Then :

Corollary 6.1 *Let Assumptions ST(2), F, M and Id hold. Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary solution of equation (6.1). If it exists a sequence $(\alpha_j(H))_{j\in\mathbb{N}^*}$ of non-negative real numbers such that for all*

$(x_j)_{j\in\mathbb{N}}$, $(y_j)_{j\in\mathbb{N}}$ *in* $(\mathbb{R}^m)^{(\mathbb{N})}$,

$$\Big\| H_\theta((x_j)_{j\in\mathbb{N}}) - H_\theta((y_j)_{j\in\mathbb{N}}) \Big\|_\Theta \leqslant \sum_{j=1}^{\infty} \alpha_j(H) \| x_j x_j' - y_j y_j' \|, \tag{6.16}$$

with

$$\alpha_j(f) + \alpha_j(H) = O\big(j^{-\ell}\big) \quad \text{with} \quad \ell > 3/2, \tag{6.17}$$

then the QMLE $\widehat{\theta}_n$ *defined by (6.7) is strongly consistent, i.e.*

$$\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta_0.$$

Classically, other conditions have to be added in order to derive the asymptotic normality of the QMLE. Thus, for $i = 1, 2$, define the following assumption :

Assumption $D^{(i)}$ **:** *The functions* $\theta \in \Theta \mapsto f_\theta((x_j)_{j\in\mathbb{N}})$ *and* $\theta \in \Theta \mapsto M_\theta((x_j)_{j\in\mathbb{N}})$ *are i-times continuously differentiable for all* $(x_j)_{j\in\mathbb{N}} \in (\mathbb{R}^m)^{(\mathbb{N})}$ *and such that :*

1. $\Big\| \dfrac{\partial^i f_\theta(0,0,\dots)}{\partial\theta_{k_1}\partial\theta_{k_i}} \Big\|_\Theta + \Big\| \dfrac{\partial^i M_\theta(0,0,\dots)}{\partial\theta_{k_1}\partial\theta_{k_i}} \Big\|_\Theta < \infty$ *for all* $k_1, k_i \in \{1, \dots, d\}$;
2. *There exist two sequences* $(\alpha_j^{(i)}(f))_{j\in\mathbb{N}^*}$ *and* $(\alpha_j^{(i)}(M))_{j\in\mathbb{N}^*}$ *of non-negative real numbers with* $\displaystyle\sum_{j=1}^{\infty} \alpha_j^{(i)}(f) + \sum_{j=1}^{\infty} \alpha_j^{(i)}(M) < \infty$ *for all* $(x_j)_{j\in\mathbb{N}}$, $(y_j)_{j\in\mathbb{N}}$ *in* $(\mathbb{R}^m)^{(\mathbb{N})}$ *and for all* $k_1, k_i \in \{1, \dots, d\}$,

$$\begin{cases} \Big\| \dfrac{\partial^i f_\theta((x_j)_{j\in\mathbb{N}})}{\partial\theta_{k_1}\partial\theta_{k_i}} - \dfrac{\partial^i f_\theta((y_j)_{j\in\mathbb{N}})}{\partial\theta_{k_1}\partial\theta_{k_i}} \Big\|_\Theta & \leqslant \displaystyle\sum_{j=1}^{\infty} \alpha_j^{(i)}(f) \| x_j - y_j \|, \\ \Big\| \dfrac{\partial^i M_\theta((x_j)_{j\in\mathbb{N}})}{\partial\theta_{k_1}\partial\theta_{k_i}} - \dfrac{\partial^i M_\theta((y_j)_{j\in\mathbb{N}})}{\partial\theta_{k_1}\partial\theta_{k_i}} \Big\|_\Theta & \leqslant \displaystyle\sum_{j=1}^{\infty} \alpha_j^{(i)}(M) \| x_j - y_j \|. \end{cases}$$

Theorem 6.2 *Let Assumptions* $ST(4)$, M , $D^{(1)}$ *and* $D^{(2)}$ *and Id hold. Let* $(X_t)_{t\in\mathbb{Z}}$ *be a stationary solution of equation (6.1) for* $\theta_0 \in \overset{\circ}{\Theta}$, *where condition (6.15) holds. Moreover, if*

$$\alpha_j^{(1)}(f) + \alpha_j^{(1)}(M) = O\big(j^{-\ell'}\big) \quad \text{with} \quad \ell' > 3/2, \tag{6.18}$$

then the QMLE $\widehat{\theta}_n$ *is consistent and asymptotically normal, i.e.,*

$$\sqrt{n}\big(\widehat{\theta}_n - \theta_0\big) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_d\big(0\ ,\ F(\theta_0)^{-1} \cdot G(\theta_0) \cdot F(\theta_0)^{-1}\big),$$

where the matrix $F(\theta_0)$ *(assumed to be invertible) and* $G(\theta_0)$ *are respectively defined in (6.51) and (6.47).*

Like for Corollary 6.1, it is also possible to establish a central limit theorem by replacing conditions on the matrix M_θ by conditions on the matrix H_θ. Thus, for $i=1,2$, define the following assumption :

Assumption $E^{(i)}$ **:** *The functions $\theta \in \Theta \mapsto f_\theta((x_j)_{j\in\mathbb{N}})$ and $\theta \in \Theta \mapsto H_\theta((x_j)_{j\in\mathbb{N}})$ are i-times continuously differentiable for all $(x_j)_{j\in\mathbb{N}} \in (\mathbb{R}^m)^{(\mathbb{N})}$ and such that :*

1. $\displaystyle \Big\|\frac{\partial^i f_\theta(0,0,\ldots)}{\partial\theta_{k_1}\partial\theta_{k_i}}\Big\|_\Theta + \Big\|\frac{\partial^i H_\theta(0,0,\ldots)}{\partial\theta_{k_1}\partial\theta_{k_i}}\Big\|_\Theta < \infty$ *for all* $k_1, k_i \in \{1,\ldots,d\}$;
2. *There exist two sequences $(\alpha_j^{(i)}(f))_{j\in\mathbb{N}^*}$ and $(\alpha_j^{(i)}(H))_{j\in\mathbb{N}^*}$ of non-negative real numbers such that* $\displaystyle\sum_{j=1}^\infty \alpha_j^{(i)}(f) + \sum_{j=1}^\infty \alpha_j^{(i)}(H) < \infty$ *with for all $(x_j)_{j\in\mathbb{N}}$, $(y_j)_{j\in\mathbb{N}}$ in $(\mathbb{R}^m)^{(\mathbb{N})}$ and*

$$\text{for all } k_1, k_i \in \{1,\ldots,d\}, \quad \begin{cases} \Big\|\dfrac{\partial^i f_\theta((x_j)_{j\in\mathbb{N}})}{\partial\theta_{k_1}\partial\theta_{k_i}} - \dfrac{\partial^i f_\theta((y_j)_{j\in\mathbb{N}})}{\partial\theta_{k_1}\partial\theta_{k_i}}\Big\|_\Theta & \leqslant \displaystyle\sum_{j=1}^\infty \alpha_j^{(i)}(f)\|x_j-y_j\|, \\ \Big\|\dfrac{\partial^i H_\theta((x_j)_{j\in\mathbb{N}})}{\partial\theta_{k_1}\partial\theta_{k_i}} - \dfrac{\partial^i H_\theta((y_j)_{j\in\mathbb{N}})}{\partial\theta_{k_1}\partial\theta_{k_i}}\Big\|_\Theta & \leqslant \displaystyle\sum_{j=1}^\infty \alpha_j^{(i)}(H)\|x_jx_j'-y_jy_j'\|. \end{cases}$$

Then,

Corollary 6.2 *Let Assumptions ST(r) with $r \geq 4$, M , $E^{(1)}$ and $E^{(2)}$ and Id. Let $(X_t)_{t\in\mathbb{Z}}$ be a stationary solution of equation (6.1) for $\theta_0 \in \overset{\circ}{\Theta}$ and where condition (6.17) holds. Moreover, if*

$$\alpha_j^{(1)}(f) + \alpha_j^{(1)}(H) = O\big(j^{-\ell'}\big) \quad \text{with} \quad \ell' > 3/2, \tag{6.19}$$

then the QMLE $\widehat{\theta}_n$ is consistent and asymptotically normal, i.e.,

$$\sqrt{n}\big(\widehat{\theta}_n - \theta_0\big) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_d\big(0\ ,\ F(\theta_0)^{-1}\cdot G(\theta_0)\cdot F(\theta_0)^{-1}\big), \tag{6.20}$$

where the matrix $F(\theta_0)$ (assumed to be invertible) and $G(\theta_0)$ are respectively defined in (6.51) and (6.47).

6.4 Examples

In this section, several univariate and multivariate processes fitting the representation of the system (6.1) are presented. In all the sequel, $\theta_0 \in \Theta$ a locally compact open subset of $\mathbb{R}^d$ and $(\xi_t)_{t\in\mathbb{Z}} = (\xi_t^{(k)})_{t\in\mathbb{Z},\ 1\leq k\leq p}$ is a sequence of $\mathbb{R}^p$-random vectors such that $(\xi_t^{(k)})$ is a family of r-order zero-mean independent and identically distributed random variables, such that $\kappa_2 = \mathbb{E}\big[|\xi_0^{(k)}|^2\big] = \mathrm{Var}\,(\xi_0^{(k)}) = 1$ (with $r \geq 2$).

The previous asymptotic results are applied to these examples. For certain of them (univariate and multivariate ARCH, GARCH, AR and GARCH-ARMA processes), the consistence and the asymptotic normality have been already obtained by other authors under different (often stronger for multivariate processes) and sometimes similar conditions than here. For the other examples (TARCH, bilinear, LARCH, NLARCH processes), the consistence and the asymptotic normality of QMLE are new results.

Processes satisfying $f \equiv 0$ are first considered and univariate processes (with $m = p = 1$) to begin with.

6.4.1 ARCH(∞) processes

The famous and from now on classical GARCH(q', q) model was introduced by Engle [53] and Bollerslev [13]. Such time series $(X_t)_{t \in \mathbb{Z}}$ is a stationary solution of relations (6.24). However, a GARCH process can be written as a particular case of ARCH(∞) model (introduced in Robinson, 1991) that satisfied :

$$X_t = \sigma_t \xi_t, \qquad \sigma_t^2 = b_0(\theta_0) + \sum_{j=1}^{\infty} b_j(\theta_0) X_{t-j}^2, \tag{6.21}$$

where, for all $\theta \in \Theta$, $(b_j(\theta))_j$ are sequences of non-negative real numbers (which depends on the families (c_j) and (d_j) in the case of GARCH(q', q) process). In such a case,

$$M_\theta(X_{t-1}, X_{t-2}, \ldots) = \Big(b_0(\theta) + \sum_{j=1}^{\infty} b_j(\theta) X_{t-j}^2\Big)^{1/2} \quad \text{and} \quad H_\theta(X_{t-1}, X_{t-2}, \ldots) = b_0(\theta) + \sum_{j=1}^{\infty} b_j(\theta) X_{t-j}^2.$$

As a consequence, $\alpha_j(M) = \sup_\theta \sqrt{b_j(\theta)}$ and Assumption ST(r) is satisfied as soon as

$$\sum_{j=1}^{\infty} \sup_{\theta \in \Theta} \sqrt{b_j(\theta)} < \Big(\mathbb{E}\big[|\xi_0|^r\big]\Big)^{-1/r}, \tag{6.22}$$

i.e. a r-order process stationary solution X of (6.21) exists. However, the specific case of stationarity of ARCH models has been intensively studied and r-order stationarity condition have been obtained by Giraitis *et al.* (2000),

$$\sum_{j=1}^{\infty} \sup_{\theta \in \Theta} b_j(\theta) < \Big(\mathbb{E}\big[|\xi_0|^r\big]\Big)^{-2/r} \tag{6.23}$$

that is weaker than the previous condition (6.22) (see also Giraitis *et al.* [59] for an excellent survey about ARCH models). Thus, straightforward applications of Corollaries 6.1 and 6.2 imply

Property 6.1 *Let $X = (X_t)_{t\in\mathbb{Z}}$ be a r-order stationary solution of (6.21) where condition (6.23) holds together with assumption Id. Moreover, assume that* $\inf_{\theta\in\Theta} b_0(\theta) > 0$, *the functions $\theta \in \Theta \mapsto b_j(\theta)$ are injective continuous functions for all $j \in \mathbb{N}$ and*

$$\sup_{\theta\in\Theta} b_j(\theta) = O\big(j^{-\ell}\big) \quad \textit{with} \quad \ell > 3/2.$$

Then,

1. *if $r = 2$, $\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta_0$.*
2. *if $r = 4$, and if the functions $\theta \in \Theta \mapsto b_j(\theta)$ are 2-times continuously differentiable for all $j \in \mathbb{N}$ and such that for all $(k, k') \in \{1, \ldots, d\}^2$,*

$$\sup_{\theta\in\Theta} \Big|\frac{\partial b_j(\theta)}{\partial\theta_k}\Big| = O\big(j^{-\ell'}\big) \quad \textit{with} \quad \ell' > 3/2 \quad \textit{and} \quad \sup_{\theta\in\Theta} \Big|\frac{\partial^2 b_j(\theta)}{\partial\theta_k\partial\theta_{k'}}\Big| = O\big(j^{-\ell''}\big) \quad \textit{with} \quad \ell'' > 1,$$

   then the QMLE $\widehat{\theta}_n$ is asymptotically normal, i.e. satisfies (6.20).

Recently, Robinson and Zaffaroni [102] provided consistency and asymptotic normality for the QMLE of univariate ARCH processes. To summarize and compare their Theorem 1 and 2 (see Robinson and Zaffaroni [102] p. 1054-1055) with previous Property 6.1 (even if our results concerning QMLE are mainly devoted to multivariate processes),

- concerning the consistency, they obtained an almost sure convergence (instead weak convergence here) under order moment $r > 2$ (instead $r = 2$ here) and a decreasing rate $j^{-\ell}$ with $\ell > 1$ (instead $\ell > 3/2$ here) for the sequence $(b_j(\theta))_j$ but also with derivative conditions (nothing like this here);
- concerning the asymptotic normality, their conditions on both the first derivatives of $\theta \mapsto b_j(\theta)$ are the same than in Property 6.1, and conditions on the third derivatives are also required (nothing like this here).

6.4.2 GARCH(q, q') models

Here X_t is the stationary solution of

$$X_t = \sigma_t\xi_t, \qquad \sigma_t^2 = c_0(\theta_0) + \sum_{j=1}^{q} c_j(\theta_0)X_{t-j}^2 + \sum_{j=1}^{q'} d_j(\theta_0)\sigma_{t-j}^2, \tag{6.24}$$

where $c_j(\theta)$ and $d_j(\theta)$ are non negative real numbers for all $\theta \in \Theta$. In such a case, Assumption ST(r) could be replaced by the conditions given by Ling and McAleer [84].

This model can be nested in the class of ARCH(∞) models (see again Giraitis *et al.* [59], as one needs to set

$$b_0(\theta) := \frac{c_0(\theta)}{1 - \sum_{j=1}^{q'} d_j(\theta)} \quad \text{and} \quad \sum_{i=1}^{\infty} b_i(\theta)z^i := \frac{\sum_{i=1}^{q} c_i(\theta)z^i}{1 - \sum_{i=1}^{q'} d_i(\theta)z^i} \quad \text{for all} \quad z \in \mathbb{C}$$

(in the last formula, both the polynoms are supposed to be coprime). As a consequence, the sequence $(\sup_{\theta\in\Theta}|b_j(\theta)|)_j$ exponentially decreases to 0. Our condition ST(r) becomes

$$\sum_{j=1}^{\infty}\sup_{\theta\in\Theta}\sqrt{b_j(\theta)}<\Big(\mathbb{E}\big[|\xi_0|^r\big]\Big)^{-2/r}. \tag{6.25}$$

In the special case of GARCH(1,1) $c_j=d_j=0$ for $j>1$ and $\theta=(c_0,c_1,d_1)$, ST(2) rewrites as $\sqrt{c_1}+\sqrt{d_1}<1$. The optimal condition of moment of order 2 write as $c_1+d_1<1$. Our approach based on ARCH(∞) representation do not allow as good results as in Berkes *et al.* [11].

Property 6.2 *Let $X=(X_t)_{t\in\mathbb{Z}}$ be a r-order stationary solution of (6.24) where condition (6.25) holds together with assumption Id. Moreover, assume that $\inf_{\theta\in\Theta}c_0(\theta)>0$ and $\theta\in\Theta\mapsto c_j(\theta)$ and $\theta\in\Theta\mapsto d_j(\theta)$ are injective continuous functions for all j. Then,*

1. *if $r=2$ then $\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta_0$.*
2. *if $r=4$, and if the functions $\theta\in\Theta\mapsto c_j(\theta)$ and $\theta\in\Theta\mapsto d_j(\theta)$ are 2-times continuously differentiable for all $j\in\mathbb{N}$ and such that then the QMLE $\widehat{\theta}_n$ is asymptotically normal, i.e. satisfies (6.20).*

The natural case where $\theta=(c_0,c_1,\dots,c_q,d_1,\dots,d_{q'})$ implies that assumptions Id, $\mathrm{D}^{(1)}$, $\mathrm{D}^{(2)}$ are checked like the following property :

Property 6.3 *Let $X=(X_t)_{t\in\mathbb{Z}}$ be a r-order stationary solution of (6.24) where condition (6.25) holds with $c_0>0$. Then,*

1. *if $r=2$ then $\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta_0$.*
2. *if $r=4$, then the QMLE $\widehat{\theta}_n$ is asymptotically normal, i.e. satisfies (6.20).*

In the framework of GARCH processes, the (strong) consistance results established by Straumann and Mikosch [105] hold under weaker assumption. The existence of a moment of order 2 on ξ_0 is even not necessary (see Pfanzagl [94] for more details). The condition (6.25) for $r=4$ implying the asymptotic normality of the QMLE (in the case $\theta=(c_0,c_1,\dots,c_q,d_1,\dots,d_{q'})$) has been previously obtained by Berkes and Horvath [10].

6.4.3 TARCH(∞) models

A real valued time series (X_t) is called a TARCH(∞) (Threshold ARCH(∞)) process if it satisfies

$$X_t=\sigma_t\xi_t,\qquad \sigma_t=b_0(\theta_0)+\sum_{j=1}^{\infty}\big(b_j^+(\theta_0)\max\{X_{t-j},0\}-b_j^-(\theta_0)\min\{X_{t-j},0\}\big), \tag{6.26}$$

where the parameters $b_0(\theta)$, $b_j^+(\theta)$ and $b_j^-(\theta)$ are assumed to be non negative real numbers. This class of processes is a generalization of the class of TGARCH(p,q) processes (introduced by Rabemananjara and Zakoïan [99] and AGARCH(p,q) processes (introduced by Ding *et al.* [32]. For such a process satisfying (6.26), Assumption ST(r) is checked when

$$\sum_{j=1}^{\infty} \sup_{\theta\in\Theta} \max\big(b_j^-(\theta), b_j^+(\theta)\big) \leq \Big(\mathbb{E}\big[|\xi_0|^r\big]\Big)^{-1/r} \tag{6.27}$$

since $\alpha_j(M) = \sup_{\theta\in\Theta} \max\big(b_j^-(\theta), b_j^+(\theta)\big)$. Consequently,

Property 6.4 *Let $X = (X_t)_{t\in\mathbb{Z}}$ be a r-order stationary solution of (6.26) where condition (6.27) holds together with assumption Id. Moreover, assume that $\inf_{\theta\in\Theta} b_0(\theta) > 0$, the functions $\theta \in \Theta \mapsto b_0(\theta)$, $\theta \in \Theta \mapsto b_j^+(\theta)$ and $\theta \in \Theta \mapsto b_j^-(\theta)$ are injective continuous functions for all j, and*

$$\sup_{\theta\in\Theta} \max\big(b_j^-(\theta), b_j^+(\theta)\big) = O\big(j^{-\ell}\big) \quad \text{with} \quad \ell > 3/2.$$

Then,

1. *if $r = 2$, $\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta_0$.*
2. *if $r = 4$, and the functions $\theta \in \Theta \mapsto b_0(\theta)$, $\theta \in \Theta \mapsto b_j^+(\theta)$ and $\theta \in \Theta \mapsto b_j^-(\theta)$ are 2-times continuously differentiable for all $j \in \mathbb{N}$ and such that for all $(k,k') \in \{1,\ldots,d\}^2$,*

$$\sup_{\theta\in\Theta} \max\Big(\Big|\frac{\partial b_j^+(\theta)}{\partial\theta_k}\Big|, \Big|\frac{\partial b_-^+(\theta)}{\partial\theta_k}\Big|\Big) = O\big(j^{-\ell'}\big) \quad \text{with} \quad \ell' > 3/2 \quad \text{and}$$
$$\sum_{j=1}^{\infty} \sup_{\theta\in\Theta} \max\Big(\Big|\frac{\partial^2 b_j^+(\theta)}{\partial\theta_k\partial\theta_{k'}}\Big|, \Big|\frac{\partial^2 b_j^-(\theta)}{\partial\theta_k\partial\theta_{k'}}\Big|\Big) < \infty, \tag{6.28}$$

then the QMLE $\widehat{\theta}_n$ is asymptotically normal, i.e. satisfies (6.20).

This result is new and seems to be very interesting for the estimation of TARCH processes parameters. We focuss now on vector valued processes such that $f \equiv 0$.

6.4.4 Multivariate ARCH(∞) processes

As for the scalar case, multivariate ARCH(∞) processes generalize multivariate GARCH(q, q') models presented below. Multivariate ARCH(∞) processes are defined from relation (6.1) with a symmetric $M_\theta = H_\theta^{1/2}$ (and therefore $M_\theta(M_\theta)' = H_\theta$) and

$$H_\theta(X_{t-1}, X_{t-2}, \ldots) := B_0(\theta_0) + \sum_{i=1}^{\infty} B_i(\theta_0) X_{t-i} X_{t-i}' B_i'(\theta_0), \tag{6.29}$$

where B_i are function with values in non-negative definite $d \times d$ matrices. In such a case, Assumptions ST(r) is satisfied when

$$\sum_{j=1}^{\infty} \|B_j\|_\Theta < \Big(\mathbb{E}\big[\|\xi_0\|^r\big]\Big)^{-1/r} \tag{6.30}$$

since $\alpha_j(M) = \|B_j\|_\Theta$. Moreover, $\alpha_j(H) = \|B_j\|_\Theta^2$. Like for univariate ARCH(∞) processes, Theorems 6.1 and 6.2 imply :

Property 6.5 *Let $X = (X_t)_{t\in\mathbb{Z}}$ be a r-order multivariate stationary $ARCH(\infty)$ with (6.29) where condition (6.30) holds together with assumption Id. Moreover, assume that* $\inf_{\theta\in\Theta} \det B_0(\theta) > 0$ *and $\theta \in \Theta \mapsto B_j(\theta)$ are injective continuous functions for all $j \in \mathbb{N}$. Then,*

1. *if $r = 2$, $\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta_0$.*
2. *if $r = 4$, and if the functions $\theta \in \Theta \mapsto b_j(\theta)$ are 2-times continuously differentiable for all $j \in \mathbb{N}$ and such that for all $(k, k') \in \{1, \ldots, d\}^2$,*

$$\Big\|\frac{\partial B_j(\theta)}{\partial \theta_k}\Big\|_\Theta = O\big(j^{-\ell'}\big) \quad \textit{with} \quad \ell' > 3/2 \quad \textit{and} \quad \sum_{j=1}^{\infty} \Big\|\frac{\partial^2 B_j(\theta)}{\partial \theta_k \partial \theta_{k'}}\Big\|_\Theta < \infty,$$

then the QMLE $\widehat{\theta}_n$ is asymptotically normal, i.e. satisfies (6.20).

As far as the authors know, such models are introduced here for the first time.

6.4.5 Multivariate GARCH(q, q') models

Classically multivariate GARCH(q, q') models are shared in two subcases, the BEKK and VEC models. The next section is devoted to ARMA-GARCH processes for which VEC models are subcases. BEKK processes are solutions of equation (6.29) or equivalently

$$vec(H_t) = vec(C_0) + \sum_{i=1}^{q} C_i(\theta_0)^* vec(X_{t-i}X'_{t-i}) + \sum_{i=1}^{p} D_i(\theta_0)^* vec(H_{t-i}),$$

with vec the operator that stacks the column of a matrix, and for any $p \times k$ matrix A : $A_i^* = \sum_{j=1}^{k} A_{i,j} \otimes A_{i,j}$ for $i = 1, \ldots, p$, $\otimes$ denoting the Kronecker product. The multivariate $ARCH(\infty)$ representation holds with B_j such as

$$B_0^* := \Big(1 - \sum_{j=1}^{q'} D_j^*\Big)^{-1} \times C_0^* \quad \text{and} \quad \sum_{i=1}^{\infty} B_i^* Z^i := \Big(1 - \sum_{i=1}^{q'} D_i^* Z^i\Big)^{-1} \times \sum_{i=1}^{q} C_i^* Z^i \quad \text{for all} \quad Z \in \mathbb{C}^m$$

(in the last formula, both the polynoms are supposed to be coprime). Here Assumptions ST(r) is satisfied when

$$\sum_{j=1}^{\infty} \|B_j\|_\Theta < \Big(\mathbb{E}\big[\|\xi_0\|^r\big]\Big)^{-1/r}. \tag{6.31}$$

since $\alpha_j(M) = \|B_j\|_\Theta$.

The natural case where $\theta = (C_0, C_1, \ldots, c_q, D_1, \ldots, D_{q'})$ implies that assumptions $\mathrm{D}^{(1)}$, $\mathrm{D}^{(2)}$ are automatically satisfied.

Property 6.6 *Let $X = (X_t)_{t\in\mathbb{Z}}$ be a r-order stationary solution of (6.29) where condition (6.31) holds together with assumption Id. Then, when* $\inf_{\theta\in\Theta} \det C_0(\theta) > 0$

1. *If $r = 2$ then $\widehat{\theta}_n \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} \theta_0$.*
2. *If $r = 4$, then the QMLE $\widehat{\theta}_n$ is asymptotically normal, i.e. satisfies (6.20).*

The asymptotic normality was settled before by Comte and Libermann under bounded moments of order 8. Our result need just a moment of order 4.

6.4.6 Multivariate LARCH(∞) and NLARCH(∞) models

In this section let $\{B_j(\theta)\}_{j\in\mathbb{N}^*}$ be a sequence of real matrices $m \times d$, and $B_0(\theta)$ be a real vector of dimension m. The LARCH(∞) process is defined in Giraitis *et al.* [58] and Doukhan *et al.* [47] as the solution to the recurrence equation :

$$X_t = \zeta_t \left(B_0(\theta_0) + \sum_{j=1}^{\infty} B_j(\theta_0) X_{t-j} \right).$$

Remark that the innovations $(\zeta_t)_{t\in\mathbb{Z}}$ are classically in that context random matrices. A generalization of LARCH(∞) models (see Doukhan and Wintenberger [49]) is given by equation

$$X_t \;=\; \zeta_t \left(B_0(\theta) + \sum_{j=1}^{\infty} B_j(\theta)(X_{t-j}) \right). \tag{6.32}$$

where now $B_j(\theta) : \mathbb{R}^m \to \mathbb{R}^p$ are $b_j(\theta)$-Lipschitz functions. When the innovations in this model are concentrated on the diagonal, it is possible to rewrite it as

$$X_t = M_{\theta_0}(X_{t-1}, \ldots)\xi_t,$$

where $(\xi_t)_i = (\zeta_t)_{i,i}$ and $(M_\theta(X_{t-1}, \ldots))_{i,i} = (B_0(\theta) + \sum_{j=1}^{\infty} B_j(\theta) X_{t-j})_i$.

A non linear case is when $(B_j(\theta)(x))_k = \sum_{i=1}^m B_{j,i,k}(\theta)(x)^+ \max\{x_i, 0\} + B_{j,i,k}(\theta)(x)^- \min\{x_i, 0\}$ for all $1 \leqslant k \leqslant p$. Those models fulfill the Assumption ST(r) as soon as

$$\sum_{j=1}^{\infty} \|b_j\|_\Theta < \Big(\mathbb{E}\big[\|\xi_0\|^r\big]\Big)^{-1/r}. \tag{6.33}$$

We derive the consistency and the asymptotic normality :

Property 6.7 *Let $X = (X_t)_{t\in\mathbb{Z}}$ be a r-order stationary solution of (6.32) where condition (6.33) holds together with assumption Id. Moreover, assume that $\theta \in \Theta \mapsto B_j(\theta)$ are injective continuous functions for all $j \in \mathbb{N}$ and that $\inf_{\theta\in\Theta} \det \Pi_V B_0(\theta)\big(\Pi_V B_0(\theta)\big)' > 0$ where Π_A is the projector on the subspace A of $\mathbb{R}^m$ and V is the orthogonal of Vect$\{B_j, j \geqslant 1\}$. When*

$$\|b_j(\theta)\|_\Theta = O\big(j^{-\ell}\big) \quad \text{with} \quad \ell > 3/2,$$

1. *if $r = 2$ the consistency $\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta_0$ holds.*
2. *if $r = 4$, the QMLE $\widehat{\theta}_n$ is asymptotically normal, i.e. satisfies (6.20) when*

$$\left\| Lip \, \frac{\partial B_j}{\partial \theta_{k_1}} \right\|_\Theta = O\big(j^{-\ell'}\big) \ \ \text{with} \ \ \ell' > 3/2 \quad \text{and} \ \sum_{j=1}^{\infty} \left\| Lip \, \frac{\partial^2 B_j}{\partial \theta_{k_1} \partial \theta_{k_2}} \right\|_\Theta < \infty \ \text{for all} \ k_1, k_2 \in \{1, \ldots, d\}.$$

In that context, the parametrization of the functions $B_j(\theta)$ with respect to θ has not yet been discussed. Moreover, the Assumption M restrict the application of our result. For example, the unidimensionnal LARCH model cannot be treated because $M_\theta = c(\theta) + \sum_{j=1}^{\infty} c_j(\theta) X_{t-j}$ can be null. The bilinear models of Giraitis and Surgailis [61]

$$X_t = \xi_t \left(a(\theta_0) + \sum_{j=1}^{\infty} a_j(\theta_0) X_{t-j} \right) + b(\theta_0) + \sum_{j=1}^{\infty} b_j(\theta_0) X_{t-j}$$

could be seen as a vector LARCH model and then could possibly be treated under restrictions that ensure Assumption M.

6.4.7 Multivariate Nonlinear AR(∞) models

Now let just focuss on one case where $M = I_m$ and $f \neq 0$. In this context, assumption M is always satisfied. The QMLE is here also the least square estimator. Here

$$f_\theta(X_{t-1}, X_{t-2}, \ldots) = A_0(\theta_0) + \sum_{i=1}^{\infty} A_i(\theta_0) X_{t-i},$$

where A_0 and A_i are Lipschitz functions with values in positive definite $d \times d$ matrices.

Assumption ST(r) writes in that context as

$$\sum_{j=1}^{\infty} \|\mathrm{Lip}\, A_j\|_\Theta < 1. \tag{6.34}$$

The consistency and the asymptotic normality follows

Property 6.8 *Let $X = (X_t)_{t\in\mathbb{Z}}$ be a nonlinear AR(∞) process where condition (6.34) holds together with assumption Id. Then,*

1. *If $r = 2$ then $\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta_0$.*
2. *If $r = 4$, then the QMLE $\widehat{\theta}_n$ is asymptotically normal, i.e. satisfies (6.20), when*

$$\Big\| Lip\, \frac{\partial A_j}{\partial \theta_k} \Big\|_\Theta = O\big(j^{-\ell'}\big) \quad \textit{with} \quad \ell' > 3/2 \quad \textit{and} \quad \sum_{j=1}^{\infty} \Big\| Lip\, \frac{\partial^2 A_j}{\partial \theta_k \theta_{k'}} \Big\|_\Theta < \infty \textit{ for all } k, k' \in \{1, \dots, d\}.$$

6.4.8 Multivariate ARMA-GARCH models

In all that section M_θ is concentrated on its diagonal. Here f is not identically null. The vector ARMA-GARCH model is introduced by Ling and McAleer [84] as the solution of

$$\Phi(L)(X_t - \mu) = \Psi(L)\varepsilon_t, \tag{6.35}$$

$$\varepsilon_t = M_\theta(X_{t-1}, X_{t-2}, \dots)\xi_t, \quad diag(H_t) = C_0 + \sum_{i=1}^{q} C_i diag(\varepsilon_{t-i}\varepsilon'_{t-i}) + \sum_{i=1}^{q'} D_i diag(H_{t-i}), \tag{6.36}$$

where $diag$ is the operator that stacks the diagonal of a matrix, $\Phi(L) = I_m - \Phi_1 L - \cdots - \Phi_s L^s$ and $\Psi(L) = I_m - \Psi_1 L - \cdots - \Psi_{s'} L^{s'}$ are polynomials in the lag operator L. Assuming that Ψ and Φ are coprime, the equation (6.35) has the AR(∞) representation $\Psi^{-1}(L)\Phi(L)(X_t - \mu) = \varepsilon_t = M_\theta(X_{t-1}, X_{t-2}, \dots)\xi_t$. The vector ARMA-GARCH model is then a subscase of 6.1 with $f_\theta(x_1, \dots) = \sum_{i=1}^{\infty} \Gamma_i(x_i - \mu)$, where $\Gamma(L) = I_m + \sum_{i=1}^{\infty} \Gamma_i L^i = \Psi^{-1}(L)\Phi(L)$. In the specific case $f \equiv 0$, the model is the classical VEC one studied by Jeantheau [74].

Assumption ST(r) is sastisfied when

$$\sum_{i=1}^{\infty} \|\Gamma_i\|_\Theta + \Big(\mathbb{E}\big[\|\xi_0\|^r\big]\Big)^{1/r} \sum_{j=1}^{\infty} \|B_j\|_\Theta < 1, \tag{6.37}$$

where the matrices B_j are such as

$$\sum_{i=1}^{\infty} B_i Z^i := \left(1 - \sum_{i=1}^{q'} D_i Z^i\right)^{-1} \times \sum_{i=1}^{q} C_i Z^i \quad \text{for all} \quad Z \in \mathbb{C}^m$$

when both the polynoms are supposed to be coprime. This Assumption for bounded moments has to be compared with the one of the Theorem 2.1. of Ling and McAleer [84]. But both of them are practically uncomputable.

The natural case where $\theta = (\mu, \Phi_1, \dots, \Phi_s, \Psi_1, \dots, \Psi_{s'}, C_0, C_1, \dots, c_q, D_1, \dots, D_{q'})$ implies that assumptions $\mathrm{D}^{(1)}$, $\mathrm{D}^{(2)}$ are automatically satisfied.

Property 6.9 *Let $X = (X_t)_{t\in\mathbb{Z}}$ be a r-order stationary solution of (6.35) and (6.36) where condition (6.37) holds together with assumption Id. Then, if* $\inf_{\theta\in\Theta} \det C_0 C_0' > 0$, $\|\Gamma_j(\theta)\|_\Theta = O\big(j^{-\ell}\big)$ *with* $\ell > 3/2$, *and*

1. $r = 2$, *he QMLE $\widehat{\theta}_n$ is consistent $\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta_0$.*
2. $r = 4$, *the QMLE $\widehat{\theta}_n$ is asymptotically normal, i.e. satisfies (6.20), when*

$$\left\|\frac{\partial \Gamma_j}{\partial \theta_k}\right\|_\Theta = O\big(j^{-\ell'}\big) \quad \text{with} \quad \ell' > 3/2 \quad \text{and} \quad \sum_{j=1}^\infty \left\|\frac{\partial^2 \Gamma_j}{\partial \theta_k \theta_{k'}}\right\|_\Theta < \infty \text{ for all } k, k' \in \{1, \dots, d\}.$$

Ling and McAleer [84] provided consistency and asymptotic normality for the QMLE of vactor ARMA-GARCH processes.

- Concerning the consistency, they obtained the consistency under bounded moments of order 2 and derivative conditions. We improve their results. In the special VEC context $f \equiv 0$, the (strong) consistency of Jeantheau [74], based on a powerful theorem, holds under a bounded moment in logarithm and without assumptions on the derivative. Our results is not as good in that special case ;
- concerning the asymptotic normality, we improve the result of Ling and McAleer [84] which needs bounded moments of order 6.

6.5 Proofs

6.5.1 Proof of Theorem 6.1

The proof of the theorem is divided in two parts. In (i) a uniform (in θ) law of large number on $(\widehat{q}_t)_{t\in\mathbb{N}^*}$ (defined in (6.5)) is established. In (ii), it is proved that $L(\theta) := -\mathbb{E}(q_t(\theta))/2$ has a unique maximum in θ_0. Those two conditions leads to the consistency of $\widehat{\theta}_n$ (see Wald for instance).

(i) Using Proposition 6.2, with $q_t = G(X_t, X_{t-1}, \cdots)$, one deduces that $(q_t)_{t\in\mathbb{Z}}$ (defined in (6.3)) is a stationary ergodic sequence. From Straumann and Mikosch [105], we know that if $(v_t)_{t\in\mathbb{Z}}$ is a stationary ergodic sequence of random elements with values in $\mathbb{C}(\Theta, \mathbb{R}^m)$, then the uniform (in $\theta \in \Theta$) law of large numbers is implied by $\mathbb{E}\|v_0\|_\Theta < \infty$. As a consequence, $(q_t)_{t\in\mathbb{Z}}$ satisfies a uniform (in $\theta \in \Theta$) strong law of large numbers as soon as $\mathbb{E}\big[\sup_\theta |q_t(\theta)|\big] < \infty$. But, from the inequality $\log(x) \le x-1$ for all $x \in]0,\infty[$ and Lemma 6.1, for all $t \in \mathbb{Z}$,

$$
\begin{aligned}
|q_t(\theta)| &\leqslant \frac{\|X_t - f_t(\theta)\|^2}{(\underline{M})^{1/m}} + m\left|\frac{1}{m}\log \underline{M} + \frac{\|M_\theta^t (M_\theta^t)'\|}{\underline{M}^{1/m}} - 1\right| \quad \text{for all } \theta \in \Theta \\
\Longrightarrow \sup_{\theta\in\Theta} |q_t(\theta)| &\leq \frac{\|X_t - f_t(\theta)\|_\Theta^2}{(\underline{M})^{1/m}} + \Big|\log \underline{M}\Big| + m \times \frac{\|M_\theta^t (M_\theta^t)'\|_\Theta}{\underline{M}^{1/m}}
\end{aligned} \tag{6.38}
$$

because $v'A^{-1}v \le \|v'v\|\|A^{-1}\|$ and $\|A^{-1}\| = 1/\|A\| \le 1/(\det A)^{1/m}$ for $A \in \mathcal{M}_m(\mathbb{R})$, invertible, and $v \in \mathbb{R}^m$. But for all $t \in \mathbb{Z}$, $\mathbb{E}\|X_t\|^r < \infty$ (see Proposition 6.1) and $\mathbb{E}\big[\|f_\theta^t\|_\Theta^r\big] < \infty$ under Assumption F (see Lemma 6.1). Then, with $r \ge 2$, $\mathbb{E}\big[\|X_t - f_t(\theta)\|_\Theta^2\big] < \infty$. Moreover, for all $t \in \mathbb{Z}$, $\mathbb{E}\big[\|M_\theta^t \big(M_\theta^t\big)'\|_\Theta\big] < \infty$ from Lemma 6.1. As a consequence the right hand side of (6.38) admits a moment of order 1 uniformly in θ and therefore

$$
\mathbb{E}\big[\sup_{\theta\in\Theta} |q_t(\theta)|\big] < \infty.
$$

The uniform strong law of large number on $(q_t(\theta))$ directly follows and thus

$$
\Big\|\frac{L_n(\theta)}{n} - L(\theta)\Big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0 \quad \text{with} \quad L(\theta) := -\frac{1}{2}\mathbb{E}\big[q_0(\theta)\big]. \tag{6.39}
$$

Now, one shows that $\dfrac{1}{n}\|\widehat{L}_n - L_n\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0$. Indeed, for all $\theta \in \Theta$ and $t \in \mathbb{N}^*$,

$$
\begin{aligned}
\big|\widehat{q}_t(\theta) - q_t(\theta)\big| =& (X_t - \widehat{f}_\theta^t)'\big(\widehat{M}_\theta^t(\widehat{M}_\theta^t)'\big)^{-1}\big(X_t - \widehat{f}_\theta^t\big) - (X_t - f_\theta^t)'\big(M_\theta^t(M_\theta^t)'\big)^{-1}\big(X_t - f_\theta^t\big) \\
&+ \Big(\log\Big(\det\big(\widehat{M}_\theta^t(\widehat{M}_\theta^t)'\big)\Big) - \log\Big(\det\big(M_\theta^t(M_\theta^t)'\big)\Big)\Big) \\
\le& (X_t - \widehat{f}_\theta^t)'\Big[\big(\widehat{M}_\theta^t(\widehat{M}_\theta^t)'\big)^{-1} - \big(M_\theta^t(M_\theta^t)'\big)^{-1}\Big]\big(X_t - \widehat{f}_\theta^t\big) \\
&+ (2X_t - \widehat{f}_\theta^t - f_\theta^t)'\big(M_\theta^t(M_\theta^t)'\big)^{-1}\big(f_\theta^t - \widehat{f}_\theta^t\big) + \frac{1}{|C|}\Big|\det\big(\widehat{M}_\theta^t(\widehat{M}_\theta^t)'\big) - \det\big(M_\theta^t(M_\theta^t)'\big)\Big| \\
\le& 2\big(\|X_t\| + \|\widehat{f}_\theta^t\|_\Theta\big)\big\|\big(\widehat{M}_\theta^t(\widehat{M}_\theta^t)'\big)^{-1} - \big(M_\theta^t(M_\theta^t)'\big)^{-1}\big\|_\Theta \\
&+ \frac{1}{\underline{M}}\Big\|\det\big(\widehat{M}_\theta^t(\widehat{M}_\theta^t)'\big) - \det\big(M_\theta^t(M_\theta^t)'\big)\Big\|_\Theta \\
&+ \Big(2\|X_t\| + \|\widehat{f}_\theta^t\|_\Theta + \|f_\theta^t\|_\Theta\Big)\big\|\big(M_\theta^t(M_\theta^t)'\big)^{-1}\big\|_\Theta \big\|f_\theta^t - \widehat{f}_\theta^t\big\|_\Theta
\end{aligned} \tag{6.40}
$$

from the mean value theorem, with $C \in [\det\big(M_\theta^t(M_\theta^t)'\big), \det\big(\widehat{M}_\theta^t(\widehat{M}_\theta^t)'\big)]$ and therefore $|C| > \underline{M}$. In one hand,

$$
\|\big(\widehat{M}_\theta^t(\widehat{M}_\theta^t)'\big)^{-1} - \big(M_\theta^t(M_\theta^t)'\big)^{-1}\|_\Theta \;\le\; \big\|\big(\widehat{M}_\theta^t(\widehat{M}_\theta^t)'\big)^{-1}\big\|_\Theta \cdot \big\|\widehat{M}_\theta^t(\widehat{M}_\theta^t)' - M_\theta^t(M_\theta^t)'\big\|_\Theta \cdot \big\|\big(M_\theta^t(M_\theta^t)'\big)^{-1}\big\|_\Theta
$$

In the other hand, for $A \in \mathcal{M}_m(\mathbb{R})$ an invertible squared matrix, and $H \in \mathcal{M}_m(\mathbb{R})$,

$$
\det(A+H) = \det(A) + \det(A)\cdot \mathrm{Tr}\big((A^{-1})'H\big) + o(\|H\|).
$$

From the assumptions, $\|\big(M_\theta^t(M_\theta^t)'\big)^{-1}\|_\Theta \leq \underline{M}^{-m}$ for all $t\in\mathbb{N}^*$. Moreover, $\big|\mathrm{Tr}\big((A^{-1})'H\big)\big| \leq \big\|A^{-1}\big\|\cdot\big\|H\big\|$. Thus, it exists $C>0$ not depending on t such that inequality (6.40) becomes for all $t\in\mathbb{N}^*$:

$$\begin{aligned}
\sup_{\theta\in\Theta}\big|\widehat{q}_t(\theta)-q_t(\theta)\big| &\leq C\big(\|X_t\|+\|\widehat{f}_\theta^t\|_\Theta+\|f_\theta^t\|_\Theta\big)\times\\
&\quad\Big(\big\|\widehat{M}_\theta^t(\widehat{M}_\theta^t)'-M_\theta^t(M_\theta^t)'\big\|_\Theta+\big\|f_\theta^t-\widehat{f}_\theta^t\big\|_\Theta\Big) \qquad (6.41)\\
\Longrightarrow\quad \mathbb{E}\big[\sup_{\theta\in\Theta}\big|\widehat{q}_t(\theta)-q_t(\theta)\big|^{2/3}\big] &\leq C'\mathbb{E}\Big[\big(\|X_t\|^{2/3}+\|\widehat{f}_\theta^t\|_\Theta^{2/3}+\|f_\theta^t\|_\Theta^{2/3}\big)\\
&\quad\big(\big\|\widehat{M}_\theta^t(\widehat{M}_\theta^t)'-M_\theta^t(M_\theta^t)'\big\|_\Theta^{2/3}+\big\|f_\theta^t-\widehat{f}_\theta^t\big\|_\Theta^{2/3}\big)\Big]\\
&\leq C''\Big(\mathbb{E}\big[\|X_t\|^2\big]+\mathbb{E}\big[\|\widehat{f}_\theta^t\|_\Theta^2\big]+\mathbb{E}\big[\|f_\theta^t\|_\Theta^2\big]\Big)^{1/3}\\
&\quad\times\Big(\mathbb{E}\big[\big\|\widehat{M}_\theta^t(\widehat{M}_\theta^t)'-M_\theta^t(M_\theta^t)'\big\|_\Theta\big]+\mathbb{E}\big[\big\|f_\theta^t-\widehat{f}_\theta^t\big\|_\Theta\big]\Big)^{2/3}\\
&\leq C'''\Big(\sum_{j\geqslant t}\big[\alpha_j(f)+\alpha_j(M)\big]\Big)^{2/3}, \qquad (6.42)
\end{aligned}$$

from Hölder and Minkovski Inequalities and thanks to $r\geq 2$ (and therefore $r/2\geq 1$), with $C'>0$ $C''>0$ and $C''>0$ not depending on θ and t (the choice of the power $2/3$ for $\big|\widehat{q}_t(\theta)-q_t(\theta)\big|$ is optimal).
Now, consider for $n\in\mathbb{N}^*$,

$$S_n := \sum_{t=1}^n \frac{1}{t}\sup_{\theta\in\Theta}\big|\widehat{q}_t(\theta)-q_t(\theta)\big|.$$

Applying the Kronecker Lemma (see for instance Stout, 1974), if $\lim_{n\to\infty} S_n<\infty$ a.s. then $\frac{1}{n}\cdot\|\widehat{L}_n - L_n\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0$. Following the Stout's arguments, it remains to show that for all $\varepsilon>0$,

$$\mathbb{P}(\forall n\in\mathbb{N},\ \ \exists m>n \text{ such that } |S_m-S_n|>\varepsilon) := \mathbb{P}(A)=0. \qquad (6.43)$$

Let $\varepsilon>0$ and denote

$$A_{m,n} := \{|S_m-S_n|>\varepsilon\}$$

for $m>n$. Remark that $A=\bigcap_{n\in\mathbb{N}}\bigcup_{m>n}A_{m,n}$. For $n\in\mathbb{N}^*$, the sequence of sets $(A_{m,n})_{m>n}$ is obviously increasing, and if $A_n:=\bigcup_{m>n}A_{m,n}$, then $\lim_{m\to\infty}\mathbb{P}(A_{m,n})=\mathbb{P}(A_n)$. Remark that $(A_n)_{n\in\mathbb{N}}$ is a decreasing sequence of sets and thus,

$$\lim_{m\to\infty}\lim_{m\to\infty}\mathbb{P}(A_{m,n})=\lim_{n\to\infty}\mathbb{P}(A_n)=\mathbb{P}(A).$$

It remains to bound $\mathbb{P}(A_{m,n})$. From the Bienaymé-Tchebytchev Inequality,

$$\begin{aligned}
\mathbb{P}(A_{m,n})=\mathbb{P}\left(\sum_{t=n+1}^m\frac{1}{t}\sup_{\theta\in\Theta}\big|\widehat{q}_t(\theta)-q_t(\theta)\big|>\varepsilon\right) &\leqslant \frac{1}{\varepsilon^{2/3}}\mathbb{E}\left[\Big(\sum_{t=n+1}^m\frac{1}{t}\sup_{\theta\in\Theta}\big|\widehat{q}_t(\theta)-q_t(\theta)\big|\Big)^{2/3}\right]\\
&\leqslant \frac{1}{\varepsilon^{2/3}}\sum_{t=n+1}^m\frac{1}{t^{2/3}}\mathbb{E}\Big[\sup_{\theta\in\Theta}\big|\widehat{q}_t(\theta)-q_t(\theta)\big|^{2/3}\Big].
\end{aligned}$$

Using (6.42) and condition (6.15), it exists $C > 0$ such that $\Big(\sum_{j=t}^{\infty} \alpha_j(f) + \alpha_j(M)\Big)^{2/3} \leq C \frac{1}{t^{2(\ell-1)/3}}$, since $\ell > 3/2$. Thus, $\frac{1}{t^{2/3}} \mathbb{E}\Big[\sup_{\theta \in \Theta} \big|\widehat{q}_t(\theta) - q_t(\theta)\big|^{2/3}\Big] = O\big(t^{-2\ell/3}\big)$ and,

$$\sum_{t=1}^{\infty} \frac{1}{t^{2/3}} \mathbb{E}\Big[\sup_{\theta \in \Theta} \big|\widehat{q}_t(\theta) - q_t(\theta)\big|^{2/3}\Big] < \infty.$$

As a consequence, for all $n \in \mathbb{N}^*$

$$\begin{aligned} & \mathbb{P}(A_n) \leqslant \frac{1}{\varepsilon^{2/3}} \sum_{t=n+1}^{\infty} \frac{1}{t^{2/3}} \mathbb{E}\Big[\sup_{\theta \in \Theta} \big|\widehat{q}_t(\theta) - q_t(\theta)\big|^{2/3}\Big] \\ \Longrightarrow \quad & \mathbb{P}(A_n) \underset{n \to \infty}{\longrightarrow} 0. \end{aligned}$$

Finally, $\mathbb{P}(A) = 0$ and then $\frac{1}{n} \cdot \|\widehat{L}_n - L_n\|_\Theta \xrightarrow[n \to \infty]{a.s.} 0$.

(ii) Using usual Kullback information for Gaussian multivariate densities (it is classical results for multivariate QMLE, see for instance Jentheau, 1998), Assumption Id implies that $\theta \mapsto L(\theta)$ has a unique maximum in θ_0. $\square$

6.5.2 Proof of Corollary 6.1

In the previous proof of Theorem 6.1, the distance $\sup_{\theta \in \Theta} \big|\widehat{q}_t(\theta) - q_t(\theta)\big|$ is depending on the distance $\big\|\widehat{M}_\theta^t (\widehat{M}_\theta^t)' - M_\theta^t (M_\theta^t)'\big\|_\Theta = \big\|\widehat{H}_\theta^t - H_\theta^t\big\|_\Theta$. Therefore, by using from Condition 6.16, one obtains also

$$\sup_{\theta \in \Theta} \big|\widehat{q}_t(\theta) - q_t(\theta)\big| \quad \leq \quad C''' \Big(\sum_{j \geqslant t} \big[\alpha_j(f) + \alpha_j(H)\big]\Big)^{2/3}.$$

It implies that the QMLE $\widehat{\theta}_n$ defined by (6.7) is consistent since Condition 6.15 yields. $\square$

6.5.3 Proof of Theorem 6.2

One begins by showing the following Lemma :

Lemma 6.2 *Let Assumptions ST(r) with $r \geq 4$, $D^{(1)}$, M and $(X_t)_{t \in \mathbb{Z}}$ be a stationary solution of equation (6.1). Then,*

$$n^{-1/2} \frac{\partial}{\partial \theta} L_n(\theta_0) \xrightarrow[n \to \infty]{\mathcal{D}} \mathcal{N}_d(0, G(\theta_0)), \tag{6.44}$$

where the matrix $G(\theta_0) = (G(\theta_0))_{1 \leq i,j \leq d}$ is given in (6.47).

Proof of Lemma 6.2

First we would like to prove that vector $\dfrac{\partial L_n}{\partial\theta} = (\dfrac{\partial L_n}{\partial\theta_i})_{1\le i\le d}$ exists and is uniformly continuous for all $\theta\in\Theta$. It requires to show the same property for $\dfrac{\partial\big(M_\theta^t(M_\theta^t)'\big)}{\partial\theta} = (\dfrac{\partial\big(M_\theta^t(M_\theta^t)'\big)}{\partial\theta_i})_{1\le i\le d}$ and $\dfrac{\partial f_\theta^t}{\partial\theta} = (\dfrac{\partial f_\theta^t}{\partial\theta_i})_{1\le i\le d}$ for all $t\in\mathbb{Z}$. Only the case $\dfrac{\partial\big(M_\theta^t(M_\theta^t)'\big)}{\partial\theta}$ is now studied, the case of $\dfrac{\partial f_\theta^t}{\partial\theta}$ may be solved similarly. Denote by $\mathcal{D}^{(1)}\mathcal{C}(\Theta,\mathcal{M}_m)$ the Banach space of continuously differentiable functions with values in $\mathcal{M}_m$, equipped with the uniform norm

$$\|g\|_{(1),\Theta} = \|g\|_\Theta + \sum_{i=1}^d \Big\|\frac{\partial g}{\partial\theta_i}\Big\|_\Theta.$$

Under Assumption M and with $\theta\to M_\theta$ satisfying Assumption $D^{(1)}$, for all sequence $(x_i)_{i\in\mathbb{N}}\in(\mathbb{R}^m)(\mathbb{N})$, $M_\theta((x_i)_{i\in\mathbb{N}})$ and therefore $M_\theta((x_i)_{i\in\mathbb{N}})(M_\theta((x_i)_{i\in\mathbb{N}}))'$ belong to $\mathcal{D}^{(1)}\mathcal{C}(\Theta,\mathcal{M}_m)$. With the same method and notations than in the proof of Lemma 6.1, for $p\in\mathbb{N}$, the function $\theta\in\Theta\to h^{t,p}(\theta)$ is proved to satisfy a Cauchy criteria in $\mathbb{L}^{r/2}\big(\mathcal{D}^{(1)}\mathcal{C}(\Theta,\mathcal{M}_m)\big)$. Indeed, previous arguments imply that $h^{t,p}$ belongs to $\mathcal{D}^{(1)}\mathcal{C}(\Theta,\mathcal{M}_m)$. Moreover, the $r/2$ moments of $\|h^{t,p}(\theta)\|_{(1),\Theta}$ exist because :

$$\begin{aligned}
\big(\|h^{t,p}\|_{(1),\Theta}\big)^{r/2} &= \Big(\|h^{t,p}\|_\Theta + \sum_{i=1}^d \Big\|\frac{\partial h^{t,p}}{\partial\theta_i}\Big\|_\Theta\Big)^{r/2} \\
&\le (d+1)^{r/2-1}\Big(\big(\|M_\theta(0,0,\ldots)\|_\Theta + \sum_{j=1}^\infty \|X_{t-j}\|\alpha_j(M)\big)^r \\
&\quad +2^{r/2}\big\|M_\theta(X_{t-1},\ldots,X_{t-p},0,\ldots)\big\|_\Theta^{r/2}\sum_{i=1}^d\Big\|\frac{\partial M_\theta(X_{t-1},\ldots,X_{t-p},0,\ldots)}{\partial\theta_i}\Big\|_\Theta^{r/2}\Big) \\
&\le C\Bigg(\Big(\|M_\theta(0,0,\ldots)\|_\Theta^r + \big(\sum_{j=1}^\infty \alpha_j(M)\|X_{t-j}\|\big)^r\Big) \\
&\quad +\Big(\|M_\theta(0,0,\ldots)\|_\Theta^{r/2} + \big(\sum_{j=1}^\infty \alpha_j(M)\|X_{t-j}\|\big)^{r/2}\Big) \\
&\quad \times\sum_{i=1}^d\Big(\Big\|\frac{\partial M_\theta(0,0,\ldots)}{\partial\theta_i}\Big\|_\Theta^{r/2} + \big(\sum_{j=1}^\infty \alpha_j^{(1)}(M)\|X_{t-j}\|\big)^{r/2}\Big)\Bigg),
\end{aligned}$$

with C depending only on d. Here again using again $\mathbb{E}\|X_0\|^r < \infty$ and Hölder and Minkowsky inequalities :

$$\mathbb{E}\big[\big(\|h^{t,p}\|_{(1),\Theta}\big)^{r/2}\big] \le C' \Bigg(\|M_\theta(0,0,\ldots)\|_\Theta^r + \mathbb{E}\big[\|X_0\|^r\big]\big(\sum_{j=1}^\infty \alpha_j(M)\big)^r + \Big(\|M_\theta(0,0,\ldots)\|_\Theta^r + \mathbb{E}\big[\|X_0\|^r\big]\big(\sum_{j=1}^\infty \alpha_j(M)\big)^r\Big)^{1/2} \times \Big(\sum_{i=1}^d \Big\|\frac{\partial M_\theta(0,0,\ldots)}{\partial \theta_i}\Big\|_\Theta^r + \mathbb{E}\big[\|X_0\|^r\big]\big(\sum_{j=1}^\infty \alpha_j^{(1)}(M)\big)^r\Big)^{1/2}\Bigg),$$

with C' depending only on d. From $\sum_j \alpha_j(M) < \infty$ and $\sum_j \alpha_j^{(1)}(M) < \infty$, it is clear that it exists B_1 not depending on t and p such that $\mathbb{E}\big[\big(\|h^{t,p}\|_{(1),\Theta}\big)^{r/2}\big] \le B_1 < \infty$. In the same way as in the proof of Lemma 6.1 it is easy to show that the sequence $(h^{t,p})_{p\in\mathbb{N}^*}$ satisfies the Cauchy criteria in the Banach space $\mathbb{L}^{r/2}(\mathcal{D}^{(1)}\mathcal{C}(\Theta,\mathcal{M}_m))$. Its limit $M_\theta^t(M_\theta^t)'$ is therefore also in $\mathbb{L}^{r/2}(\mathcal{D}^{(1)}\mathcal{C}(\Theta,\mathcal{M}_m))$ and is a.e. uniformly continuously differentiable in θ. Moreover, always like in in the proof of Lemma 6.1, one obtains for all $1 \le i \le d$,

$$\mathbb{E}\Big[\Big\|\frac{\partial\big(M_\theta^t(M_\theta^t)'\big)}{\partial\theta_i}\Big\|_\Theta^{r/2}\Big] \le \mathbb{E}\big[\big(\|M_\theta^t(M_\theta^t)'\|_{(1),\Theta}\big)^{r/2}\big] \le B_1 < \infty.$$

In the same way (but more shortly), one obtains $f_\theta^t \in \mathbb{L}^r(\mathcal{D}^{(1)}\mathcal{C}(\Theta,\mathbb{R}^m))$.

Secondly, its is now possible to compute partial derivatives of $L_n(\theta)$. Indeed, using $L_n(\theta) = -\frac{1}{2}\sum_{t=1}^n q_t(\theta)$, for $1 \le k \le d$,

$$\begin{aligned}\frac{\partial q_t(\theta)}{\partial\theta_k} &= -2\Big(\frac{\partial f_\theta^t}{\partial\theta_k}\Big)'\big(M_\theta^t(M_\theta^t)'\big)^{-1}(X_t - f_\theta^t) + (X_t - f_\theta^t)'\frac{\partial\big(M_\theta^t(M_\theta^t)'\big)^{-1}}{\partial\theta_k}(X_t - f_\theta^t) \\ &\quad + \frac{1}{\det\big(M_\theta^t(M_\theta^t)'\big)}\cdot\frac{\partial\det\big(M_\theta^t(M_\theta^t)'\big)}{\partial\theta_k} \qquad (6.45)\end{aligned}$$

In one hand, $\theta \in \Theta \mapsto \dfrac{\partial q_t(\theta)}{\partial\theta_k}$ is a measurable function because :

- the function $\theta \in \Theta \mapsto \dfrac{\partial\big(M_\theta^t(M_\theta^t)'\big)^{-1}}{\partial\theta_k} = -\big(M_\theta^t(M_\theta^t)'\big)^{-1}\dfrac{\partial\big(M_\theta^t(M_\theta^t)'\big)}{\partial\theta_k}\big(M_\theta^t(M_\theta^t)'\big)^{-1}$ is measurable ;
- the function $\theta \in \Theta \mapsto \dfrac{\partial\det\big(M_\theta^t(M_\theta^t)'\big)}{\partial\theta_k} = \det\big(M_\theta^t(M_\theta^t)'\big) \times \mathrm{Tr}\Big(\big(M_\theta^t(M_\theta^t)'\big)^{-1}\dfrac{\partial\big(M_\theta^t(M_\theta^t)'\big)}{\partial\theta_k}\Big)$ is measurable.

In another hand, $\mathbb{E}\Big(\dfrac{\partial q_t(\theta_0)}{\partial \theta_k} \mid X_{t-1},\dots\Big)=0$. Indeed, by definition for all $t\in\mathbb{Z}$, $\mathbb{E}(X_t \mid X_{t-1},\dots)=f^t_{\theta_0}$ and $\mathbb{E}\Big(\big(X_t-f^t_{\theta_0}\big)\big(X_t-f^t_{\theta_0}\big)' \mid X_{t-1},\dots\Big)=M^t_{\theta_0}(M^t_{\theta_0})'$. Therefore,

$$\begin{aligned}
\mathbb{E}\Big(\frac{\partial q_t(\theta_0)}{\partial \theta_k} \mid X_{t-1},\dots\Big) &= -2\Big(\frac{\partial f^t_{\theta_0}}{\partial \theta_k}\Big)'\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}\mathbb{E}\big(X_t-f^t_{\theta_0} \mid X_{t-1},\dots\big) \\
&\quad -\mathbb{E}\Big(\big(X_t-f^t_{\theta_0}\big)'\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial\theta_k}\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}\big(X_t-f^t_{\theta_0}\big)\Big) \\
&\quad +\frac{1}{\det\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}\times\det\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)\times\mathrm{Tr}\Big(\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial\theta_k}\Big) \\
&= 0-\mathbb{E}\Big[\mathrm{Tr}\Big(\big(X_t-f^t_{\theta_0}\big)'\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial\theta_k} \\
&\quad \big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}\big(X_t-f^t_{\theta_0}\big)\Big)\Big]+\mathbb{E}\Big[\mathrm{Tr}\Big(\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial\theta_k}\Big)\Big] \\
&= \mathrm{Tr}\Big(\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial\theta_k}\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1} \\
&\quad \mathbb{E}\Big[\big(X_t-f^t_{\theta_0}\big)\Big)\big(X_t-f^t_{\theta_0}\big)'\Big]\Big)+\mathrm{Tr}\Big(\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial\theta_k}\Big) \\
&= 0,
\end{aligned}$$

always using the commutativity of products for the linear application Trace. Then $\Big(\dfrac{\partial q_t(\theta_0)}{\partial\theta},\mathcal{F}_t\Big)_{t\in\mathbb{Z}}$ is a $\mathbb{R}^m$-valued martingale difference, where $\mathcal{F}_t=\sigma(X_t,X_{t-1},\dots)$. Remark also that arguments of Proposition 6.2 proves the stationary ergodicity of this martingale difference. To conclude with the Billingsley-Ibragimov Central Limit Theorem for martingale-difference random fields (see Billingsley, 1995), we have to prove that $\mathbb{E}\Big[\Big\|\dfrac{\partial q_t(\theta_0)}{\partial\theta}\Big\|^2\Big]<\infty$. But,

$$\begin{aligned}
\Big(\frac{\partial q_t(\theta_0)}{\partial\theta_k}\Big)^2 &\le C\left(\Big\|\Big(\frac{\partial f^t_{\theta_0}}{\partial\theta_k}\Big)'\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}\big(X_t-f^t_{\theta_0}\big)\Big\|^2+\Big\|\big(X_t-f^t_{\theta_0}\big)'\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}}{\partial\theta_k}\big(X_t-f^t_{\theta_0}\big)\Big\|^2\right. \\
&\qquad \left.+\Big\|\mathrm{Tr}\Big(\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial\theta_k}\Big)\Big\|^2\right) \\
&\le C\left(\underline{M}^{-2/m}\Big(\Big\|\frac{\partial f^t_{\theta_0}}{\partial\theta_k}\Big\|^2\Big\|X_t-f^t_{\theta_0}\Big\|^2+\Big\|\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial\theta_k}\Big\|^2\Big)\right. \\
&\qquad \left.+\Big\|\big(X_t-f^t_{\theta_0}\big)'\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial\theta_k}\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}\big(X_t-f^t_{\theta_0}\big)\Big\|^2\right),
\end{aligned}$$

using usual inequalities between the norms and the traces of the product of matrix and with the bound $\big(\underline{M}\big)^{1/m}\le\inf_{\theta\in\Theta}\|M^t_{\theta_0}\big(M^t_{\theta_0}\big)'\|$ given in Lemma 6.1. Moreover, since X satisfies relation (6.1), it is clear that $X_t - f^t_{\theta_0} = M^t_{\theta_0}\xi_t$ and therefore,

$$\Big(\frac{\partial q_t(\theta_0)}{\partial \theta_k}\Big)^2 \leq C\underline{M}^{-2/m}\left(\Big\|\frac{\partial f^t_{\theta_0}}{\partial \theta_k}\Big\|^2\Big\|M^t_{\theta_0}(M^t_{\theta_0})'\Big\|\|\xi_t\|^2 + \Big\|\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial \theta_k}\Big\|^2 + \|\xi'_t\xi_t\|^2\Big\|\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial \theta_k}\Big\|^2\right)$$

$$\Longrightarrow \quad \mathbb{E}\Big[\Big(\frac{\partial q_t(\theta_0)}{\partial \theta_k}\Big)^2\Big] \leq C\underline{M}^{-2/m}\left(\mathbb{E}\Big[\Big\|\frac{\partial f^t_{\theta_0}}{\partial \theta_k}\Big\|^2\Big\|M^t_{\theta_0}(M^t_{\theta_0})'\Big\|\Big] \times \mathbb{E}\big[\|\xi_t\|^2\big] + \mathbb{E}\Big[\Big\|\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial \theta_k}\Big\|^2\Big]\right.$$
$$\left. + \,\mathbb{E}\big[\|\xi'_t\xi_t\|^2\big] \times \mathbb{E}\Big[\Big\|\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial \theta_k}\Big\|^2\Big]\right)$$

Therefore,

$$\mathbb{E}\Big[\Big\|\frac{\partial q_t(\theta_0)}{\partial \theta}\Big\|^2\Big] = \sum_{k=1}^d \mathbb{E}\Big[\Big(\frac{\partial q_t(\theta_0)}{\partial \theta_k}\Big)^2\Big] < \infty. \tag{6.46}$$

since

– $\mathbb{E}\big[\|\xi'_t\xi_t\|^2\big] < \infty$ (recall $r \geq 4$) ;

– $\mathbb{E}\Big[\Big\|\frac{\partial f^t_{\theta_0}}{\partial \theta_k}\Big\|^2\Big\|M^t_{\theta_0}(M^t_{\theta_0})'\Big\|\Big] \leq \Big(\mathbb{E}\Big[\Big\|\frac{\partial f^t_{\theta_0}}{\partial \theta_k}\Big\|^4\Big] \times \mathbb{E}\Big[\Big\|M^t_{\theta_0}(M^t_{\theta_0})'\Big\|^2\Big]\Big)^{1/2} < \infty$ from the Cauchy-Schwarz inequality for $f^t_\theta \in \mathbb{L}^r(\mathcal{D}^{(1)}\mathcal{C}(\Theta,\mathbb{R}^m))$ and $\mathbb{E}\Big[\Big\|M^t_{\theta_0}(M^t_{\theta_0})'\Big\|^2\Big] < \infty$ from Lemma 6.1 ;

– $\mathbb{E}\Big[\Big\|\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial \theta_k}\Big\|^2\Big] < \infty$ because $M^t_\theta(M^t_\theta)' \in \mathbb{L}^{r/2}(\mathcal{D}^{(1)}\mathcal{C}(\Theta,\mathcal{M}_m))$.

To conclude, the asymptotic covariance of $\dfrac{\partial q_t(\theta_0)}{\partial \theta}$ has to be computed. Thus, since $X_t - f^t_{\theta_0} = M^t_{\theta_0}\xi_t$ and ξ_t independent to $\sigma(X_{t-1}, X_{t-2}, \ldots)$, with $\mathrm{Tr}(ABC) = \mathrm{Tr}(CAB) = \mathrm{Tr}(ACB)$ for symmetric matrix A, B and C, after computations,

$$\big(G(\theta_0)\big)_{ij} = \mathbb{E}\Big[\frac{\partial q_t(\theta_0)}{\partial \theta_i}\frac{\partial q_t(\theta_0)}{\partial \theta_j}\Big] = \mathbb{E}\left[-\mathrm{Tr}\Big(\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial \theta_i}\Big)\mathrm{Tr}\Big(\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial \theta_j}\Big)\right.$$
$$\left. +4\Big(\frac{\partial f^t_{\theta_0}}{\partial \theta_i}\Big)'\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-1}\Big(\frac{\partial f^t_{\theta_0}}{\partial \theta_j}\Big) + \big(\kappa_4 + (p-1)\big)\mathrm{Tr}\Big(\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)^{-2}\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial \theta_i}\frac{\partial\big(M^t_{\theta_0}(M^t_{\theta_0})'\big)}{\partial \theta_j}\Big)\right] \tag{6.47}$$

because we also assume that ξ_t and $-\xi_t$ have the same distribution $\Longrightarrow \quad \mathbb{E}\big[\xi_t\xi'_tA\xi_t\big] = 0$ for A a matrix, and $\mathbb{E}\big[\xi_t\xi'_t\xi_t\xi'_t\big] = \big(\kappa_4 + (m-1)\big)I_p$ with I_p the $p \times p$ identity matrix. $\square$

Proof of the Theorem 6.2

A uniform law of large numbers (ULLN) for the second derivative of $q_t(\theta)$, implying a uniform a.s. convergence of the second derivative of $L_n(\theta)$, is established using the following arguments : the proof of Proposition 6.2 insures the stationary ergodicity of the second derivative of $q_t(\theta)$ (because it is a measurable function of $X_t, X_{t-1}, \ldots$) and therefore it satisfies a ULLN after proving that its first uniform moment is bounded.

But first, the a.s. existence of $\dfrac{\partial^2 q_t(\theta)}{\partial\theta^2} = \Big(\dfrac{\partial^2 q_t(\theta)}{\partial\theta_i\partial\theta_j}\Big)_{1\le i,j\le d}$ has to be proved. Thus, f_θ^t and $M_\theta^t(M_\theta^t)'$ are shown to be a.s. twice continuously differentiable functions. As previously in the proof of Lemma 6.2, only the case of $M_\theta^t(M_\theta^t)'$ is going to be studied, the case of f_θ^t being quite the same but simpler. Indeed, denote by $\mathcal{D}^{(2)}\mathcal{C}(\Theta,\mathcal{M}_m)$ the Banach space of continuously twice differentiable functions with values in $\mathcal{M}_m$, equipped with the uniform norm

$$\|g\|_{(2),\Theta} = \|g\|_\Theta + \sum_{i=1}^d \Big\|\frac{\partial g}{\partial\theta_i}\Big\|_\Theta + \sum_{i=1}^d\sum_{j=1}^d \Big\|\frac{\partial^2 g}{\partial\theta_i\partial\theta_j}\Big\|_\Theta.$$

Under Assumptions M, $D^{(0)}$, $D^{(1)}$ and $D^{(2)}$, for all sequence $(x_i)_{i\in\mathbb{N}} \in (\mathbb{R}^m)(\mathbb{N})$, $M_\theta((x_i)_{i\in\mathbb{N}})$ and therefore $M_\theta((x_i)_{i\in\mathbb{N}})(M_\theta((x_i)_{i\in\mathbb{N}}))'$ belong to $\mathcal{D}^{(2)}\mathcal{C}(\Theta,\mathcal{M}_m)$. With the same method and notations than in the proof of Lemma 6.2, for $p\in\mathbb{N}$, the function $\theta\in\Theta\to h^{t,p}(\theta)$ is proved to satisfy a Cauchy criteria in the Banach space $\mathbb{L}^{r/2}\big(\mathcal{D}^{(1)}\mathcal{C}(\Theta,\mathcal{M}_m)\big)$. Moreover, one obtains

$$\mathbb{E}\Big[\Big\|\frac{\partial^2\big(M_\theta^t(M_\theta^t)'\big)}{\partial\theta_i\partial\theta_j}\Big\|_\Theta^{r/2}\Big] \le \mathbb{E}\big[\|M_\theta^t(M_\theta^t)'\|_{(2),\Theta}^{r/2}\big] \le B_2 < \infty,$$

with $B_2\ge 0$ not depending on t. One obtains also $\mathbb{E}\big[\|f_\theta^t\|_{(2),\Theta}^r\big] < \infty$.

Secondly, it is now possible to compute second partial derivatives of $L_n(\theta)$, *i.e.* second partial derivatives of $q_t(\theta)$. Indeed, for $1\le i,j\le d$,

$$\begin{aligned}\frac{\partial^2 q_t(\theta)}{\partial\theta_i\partial\theta_j} &= -2\Big(\frac{\partial^2 f_\theta^t}{\partial\theta_i\partial\theta_j}\Big)'\big(M_\theta^t(M_\theta^t)'\big)^{-1}\big(X_t-f_\theta^t\big) + \big(X_t-f_\theta^t\big)'\frac{\partial^2\big(M_\theta^t(M_\theta^t)'\big)^{-1}}{\partial\theta_i\partial\theta_j}\big(X_t-f_\theta^t\big)\\ &+2\Big(\frac{\partial f_\theta^t}{\partial\theta_i}\Big)'\big(M_\theta^t(M_\theta^t)'\big)^{-1}\Big(\frac{\partial f_\theta^t}{\partial\theta_i}\Big) - 2\Big(\Big(\frac{\partial f_\theta^t}{\partial\theta_i}\Big)'\frac{\partial\big(M_\theta^t(M_\theta^t)'\big)^{-1}}{\partial\theta_j} + \Big(\frac{\partial f_\theta^t}{\partial\theta_j}\Big)'\frac{\partial\big(M_\theta^t(M_\theta^t)'\big)^{-1}}{\partial\theta_i}\Big)\big(X_t-f_\theta^t\big)\\ &+\mathrm{Tr}\Big(\Big(\frac{\partial\big(M_\theta^t(M_\theta^t)'\big)^{-1}}{\partial\theta_j}\Big)\Big(\frac{\partial\big(M_\theta^t(M_\theta^t)'\big)}{\partial\theta_i}\Big)\Big) + \mathrm{Tr}\Big(\big(M_\theta^t(M_\theta^t)'\big)^{-1}\Big(\frac{\partial^2\big(M_\theta^t(M_\theta^t)'\big)}{\partial\theta_i\partial\theta_j}\Big)\Big)\end{aligned}$$

Therefore, using the bound $\big\|\big(H_\theta^t\big)^{-1}\big\|_\Theta \le \underline{M}^{1/m}$ of Lemma 6.1 and usual relations between norms and traces of matrix,

$$\begin{aligned}\Big\|\frac{\partial^2 q_t(\theta)}{\partial\theta_i\partial\theta_j}\Big\|_\Theta \le C\Bigg(&\Big(\Big\|\frac{\partial^2 f_{\theta_0}^t}{\partial\theta_i\partial\theta_j}\Big\|_\Theta + \Big\|\frac{\partial H_{\theta_0}^t}{\partial\theta_j}\Big\|_\Theta\Big\|\frac{\partial f_{\theta_0}^t}{\partial\theta_i}\Big\|_\Theta + \Big\|\frac{\partial H_{\theta_0}^t}{\partial\theta_i}\Big\|_\Theta\Big\|\frac{\partial f_{\theta_0}^t}{\partial\theta_j}\Big\|\Big)\|X_t-f_{\theta_0}^t\|_\Theta\\ &+\Big(\Big\|\frac{\partial\big(H_\theta^t\big)^{-1}}{\partial\theta_i}\Big\|_\Theta\Big\|\frac{\partial\big(H_\theta^t\big)^{-1}}{\partial\theta_j}\Big\|_\Theta + \Big\|\frac{\partial^2 H_{\theta_0}^t}{\partial\theta_i\partial\theta_j}\Big\|_\Theta\Big)\|X_t-f_{\theta_0}^t\|_\Theta^2 + \Big\|\frac{\partial f_{\theta_0}^t}{\partial\theta_i}\Big\|_\Theta\Big\|\frac{\partial f_{\theta_0}^t}{\partial\theta_j}\Big\|_\Theta\Bigg)\end{aligned}$$

Since for all $t\in\mathbb{Z}$, $1\le i,j\le d$,

– $\mathbb{E}\Big[\Big\|H_\theta^t\Big\|_\Theta^{r/2}\Big] < \infty$, $\mathbb{E}\Big[\Big\|\dfrac{\partial H_\theta^t}{\partial \theta_i}\Big\|_\Theta^{r/2}\Big] < \infty$, $\mathbb{E}\Big[\Big\|\dfrac{\partial^2 H_\theta^t}{\partial \theta_i \partial \theta_j}\Big\|_\Theta^{r/2}\Big] < \infty$;

– $\mathbb{E}\big[\|f_\theta^t\|_\Theta^r\big] < \infty$, $\mathbb{E}\Big[\Big\|\dfrac{\partial f_\theta^t}{\partial \theta_i}\Big\|_\Theta^r\Big] < \infty$ and $\mathbb{E}\Big[\Big\|\dfrac{\partial^2 f_\theta^t}{\partial \theta_i \partial \theta_j}\Big\|_\Theta^r\Big] < \infty$;

– ξ_t is independent of M_{θ_0} and $f_{\theta_0}^t$, with $\mathbb{E}\big[\|\xi_t\|^r\big] < \infty$.

To conclude that $\mathbb{E}\Big\|\dfrac{\partial^2 q_t(\theta)}{\partial \theta_i \partial \theta_j}\Big\|_\Theta^{r/4} < \infty$ with Cauchy-Schwartz inequalities in the case $r \geq 4$, it remains to show that $\mathbb{E}\Big\|\dfrac{\partial \big(H_\theta^t\big)^{-1}}{\partial \theta_j}\Big\|_\Theta^r < \infty$ for all j. This derivative is equal to $-\big(H_\theta^t\big)^{-1}\dfrac{\partial H_\theta^t}{\partial \theta_k}\big(H_\theta^t\big)^{-1}$. Replacing $H_\theta^t = M_\theta^t M_\theta^{t\prime}$ we obtain

$$\Big\|\frac{\partial \big(H_\theta^t\big)^{-1}}{\partial \theta_j}\Big\|_\Theta \leqslant 2\underline{M}^{3/2}\Big\|\frac{\partial M_\theta^t}{\partial \theta_j}\Big\|_\Theta.$$

We conclude using the fact that $\mathbb{E}\Big\|\dfrac{\partial M_\theta^t}{\partial \theta_j}\Big\|_\Theta^r < \infty$. As a consequence, the uniform strong law of large number on $\frac{\partial^2 q_t(\theta)}{\partial \theta_i \partial \theta_j}$ holds if $r \geqslant 4$. A similar ergodic argument as in Proposition 6.2 yields

$$\Big\|\frac{1}{n}\frac{\partial^2 L_n}{\partial \theta^2} - \frac{\partial^2 L}{\partial \theta^2}\Big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0 \quad \text{with} \quad \frac{\partial^2 L}{\partial \theta^2}(\theta) := -\frac{1}{2}\mathbb{E}\Big[\frac{\partial^2 q_0}{\partial \theta^2}(\theta)\Big]. \tag{6.48}$$

The end of the proof is very classical. All the assumptions of Theorem 6.1 are satisfied and

$$\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta_0. \tag{6.49}$$

Consequently, a Taylor expansion can be established for the first derivative of L_n and therefore for all $n \geq 1$,

$$\frac{\partial L_n(\widehat{\theta}_n)}{\partial \theta_i} = \frac{\partial L_n(\theta_0)}{\partial \theta_i} + \frac{\partial^2 L_n(\overline{\theta}_{n,i})}{\partial \theta_i \partial \theta}(\widehat{\theta}_n - \theta_0), \tag{6.50}$$

where $\|\overline{\theta}_{n,i} - \theta_0\| \leqslant \|\widehat{\theta}_n - \theta_0\|$ for all i. Using equations (6.48) and (6.49), we conclude with the uniform convergence theorem that

$$F_n := \Big(\frac{1}{n}\frac{\partial^2 L_n(\overline{\theta}_{n,i})}{\partial \theta_i \partial \theta}\Big)_{1\leqslant i\leqslant d} \xrightarrow[n\to\infty]{a.s.} F(\theta_0).$$

Using expressions of the derivatives of H_θ^t and since $X_t - f_{\theta_0}^t = M_{\theta_0}\xi_t$, with ξ_t and $\sigma(X_{t-1}, X_{t-2}, \ldots)$ independent, we obtain the explicit expression

$$\begin{aligned}\big(F(\theta_0)\big)_{ij} = -\frac{1}{2}\mathbb{E}\Big[\frac{\partial^2 q_0(\theta_0)}{\partial \theta_i \partial \theta_j}\Big] &= -\frac{1}{2}\mathbb{E}\Bigg[2\Big(\frac{\partial f_{\theta_0}^t}{\partial \theta_i}\Big)'\big(H_{\theta_0}^t\big)^{-1}\Big(\frac{\partial f_{\theta_0}^t}{\partial \theta_i}\Big) \\ &\quad +3\mathrm{Tr}\left(\big(H_{\theta_0}^t\big)^{-2}\Big(\frac{\partial H_{\theta_0}^t}{\partial \theta_j}\Big)\Big(\frac{\partial H_{\theta_0}^t}{\partial \theta_i}\Big)\right)\Bigg].\end{aligned} \tag{6.51}$$

If $F(\theta_0)$ is an invertible $d \times d$ matrix, it exists n large enough such as F_n is an invertible matrix. Moreover, (6.50) implies for n large enough,

$$n(\widehat{\theta}_n - \theta_0) = F_n^{-1}\left(\frac{\partial L_n(\widehat{\theta}_n)}{\partial \theta} - \frac{\partial L_n(\theta_0)}{\partial \theta}\right).$$

Therefore, if $\frac{1}{\sqrt{n}}\Big\|\frac{\partial L_n(\widehat{\theta}_n)}{\partial \theta}\Big\| \xrightarrow[n\to\infty]{\mathcal{P}} 0$, using Lemma 6.2 and the Slutsky Theorem, one obtains the result of Theorem 6.2. Since $\frac{\partial \widehat{L}_n(\tilde{\theta}_n)}{\partial \theta} = 0$ ($\widehat{\theta}_n$ is a local extremum for $\widehat{L}_n$), it is sufficient to show that

$$\mathbb{E}\Big[\frac{1}{\sqrt{n}}\Big\|\frac{\partial L_n}{\partial \theta} - \frac{\partial \widehat{L}_n}{\partial \theta}\Big\|_\Theta\Big] \underset{n\to\infty}{\longrightarrow} 0. \tag{6.52}$$

Therefor, we evaluate $\mathbb{E}\Big[\frac{1}{\sqrt{n}}\Big\|\frac{\partial q_t(\theta)}{\partial \theta_i} - \frac{\partial \widehat{q}_t(\theta)}{\partial \theta_i}\Big\|_\Theta\Big]$ for all $t \in \mathbb{N}^*$. Using the inequality $|a_1b_1c_1 - a_2b_2c_2| \leq |a_1 - a_2||b_2||c_2| + |a_1||b_1 - b_2||c_2| + |a_1||b_1||c_1 - c_2|$ and the relation (6.45), one obtains :

$$\begin{aligned}
\Big\|\frac{\partial q_t(\theta)}{\partial \theta_i} - \frac{\partial \widehat{q}_t(\theta)}{\partial \theta_i}\Big\|_\Theta \;\; &\leq \;\; 2\Big\|\frac{\partial \widehat{f}_\theta^t}{\partial \theta_i} - \frac{\partial f_\theta^t}{\partial \theta_i}\Big\|_\Theta \Big\|\big(\widehat{H}_\theta^t\big)^{-1}\Big\|_\Theta \|X_t - \widehat{f}_\theta^t\|_\Theta \\
&+2\Big\|\frac{\partial f_\theta^t}{\partial \theta_i}\Big\|_\Theta \Big\|\big(H_\theta^t\big)^{-1}\Big\|_\Theta \|\widehat{f}_\theta^t - f_\theta^t\|_\Theta + 2\Big\|\frac{\partial f_\theta^t}{\partial \theta_i}\Big\|_\Theta \Big\|\big(H_\theta^t\big)^{-1} - \big(\widehat{H}_\theta^t\big)^{-1}\Big\|_\Theta \|X_t - \widehat{f}_\theta^t\|_\Theta \\
&+\|\widehat{f}_\theta^t - f_\theta^t\|_\Theta \Big\|\frac{\partial \big(\widehat{H}_\theta^t\big)^{-1}}{\partial \theta_i}\Big\|_\Theta \|X_t - \widehat{f}_\theta^t\|_\Theta + \|X - f_\theta^t\|_\Theta \|X_t - \widehat{f}_\theta^t\|_\Theta \Big\|\frac{\partial \big(H_\theta^t\big)^{-1}}{\partial \theta_i} - \frac{\partial \big(\widehat{H}_\theta^t\big)^{-1}}{\partial \theta_i}\Big\|_\Theta \\
&+\|\widehat{f}_\theta^t - f_\theta^t\|_\Theta \|X_t - f_\theta^t\|_\Theta \Big\|\frac{\partial \big(H_\theta^t\big)^{-1}}{\partial \theta_i}\Big\|_\Theta + \Big\|\big(\widehat{H}_\theta^t\big)^{-1}\Big\|_\Theta \Big\|\frac{\partial H_\theta^t}{\partial \theta_i} - \frac{\partial \widehat{H}_\theta^t}{\partial \theta_i}\Big\|_\Theta \\
&+\Big\|\big(H_\theta^t\big)^{-1} - \big(\widehat{H}_\theta^t\big)\Big\|_\Theta \Big\|\frac{\partial \big(H_\theta^t\big)^{-1}}{\partial \theta_i}\Big\|_\Theta
\end{aligned}$$

It was seen previously that for all $t \in \mathbb{Z}$, $1 \leq i \leq d$,
- $\Big\|\big(\widehat{H}_\theta^t\big)^{-1}\Big\|_\Theta \leq \big(\underline{M}\big)^{-1/m}$, $\Big\|\big(H_\theta^t\big)^{-1}\Big\|_\Theta \leq \big(\underline{M}\big)^{-1/m}$;
- $\mathbb{E}\Big[\Big\|H_\theta^t\Big\|_\Theta^{r/2}\Big] < \infty$, $\mathbb{E}\Big[\Big\|\frac{\partial H_\theta^t}{\partial \theta_i}\Big\|_\Theta^{r/2}\Big] < \infty$;
- $\mathbb{E}\big[\|f_\theta^t\|_\Theta^r\big] < \infty$, $\mathbb{E}\Big[\Big\|\frac{\partial f_\theta^t}{\partial \theta_i}\Big\|_\Theta^r\Big] < \infty$ and $\mathbb{E}\big[\|X_t - f_\theta^t\|_\Theta^r\big] < \infty$.

Moreover, using Assumptions ST(r) and D$^{(1)}$,
- $\mathbb{E}\Big[\|f_\theta^t - \widehat{f}_\theta^t\|_\Theta^r\Big] \leq \mathbb{E}\big[\|X_t\|^r\big]\Big(\sum_{j\geq t}\alpha_j(f)\Big)^r$ and $\mathbb{E}\Big[\Big\|\frac{\partial f_\theta^t}{\partial \theta_i} - \frac{\partial \widehat{f}_\theta^t}{\partial \theta_i}\Big\|_\Theta^r\Big] \leq \mathbb{E}\big[\|X_t\|^r\big]\Big(\sum_{j\geq t}\alpha_j^{(1)}(f)\Big)^r$;

– $\mathbb{E}\Big[\big\|H_\theta^t-\widehat{H}_\theta^t\big\|_\Theta^{r/2}\Big]\leq C\big(\mathbb{E}\big[\|X_t\|^r\big]\big)^{1/2}\Big(\sum_{j\geq t}\alpha_j(M)\Big)^{r/2}$,

$\mathbb{E}\Big[\Big\|\frac{\partial H_\theta^t}{\partial\theta_i}-\frac{\partial\widehat{H}_\theta^t}{\partial\theta_i}\Big\|_\Theta^{r/2}\Big]\leq C\big(\mathbb{E}\big[\|X_t\|^r\big]\big)^{1/2}\Big(\Big(\sum_{j\geq t}\alpha_j(M)\Big)^{r/2}+\Big(\sum_{j\geq t}\alpha_j^{(1)}(M)\Big)^{r/2}\Big)$,

$\mathbb{E}\Big[\Big\|\frac{\partial\big(H_\theta^t\big)^{-1}}{\partial\theta_i}-\frac{\partial\big(\widehat{H}_\theta^t\big)^{-1}}{\partial\theta_i}\Big\|_\Theta^{r/2}\Big]\leq C\big(\mathbb{E}\big[\|X_t\|^r\big]\big)^{1/2}\Big(\Big(\sum_{j\geq t}\alpha_j(M)\Big)^{r/2}+\Big(\sum_{j\geq t}\alpha_j^{(1)}(M)\Big)^{r/2}\Big)$;

Using Hölder inequalities, for $r\geq 4$, it exists $C\geq 0$ such that

$$\Big\|\frac{\partial q_t(\theta)}{\partial\theta_i}-\frac{\partial\widehat{q}_t(\theta)}{\partial\theta_i}\Big\|_\Theta\leq C\sum_{j\geq t}\big(\alpha_j(f)+\alpha_j(M)+\alpha_j^{(1)}(f)+\alpha_j^{(1)}(M)\big).$$

Consequently $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathbb{E}\Big[\Big\|\frac{\partial q_t(\theta)}{\partial\theta_i}-\frac{\partial\widehat{q}_t(\theta)}{\partial\theta_i}\Big\|_\Theta \underset{n\to\infty}{\longrightarrow} 0$ if (6.18) is satisfied. □

Proof of the Corollary 6.2

The proofs of Theorem 6.2 and Lemmas 6.2 can be easily translated using Assumptions $\mathrm{E}^{(1)}$ and $\mathrm{E}^{(2)}$ instead of $\mathrm{D}^{(1)}$ and $\mathrm{D}^{(2)}$. The key point of this replacement are the following properties :

- $$\begin{aligned}\mathbb{E}\Big[\Big\|H_\theta^t-\widehat{H}_\theta^t\Big\|_\Theta^{r/2}\Big] &\leq \mathbb{E}\Big[\Big\|\sum_{j\geq t}\alpha_j(N)X_{t-j}X'_{t-j}\Big\|_\Theta^{r/2}\Big]\\ &\leq \big(\mathbb{E}\big[\|X_tX'_t\|^{r/2}\big]\big)^{1/2}\Big(\sum_{j\geq t}\alpha_j(H)\Big)^{r/2}\end{aligned}$$
- $$\begin{aligned}\mathbb{E}\Big[\Big\|\frac{\partial H_\theta^t}{\partial\theta_i}-\frac{\partial\widehat{H}_\theta^t}{\partial\theta_i}\Big\|_\Theta^{r/2}\Big] &\leq \mathbb{E}\Big[\Big\|\sum_{j\geq t}\alpha_j^{(1)}(H)X_{t-j}X'_{t-j}\Big\|_\Theta^{r/2}\Big]\\ &\leq \big(\mathbb{E}\big[\|X_tX'_t\|^{r/2}\big]\big)^{1/2}\Big(\sum_{j\geq t}\alpha_j^{(1)}(H)\Big)^{r/2}.\ \square\end{aligned}$$

Bibliographie

[1] ANDREWS, D. W. K. Nonstrong mixing autoregressive processes. *J. Appl. Probab. 21*, 4 (1984), 930–934.

[2] ANGO NZE, P., BÜHLMANN, P., AND DOUKHAN, P. Weak dependence beyond mixing and asymptotics for nonparametric regression. *Annals of Statistics 30* (2002), 397–430.

[3] ANGO NZE, P., AND DOUKHAN, P. Functional estimation for time series : uniform convergence properties. *Journal of Statistical Planning and Inference 68* (1998), 5–29.

[4] ANGO NZE, P., AND DOUKHAN, P. Weak dependence : Models and applications to econometrics. *Econometric Theory 20* (2003), 995–1045.

[5] ANTONIADIS, A., AND FAN, J. Regularization of wavelet approximations. *J. Am. Stat. Assoc. 96*, 455 (2001), 939–967.

[6] BARBOUR, A., GERRARD, R., AND REINERT, G. Iterates of expanding maps. *Probab. theory relat. Fields 116* (2000), 151–180.

[7] BARDET, J.-M., DOUKHAN, P., AND LEÓN, J. R. Uniform limit theorems for the periodogram of a weakly dependent time series and their applications to Whittle's estimate. Preprint.

[8] BAUWENS, L., LAURENT, S., AND ROMBOUTS, J. Multivariate garch models : A survey. *J. Appl. Econ. 21* (2006), 79–109.

[9] BERBEE, H. Chains with infinite connections : uniqueness and Markov representation. *Probab. Theory Related Fields 76*, 2 (1987), 243–253.

[10] BERKES, I., AND HORVÁTH, L. The efficiency of the estimators of the parameters in garch processes. *Ann. Statist. 32*, 2 (2004), 633–655.

[11] BERKES, I., HORVÁTH, L., AND KOKOSZKA, P. Garch processes : structure and estimation. *Bernoulli 9* (2003), 201–227.

[12] BILLINGSLEY. *Convergence of probability measures.* John Wiley & Sons Inc., New York, 1968.

[13] BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. *J. Econometrics 31* (1986), 307–327.

[14] Bollerslev, T. Modelling the coherence in short-run nominal exchange rates : a multivariate generalized arch approach. *Review of Economics and Statistics 72* (1990), 498–505.

[15] Bollerslev, T., and Wooldridge, J. M. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Rev. 11*, 2 (1992), 143–172.

[16] Borovkova, S., Burton, R., and Dehling, H. Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Trans. Amer. Math. Soc. 353*, 11 (2001), 4261–4318.

[17] Bougerol, P. Kalman filtering with random coefficients and contractions. *Probab. Theory Related Fields 31* (1993), 942–959.

[18] Bühlmann, P., and Wyner, A. J. Variable length Markov chains. *Ann. Statist. 27*, 2 (1999), 480–513.

[19] Bulinski, A., and Shashkin, A. Strong invariance principle for dependent random fields. Preprint.

[20] Collet, P., Martínez, S., and Schmitt, B. Exponential inequalities for dynamical measures of expanding maps of the interval. *Probab. theory relat. Fields 123* (2002), 301–322.

[21] Comets, F., Fernández, R., and Ferrari, P. A. Processes with long memory : regenerative construction and perfect simulation. *Ann. Appl. Probab. 12*, 3 (2002), 921–943.

[22] Comte, F., and Lieberman, O. Asymptotic theory for multivariate garch processes. *J. Multivariate Anal. 84*, 1 (2003), 61–84.

[23] Comte, F., and Merlevède, F. Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM P&S 6* (2002), 211–238.

[24] Daubechies, I. *Ten lectures on wavelets*, vol. 61. SIAM press, 1992.

[25] Dedecker, J. A central limit theorem for random fields. *Probability Theory and Related Fields 110* (1998), 397–426.

[26] Dedecker, J., and Doukhan, P. A new covariance inequality and applications. *Stochastic processes and their applications 106*, 1 (2003), 63–80.

[27] Dedecker, J., Doukhan, P., Lang, G., Leon, J. R., Louhichi, S., and Prieur, C. *Weak dependence : models, theory and applications.* Springer-Verlag, 2007.

[28] Dedecker, J., and Prieur, C. Coupling for τ-dependent sequences and applications. *J. theor. probab. 17*, 4 (2004), 861–855.

[29] Dedecker, J., and Prieur, C. New dependence coefficients : Examples and applications to statistics. *Probab. theory relat. Fields 132* (2005), 203–235.

[30] Dedecker, J., and Rio, E. On the functional central limit theorem for stationary process. *Annales de l'Institut Henri-Poincaré, Probabilités et statisitque 36* (2000), 1–34.

[31] Diaconis, P., and Freedman, D. Iterated random functions. *SIAM Rev. 41*, 1 (1999), 45–76.

[32] Ding, Z., Granger, C. W., and Engle, R. A long memory property of stock market returns and a new model. *J. Empirical Finance 1* (1993), 83–106.

[33] Dobrushin, R. L. Prescribing a system of random variables by conditional distributions. *Theory Probab. Applic. 15* (1970), 458–486.

[34] Dobrushin, R. L., and Kusuoka, S. *Statistical mechanics and fractals*, vol. 1567 of *Lecture Notes in Mathematics*. Springer-Verlag, New York, 1993.

[35] Donoho, D., Johnstone, I., Kerkyacharian, G., and Picard, D. Density estimation by wavelet thresholding. *Ann. Stat. 24*, 2 (1996), 508–539.

[36] Donoho, D. L., and Johnstone, I. A. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Assoc. 90*, 432 (1995), 1200–1224.

[37] Doukhan, P. Consistency of delta-sequence estimates of a density or of a regression function for a weakly dependent stationary sequence. In *Estimation Fonctionnelle*. Séminaire de statistique d'Orsay, 1991, pp. 91–55.

[38] Doukhan, P. *Mixing*, vol. 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994.

[39] Doukhan, P., and Lang, G. Rates of convergence in the weak invariance principle for the empirical repartition process of weakly dependent sequences. *Stat. inference stoch. process. 5* (2002), 199–228.

[40] Doukhan, P., Lang, G., Surgailis, D., and Viano, M.-C. Functional limit theorem for the empirical process of a class of Bernoulli shifts with long memory. *J. Theoret. Probab. 18*, 1 (2005), 161–186.

[41] Doukhan, P., and Louhichi, S. A new weak dependence condition and applications to moment inequalities. *Stoch. process. their appl. 84* (1999), 313–342.

[42] Doukhan, P., and Louhichi, S. Functional estimation for weakly dependent stationary time series. *Scandinavian Journal of Statistics 28* (2001), 325–342.

[43] Doukhan, P., Madre, H., and Rosenbaum, M. ARCH type bilinear weakly dependent models. To appear in Statistics.

[44] Doukhan, P., and Neumann, M. A Bernstein type inequality for times series. unpublished, available at http ://www.crest.fr/pageperso/doukhan/publications.htm.

[45] Doukhan, P., and Portal, F. Principe d'invariance faible pour la fonction de répartition empirique dans un cadre multidimensionnel et mélangeant. *Probab. Math. Statist. 8* (1987), 117–132.

[46] Doukhan, P., and Sifre, J. *Cours d'analyse - Analyse réelle et intégration.* Dunod, 2001.

[47] Doukhan, P., Teyssière, G., and Winant, P. A LARCH(∞) vector valued process. In *Dependence in Probability and Statistics* (New York, 2006), P. Bertail, P. Doukhan, and P. Soulier, Eds., vol. 187 of *Lectures Notes in Statistics*, Springer, pp. 245–258.

[48] Doukhan, P., and Truquet, L. Weakly dependent random fields with infinite interactions. Submitted.

[49] Doukhan, P., and Wintenberger, O. Weakly dependent chains with infinite memory. Submitted.

[50] Duflo, M. *Algorithmes stochastiques*, vol. 23 of *Mathématiques & Applications (Berlin) [Mathematics & Applications].* Springer-Verlag, Berlin, 1986.

[51] Duflo, M. *Random iterative models*, vol. 34 of *Applications of Mathematics (New York).* Springer-Verlag, Berlin, 1997.

[52] Dunsmuir, W., and Hannan, E. J. Vector linear time series models. *Advances in Appl. Probability 8*, 2 (1976), 339–364.

[53] Engle, R. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50* (1982), 987–1007.

[54] Engle, R., and Kroner, K. Multivariate simultaneous generalized arch. *Econometric Theory 11* (1995), 122–150.

[55] Fernández, R., and Maillard, G. Chains with complete connections and one-dimensional Gibbs measures. *Electron. J. Probab. 9*, 6 (2004), 145–176.

[56] Georgii, H.-O. *Gibbs measures and phase transitions*, vol. 9 of *de Gruyter Studies in Mathematics.* Walter de Gruyter and Co., Berlin, 1988.

[57] Giraitis, L., Koul, H., and Surgailis, D. Asymptotic normality of regression estimators with long memory errors. *Statistics & Probability Letters 29* (1996), 317–335.

[58] Giraitis, L., Leipus, R., Robinson, P. M., and Surgailis, D. Larch, leverage and long memory. *Journal of Financial Econometrics 2*, 2 (2004), 177–210.

[59] Giraitis, L., Leipus, R., and Surgailis, D. *Long-Memory in Economics.* Springer Verlag., 2006, ch. Recent advances in ARCH modelling, pp. 3–38.

[60] Giraitis, L., and Robinson, P. Whittle estimation of arch models. *Econometric Theory 17* (2001), 608–631.

[61] Giraitis, L., and Surgailis, D. Arch-type bilinear models with double long memory. *Stoch. Proc. and their Appl. 100* (2002), 275–300.

[62] Gordin, M. I. The central limit theorem for stationary processes. *Dokl. Akad. Nauk SSSR 188* (1969), 739–741.

[63] GUYON, X. *Random fields on a network : modeling, statistics, and applications.* New York : Springer-Verlag, 1995.

[64] HALL, P., LAHIRI, S. N., AND K., T. Y. On bandwidth choice for density estimation with dependent data. *Ann. stat. 23*, 6 (1995), 2241–2263.

[65] HALLIN, M., LU, Z., AND TRAN, L. Density estimation for spatial linear processes. *Bernoulli 7* (2001), 657–668.

[66] HANNAN, E. J. The asymptotic theory of linear time series models. *J. Appl. Probab. 10* (1973), 130–145.

[67] HARDLE, W., KERKYACHARIAN, G., PICARD, D., AND TSYBAKOV, A. *Wavelets Approximation and Statistical Applications*, vol. 129. Springer Verlag, Lecture Notes in Statistics, 1998.

[68] HART, J. D., AND VIEU, P. Data-driven bandwidth choice for density estimation based on dependent data. *Ann. stat. 18*, 2 (1990), 873–890.

[69] HEYDE, C. C. On the central limit theorem and iterated logarithm law for stationary processes. *Bull. Austral. Math. Soc. 12* (1975), 1–8.

[70] IBRAGIMOV, I. A. Some limit theorems for stationary processes. *Theory Probab. Appl. 7* (1962), 349–382.

[71] IBRAGIMOV, I. A. A note on the central limit theorem for dependent random variables. *Theory Probab. Appl. 20* (1975), 135–141.

[72] IOSIFESCU, M., AND GRIGORESCU, S. *Dependence with complete connections and its applications*, vol. 96 of *Cambridge Tracts in Mathematics.* Cambridge University Press, Cambridge, 1990.

[73] IOSIFESCU, M., AND THEODORESCU, R. *Random processes and learning.* Springer-Verlag, New York, 1969.

[74] JEANTHEAU, T. Strong consistency of estimators for multivariate arch models. *Econometric Theory 14*, 1 (1998), 70–86.

[75] JOHNSTONE, I., KERKACHARYAN, G., PICARD, D., AND M., R. Wavelet deconvolution in a periodic setting. *Journal of Royal Society Society 3*, 66 (2004), 1–27.

[76] KAC, M. *Probability and related topics in physical sciences*, vol. 1a of *Lectures in Applied Mathematics Series.* American Mathematical Society, London, 1959.

[77] KALLENBERG, O. *Foundations of modern probability.* Probability and its Applications. Springer-Verlag, New York, 1997.

[78] KERKYACHARIAN, G., AND PICARD, D. Thresholding algorithms, maxisets and well-concentrated bases. *Test 9* (2000), 283–344.

[79] Lahiri, S. N. *Resampling Methods for Dependent Data.* Springer Series in Statistics, 2003.

[80] Landau, L. D., and Lifshitz, E. M. *Statistical Physics*, 3 ed., vol. 5 of *Course of Theoretical Physics.* Butterworth-Heinemann, Oxford, 1980.

[81] Lang, G., and Soulier, P. Convergence de mesures spectrales aléatoires et applications à des principes d'invariance. *Statistical inference for stochastic processes 3*, 1-2 (2000), 41–51.

[82] Lasota, A., and Yorke, J. On the existence of invariant measure for piecewise monotonic transformations. *Trans. Amer. Math. Soc. 186* (1974), 481–488.

[83] Liebscher, E. Estimation of the density and the regression function under mixing conditions. *Statist. Decisions 19* (2001), 9–26.

[84] Ling, S., and McAleer, M. Asymptotic theory for a vector arma-garch model. *Econometric Theory 19* (2003), 280–310.

[85] Louhichi, S. Rates of convergence in the CLT for some weakly dependent random variables. *Teor. Veroyatnost. i Primenen 46*, 2 (2001), 345–364.

[86] Louhichi, S., and Soulier, P. Marcinkiewicz-Zygmund strong laws for infinite variance time series. *Stat. Inference Stoch. Process. 1*, 2 (2000), 31–40.

[87] Mauricio, J. A. Exact maximum likelihood estimation of stationary vector arma models. *J. Amer. Statist. Assoc. 90* (1995), 282–291.

[88] Merlevède, F., Peligrad, M., and Utev, S. Recent advances in invariance principles for stationary sequences. *Probability Surveys 3* (2006), 1–36.

[89] Meyer, Y. *Wavelets and operators.* Cambridge University Press, 1992.

[90] Newman, C. M., and Wright, A. L. An invariance principle for certain dependent sequences. *Ann. Probab. 9*, 4 (1981), 671–675.

[91] Peligrad, M., and Utev, S. Central limit theorem for stationary linear processes. *Ann. Probab. 34*, no. 4 (2006), 1608Ũ1622.

[92] Peligrad, M., and Utev, S. Invariance principle for stochastic processes with short memory. In *High Dimensional Probability*, IMS Lecture Notes Monograph Series. IMS, 2006.

[93] Petrov, V. *Limit Theorems of Probability Theory.* Oxford University Press, Oxford (UK), 1995.

[94] Pfanzagl, J. On the mesurability and consistency of minimum contrast estimates. *Metrika 14* (1969), 249–334.

[95] Pitt, L. D. Positively correlated normal variables are associated. *Ann. Probab. 10*, 2 (1982), 496–499.

[96] Polya, G., and Szegö, G. *Problems and theorems in Analysis 1.* Classics in Mathematics. Springer-Verlag, Berlin, 1978.

[97] Prieur, C. *Applications statistiques de suites faiblement dépendantes et de systèmes dynamiques.* PhD thesis, CREST, 2001.

[98] Prieur, C. Density estimation for one-dimensional dynamical systems. *ESAIM P & S 5* (2001), 51–76.

[99] Rabemananjara, R., and Zakoian, J. Threshold arch models and asymmetries in volatility. *Journal of Applied Econometrics 8* (1993), 31–49.

[100] Ragache, N., and Wintenberger, O. Convergence rates for density estimators of weakly dependent time series. In *Dependence in Probability and Statistics* (New York, 2006), P. Bertail, P. Doukhan, and P. Soulier, Eds., vol. 187 of *Lectures Notes in Statistics*, Springer, pp. 349–372.

[101] Rio, E. *Théorie asymptotique des processus aléatoires faiblement dépendants*, vol. 31 of *Mathématiques & Applications (Berlin) [Mathematics & Applications].* Springer-Verlag, Berlin, 2000.

[102] Robinson, P., and Zaffaroni, P. Pseudo-maximum likelihood estimation of arch(∞) models. *Ann. Statist. 34* (2006), 1049–1074.

[103] Rosenblatt, M. A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A. 42* (1956), 43–47.

[104] Rozanov, Y., and Volkonskii, V. Some limit theorems for random functions i. *Theory Probab. Appl. 4* (1959), 178–197.

[105] Straumann, D., and Mikosch, T. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series : A stochastic recurrence equations approach. *Ann. Statist. 34*, 5 (2006), 2449Ű2495.

[106] Szegö, G. *Orthogonal polynomials.* American Mathematical Society Colloquium Publication, 1933.

[107] Tribouley, K., and Viennet, G. l_p adaptive density estimation in a β mixing framework. *Ann. Inst. Henri Poincaré, B calc. probab. stat. 34*, 2 (1998), 179–208.

[108] Tsybakov, A. *Introduction à l'estimation non-paramétrique*, vol. 41. Springer, 2004.

[109] Vannucci, M. Nonparametric density estimation using wavelets. Tech. rep., Texas A and M university, 1998.

[110] Viana, M. *Stochastic dynamics of deterministic systems.* PhD thesis, Instituto de Matematica Pura e Aplicada, Rio de Janeiro, Brazil., 1997. available at http ://w3.impa.br/~viana.

[111] Volný, D. Approximating martingales and the central limit theorem for strictly stationary processes. *Stoch. Processes Appl. 44*, 41-74 (1993), 41–74.

[112] Wavelab. http ://www-stat.stanford.edu/~wavelab/.

[113] WU, W. B., AND MIN, W. On linear processes with dependent innovations. *Stochastic Process. Appl. 115*, 6 (2005), 939–958.