

Estimation récursive dans certains modèles de déformation

Philippe Fraysse

► **To cite this version:**

Philippe Fraysse. Estimation récursive dans certains modèles de déformation. Mathématiques générales [math.GM]. Université Sciences et Technologies - Bordeaux I, 2013. Français. NNT : 2013BOR14797 . tel-00844393

HAL Id: tel-00844393

<https://tel.archives-ouvertes.fr/tel-00844393>

Submitted on 15 Jul 2013

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

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

N° ordre : 4797



Thèse

présentée à

l'Université Bordeaux 1

ECOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE

par

Philippe Fraysse

pour obtenir le grade de

Docteur

SPÉCIALITÉ : MATHÉMATIQUES APPLIQUÉES

Estimation récursive dans certains modèles de déformation

Soutenue le 4 Juillet 2013 devant la commission d'examen composée de

M. Marc ARNAUDON	PROF.	Université de Bordeaux 1	Examineur
M. Bernard BERCU	PROF.	Université de Bordeaux 1	Directeur de thèse
M. Bernard DELYON	PROF.	Université de Rennes 1	Rapporteur
M. Jean-Claude FORT	PROF.	Université Paris Descartes	Président du jury
M. Yuri GOLUBEV	D.R.	CNRS Aix-Marseille	Examineur
M. Anatoli JUDITSKY	PROF.	Université Joseph Fourier	Rapporteur
M. Jean-Michel LOUBÈS	PROF.	Université Paul Sabatier	Examineur

INSTITUT DE MATHÉMATIQUES DE BORDEAUX

Université Bordeaux 1

351, Cours de la Libération,

33 405 TALENCE Cedex

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE DE BORDEAUX

U.F.R. de Mathématiques et Informatique

Bâtiment A33 – 1^{er} étage – Porte 112

351, Cours de la Libération

33 405 TALENCE Cedex

CENTRE DE RECHERCHE INRIA BORDEAUX – SUD-OUEST

200, Avenue de la Vieille Tour,

33 405 TALENCE

Remerciements

En premier lieu, je remercie mon directeur de thèse, Bernard Bercu, pour m'avoir encadré pendant cette thèse. Je tiens à lui exprimer toute ma reconnaissance et le remercie pour m'avoir toujours fait confiance.

Ensuite, je remercie Bernard Delyon et Anatoli Iouditski pour avoir accepté de rapporter cette thèse. En particulier, merci pour les remarques pertinentes qui ont permis d'améliorer le manuscrit.

Je remercie Jean-Michel Loubès d'avoir accepté de faire partie de mon jury. J'ai travaillé avec lui et Hélène Lescornel sur le chapitre 6 de cette thèse. Je tiens à le remercier pour son accueil à Toulouse et sa gentillesse.

Je remercie enfin Marc Arnaudon, Jean-Claude Fort et Yuri Golubev pour avoir accepté d'être dans le jury de cette thèse.

Une pensée pour tous les gens du bureau de l'INRIA ou de l'IMB. J'ai évidemment une pensée plus particulière pour François et Romain.

Ces trois années de thèse m'ont aussi permis d'assouvir ma passion infinie pour le tennis. J'espère à l'avenir pouvoir m'investir dans ce milieu. Enfin, même si ma motivation et ma passion ne sont plus aussi fortes qu'avant, je remercie le Paris Saint-Germain d'avoir enfin réussi à être champion de France.

Estimation récursive dans certains modèles de déformation

Cette thèse est consacrée à l'étude de certains modèles de déformation semi-paramétriques. Notre objectif est de proposer des méthodes récursives, issues d'algorithmes stochastiques, pour estimer les paramètres de ces modèles. Dans la première partie, on présente les outils théoriques existants qui nous seront utiles dans la deuxième partie. Dans un premier temps, on présente un panorama général sur les méthodes d'approximation stochastique, en se focalisant en particulier sur les algorithmes de Robbins-Monro et de Kiefer-Wolfowitz. Dans un second temps, on présente les méthodes à noyaux pour l'estimation de fonction de densité ou de régression. On s'intéresse plus particulièrement aux deux estimateurs à noyaux les plus courants qui sont l'estimateur de Parzen-Rosenblatt et l'estimateur de Nadaraya-Watson, en présentant les versions récursives de ces deux estimateurs. Dans la seconde partie, on présente tout d'abord une procédure d'estimation récursive semi-paramétrique du paramètre de translation et de la fonction de régression pour le modèle de translation dans la situation où la fonction de lien est périodique. On généralise ensuite ces techniques au modèle vectoriel de déformation à forme commune en estimant les paramètres de moyenne, de translation et d'échelle, ainsi que la fonction de régression. On s'intéresse finalement au modèle de déformation paramétrique de variables aléatoires dans le cadre où la déformation est connue à un paramètre réel près. Pour ces trois modèles, on établit la convergence presque sûre ainsi que la normalité asymptotique des estimateurs paramétriques et non paramétriques proposés. Enfin, on illustre numériquement le comportement de nos estimateurs sur des données simulées et des données réelles.

Mots-clés : estimation récursive, modèles semi-paramétriques, algorithmes stochastiques, estimateurs à noyaux, martingales.

Recursive estimation for some deformation models

This thesis is devoted to the study of some semi-parametric deformation models. Our aim is to provide recursive methods, related to stochastic algorithms, in order to estimate the different parameters of the models. In the first part, we present the theoretical tools which we will use in the next part. On the one hand, we focus on stochastic approximation methods, in particular the Robbins-Monro algorithm and the Kiefer-Wolfowitz algorithm. On the other hand, we introduce kernel estimators in order to estimate a probability density function and a regression function. More particularly, we present the two most famous kernel estimators which are the one of Parzen-Rosenblatt and the one of Nadaraya-Watson. We also present their recursive version. In the second part, we present the results we obtained in this thesis. Firstly, we provide a recursive estimation method of the shift parameter and the regression function for the translation model in which the regression function is periodic. Secondly, we extend this estimation procedure to the shape invariant model, providing estimation of the height parameter, the translation parameter and the scale parameter, as well as the common shape function. Thirdly, we are interested in the parametric deformation model of random variables where the deformation is known and depending on an unknown parameter. For these three models, we establish the almost sure convergence and the asymptotic normality of each estimator. Finally, we numerically illustrate the asymptotic behaviour of our estimators on simulated data and on real data.

Keywords : recursive estimation, semi-parametric models, stochastic algorithms, kernel estimators, martingales.

Table des matières

Introduction	11
I Approximation stochastique et estimation fonctionnelle	15
1 Approximation stochastique	17
1.1 Cadre	17
1.2 Une approche déterministe	18
1.3 L'algorithme de Robbins-Monro	18
1.3.1 Hypothèses	19
1.3.2 Propriétés presque sûres de l'algorithme	19
1.3.3 Convergence en loi et efficacité asymptotique	22
1.3.4 Convergence quadratique et bornes	23
1.3.5 Un exemple d'application	24
1.3.6 Projection	26
1.4 L'algorithme de Kiefer-Wolfowitz	28
1.4.1 Hypothèses	28
1.4.2 Propriétés presque sûres de l'algorithme	29
1.4.3 Convergence en loi	31
1.4.4 Projection	31
1.5 Un cadre non martingale	32
1.5.1 Innovations markoviennes	32
1.5.2 Méthode de l'équation différentielle ordinaire	33
2 Estimation fonctionnelle	35
2.1 Introduction	35
2.1.1 Estimation non paramétrique	35
2.1.2 Noyaux	36
2.2 Estimation d'une densité de probabilité	36
2.2.1 Intuition et construction de l'estimateur	36
2.2.2 L'estimateur de Parzen-Rosenblatt non-récuratif	38
2.2.3 L'estimateur de Parzen-Rosenblatt récursif	40
2.2.4 L'estimateur de Deheuvels	41
2.3 Estimation d'une fonction de régression	41
2.3.1 L'estimateur de Nadaraya-Watson non-récuratif	42
2.3.2 L'estimateur de Nadaraya-Watson récursif	43

2.3.3	L'estimateur de Révész	44
II	Estimation semi-paramétrique pour des modèles de déformation	45
3	Présentation des modèles de déformation	47
3.1	Les modèles que nous étudions	47
3.2	Conditions d'identifiabilité	48
3.2.1	Modèle $Y_n = f(X_n - \theta) + \varepsilon_n$	48
3.2.2	Modèle $Y_{i,j} = a_j f(X_i - \theta_j) + v_j + \varepsilon_{i,j}$	50
3.2.3	Deux contre exemples	51
4	Le modèle de translation	57
4.1	Procédure d'estimation	57
4.1.1	Estimation de θ	57
4.1.2	Estimation de f	59
4.2	Résultats de convergence paramétrique	60
4.3	Résultats de convergence non paramétrique	62
4.4	Preuves des résultats paramétriques	62
4.4.1	Preuve du Théorème 4.2.1.	62
4.4.2	Preuve du Théorème 4.2.2.	64
4.4.3	Preuve du Théorème 4.2.3.	66
4.5	Preuves des résultats non paramétriques	66
4.5.1	Preuve du Théorème 4.3.1.	66
4.5.2	Preuve du Théorème 4.3.2.	70
4.6	Extension au cas où la densité g est inconnue	74
4.6.1	Preuve du Théorème 4.6.1	75
5	Modèle de déformation à forme commune	81
5.1	Procédure d'estimation	81
5.1.1	Estimation de v	83
5.1.2	Estimation de θ	83
5.1.3	Estimation de a	84
5.1.4	Estimation de f	84
5.2	Résultats de convergence paramétrique	85
5.2.1	Estimation de v	85
5.2.2	Estimation de θ	86
5.2.3	Estimation de a	89
5.3	Résultats de convergence non paramétrique	90
5.4	Preuves des résultats paramétriques	91
5.4.1	Preuve du Théorème 5.2.1.	91
5.4.2	Preuve du Théorème 5.2.2.	91
5.4.3	Preuve du Théorème 5.2.3.	91
5.4.4	Preuve du Théorème 5.2.4.	94
5.4.5	Preuve du Théorème 5.2.5.	94
5.5	Preuves des résultats non paramétriques	98

5.5.1	Preuve du Théorème 5.3.1.	98
5.5.2	Preuve du Théorème 5.3.2.	99
6	Déformation de variables aléatoires	103
6.1	Procédure d'estimation	104
6.1.1	Estimation de θ	107
6.1.2	Estimation de la loi de ε	108
6.2	Résultats de convergence paramétrique	109
6.3	Résultats de convergence non paramétrique	110
6.4	Preuves des convergences paramétriques	110
6.4.1	Preuve du Théorème 6.2.1.	110
6.4.2	Preuve du Théorème 6.2.2.	112
6.5	Preuves des convergences non paramétriques	113
6.5.1	Preuve du Théorème 6.3.1	114
6.5.2	Preuve du Théorème 6.3.2	116
6.6	Extension au cas où φ_θ n'est pas inversible	119
7	Simulations	121
7.1	Le modèle de translation	121
7.1.1	Données simulées	121
7.1.2	Données ECG	125
7.2	Modèle à forme commune	127
7.2.1	Données simulées	127
7.2.2	Données ECG	131
7.3	Déformation de variables aléatoires	133
	Conclusion et perspectives	137
	Bibliographie	138

Introduction

Les modèles de déformation périodiques sont souvent très utiles pour modéliser des phénomènes réels. Dans les domaines de la médecine, de l'économétrie ou encore de la météorologie, l'étude de ce type de modèles permet d'avoir une bonne approximation de la réalité. Dans la littérature statistique, deux grandes classes de modèles sont omniprésentes : les modèles paramétriques et les modèles non paramétriques. Ces deux types de modèles ont leurs avantages et leurs défauts spécifiques.

Afin de conjuguer les avantages des approches paramétriques et non paramétriques, à savoir la capacité d'interprétation des modèles paramétriques et la souplesse des modèles non paramétriques, les modèles semi-paramétriques ont été développés. Ces modèles dépendent généralement d'un ou de plusieurs paramètres ainsi que d'une fonction de lien à estimer.

Dans cette thèse, on étudie tout d'abord le modèle semi-paramétrique de translation dans la situation où la fonction de lien est périodique. On s'intéresse ensuite à son extension par le modèle vectoriel de déformation à forme commune. Notre objectif est la mise en oeuvre de nouvelles procédures d'estimation récursive semi-paramétrique des paramètres de moyenne, de translation et d'échelle, ainsi que de la fonction de lien périodique. Les méthodes d'estimation récursive n'ont jamais été développées pour ce type de modèles semi-paramétriques. On s'intéresse finalement au modèle de déformation paramétrique de variables aléatoires dans le cadre où la déformation est connue à un paramètre réel près qui est estimé par un algorithme stochastique. D'une manière générale, l'intérêt des méthodes récursives est de prendre en compte l'arrivée temporelle des informations et de gagner ainsi du temps dans les algorithmes d'estimation mis en oeuvre. L'immense avantage de ces méthodes est qu'il n'est pas nécessaire de relancer le calcul de l'estimateur sur toutes les données à chaque fois qu'une nouvelle donnée vient compléter les anciennes. L'idée est d'utiliser l'estimation calculée sur les anciennes données et de les remettre à jour en tenant uniquement compte des nouvelles données. L'implémentation de ces algorithmes est alors très simple, et les applications d'une telle approche sont nombreuses. Les algorithmes stochastiques, en particulier ceux de Robbins-Monro et de Kiefer-Wolfowitz, font partie de ces méthodes récursives. L'algorithme de Robbins-Monro est au coeur de la thèse. Plus précisément, pour les deux modèles semi-paramétriques évoqués précédemment, nous avons développé une procédure d'estimation des paramètres basée sur l'algorithme de Robbins-Monro qui ne nécessite aucune connaissance préalable sur la fonction de lien. On a également utilisé cet algorithme pour l'estimation du paramètre inconnu dans le modèle de déformation de variables aléatoires. La thèse s'articule en deux parties. La première partie, contenant les chapitres 1 et 2, porte sur les outils théoriques connus qui seront utilisés dans la thèse. La seconde partie, contenant les chapitres 3, 4, 5, 6 et 7, concerne les ré-

sultats obtenus pendant la thèse sur les trois modèles de déformation semi-paramétriques. On va maintenant préciser les grandes lignes de la thèse.

La première partie est scindée en deux chapitres.

Dans le chapitre 1, on présente un panorama général sur les méthodes d'approximation stochastique en nous intéressant principalement aux algorithmes de Robbins-Monro et de Kiefer-Wolfowitz. On donne les résultats connus de convergence presque sûre, de vitesse de convergence ainsi que de normalité asymptotique.

Le chapitre 2 traite d'estimation non paramétrique, et plus particulièrement d'estimateurs à noyaux de densité ou de fonction de régression. On présente notamment les estimateurs non-récurrents et récurrents de Parzen-Rosenblatt et de Nadaraya-Watson.

La seconde partie est composée de cinq chapitres.

Dans le chapitre 3, on présente de manière précise les modèles de déformation qui seront au coeur de la thèse. On s'attarde notamment sur les conditions d'identifiabilité de ces modèles, essentielles à leur étude rigoureuse.

Le chapitre 4 est consacré à l'étude du modèle semi-paramétrique de translation défini, pour tout $n \geq 0$, par

$$Y_n = f(X_n - \theta) + \varepsilon_n.$$

On propose un algorithme stochastique de type Robbins-Monro pour l'estimation du paramètre de translation θ . Cet algorithme ne nécessite pas une évaluation de la fonction de lien périodique f . Ensuite, on présente un algorithme récursif de type Nadaraya-Watson pour l'estimation de la fonction f , cet estimateur prenant en compte l'estimation préalable de θ . On met donc en place un algorithme stochastique double pour l'estimation de θ et f . On établit la convergence presque sûre ainsi que la normalité asymptotique de chacun des estimateurs. En fin de chapitre, on généralise l'estimation de θ au cas où la loi commune des temps d'observation aléatoire (X_n) est inconnue.

Le chapitre 5 est dédié à l'extension du travail précédent au modèle vectoriel de déformation à forme commune défini, pour tout $1 \leq i \leq n$ et pour tout $1 \leq j \leq p$, par

$$Y_{i,j} = a_j f(X_i - \theta_j) + v_j + \varepsilon_{i,j}.$$

On propose de nouvelles procédures d'estimation récursive semi-paramétrique des vecteurs des moyennes v , des translations θ et des échelles a ainsi que de la fonction de lien périodique f . On établit la convergence presque sûre de tous nos estimateurs ainsi que leur normalité asymptotique. L'apport essentiel de ce chapitre par rapport au précédent est que l'on est en mesure d'estimer le vecteur des échelles a qui est crucial dans l'étude de certains phénomènes réels comme la prévision des données de température.

Le chapitre 6 porte sur l'étude du modèle de déformation paramétrique de variables aléatoires défini, pour tout $n \geq 0$, par

$$X_n = \varphi_\theta(\varepsilon_n).$$

On présente un algorithme stochastique de type Robbins-Monro basé sur une fonction de contraste donnée par la distance de Wasserstein, pour l'estimation du paramètre inconnu θ . On montre la convergence presque sûre de notre estimateur ainsi que la normalité asymptotique dès que l'estimateur est excité par une perturbation exogène.

Finalement, le chapitre 7 est consacré à l'implémentation et à l'étude des performances numériques de chacune de nos procédures d'estimation pour les trois modèles

semi-paramétriques de déformation. Tout d'abord, on illustre les comportements asymptotiques de nos estimateurs sur des données simulées. Ensuite, on applique nos procédures d'estimation à des données réelles d'électrocardiogramme.

La thèse s'achève par une conclusion générale ainsi que quelques perspectives de recherche.

Première partie

Approximation stochastique et
estimation fonctionnelle

Chapitre 1

Approximation stochastique

1.1 Cadre

Les méthodes d'approximation stochastique sont une famille d'algorithmes stochastiques itératifs qui permettent d'approcher les zéros ou les extrema de fonctions ne pouvant être calculées directement, mais seulement estimées à partir d'observations bruitées. Plus précisément, soit ϕ une fonction inconnue définie sur \mathbb{R}^d pour $d \geq 1$, pouvant s'écrire, pour tout $x \in \mathbb{R}^d$, sous la forme

$$\phi(x) = \mathbb{E}[\Phi(x, \varepsilon)], \quad (1.1.1)$$

où l'on sait évaluer $\Phi(x, \varepsilon)$ qui dépend d'un vecteur aléatoire ε . Alors, les méthodes d'approximation stochastique sont des algorithmes de la forme

$$X_{n+1} = X_n + \gamma_n \Phi(X_n, \varepsilon_{n+1}) \quad (1.1.2)$$

où (ε_n) est une suite de vecteurs aléatoires et (γ_n) est une suite de pas déterministe que l'on explicitera dans la suite. Pour une référence générale sur le sujet, on peut consulter l'article de Delyon [20]. Les algorithmes stochastiques permettent en particulier de résoudre deux types de problèmes qui sont, d'une part, la recherche de points x^* pour lesquels une fonction inconnue ϕ de \mathbb{R}^d dans \mathbb{R}^d atteint un niveau donné α , c'est-à-dire la recherche de

$$\{x \in \mathbb{R}^d : \phi(x) = \alpha\},$$

et, d'autre part, la recherche de maximum d'une fonction inconnue ϕ à valeurs dans \mathbb{R} ,

$$\arg \max_{x \in \mathbb{R}^d} \phi(x).$$

L'algorithme stochastique le plus célèbre pour résoudre le premier problème est l'algorithme de Robbins-Monro, tandis que le second se résout grâce à l'algorithme de Kiefer-Wolfowitz. Ce chapitre est consacré à la présentation de ces deux algorithmes stochastiques et nous verrons dans la suite de la thèse des exemples d'application de ces algorithmes pour l'estimation dans des modèles de déformation. Cependant, avant de détailler la présentation de ces deux algorithmes, nous présentons un algorithme déterministe dont les algorithmes stochastiques sont des versions perturbées.

1.2 Une approche déterministe

Pour ϕ une fonction réelle et $\alpha \in \mathbb{R}$, on souhaite approcher le point x^* satisfaisant $\phi(x^*) = \alpha$. La proposition qui suit est une version simplifiée de la Proposition 1.2.3 page 11 du livre de Dufflo [23].

Proposition 1.2.1 *Soit ϕ une fonction continue satisfaisant*

- *Il existe $x^* \in \mathbb{R}$ tel que $\phi(x^*) = \alpha$.*
- *Pour tout $x \neq x^*$, $(x - x^*)(\phi(x) - \alpha) < 0$.*
- *Il existe $C > 0$ telle que, pour tout $x \in \mathbb{R}$, $|\phi(x)| \leq C(1 + |x|)$.*

De plus, soit (γ_n) une suite positive déterministe décroissante vers 0 et qui vérifie

$$\sum_{n=0}^{+\infty} \gamma_n = +\infty.$$

Alors, l'algorithme déterministe défini récursivement par $x_0 \in \mathbb{R}$ et pour tout $n \geq 0$,

$$x_{n+1} = x_n + \gamma_n (\phi(x_n) - \alpha) \tag{1.2.1}$$

converge vers x^ .*

Très souvent, la fonction ϕ dont on cherche le x^* tel que $\phi(x^*) = \alpha$ n'est connue qu'à une perturbation près. Il est alors difficile d'utiliser des algorithmes déterministes du type (1.2.1) pour la recherche du point x^* . Dans ces nombreux cas, on a alors recours à des algorithmes stochastiques dont celui de Robbins-Monro [56] est le plus célèbre d'entre eux.

1.3 L'algorithme de Robbins-Monro

Dans cette section, pour $\alpha \in \mathbb{R}^d$ et pour une fonction ϕ inconnue de \mathbb{R}^d dans \mathbb{R}^d , on s'intéresse au problème de chercher x^* satisfaisant

$$\phi(x^*) = \alpha. \tag{1.3.1}$$

Commençons par une heuristique de cet algorithme. On suppose que $d = 1$ et que la fonction ϕ est décroissante de \mathbb{R} dans \mathbb{R} . Soit X_0 arbitrairement choisi. A l'instant $n \geq 1$, on détermine X_n en fonction des observations antérieures, en supposant qu'on observe $\phi(X_n)$ à travers une variable aléatoire T_{n+1} qui vérifie $\mathbb{E}[T_{n+1} | \mathcal{F}_n] = \phi(X_n)$. Si $T_{n+1} \geq \alpha$, alors $\phi(X_n) \geq \alpha$ donc $X_n \leq x^*$. Pour approcher x^* il faudra ajouter une quantité positive à X_n . Au contraire, si $T_{n+1} \leq \alpha$, il sera nécessaire de soustraire une quantité positive à X_n . Ainsi, l'algorithme de Robbins-Monro est de la forme, pour tout $n \geq 0$,

$$X_{n+1} = X_n + \gamma_n (T_{n+1} - \alpha).$$

A partir de maintenant, la fonction ϕ est une fonction qui va de \mathbb{R}^d dans \mathbb{R}^d .

1.3.1 Hypothèses

Nous allons étudier l'algorithme défini par $X_0 \in \mathbb{R}^d$ et, pour tout $n \geq 0$,

$$X_{n+1} = X_n + \gamma_n (T_{n+1} - \alpha). \quad (1.3.2)$$

On note \mathcal{F}_n la σ -algèbre des événements s'étant produits avant l'instant n ,

$$\mathcal{F}_n = \sigma(X_0, T_1, \dots, X_{n-1}, T_n).$$

Afin d'établir la convergence de cet algorithme, nous faisons les hypothèses suivantes.

Sur la fonction ϕ :

$$(RM_1) \quad \phi \text{ est continue,}$$

$$(RM_2) \quad \forall x \neq x^*, (x - x^*)^T (\phi(x) - \alpha) < 0.$$

Sur la suite d'observations (T_n) :

$$(RM_3) \quad \forall n \geq 0, \quad \mathbb{E}[T_{n+1} | \mathcal{F}_n] = \phi(X_n) \text{ p.s.}$$

$$(RM_4) \quad \text{Il existe } \mu > 0 \text{ telle que } \mathbb{E}[\|T_{n+1}\|^2 | \mathcal{F}_n] \leq \mu (1 + \|X_n - x^*\|^2) \text{ p.s.}$$

Sur la suite de pas (γ_n) décroissante vers 0 :

$$(RM_5) \quad \sum_{n=0}^{+\infty} \gamma_n = +\infty \quad \text{et} \quad \sum_{n=0}^{+\infty} \gamma_n^2 < +\infty.$$

L'hypothèse (RM_2) est une hypothèse de « monotonie » de la fonction ϕ . Avec l'hypothèse (RM_1) , elle garantit en particulier l'unicité de x^* . Les hypothèses (RM_3) et (RM_4) sont essentielles, l'hypothèse (RM_3) assurant qu'on observe une valeur « pas très loin » de la vraie fonction ϕ , et l'hypothèse (RM_4) qui impose que le moment conditionnel de $\|T_{n+1}\|^2$ soit à croissance polynômiale de degré au plus 2. Si l'hypothèse (RM_4) n'est pas vérifiée, des versions modifiées de l'algorithme (1.3.2) ont été proposées par Chen et Zhu [12]. On pourra aussi consulter les travaux intéressants de Lelong [44] [45] sur cet algorithme modifié. Un exemple de suite (T_{n+1}) est donné par $T_{n+1} = \Phi(X_n, \varepsilon_{n+1})$ où (ε_{n+1}) est une suite de vecteurs aléatoires indépendants et de même loi et indépendante de (X_n) . Pour cette suite, on est bien dans le cas de (1.1.2) décrit précédemment. De plus, pour $\alpha = 0$, si ϕ est différentiable et que $T_{n+1} = \nabla \phi(X_n) + \varepsilon_{n+1}$, alors on est dans le cas particulier d'un algorithme de gradient stochastique. Les algorithmes de gradient stochastique sont donc des cas particuliers de l'algorithme de Robbins-Monro.

1.3.2 Propriétés presque sûres de l'algorithme

Dans le papier original de Robbins et Monro [56], les auteurs montrent la convergence en probabilité de l'algorithme dans le cas unidimensionnel. Depuis, la version multidimensionnelle de l'algorithme de Robbins-Monro ainsi que ses propriétés presque sûres ont été largement étudiées ([7], [51]). Le premier à s'être intéressé au cadre multidimensionnel est Blum [8]. Nous commençons par énoncer un lemme central dans la démonstration de la convergence presque sûre de l'algorithme, lemme qui nous sera utile par la suite. Ce lemme est dû à Robbins et Siegmund [57]. On pourra aussi trouver ce lemme ainsi que des corollaires dans le livre de Bercu et Chafaï [3] à partir de la page 104.

Lemme 1.3.1 (Robbins-Siegmund) Soient (V_n) , (A_n) , (B_n) et (a_n) quatre suites positives adaptées à \mathcal{F}_n . On suppose que V_0 est intégrable et pour tout $n \geq 0$,

$$\mathbb{E}[V_{n+1}|\mathcal{F}_n] \leq V_n(1 + a_n) + A_n - B_n \quad p.s.$$

Alors, si

$$\sum_{n=0}^{+\infty} a_n < +\infty \quad \text{et} \quad \sum_{n=0}^{+\infty} A_n < +\infty \quad p.s.$$

(V_n) converge p.s. vers une variable aléatoire V_∞ finie p.s. et

$$\sum_{n=0}^{+\infty} B_n < +\infty \quad p.s.$$

On donne aussi un lemme très utile dans les preuves, lemme dû à Toeplitz, qui est une généralisation du lemme de Césaro. On pourra le trouver par exemple dans le livre de Bercu et Chafaï [3] page 103.

Lemme 1.3.2 (Toeplitz) Soit (a_n) une suite de \mathbb{R}_+^* qui vérifie

$$\sum_{n=1}^{+\infty} a_n = +\infty.$$

Soit de plus une suite (x_n) de \mathbb{R} qui converge vers $x_\infty \in \mathbb{R}$. Alors, on a

$$\lim_{n \rightarrow +\infty} \frac{\sum_{k=1}^n a_k x_k}{\sum_{k=1}^n a_k} = x_\infty.$$

Le théorème suivant assure la convergence presque sûre de X_n vers x^* . On pourra en trouver une démonstration dans le livre de Duflo [23] page 29.

Théorème 1.3.1 (Robbins-Monro) Sous les hypothèses (RM_{1-5}) , la suite (X_n) définie par (1.3.2) converge presque sûrement vers x^* .

Les deux résultats suivants, précisent la vitesse de convergence presque sûre de l'algorithme en donnant une loi du log-itéré et une loi forte quadratique. Le premier à avoir obtenu une loi du log-itéré dans le cas uni-dimensionnel est Ruppert [58] alors que Le Breton et Novikov [43] se sont intéressés au cas multidimensionnel et à la loi forte quadratique. Pour ce faire, on a besoin de rajouter des hypothèses de moments un peu plus fortes sur la suite d'observations (T_n) . Pour tout $n \geq 0$, on note $\Delta_{n+1} = T_{n+1} - \phi(X_n)$ et on suppose que

(RM_6) ϕ est deux fois continûment différentiable.

(RM_7) Il existe une matrice C symétrique semi-définie positive telle que

$$\lim_{n \rightarrow +\infty} \mathbb{E}[\Delta_{n+1} \Delta_{n+1}^T | \mathcal{F}_n] = C \quad p.s.$$

(RM_8) Il existe $a > 2$ tel que

$$\sup_{n \geq 0} \mathbb{E}[\|\Delta_{n+1}\|^a | \mathcal{F}_n] < +\infty \quad p.s.$$

Il est possible de remplacer l'hypothèse (RM_6) par une hypothèse (RM'_6) un peu plus faible, qu'on peut trouver par exemple dans le papier [51]. Cette hypothèse est la suivante.

(RM'_6) Il existe une constante $b > 1$ telle que, au voisinage de x^* ,

$$\phi(x) = \alpha + H(x - x^*) + \mathcal{O}(\|x - x^*\|^b).$$

Les résultats qui suivent sont alors valables avec cette hypothèse. Néanmoins, on se place dans le cadre habituel de (RM_6) pour énoncer les Théorèmes. Dans toute la suite, on note H la différentielle de ϕ en x^* , I_d la matrice identité de taille d et on suppose que $\gamma_n = 1/n$. On rappelle qu'une matrice de Hurwitz est une matrice carrée dont toutes les valeurs propres ont une partie réelle strictement négative.

Théorème 1.3.2 *Supposons (RM_{1-8}) . Si de plus, la matrice $H + \frac{1}{2}I_d$ est une matrice de Hurwitz, alors pour tout $v \in \mathbb{R}^p$ tel que v est un vecteur propre de H^T , on a la loi du log-itéré*

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left(\frac{n}{2 \log(\log(n))} \right)^{1/2} v^T (X_n - x^*) &= - \liminf_{n \rightarrow \infty} \left(\frac{n}{2 \log(\log(n))} \right)^{1/2} v^T (X_n - x^*) \\ &= (v^T \Sigma_{RM} v)^{1/2} \quad \text{p.s.} \end{aligned} \quad (1.3.3)$$

En particulier,

$$\limsup_{n \rightarrow \infty} \left(\frac{n}{2 \log(\log(n))} \right) \|X_n - x^*\|^2 \leq \text{Tr}(P^T \Sigma_{RM} P) \quad \text{p.s.} \quad (1.3.4)$$

où P est la matrice orthogonale dont les colonnes sont les vecteurs propres de H^T et où

$$\Sigma_{RM} = \int_0^{+\infty} \exp\left(t\left(H + \frac{1}{2}I_d\right)\right) C \exp\left(t\left(H^T + \frac{1}{2}I_d\right)\right) dt. \quad (1.3.5)$$

Théorème 1.3.3 *Supposons (RM_{1-8}) . Si de plus, la matrice $H + \frac{1}{2}I_d$ est une matrice de Hurwitz, alors on a la loi forte quadratique*

$$\lim_{n \rightarrow \infty} \frac{1}{\log(n)} \sum_{k=1}^n (X_k - x^*)(X_k - x^*)^T = \Sigma_{RM} \quad \text{p.s.} \quad (1.3.6)$$

Remarque 1.3.1 *La matrice de covariance limite Σ_{RM} est la solution Σ de l'équation de Lyapunov*

$$\left(H + \frac{1}{2}I_d\right) \Sigma + \Sigma \left(H^T + \frac{1}{2}I_d\right) = -C.$$

Remarque 1.3.2 *Les Théorèmes 1.3.2 et 1.3.3 restent valables pour un pas $\gamma_n = 1/n^\alpha$ avec $1/2 < \alpha < 1$ et $\alpha a > 2$. Cette dernière condition est valable si on suppose que $a \geq 4$. Cependant, comme précisé dans [51], il est plus approprié de choisir le pas $\gamma_n = 1/n$. En effet, si $\gamma_n = 1/n^\alpha$, la loi forte quadratique (1.3.6) devient*

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1-\alpha}} \sum_{k=1}^n (X_k - x^*)(X_k - x^*)^T = \frac{1}{1-\alpha} \Sigma_{RM} \quad \text{p.s.} \quad (1.3.7)$$

Puisque $\log(n)$ est négligeable devant $n^{1-\alpha}$, le choix de $\gamma_n = 1/n$ est donc meilleur que $\gamma_n = 1/n^\alpha$.

1.3.3 Convergence en loi et efficacité asymptotique

Convergence en loi

Sous les hypothèses précédentes, nous avons également un théorème central limite, voir par exemple [52].

Théorème 1.3.4 *Supposons (RM₁₋₈). Si de plus, la matrice $H + \frac{1}{2}I_d$ est une matrice de Hurwitz, alors on a le théorème central limite*

$$\sqrt{n}(X_n - x^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_{RM}) \quad (1.3.8)$$

où Σ_{RM} est donnée par (1.3.5).

Remarque 1.3.3 *Dans le cas unidimensionnel, le Théorème 1.3.4 s'adapte de la manière suivante.*

1. Si $H < -1/2$, alors

$$\sqrt{n}(X_n - x^*) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{C}{|2H + 1|}\right).$$

2. Si $H = -1/2$, alors

$$\sqrt{\frac{n}{\log(n)}}(X_n - x^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, C).$$

3. Si $-1/2 < H < 0$, alors

$$n^{-H}(X_n - x^*) \xrightarrow{\mathcal{L}} Z,$$

où Z est une variable aléatoire finie p.s.

Pour une référence précise, on pourra consulter le Theorem 2.2.12 page 52 du livre de Dufo [23].

Efficacité asymptotique

Il est possible, via une petite modification de (1.3.2), de construire un estimateur asymptotiquement efficace X_n de x^* . Plus précisément, il suffit de remplacer $\gamma_n = 1/n$ dans (1.3.2) par $\gamma_n = -H^{-1}/n$ où H^{-1} est l'inverse de la différentielle H de ϕ en x^* . Dans le cas uni-dimensionnel, Fabian [25] a alors établi que X_n est un estimateur asymptotique efficace de x^* . Dans le cas multi-dimensionnel, c'est encore le cas et on a plus précisément le théorème central limite suivant

$$\sqrt{n}(X_n - x^*) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, H^{-1}C(H^{-1})^T\right), \quad (1.3.9)$$

où $H^{-1}C(H^{-1})^T$ est la matrice de covariance asymptotique optimale. Cependant, il se peut que H soit inconnue et donc qu'on ne puisse pas considérer le pas γ_n fournissant un estimateur asymptotiquement efficace. Pour palier ce problème, Ruppert [59] et Polyak [54] ont proposé un algorithme de moyennisation. Plus précisément, l'algorithme moyennisé de Robbins-Monro consiste à faire la moyenne des X_n définis par (1.3.2) à chaque étape de

l'algorithme. Formellement, on construit la suite \bar{X}_n en considérant $\bar{X}_0 = X_0$ et pour tout $n \geq 1$,

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k. \quad (1.3.10)$$

Cette suite vérifie alors la relation de récurrence pour tout $n \geq 0$,

$$\bar{X}_{n+1} = \bar{X}_n + \frac{1}{n+1} (X_{n+1} - \bar{X}_n)$$

et \bar{X}_n est un estimateur asymptotiquement efficace de x^* pour un pas $\gamma_n = 1/n^\alpha$, avec $1/2 < \alpha < 1$. On trouvera dans le papier de Pelletier [53] les résultats de convergence presque sûres, loi du log-itéré et loi forte quadratique, pour l'algorithme (1.3.10).

1.3.4 Convergence quadratique et bornes

Outre la convergence presque sûre et la normalité asymptotique, il est possible d'établir la convergence en moyenne quadratique de l'algorithme de Robbins-Monro. Pour cela, on considère l'algorithme défini par (1.3.2) et on pose, pour tout $n \geq 0$,

$$b_n = \mathbb{E} [\|X_n - x^*\|^2]. \quad (1.3.11)$$

Le premier à s'être intéressé à ce problème est Kallianpur [38]. Le résultat suivant donne une borne pour la convergence en moyenne quadratique sous une hypothèse de minoration de la différentielle de ϕ .

Théorème 1.3.5 *Supposons (RM₁₋₅). De plus, supposons que ϕ soit différentiable et vérifie, pour tout $x \in \mathbb{R}^d$ et pour tout $y \in \mathbb{R}^d$, l'inégalité d'attractivité uniforme,*

$$x^T D\phi(y)x \leq -\lambda \|x\|^2$$

avec $\lambda > 0$. Alors, pour tout $n \geq 1$, on a

$$b_n \leq \beta_{n-1} \left(\|X_0 - x^*\|^2 + \mu \sum_{k=0}^{n-1} \frac{\gamma_k^2}{\beta_k} \right) \quad (1.3.12)$$

où

$$\beta_{n-1} = \prod_{k=0}^{n-1} (1 - 2\lambda\gamma_k + \mu\gamma_k^2),$$

μ étant la constante donnée dans (RM₄).

On pourra également consulter le travail récent de Frikha et Menozzi [29] dans lequel ils établissent des inégalités de concentration pour les algorithmes stochastiques, le résultat pour Robbins-Monro étant le Théorème 2.2 du papier.

1.3.5 Un exemple d'application

Un premier exemple d'application de l'algorithme de Robbins-Monro est la recherche de quantiles d'une fonction de répartition inconnue. En effet, soit Z une variable aléatoire réelle de fonction de répartition F inconnue. Pour $0 < \alpha < 1$, on définit le quantile d'ordre α de Z comme le plus petit réel z_α satisfaisant

$$F(z_\alpha) = \mathbb{P}(Z \leq z_\alpha) = \alpha.$$

Par définition, F est croissante, continue à droite avec limite à gauche. Afin de garantir l'unicité de z_α , on suppose que F est continue et strictement croissante. Ainsi F vérifie (RM_1) et pour tout $z \in \mathbb{R}$ avec $z \neq z_\alpha$,

$$(z - z_\alpha)(F(z) - \alpha) > 0.$$

De plus, on suppose qu'on dispose d'une suite de variables aléatoires (Z_n) , indépendantes et de même loi que Z . Alors, pour estimer z_α , on définit l'algorithme de Robbins-Monro par $X_0 \in \mathbb{R}$ et, pour tout $n \geq 0$,

$$X_{n+1} = X_n - \gamma_n (T_{n+1} - \alpha) \tag{1.3.13}$$

où

$$T_{n+1} = \mathbb{I}_{\{Z_{n+1} \leq X_n\}}. \tag{1.3.14}$$

Les hypothèses (RM_3) et (RM_4) sont facilement vérifiées pour la suite (T_n) définie par (1.3.14). Nous prenons l'exemple d'une variable aléatoire Z de loi exponentielle de paramètre $\lambda > 0$ et nous allons notamment voir l'influence de la croissance de F et du choix du pas γ_n sur la convergence numérique de l'algorithme. La fonction de répartition de Z est définie, pour tout $x \in \mathbb{R}$, par

$$F_\lambda(x) = (1 - \exp(-\lambda x)) \mathbb{I}_{x>0}. \tag{1.3.15}$$

On a tracé sur la Figure 1.1 les fonctions de répartition pour les paramètres $\lambda_1 = 0.5$, $\lambda_2 = 1$, $\lambda_3 = 3$ et $\lambda_4 = 10$. Pour $0 < \alpha < 1$, un calcul simple entraîne que

$$z_\alpha = -\frac{\log(1 - \alpha)}{\lambda}. \tag{1.3.16}$$

Par exemple, pour $\alpha = 1/2$, z_α correspond à la médiane de Z et dans ce cas

$$z_{1/2} = \frac{\log(2)}{\lambda}. \tag{1.3.17}$$

Influence de la croissance de F_λ

On a lancé l'algorithme de Robbins-Monro pour un pas $\gamma_n = 1/n$, pour différentes valeurs de n et différentes valeurs du niveau α . On a répertorié dans le tableau ci-dessous les résultats obtenus pour la valeur absolue de la différence entre X_n et z_α .

$\lambda_1 = 0.5$	$n = 200$	$n = 1000$	$n = 5000$	$\lambda_2 = 1$	$n = 200$	$n = 1000$	$n = 5000$
$\alpha = 0.1$	0.0894	0.0056	0.0088	$\alpha = 0.1$	0.0353	0.0229	0.0039
$\alpha = 0.5$	0.3892	0.0452	0.1253	$\alpha = 0.5$	0.0507	0.0358	0.0099
$\alpha = 0.9$	1.8367	1.6953	1.6676	$\alpha = 0.9$	0.3193	0.1929	0.4892

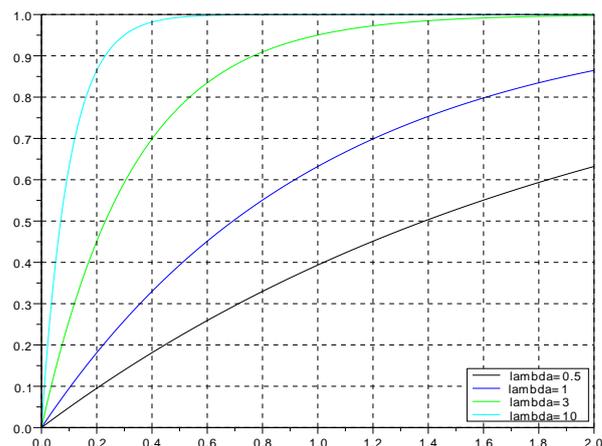


FIGURE 1.1 – Fonction de répartition de la loi exponentielle

$\lambda_3 = 3$	$n = 200$	$n = 1000$	$n = 5000$	$\lambda_4 = 10$	$n = 200$	$n = 1000$	$n = 5000$
$\alpha = 0.1$	0.1260	0.0049	0.0032	$\alpha = 0.1$	0.0044	0.0197	0.0015
$\alpha = 0.5$	0.0511	0.0138	0.0033	$\alpha = 0.5$	0.0100	10^{-5}	0.0005
$\alpha = 0.9$	0.1015	0.1775	0.0198	$\alpha = 0.9$	0.0031	0.0018	0.0035

Il y a essentiellement deux choses à noter au regard de ces résultats numériques. La première est que plus le paramètre λ est grand, meilleure est l'estimation de z_α . La deuxième est que plus le niveau α est grand, moins l'algorithme de Robbins-Monro est performant. Cela signifie que l'estimation de z_α par l'algorithme de Robbins-Monro est meilleure quand la croissance de la fonction F_λ est forte et quand la valeur à estimer se trouve à un endroit où F_λ est fortement croissante. Ainsi, pour des fonctions à croissance lente, l'algorithme va mettre du temps à converger vers une valeur convenable.

Influence du choix de γ_n

Pour palier ce problème, on peut faire varier le pas γ_n . Jusqu'à présent, nous avons considéré un pas $\gamma_n = 1/n$ ou $\gamma_n = 1/n^\alpha$ avec $1/2 < \alpha < 1$ qui sont les choix naturels pour vérifier l'hypothèse (RM_5) . Néanmoins, comme nous allons le voir dans le tableau suivant, le choix du pas a une véritable influence sur la convergence numérique de X_n vers z_α . Aussi, nous avons choisi trois types de pas vérifiant (RM_5) qui sont $\gamma_n = 1/n$, $\gamma_n = 1/n^{3/4}$ et $\gamma_n = 1/n^{1.1/2}$ et nous nous sommes placés dans le cas $\lambda_1 = 0.5$ et $\alpha = 0.9$ qui mettait le plus en difficulté l'algorithme de Robbins-Monro. Pour chaque γ_n , on a lancé 100 fois l'algorithme de Robbins-Monro et on a fait la moyenne des valeurs obtenues, afin que le résultat soit plus robuste. On a alors répertorié dans le tableau ci-dessous les résultats

obtenus pour la valeur absolue de la différence entre X_n et z_α pour chaque choix de pas.

	$n = 200$	$n = 1000$	$n = 5000$
$\gamma_n = 1/n$	2.0182	1.0806	0.2314
$\gamma_n = 1/n^{3/4}$	1.8082	0.9472	0.3080
$\gamma_n = 1/n^{1.1/2}$	1.3014	0.2860	0.0275

Il est donc clair au vu des résultats numériques que le meilleur pas pour ces simulations est $\gamma_n = 1/n^{1.1/2}$ et que les différences de convergence entre les algorithmes dépendent fortement du pas choisi. D'une manière plus générale, le choix du pas se fait en ajustant les valeurs c , d et α pour

$$\gamma_n = \frac{c}{(d+n)^\alpha}. \quad (1.3.18)$$

Pour conclure, nous avons illustré ci-dessous la normalité asymptotique de $\sqrt{n}(X_n - x^*)$ pour le choix de $\lambda_3 = 3$.

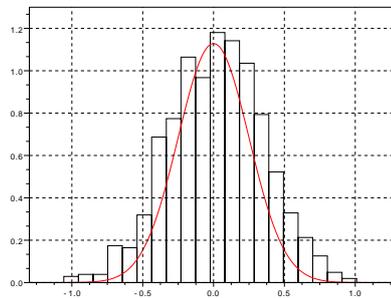


FIGURE 1.2 – Illustration de la normalité asymptotique pour $\lambda_3 = 3$

1.3.6 Projection

Dans plusieurs situations, on sait que le paramètre x^* est situé à l'intérieur d'un convexe fermé K de \mathbb{R}^d . Dans ce cas, on aimerait assurer que la suite (X_n) définie par l'algorithme de Robbins-Monro (1.3.2) appartienne bien à K . Pour cela, on peut considérer une version projetée de l'algorithme (1.3.2). Plus précisément, on définit une nouvelle suite (X_n) par $X_0 \in K$ et pour tout $n \geq 0$,

$$X_{n+1} = \pi_K(X_n + \gamma_n(T_{n+1} - \alpha)), \quad (1.3.19)$$

où π_K est la projection euclidienne sur K .

Propriétés presque sûres

Pour tout $n \geq 0$, X_n appartient bien à K , et si (X_n) converge, alors (X_n) va converger vers un élément de K puisque K est fermé. Le résultat suivant assure la convergence presque sûre de X_n vers x^* . On pourra le trouver sous une forme générale dans le livre de Kushner et Yin [40] page 127.

Théorème 1.3.6 *Supposons (RM_{1-5}) . Alors la suite (X_n) définie par (1.3.19) converge presque sûrement vers x^* .*

De plus, comme pour l'algorithme standard de Robbins-Monro, on peut préciser des vitesses de convergence presque sûre pour la version projetée. Elles se déduisent très facilement des résultats précédents, en utilisant le fait que le nombre de fois que $X_n + \gamma_n(T_{n+1} - \alpha)$ sort de K est fini presque sûrement. Les Théorèmes 1.3.2 et 1.3.3 sont alors valables pour la suite projetée.

Convergence en loi

Alors que le comportement presque sûr de la suite projetée (X_n) définie par (1.3.19) se déduit facilement de celui de l'algorithme standard de Robbins-Monro, il est un peu plus délicat de déduire du théorème central limite 1.3.4 un résultat équivalent pour l'algorithme projeté. En effet, il pourrait être tentant de dire que comme X_n converge presque sûrement vers x^* , alors « à partir d'un certain rang », l'algorithme projeté a le même comportement en loi que l'algorithme non projeté. Le problème majeur est que le nombre de fois que $X_n + \gamma_n(T_{n+1} - \alpha)$ sort de K n'est pas borné mais seulement fini presque sûrement et c'est à cause de cela qu'on ne peut pas déduire immédiatement des résultats relatifs à la convergence en loi pour l'algorithme projeté de Robbins-Monro à partir de résultats sur l'algorithme de Robbins-Monro standard. En particulier, le point central à vérifier pour la suite projetée est la tension de la suite (W_n) définie, pour tout $n \geq 0$, par

$$W_n = \frac{X_n - x^*}{\sqrt{\gamma_n}}. \quad (1.3.20)$$

On rappelle que (W_n) soit tendue, si pour tout $\varepsilon > 0$, il existe un compact A de \mathbb{R}^d telle que

$$\sup_{n \geq 0} \mathbb{P}(W_n \in A^c) < \varepsilon.$$

On a alors le résultat suivant qui est une version simplifiée du Théorème 2.1 page 330 du livre de Kushner et Yin [40].

Théorème 1.3.7 *Supposons (RM_{1-8}) et que la suite (W_n) soit tendue. Si de plus, la matrice $H + \frac{1}{2}I_d$ est une matrice de Hurwitz, alors on a le théorème central limite*

$$\sqrt{n}(X_n - x^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_{RM}) \quad (1.3.21)$$

où Σ_{RM} est donnée par (1.3.5).

On pourra trouver dans le livre de Kushner et Yin [40] des conditions sur la suite (T_n) qui assurent la tension de la suite (W_n) , notamment dans le Théorème 4.1 page 341.

Remarque 1.3.4 *Pour l'algorithme projeté que nous avons présenté, l'hypothèse (RM_2) peut être affaiblie. En effet, il suffit de supposer que (RM_2) est vérifiée seulement sur le convexe fermé K de \mathbb{R}^d , c'est-à-dire qu'il suffit de supposer que pour tout $x \neq x^*$ et $x \in K$,*

$$(x - x^*)^T(\phi(x) - \alpha) < 0.$$

Tous les théorèmes présentés précédemment restent alors valables sous cette hypothèse.

1.4 L'algorithme de Kiefer-Wolfowitz

On s'intéresse maintenant au deuxième problème évoqué au début du chapitre qui est la recherche de x^* , pour une fonction inconnue ϕ , défini par

$$x^* = \arg \max_{x \in \mathbb{R}^d} \phi(x). \quad (1.4.1)$$

Comme dans le cas de l'algorithme de Robbins-Monro, commençons par une heuristique de l'algorithme. Soit ϕ une fonction qui va de \mathbb{R} dans \mathbb{R} . Si ϕ est dérivable, alors par définition de x^* , on a $\phi'(x^*) = 0$. On sait alors que si on dispose d'observations T_{n+1} telle que $\mathbb{E}[T_{n+1} | \mathcal{F}_n] = \phi'(X_n)$, l'algorithme de Robbins-Monro défini par $X_0 \in \mathbb{R}$ et pour tout $n \geq 0$,

$$X_{n+1} = X_n + \gamma_n T_{n+1}$$

va converger p.s. vers x^* sous de bonnes hypothèses sur ϕ . Néanmoins, il se peut que nous n'ayons que des observations T_{n+1} nous permettant d'approcher ϕ et non ϕ' . On a alors recours à l'algorithme de Kiefer-Wolfowitz. L'idée est que même si on ne peut pas approcher directement ϕ' , on peut quand même approcher la dérivée par différentiation de ϕ qui sera une bonne approximation de ϕ' . En particulier, la suite d'observations T_{n+1} définie par

$$T_{n+1} = \frac{T_{n+1}^+ - T_{n+1}^-}{2c_n}$$

où les suites (T_{n+1}^+) et (T_{n+1}^-) sont telles que

$$\begin{aligned} \mathbb{E}[T_{n+1}^+ | \mathcal{F}_n] &= \phi(X_n + c_n) && \text{p.s.} \\ \mathbb{E}[T_{n+1}^- | \mathcal{F}_n] &= \phi(X_n - c_n) && \text{p.s.} \end{aligned}$$

avec (c_n) une suite déterministe à préciser, est une bonne approximation de ϕ' . L'algorithme de Kiefer-Wolfowitz est donc de la forme

$$X_{n+1} = X_n + \frac{\gamma_n}{2c_n} (T_{n+1}^+ - T_{n+1}^-).$$

Dans la suite de cette section, on considère dorénavant que ϕ est une fonction qui va de \mathbb{R}^d dans \mathbb{R} .

1.4.1 Hypothèses

Nous allons étudier l'algorithme de Kiefer-Wolfowitz défini par $X_0 \in \mathbb{R}^d$ et pour tout $n \geq 0$,

$$X_{n+1} = X_n + \frac{\gamma_n}{2c_n} (T_{n+1}^+ - T_{n+1}^-). \quad (1.4.2)$$

On note \mathcal{F}_n la σ -algèbre des événements s'étant produits avant l'instant n ,

$$\mathcal{F}_n = \sigma(X_0, T_1^+, T_1^-, \dots, X_{n-1}, T_n^+, T_n^-).$$

Afin d'établir la convergence de cet algorithme, en notant e_j est le j -ème vecteur de la base canonique de \mathbb{R}^d , nous faisons les hypothèses suivantes.

Sur la fonction ϕ :

(KW₁) ϕ est deux fois continûment différentiable et strictement concave,

(KW₂) Il existe $C_1 > 0$ telle que pour tout $x \in \mathbb{R}^d$, $\|D^2\phi(x)\| \leq C_1(1 + \|x\|)$.

Sur les vecteurs d'observations T_{n+1}^+ et T_{n+1}^- :

(KW₃) $\forall n \geq 0$, et pour tout $1 \leq j \leq d$,

$$\mathbb{E}[T_{n+1,j}^+ | \mathcal{F}_n] = \phi(X_n + c_n e_j) \quad \text{et} \quad \mathbb{E}[T_{n+1}^- | \mathcal{F}_n] = \phi(X_n - c_n e_j) \quad \text{p.s.}$$

(KW₄) Il existe $C_2 > 0$ telle que pour tout $n \geq 0$,

$$\mathbb{E}[\|T_{n+1}^+\|^2 | \mathcal{F}_n] \leq C_2(1 + \|X_n - x^*\|^2) \quad \text{et} \quad \mathbb{E}[\|T_{n+1}^-\|^2 | \mathcal{F}_n] \leq C_2(1 + \|X_n - x^*\|^2) \quad \text{p.s.}$$

Sur les suites (γ_n) et (c_n) décroissantes vers 0 :

$$(KW_5) \quad \sum_{n=0}^{+\infty} \gamma_n = +\infty, \quad \sum_{n=0}^{+\infty} \gamma_n c_n < +\infty \quad \text{et} \quad \sum_{n=0}^{+\infty} \frac{\gamma_n^2}{c_n^2} < +\infty.$$

L'hypothèse (KW₁) garantit l'unicité de x^* alors que l'hypothèse (KW₂) impose un contrôle de la hessienne de ϕ . L'hypothèse (KW₁) peut être remplacée par une hypothèse locale. En effet, on peut supposer que x^* est l'unique maximum de ϕ et imposer la stricte concavité de ϕ seulement au voisinage de x^* et non sur \mathbb{R}^d tout entier. Pour ce type d'hypothèse, on pourra par exemple consulter le livre de Hall et Heyde [33] page 238. Les hypothèses (KW₃) et (KW₄), semblables à (RM₃) et (RM₄), sont essentielles. L'hypothèse (KW₃) permet de dire que la quantité

$$\frac{T_{n+1}^+ - T_{n+1}^-}{2c_n}$$

est un estimateur asymptotiquement non biaisé de la dérivée par différentiation de ϕ alors que l'hypothèse (KW₄) impose que le moment conditionnel de $\|T_{n+1}^+\|^2$ et de $\|T_{n+1}^-\|^2$ soit à croissance polynômiale de degré au plus 2.

1.4.2 Propriétés presque sûres de l'algorithme

Comme pour l'algorithme de Robbins-Monro, dans le papier original de Kiefer et Wolfowitz [39], les auteurs montrent la convergence en probabilité dans le cas unidimensionnel. Depuis, le cas multidimensionnel de l'algorithme de Kiefer-Wolfowitz ainsi que ses propriétés presque sûres ont été étudiés par les mêmes auteurs que pour Robbins-Monro ([7], [51]). Le théorème suivant assure la convergence presque sûre de X_n vers x^* . On pourra en trouver une démonstration dans le livre de Duflo [23] page 31.

Théorème 1.4.1 *Sous les hypothèses (KW₁₋₅), la suite (X_n) définie par (1.4.2) converge presque sûrement vers x^* .*

Comme pour l'algorithme de Robbins-Monro, une loi du log-itéré ainsi qu'une loi forte quadratique existent pour l'algorithme de Kiefer-Wolfowitz. Pour ce faire, on a besoin de rajouter des hypothèses de moments un peu plus fortes sur les suites d'observations (T_{n+1}^+) et (T_{n+1}^-) . Pour cela, pour tout $n \geq 0$, on note $\Delta_{n+1} = \Delta_{n+1}^+ - \Delta_{n+1}^-$ où pour tout $1 \leq j \leq p$, $\Delta_{n+1,j}^+ = T_{n+1,j}^+ - \phi(X_n + c_n e_j)$ et $\Delta_{n+1,j}^- = T_{n+1,j}^- - \phi(X_n - c_n e_j)$ et on fait des hypothèses semblables à (RM_6) , (RM_7) et (RM_8) .

(KW_6) ϕ est trois fois continûment différentiable.

(KW_7) Il existe une matrice C symétrique semi-définie positive telle que

$$\lim_{n \rightarrow +\infty} \mathbb{E} [\Delta_{n+1} \Delta_{n+1}^T | \mathcal{F}_n] = C \text{ p.s.}$$

(KW_8) Il existe $a > 2$ tel que

$$\sup_{n \geq 0} \mathbb{E} [\|\Delta_{n+1}\|^a | \mathcal{F}_n] < +\infty \text{ p.s.}$$

On note H la matrice hessienne de ϕ en x^* et on suppose que $\gamma_n = 1/n$ et $c_n = 1/n^\gamma$ avec $0 < \gamma < 1/2$.

Théorème 1.4.2 *Supposons (KW_{1-8}) . Si de plus, la matrice $H + (\frac{1}{2} - \gamma) I_d$ est une matrice de Hurwitz, alors pour tout $v \in \mathbb{R}^p$ tel v est un vecteur propre de H^T , on a la loi du log-itéré*

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left(\frac{n^{1-2\gamma}}{2 \log(\log(n))} \right)^{1/2} v^T (X_n - x^*) &= - \liminf_{n \rightarrow \infty} \left(\frac{n^{1-2\gamma}}{2 \log(\log(n))} \right)^{1/2} v^T (X_n - x^*) \\ &= (v^T \Sigma_{KW} v)^{1/2} \quad \text{p.s.} \end{aligned} \quad (1.4.3)$$

En particulier,

$$\limsup_{n \rightarrow \infty} \left(\frac{n^{1-2\gamma}}{2 \log(\log(n))} \right) \|X_n - x^*\|^2 \leq \text{Tr}(P^T \Sigma_{KW} P) \quad \text{p.s.} \quad (1.4.4)$$

où P est la matrice orthogonale dont les colonnes sont les vecteurs propres de H^T et Σ_{KW} est donnée par

$$\Sigma_{KW} = \frac{1}{4} \int_0^{+\infty} \exp\left(t \left(H + \left(\frac{1}{2} - \gamma \right) I_d \right)\right) C \exp\left(t \left(H^T + \left(\frac{1}{2} - \gamma \right) I_d \right)\right) dt. \quad (1.4.5)$$

Théorème 1.4.3 *Supposons (KW_{1-8}) . Si de plus, la matrice $H + (\frac{1}{2} - \gamma) I_d$ est une matrice de Hurwitz, alors on a la loi forte quadratique*

$$\lim_{n \rightarrow \infty} \frac{1}{\log(n)} \sum_{k=1}^n k^{-2\gamma} (X_k - x^*) (X_k - x^*)^T = \Sigma_{KW} \quad \text{p.s.} \quad (1.4.6)$$

Remarque 1.4.1 *En principe, les constantes γ_n et c_n peuvent être choisies afin d'obtenir une vitesse de convergence optimale en un certain sens. On peut notamment consulter le travail de Sacks [60] sur le sujet. Cependant, pour simplifier les choses, nous ne rentrerons pas dans les détails de ces choix.*

1.4.3 Convergence en loi

Afin d'établir la normalité asymptotique de X_n , on note

$$V(x^*) = \left(\partial_1^3 \phi(x^*), \dots, \partial_d^3 \phi(x^*) \right)^T.$$

Sous les hypothèses précédentes, on a alors un théorème central limite, voir par exemple [52].

Théorème 1.4.4 *Supposons (KW₁₋₈). Si de plus, la matrice $H + \left(\frac{1}{2} - \gamma\right) I_d$ est une matrice de Hurwitz, alors on a le théorème central limite*

$$n^{1/2-\gamma}(X_n - x^*) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\left(H + \left(\frac{1}{2} - \gamma\right) I_d\right)^{-1} V(x^*), \Sigma_{KW}\right) \quad (1.4.7)$$

où Σ_{KW} est donnée par (1.4.5).

Remarque 1.4.2 *Dans le cas unidimensionnel, on peut voir dans le livre de Hall et Heyde [33] page 242 que la moyenne donnée dans le théorème central limite (1.4.7) est nulle. Ceci est dû au fait qu'ils font l'hypothèse que pour tout $c_0 > 0$, il existe $K_1 > 0$ et $K_2 > 0$ telles que pour tout $0 < c < c_0$ et pour tout x tel que $|x - x^*| < c$, alors*

$$K_1 (x - x^*)^2 \leq (x - x^*) \left(\frac{\phi(x+c) - \phi(x-c)}{2c} \right) \leq K_2 (x - x^*)^2.$$

Cette hypothèse implique que ϕ est symétrique dans un voisinage de x^ et plus précisément que pour tout $c < c_0$, $\phi(x^* + c) = \phi(x^* - c)$. Ainsi, la fonction $\phi^{(3)}$ est impaire et $V(x^*) = 0$.*

1.4.4 Projection

Exactement comme pour l'algorithme de Robbins-Monro, on peut considérer une version projetée de l'algorithme (1.4.2) dans le cas où l'on sait que l'argument maximal x^* est situé à l'intérieur d'un convexe fermé K de \mathbb{R}^d . Plus précisément, on définit une nouvelle suite (X_n) par $X_0 \in K$ et pour tout $n \geq 0$,

$$X_{n+1} = \pi_K \left(X_n + \frac{\gamma_n}{2c_n} (T_{n+1}^+ - T_{n+1}^-) \right), \quad (1.4.8)$$

où π_K est la projection euclidienne sur K .

Propriétés presque sûres

On pourra trouver sous une forme générale, dans le livre de Kushner et Yin [40] page 143, le résultat suivant.

Théorème 1.4.5 *Supposons (KW₁₋₅). Alors, la suite (X_n) définie par (1.4.8) converge presque sûrement vers x^* .*

De plus, on peut préciser des vitesses de convergence presque sûre pour la version projetée puisque le nombre de fois que $\frac{\gamma_n}{2c_n} (T_{n+1}^+ - T_{n+1}^-)$ sort de K est fini presque sûrement. Les Théorèmes 1.4.2 et 1.4.3 sont alors valables pour la suite projetée.

Convergence en loi

Pour les raisons que nous avons détaillées à la section 1.3.6, afin d'obtenir un résultat de convergence en loi pour l'algorithme projeté (1.4.8) pour le choix de pas $\gamma_n = 1/n$ et $c_n = 1/n^\gamma$, on a besoin de vérifier la tension de la suite (W_n) définie, pour tout $n \geq 0$, par

$$W_n = \frac{X_n - x^*}{n^{\gamma-1/2}}. \quad (1.4.9)$$

Plus précisément, on a alors le résultat suivant qui est une version simplifiée du Théorème 3.1 page 334-335 du livre de Kushner et Yin [40].

Théorème 1.4.6 *Supposons (KW_{1-8}) et que la suite (W_n) est tendue. Si de plus, la matrice $H + (\frac{1}{2} - \gamma)I_d$ est une matrice de Hurwitz, alors on a le théorème central limite*

$$n^{1/2-\gamma}(X_n - x^*) \xrightarrow{\mathcal{L}} \mathcal{N}\left(\left(H + \left(\frac{1}{2} - \gamma\right)I_d\right)^{-1} V(x^*), \Sigma_{KW}\right) \quad (1.4.10)$$

où Σ_{KW} est donnée par (1.4.5).

1.5 Un cadre non martingale

Revenons au cas général des algorithmes stochastiques satisfaisant la relation (1.1.2)

$$X_{n+1} = X_n + \gamma_n \Phi(X_n, \varepsilon_{n+1}).$$

Comme nous l'avons vu, le cas où la suite (ε_n) est constituée de vecteurs aléatoires indépendants et de même loi est le cadre de l'algorithme de Robbins-Monro étudié précédemment. Pour cet algorithme, on peut alors écrire

$$X_{n+1} = X_n + \gamma_n (\Phi(X_n, \varepsilon_{n+1}) - \phi(X_n)) + \gamma_n \phi(X_n)$$

et $\delta_{n+1} = \Phi(X_n, \varepsilon_{n+1}) - \phi(X_n)$ correspond à un accroissement de martingale. Cette propriété facilite bien les choses et on peut notamment utiliser les nombreux théorèmes de convergence de martingales pour déduire des résultats de convergence pour l'algorithme. Cependant, il peut exister des cas où la suite (ε_n) n'est pas constituée de vecteurs aléatoires, indépendants et de même loi et donc où la théorie des martingales ne s'applique pas directement. Nous présentons dans un premier temps le cas où la suite (ε_n) admet une dépendance markovienne, puis nous présentons brièvement la méthode de l'ODE qui est une méthode générale pour l'étude des algorithmes de la forme (1.1.2).

1.5.1 Innovations markoviennes

On se place ici dans le cas où la suite ε_n possède une dépendance markovienne. Plus précisément, on suppose que les variables ε_n vérifient la relation

$$\mathbb{P}(\varepsilon_{n+1} \in B | \mathcal{F}_n) = \Pi_{X_n}(\varepsilon_n, B),$$

où pour chaque x , la probabilité de transition Π_x admet une unique mesure de probabilité invariante Γ_x . De plus, on note

$$\phi(x) = \mathbb{E}_{\Gamma_x}[\Phi(x, \varepsilon)] = \int \Phi(x, y) \Gamma_x(dy).$$

Nous pouvons noter ici l'analogie avec le cadre de l'algorithme de Robbins-Monro où la fonction ϕ vérifiait $\phi(x) = \mathbb{E}[\Phi(x, \varepsilon)]$. Alors, on peut écrire

$$X_{n+1} = X_n + \gamma_n \phi(X_n) + \gamma_n (\Phi(X_n, \varepsilon_{n+1}) - \phi(X_n)),$$

et $\delta_{n+1} = \Phi(X_n, \varepsilon_{n+1}) - \phi(X_n)$ n'est plus un accroissement de martingale. En revanche, l'équation de Poisson

$$\Phi(X_n, \varepsilon_{n+1}) - \phi(X_n) = v(X_n, \varepsilon_{n+1}) - \Pi_{X_n} v(X_n, \varepsilon_{n+1})$$

aussi égale à

$$\Phi(X_n, \varepsilon_{n+1}) - \int \Phi(X_n, y) \Gamma_{X_n}(dy) = v(X_n, \varepsilon_{n+1}) - \int v(X_n, y) \Pi_{X_n}(\varepsilon_{n+1}, dy)$$

admet une solution v si Φ vérifie une hypothèse de type mélange-Lipchitz que nous ne détaillons pas ici. Pour plus de détails, on pourra consulter le livre de Duflo [23] page 200. De plus, cette solution s'écrit

$$v(X_n, \varepsilon_{n+1}) = \sum_{k=0}^{+\infty} \left(\Pi^k \Phi(X_n, \varepsilon_{n+1}) - \int \Phi(X_n, y) \Gamma_{X_n}(dy) \right).$$

On peut alors écrire la décomposition

$$\begin{aligned} X_{n+1} &= X_n + \gamma_n \phi(X_n) + \gamma_n (v(X_n, \varepsilon_{n+1}) - \Pi_{X_n} v(X_n, \varepsilon_{n+1})) \\ &= X_n + \gamma_n \phi(X_n) + \gamma_n (\delta_{n+1} + r_n^1 + r_n^2) \end{aligned}$$

où

$$\begin{aligned} \delta_{n+1} &= v(X_n, \varepsilon_{n+1}) - \Pi_{X_n} v(X_n, \varepsilon_n), \\ r_n^1 &= \Pi_{X_n} v(X_n, \varepsilon_n) - \Pi_{X_{n-1}} v(X_{n-1}, \varepsilon_n), \\ r_n^2 &= \Pi_{X_{n-1}} v(X_{n-1}, \varepsilon_n) - \Pi_{X_n} v(X_n, \varepsilon_{n+1}). \end{aligned}$$

On peut alors remarquer que δ_{n+1} est un accroissement de martingale. Par rapport à l'algorithme de Robbins-Monro, on est en présence de deux termes de reste en plus à étudier qui sont r_n^1 et r_n^2 . On renvoie au livre de Benveniste, Métivier et Priouret [2] à partir de la page 220 pour une étude précise de cet algorithme. Une introduction à cet algorithme se trouve également dans le livre de Duflo [24] page 131.

1.5.2 Méthode de l'équation différentielle ordinaire

Une façon générale d'étudier le comportement asymptotique de l'algorithme (1.1.2)

$$X_{n+1} = X_n + \gamma_n \Phi(X_n, \varepsilon_{n+1})$$

est de le comparer à celui de l'équation différentielle

$$\frac{dx}{dt} = \bar{H}(x) \tag{1.5.1}$$

où

$$\bar{H}(x) = \lim_{n \rightarrow +\infty} \mathbb{E}[\Phi(x, \varepsilon_n)].$$

Cette méthode, appelée méthode de l'équation différentielle ordinaire (en anglais, *ordinary differential equation*, abrégé en ODE) a été introduite par Ljung [47] et Kushner Clark [41] et a été ensuite largement étudiée. Les livres de Kushner et Clark [41] et de Benveniste, Métivier et Priouret [2] sont d'excellentes références sur le sujet. Sans rentrer dans les détails de la méthode, l'idée principale est de décrire le comportement asymptotique de l'algorithme en fonction de celui de l'équation différentielle (1.5.1). Pour les algorithmes stochastiques à pas γ_n décroissant, le comportement de l'équation différentielle (1.5.1) est relié à celui de l'algorithme (1.1.2) de la façon suivante : si x^* est un équilibre stable de (1.5.1) (c'est à dire que toute trajectoire de l'équation différentielle qui rentre dans un bassin d'attraction de x^* à t_0 , y reste pour $t \geq t_0$) et que la suite (X_n) rentre infiniment souvent dans un compact inclus dans le bassin d'attraction de l'équilibre x^* , alors (X_n) converge presque sûrement vers x^* . Plus tard, Benaïm [1] a décrit le comportement asymptotique de l'algorithme (1.1.2) en fonction de celui de l'équation différentielle (1.5.1). On pourra également consulter le travail de Fort et Pages [26] sur le sujet.

Chapitre 2

Estimation fonctionnelle

Ce chapitre est consacré à la présentation de méthodes d'estimation fonctionnelle et plus particulièrement des méthodes à noyaux. Un excellent livre sur l'estimation non paramétrique est le livre d'Alexandre Tsybakov dans sa version française [64] ou anglaise [65].

2.1 Introduction

2.1.1 Estimation non paramétrique

Contrairement aux modèles paramétriques, les modèles non paramétriques sont des modèles dont la structure n'est pas spécifiée *a priori* mais déterminée à partir des données. Nous allons nous intéresser à deux types d'estimation non paramétrique qui sont, d'une part, l'estimation d'une densité de probabilité et, d'autre part, l'estimation d'une fonction de régression. Précisément, en ce qui concerne l'estimation d'une densité de probabilité, supposons que l'on dispose d'une suite de variables aléatoires (X_n) indépendantes et de même loi qui possède une densité inconnue f par rapport à la mesure de Lebesgue sur \mathbb{R} . Si l'on sait *a priori* que f appartient à une famille paramétrée $\{f_\theta(x) : \theta \in \Theta\}$ avec $\Theta \subset \mathbb{R}^d$, $d \geq 1$ où pour tout $t \in \Theta$, f_t est une fonction connue, alors estimer f est équivalent à estimer le paramètre θ . On est donc dans un cadre d'estimation paramétrique. En revanche, si $f \in \mathcal{F}$ où la classe \mathcal{F} n'est pas en bijection avec un sous-ensemble d'un espace fini-dimensionnel, alors on a affaire à un problème d'estimation non paramétrique. Un exemple de classe \mathcal{F} est l'ensemble des fonctions de densité de probabilité continues sur \mathbb{R} .

Ensuite, pour l'estimation d'une fonction de régression, supposons que l'on dispose d'une suite (X_n, Y_n) de couples aléatoires indépendants liés, via la fonction de lien f ,

$$Y_n = f(X_n) + \varepsilon_n \tag{2.1.1}$$

où la suite de variables aléatoires (ε_n) est une suite de variables aléatoires indépendantes, centrées et de même loi, et où la fonction de régression f est inconnue. Le problème de régression non paramétrique est celui de l'estimation de f lorsque l'on sait que f appartient à un ensemble non paramétrique infini-dimensionnel \mathcal{F} . \mathcal{F} peut être par exemple l'ensemble des fonctions continues de $[0; 1]$ dans \mathbb{R} ou encore l'ensemble des fonctions lipschitziennes de constantes de Lipschitz plus petites que 1. Il existe plusieurs méthodes d'estimation non paramétrique, les méthodes de splines, les histogrammes, les estimateurs par polynômes

locaux etc. Dans ce chapitre, nous nous focalisons sur l'une d'entre elles, l'estimation par noyaux.

2.1.2 Noyaux

Un noyau K est une fonction définie sur \mathbb{R} et à valeurs dans \mathbb{R}^+ qui est intégrable et vérifie pour tout $x \in \mathbb{R}$, $K(x) = K(-x)$ et

$$\int_{-\infty}^{+\infty} K(x) dx = 1.$$

En particulier, un noyau est une densité de probabilité. En général, la définition d'un noyau ne nécessite pas d'hypothèse de symétrie. Cependant, dans la pratique, c'est quasiment toujours le choix qui est fait, c'est pourquoi nous faisons cette hypothèse. Il est également assez fréquent de choisir un noyau à support compact. De plus, on peut remarquer que pour tout $h > 0$, $K_h = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$ est encore un noyau. En particulier, cette propriété permet de choisir une échelle h adaptée aux données. Plusieurs types de noyaux sont couramment utilisés, notamment le noyau uniforme,

$$K(x) = \frac{1}{2} \mathbb{I}_{|x| \leq 1}, \quad (2.1.2)$$

le noyau d'Epanechnikov,

$$K(x) = \frac{3}{4} (1 - x^2) \mathbb{I}_{|x| \leq 1}, \quad (2.1.3)$$

ou encore le noyau gaussien,

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2). \quad (2.1.4)$$

On peut également citer le noyau triangle, le noyau circulaire, le noyau quadratique ou le noyau cubique. La suite du chapitre est consacrée à la présentation des estimateurs de Parzen-Rosenblatt et de Nadaraya-Watson non-récurrents et récurrents ainsi que leurs propriétés asymptotiques.

2.2 Estimation d'une densité de probabilité

Dans cette partie, on s'intéresse au premier problème évoqué en début de chapitre. On suppose que l'on dispose d'une suite (X_n) de variables aléatoires indépendantes et de même loi de densité f .

2.2.1 Intuition et construction de l'estimateur

Si on note F la fonction de répartition associée à f , alors par la loi forte des grands nombres, l'estimateur \widehat{F}_n défini, pour tout $x \in \mathbb{R}$, par

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{I}_{\{X_k \leq x\}} \quad (2.2.1)$$

converge p.s. vers $F(x)$. De plus, comme $F(x) = \int_{-\infty}^x f(t)dt$, pour tout $h > 0$ suffisamment petit, la dérivée par différentiation

$$\frac{F(x+h) - F(x-h)}{2h}$$

est une bonne approximation de $f(x)$. Ainsi, l'estimateur $\widehat{f}_n(x)$ défini, pour tout $x \in \mathbb{R}$, par

$$\begin{aligned}\widehat{f}_n(x) &= \frac{\widehat{F}_n(x+h) - \widehat{F}_n(x-h)}{2h} \\ &= \frac{1}{2nh} \sum_{k=1}^n \mathbf{I}_{\{x-h < X_k \leq x+h\}}\end{aligned}$$

est un estimateur fortement consistant de $f(x)$. L'estimateur \widehat{f}_n est appelé l'estimateur de Rosenblatt. On peut le réécrire sous la forme

$$\widehat{f}_n(x) = \frac{1}{nh} \sum_{k=1}^n K_U \left(\frac{X_k - x}{h} \right) \quad (2.2.2)$$

où K_U est le noyau uniforme défini par (2.1.2). Une simple généralisation de l'estimateur de Rosenblatt est l'estimateur de Parzen-Rosenblatt défini par

$$\widehat{f}_n(x) = \frac{1}{nh} \sum_{k=1}^n K \left(\frac{X_k - x}{h} \right) \quad (2.2.3)$$

où K est une fonction noyau quelconque. De plus, le paramètre h est appelée fenêtre de lissage de l'estimateur. Dans tout ce qui suit, nous allons considérer une fenêtre de lissage h dépendant de n et notée h_n et nous supposerons que h_n tend vers 0 quand n tend vers l'infini. Enfin, notons que \widehat{f}_n est une densité de probabilité.

Remarque 2.2.1 *L'estimateur de Parzen-Rosenblatt se généralise au cadre multidimensionnel. Tout d'abord, supposons que l'on dispose de n couples de variables aléatoires $(X_1, Y_1), \dots, (X_n, Y_n)$ indépendants et de même loi de densité $f(x, y)$ définie sur \mathbb{R}^2 . Alors, un estimateur à noyau $\widehat{f}_n(x, y)$ de $f(x, y)$, est donné par*

$$\widehat{f}_n(x, y) = \frac{1}{nh_1 h_2} \sum_{k=1}^n K_1 \left(\frac{X_k - x}{h_1} \right) K_2 \left(\frac{Y_k - y}{h_2} \right) \quad (2.2.4)$$

où K_1 et K_2 sont deux noyaux possiblement identiques et h_1 et h_2 sont les fenêtres de lissage associées à K_1 et K_2 . Ensuite, supposons que l'on dispose de n vecteurs aléatoires de \mathbb{R}^d , X_1, \dots, X_n indépendants et de même loi à densité f alors, un estimateur à noyau $\widehat{f}_n(x)$ de $f(x)$ est donné, pour tout $x \in \mathbb{R}^d$, par

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{k=1}^n K_h(X_k - x), \quad (2.2.5)$$

avec $K_h(x) = \prod_{i=1}^d \frac{1}{h_i} K_i \left(\frac{x_i}{h_i} \right)$.

2.2.2 L'estimateur de Parzen-Rosenblatt non-récuratif

On commence par s'intéresser à l'estimateur de Parzen-Rosenblatt, défini, pour tout $n \geq 1$ et pour tout $x \in \mathbb{R}$, par

$$\widehat{f}_n(x) = \frac{1}{nh_n} \sum_{k=1}^n K\left(\frac{X_k - x}{h_n}\right) \quad (2.2.6)$$

MSE et MISE

Une mesure standard de l'efficacité de l'estimateur \widehat{f}_n est son erreur quadratique moyenne (en anglais Mean Square Error abrégée en MSE) définie, pour tout $x \in \mathbb{R}$, par

$$\text{MSE}(x) = \mathbb{E}\left[\left(\widehat{f}_n(x) - f(x)\right)^2\right]. \quad (2.2.7)$$

Alors que le MSE est une mesure locale à x fixé, une mesure globale de l'efficacité de l'estimateur \widehat{f}_n est son erreur quadratique moyenne intégrée (en anglais Mean Integrated Square Error abrégée en MISE) définie, par

$$\text{MISE} = \int_{-\infty}^{+\infty} \mathbb{E}\left[\left(\widehat{f}_n(x) - f(x)\right)^2\right] dx. \quad (2.2.8)$$

L'étude de ces deux quantités se fait en décomposant le MSE en la somme de deux termes

$$\text{MSE}(x) = B_n^2(x) + V_n(x) \quad (2.2.9)$$

où $B_n(x)$ est le biais de l'estimateur $\widehat{f}_n(x)$

$$B_n(x) = \mathbb{E}\left[\widehat{f}_n(x)\right] - f(x) \quad (2.2.10)$$

alors que $V_n(x)$ est sa variance

$$V_n(x) = \mathbb{E}\left[\left(\widehat{f}_n(x) - \mathbb{E}\left[\widehat{f}_n(x)\right]\right)^2\right]. \quad (2.2.11)$$

Nous avons alors le résultat suivant pour le MSE. On pourra consulter la Proposition 1.1 page 4 et la Proposition 1.2 page 6 de [65].

Théorème 2.2.1 *Soit f bornée, deux fois dérivable à dérivées bornées. De plus, supposons que*

$$\int_{-\infty}^{+\infty} K^2(x) dx = \mu^2 < +\infty \quad \text{et} \quad \frac{1}{2} \int_{-\infty}^{+\infty} x^2 K(x) dx = \nu^2 < +\infty.$$

Alors, pour tout $n \geq 1$ et pour tout $x \in \mathbb{R}$,

$$V_n(x) \leq \mu^2 \|f\|_\infty \frac{1}{nh_n} \quad \text{et} \quad B_n^2(x) \leq \nu^4 \|f''\|_\infty^2 h_n^4.$$

En particulier,

$$\text{MSE}(x) \leq \mu^2 \|f\|_\infty \frac{1}{nh_n} + \nu^4 \|f''\|_\infty^2 h_n^4 \quad (2.2.12)$$

On déduit du résultat précédent qu'il est nécessaire d'avoir $nh_n \rightarrow +\infty$ quand $n \rightarrow +\infty$. Ainsi, la fenêtre de lissage sera toujours choisie de telle sorte que

$$h_n \xrightarrow{n \rightarrow +\infty} 0 \quad \text{et} \quad nh_n \xrightarrow{n \rightarrow +\infty} +\infty. \quad (2.2.13)$$

De plus, le MISE admet un résultat équivalent, voir par exemple la Proposition 1.4 page 12 et la Proposition 1.5 page 13 de [65].

Théorème 2.2.2 *Supposons que f est deux fois dérivable et que*

$$\int_{-\infty}^{+\infty} |f''(x)|^2 dx < +\infty.$$

De plus, supposons que

$$\int_{-\infty}^{+\infty} K^2(x) dx = \mu^2 < +\infty \quad \text{et} \quad \frac{1}{2} \int_{-\infty}^{+\infty} x^2 K(x) dx = \nu^2 < +\infty.$$

Alors, pour tout $n \geq 1$ et pour tout $x \in \mathbb{R}$,

$$\int_{-\infty}^{+\infty} V_n(x) dx \leq \mu^2 \frac{1}{nh_n} \quad \text{et} \quad \int_{-\infty}^{+\infty} B_n^2(x) \leq \nu^4 \left(\int_{-\infty}^{+\infty} |f''(x)|^2 dx \right) h_n^4.$$

En particulier,

$$\text{MISE} \leq \mu^2 \frac{1}{nh_n} + \nu^4 \left(\int_{-\infty}^{+\infty} |f''(x)|^2 dx \right) h_n^4. \quad (2.2.14)$$

Remarque 2.2.2 *Concernant le choix de la fenêtre de lissage h_n , d'après les deux théorèmes précédents, il est immédiat que*

$$\sup_{x \in \mathbb{R}} \text{MSE}(x) = \mathcal{O} \left(\frac{1}{nh_n} + h_n^4 \right)$$

et

$$\text{MISE} = \mathcal{O} \left(\frac{1}{nh_n} + h_n^4 \right).$$

Le minimum par rapport à h_n de l'expression $\frac{1}{nh_n} + h_n^4$ est atteint, pour $\gamma > 0$, par

$$h_n = \gamma n^{-1/5}. \quad (2.2.15)$$

Pour ce choix optimal de fenêtre de lissage, on obtient donc

$$\sup_{x \in \mathbb{R}} \text{MSE}(x) = \mathcal{O} \left(n^{-4/5} \right)$$

et

$$\text{MISE} = \mathcal{O} \left(n^{-4/5} \right).$$

Pour conclure cette partie, il existe une autre mesure, stochastique celle-ci, de l'efficacité de l'estimateur \widehat{f}_n . Elle est appelée erreur quadratique intégrée (en anglais Integrated Square Error abrégée en ISE). Elle est définie par

$$\text{ISE} = \int_{-\infty}^{+\infty} (\widehat{f}_n(x) - f(x))^2 dx. \quad (2.2.16)$$

Peter Hall a étudié en profondeur le comportement asymptotique du ISE. Il a notamment démontré dans [34] l'équivalence asymptotique du ISE et du MISE dans le sens où

$$\frac{\text{ISE}}{\overline{\text{MISE}}}$$

tend vers 1 en probabilité quand n tend vers $+\infty$. En particulier, il a établi le développement asymptotique du ISE suivant.

$$\int_{-\infty}^{+\infty} (\widehat{f}_n(x) - f(x))^2 dx = \frac{\mu^2}{nh_n} + \nu^4 h_n^4 \int_{-\infty}^{+\infty} f''(x)^2 dx + o_{\mathbb{P}}\left(\frac{1}{nh_n} + h_n^4\right) \quad (2.2.17)$$

De plus, dans [35], il a établi un théorème central limite multivarié. Plus précisément, il a étudié le comportement asymptotique en loi de

$$d(n) (\text{ISE} - \text{MISE})$$

où $d(n)$ est une vitesse qui dépend du comportement asymptotique de nh_n^{d+4} quand n tend vers l'infini. L'idée originale de la preuve est d'utiliser le comportement asymptotique des U -statistiques.

Inconvénient de l'estimateur

Le principal problème de l'estimateur de Parzen-Rosenblatt défini par (2.2.6) est qu'il ne s'adapte pas bien au cas où les variables aléatoires X_1, \dots, X_n nous arrivent de manière séquentielle. En effet, supposons que l'on dispose des variables X_1, \dots, X_n à un certain temps, et qu'un peu plus tard on reçoit une $(n+1)$ -ème variable X_{n+1} . Pour estimer f , on aimerait utiliser cette nouvelle observation. Néanmoins, cela impliquerait de calculer l'estimateur \widehat{f}_{n+1} en réutilisant toutes les variables, ce qui n'est pas très pratique. Une façon de palier ce problème est d'utiliser une version récursive de l'estimateur (2.2.6).

2.2.3 L'estimateur de Parzen-Rosenblatt récursif

Convergence

La version récursive de l'estimateur (2.2.6), introduit par Wolverton et Wagner [69], est définie, pour tout $n \geq 1$ et pour tout $x \in \mathbb{R}$, par

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{k=1}^n \frac{1}{h_k} K\left(\frac{X_k - x}{h_k}\right). \quad (2.2.18)$$

Le premier à s'être intéressé aux propriétés de cet estimateur est Yamato [70]. Un peu plus tard, Deheuvels [19] a établi le résultat de convergence suivant concernant la convergence presque sûre et la normalité asymptotique.

Théorème 2.2.3 *Supposons que f soit dérivable à dérivée bornée. Si la fenêtre de lissage h_n est telle que $h_n = n^{-\alpha}$ avec $0 < \alpha < 1$, alors, pour tout $x \in \mathbb{R}$,*

$$\lim_{n \rightarrow +\infty} \widehat{f}_n(x) = f(x) \quad \text{p.s.}$$

De plus, pour tout $1/5 < \alpha < 1$, on a la normalité asymptotique

$$\sqrt{nh_n} (\widehat{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\mu^2 f(x)}{1 + \alpha}\right).$$

MSE, MISE et ISE

Pour l'estimateur de Parzen-Rosenblatt récursif, nous pouvons établir des résultats équivalents aux Théorèmes 2.2.1 et 2.2.2 pour le MSE et MISE, en remplaçant dans les équations (2.2.12) et (2.2.14) la suite $\frac{1}{nh_n}$ par

$$\frac{1}{n^2} \sum_{k=1}^n \frac{1}{h_k}$$

et la suite h_n^4 par

$$\left(\frac{1}{n} \sum_{k=1}^n h_k^2 \right)^2.$$

Pour cet estimateur récursif, il serait également intéressant d'établir un développement asymptotique du ISE équivalent au résultat de Hall pour l'estimateur non-récursif.

2.2.4 L'estimateur de Deheuvels

Il existe d'autres estimateurs à noyau pour estimer une densité de probabilité. On peut notamment citer l'estimateur de Deheuvels [17] qui est défini, pour tout $n \geq 1$ et pour tout $x \in \mathbb{R}$, par

$$\check{f}_n(x) = \frac{1}{S_n} \sum_{k=1}^n K\left(\frac{X_k - x}{h_k}\right).$$

où

$$S_n = \sum_{k=1}^n h_k.$$

Deheuvels [18] a montré que son estimateur $\check{f}_n(x)$ converge en moyenne quadratique vers $f(x)$ ainsi que sa convergence presque sûre et la normalité asymptotique.

Théorème 2.2.4 *Si f est dérivable à dérivée bornée et que la fenêtre h_n est choisie telle que $h_n = n^{-\alpha}$ pour $0 < \alpha < 1$, alors pour tout $x \in \mathbb{R}$,*

$$\lim_{n \rightarrow +\infty} \check{f}_n(x) = f(x) \quad \text{p.s.}$$

De plus si $1/5 < \alpha < 1$, on a

$$\sqrt{nh_n} (\check{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mu^2 f(x)).$$

2.3 Estimation d'une fonction de régression

Dans cette partie, on s'intéresse au deuxième problème mentionné au début du chapitre. On suppose qu'on dispose d'une suite (X_n, Y_n) couples aléatoires indépendants liés par la relation (2.1.1)

$$Y_n = f(X_n) + \varepsilon_n$$

où la suite de variables aléatoires (ε_n) est une suite de variables aléatoires indépendantes, centrées et de même loi, et où la fonction de régression f est inconnue. On définit alors un estimateur de f comme une somme pondérée de la suite d'observations (Y_n)

$$\widehat{f}_n(x) = \sum_{k=1}^n W_{n,k}(x) Y_k \tag{2.3.1}$$

où les poids $W_{n,k}(x)$ dépendent seulement de n , k et x et des variables aléatoires X_1, \dots, X_n et satisfont, pour tout $x \in \mathbb{R}$,

$$\sum_{k=1}^n W_{n,k}(x) = 1.$$

2.3.1 L'estimateur de Nadaraya-Watson non-récurrent

On se donne un noyau K et une fenêtre de lissage (h_n) . Comme pour l'estimation d'une densité de probabilité, on peut construire des estimateurs à noyau pour la régression non paramétrique. Le plus célèbre d'entre eux est l'estimateur \widehat{f}_n de Nadaraya-Watson [49] [68] défini, pour tout $n \geq 1$ et pour tout $x \in \mathbb{R}$, par

$$\widehat{f}_n(x) = \begin{cases} \frac{\sum_{k=1}^n K\left(\frac{X_k-x}{h_n}\right)Y_k}{\sum_{k=1}^n K\left(\frac{X_k-x}{h_n}\right)} & \text{si } \sum_{k=1}^n K\left(\frac{X_k-x}{h_n}\right) \neq 0, \\ 0 & \text{si } \sum_{k=1}^n K\left(\frac{X_k-x}{h_n}\right) = 0. \end{cases} \quad (2.3.2)$$

L'estimateur de Nadaraya-Watson correspond au choix de

$$W_{n,k}(x) = \frac{K\left(\frac{X_k-x}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{X_j-x}{h_n}\right)} \mathbb{I}_{\{\sum_{j=1}^n K\left(\frac{X_j-x}{h_n}\right) \neq 0\}}$$

dans l'équation (2.3.1). Intuitivement, si on note X une variable aléatoire de même loi que (X_n) et Y une variable aléatoire de même loi que (Y_n) et si on suppose que le couple (X,Y) admet une densité $g(x,y)$ par rapport à la mesure de Lebesgue sur \mathbb{R}^2 , alors la densité marginale de X est donnée, pour tout $x \in \mathbb{R}$, par $g(x) = \int_{\mathbb{R}} g(x,y)dy > 0$. On peut alors écrire, pour tout $x \in \mathbb{R}$ tel que $g(x) \neq 0$,

$$f(x) = \mathbb{E}[Y|X=x] = \frac{\int_{\mathbb{R}} yg(x,y)dy}{\int_{\mathbb{R}} g(x,y)dy} = \frac{\int_{\mathbb{R}} yg(x,y)dy}{g(x)}.$$

En remplaçant $g(x)$ et $g(x,y)$ par leurs estimateurs respectifs de Parzen-Rosenblatt (2.2.6) et (2.2.4), on obtient l'estimateur de Nadaraya-Watson défini par (2.3.2).

Consistance et Normalité asymptotique

Afin d'obtenir la convergence presque sûre de l'estimateur de Nadaraya-Watson, on utilise à nouveau la décomposition biais-variance évoquée lors de l'estimation d'une densité de probabilité. Pour un développement asymptotique du biais et de la variance, on pourra consulter les travaux de Collomb [14] ainsi que la thèse Blondin [6]. En ce qui concerne la normalité asymptotique de l'estimateur de Nadaraya-Watson, la première démonstration est due à Schuster [61]. On se réfère également aux théorèmes 1.3 et 1.4 p. 117-120 de Nadaraya [50] et au théorème 4.2.1 p. 99 de Härdle [36], qui reposent sur d'autres méthodes de démonstration. On donne ici le théorème pour une fenêtre de lissage $h_n = n^{-\alpha}$.

Théorème 2.3.1 *Soit $0 < \alpha < 1$ et $h_n = n^{-\alpha}$. On suppose que f est bornée, deux fois dérivable à dérivées bornées. De plus, on suppose que X admet une densité g qui est bornée, deux fois dérivable à dérivées bornées et que le noyau K vérifie*

$$\int_{-\infty}^{+\infty} K^2(x)dx = \mu^2 < +\infty \quad \text{et} \quad \frac{1}{2} \int_{-\infty}^{+\infty} x^2 K(x)dx = \nu^2 < +\infty.$$

Si le bruit est de carré intégrable et de variance $\sigma^2 > 0$, alors pour tout $x \in \mathbb{R}$,

$$\lim_{n \rightarrow +\infty} \widehat{f}_n(x) = f(x) \quad \text{p.s.}$$

De plus, si $1/5 < \alpha < 1$, alors, pour tout $x \in \mathbb{R}$ tel que $g(x) \neq 0$,

$$\sqrt{nh_n} (\widehat{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 \mu^2}{g(x)}\right).$$

2.3.2 L'estimateur de Nadaraya-Watson récursif

Comme dans le cas de l'estimateur de Parzen-Rosenblatt, si les couples (X_n, Y_n) nous arrivent séquentiellement, il est préférable d'utiliser une version récursive de l'estimateur de Nadaraya-Watson (2.3.2). Cet estimateur récursif est défini, pour tout $n \geq 1$ et pour tout $x \in \mathbb{R}$, par

$$\widehat{f}_n(x) = \begin{cases} \frac{\sum_{k=1}^n \frac{1}{h_k} K\left(\frac{X_k - x}{h_k}\right) Y_k}{\sum_{k=1}^n \frac{1}{h_k} K\left(\frac{X_k - x}{h_k}\right)} & \text{si } \sum_{k=1}^n \frac{1}{h_k} K\left(\frac{X_k - x}{h_k}\right) \neq 0, \\ 0 & \text{si } \sum_{k=1}^n K\left(\frac{X_k - x}{h_n}\right) = 0. \end{cases} \quad (2.3.3)$$

Convergence

Le résultat qui suit se trouve dans le livre de Dufflo [23].

Théorème 2.3.2 *Soit $0 < \alpha < 1$ et $h_n = n^{-\alpha}$. On suppose que f est bornée, deux fois dérivable à dérivées bornées. De plus, on suppose que X admet une densité g qui est bornée, deux fois dérivable à dérivées bornées et que le noyau vérifie*

$$\int_{-\infty}^{+\infty} K^2(x)dx = \mu^2 < +\infty \quad \text{et} \quad \int_{-\infty}^{+\infty} x^2 K(x)dx = \nu^2 < +\infty.$$

Si le bruit est de carré intégrable et de variance $\sigma^2 > 0$, alors pour tout $x \in \mathbb{R}$,

$$\lim_{n \rightarrow +\infty} \widehat{f}_n(x) = f(x) \quad \text{p.s.}$$

De plus, si $1/3 < \alpha < 1$ et si (ε_n) admet un moment d'ordre > 2 , alors, pour tout $x \in \mathbb{R}$ tel que $g(x) \neq 0$,

$$\sqrt{nh_n} (\widehat{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 \mu^2}{g(x)(1 + \alpha)}\right).$$

2.3.3 L'estimateur de Révész

Un autre estimateur récursif de la fonction de régression f est l'estimateur de Révész [55]. Il a été relativement peu utilisé malgré son très bon comportement asymptotique. On peut toutefois citer la thèse de Slaoui et la récente contribution de Mokkadem *et al.* [48] sur le sujet. Cet estimateur est défini, pour tout $x \in \mathbb{R}$ et pour tout $n \geq 1$, par

$$\check{f}_n(x) = \check{f}_{n-1}(x) + \frac{1}{nh_n} K\left(\frac{x - X_n}{h_n}\right) (Y_n - \check{f}_{n-1}(x)).$$

Révész [55] a étudié les propriétés asymptotiques de cet estimateur pour une fenêtre $h_n = n^{-\alpha}$. Le résultat est le suivant.

Théorème 2.3.3 *Pour tout $1/2 < \alpha < 1$ et pour tout $x \in \mathbb{R}$ tel que $2g(x) > 1 - \alpha$, on a*

$$\lim_{n \rightarrow +\infty} \check{f}_n(x) = f(x) \quad \text{p.s.}$$

et

$$\sqrt{nh_n} (\check{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 \mu^2 g(x)}{2g(x) - (1 - \alpha)}\right).$$

Deuxième partie

Estimation semi-paramétrique pour des modèles de déformation

Chapitre 3

Présentation des modèles de déformation

3.1 Les modèles que nous étudions

Un grand nombre de phénomènes naturels se répètent de manière périodique. On peut par exemple penser aux cycles annuels de température en météorologie, à l'astronomie avec les fameux cycles de 11 ans de l'activité solaire géomagnétique, à la médecine et le rythme cardiaque humain mesuré par les signaux électrocardiogrammes (ECG) ou encore aux signaux électroencéphalogrammes (EEG). L'analyse statistique de données périodiques est donc très importante pour élaborer des modèles satisfaisants de telles situations. Les modèles de déformation de forme commune périodique (Shape Invariant Models en anglais, abrégé en SIM) introduits par Lawton, Sylvestre et Maggio [42] sont parmi les modèles les mieux adaptés pour modéliser ce type de phénomènes. Ces modèles sont définis, pour tout $n \geq 0$, par

$$Y_n = h(X_n) + \varepsilon_n \quad (3.1.1)$$

où les variables (X_n) sont des temps d'observations connus et les variables (Y_n) sont les observations alors que les variables (ε_n) sont des erreurs aléatoires inconnues. La fonction h est périodique et s'écrit sous la forme

$$h(x) = m + \sum_{k=1}^p a_k f(x - \theta_k)$$

où f est la forme commune inconnue du modèle, m est la moyenne globale inconnue, tandis que $\theta = (\theta_1, \dots, \theta_p)$ et $a = (a_1, \dots, a_p)$ sont des paramètres de translation et d'échelle inconnus.

Dans cette partie, nous allons tout d'abord étudier le modèle où $p = 1$, $m = 0$ et $a = 1$ défini, pour tout $n \geq 0$, par

$$Y_n = f(X_n - \theta) + \varepsilon_n \quad (3.1.2)$$

où (X_n) et (ε_n) sont deux suites indépendantes de variables aléatoires indépendantes et de même loi. Puis, nous étudierons le modèle plus général, défini, pour tout $1 \leq i \leq n$ et pour tout $1 \leq j \leq p$, par

$$Y_{i,j} = a_j f(X_i - \theta_j) + v_j + \varepsilon_{i,j}. \quad (3.1.3)$$

Enfin, nous nous intéresserons à un modèle de déformation de variables aléatoires défini, pour tout $n \geq 0$, par

$$X_n = \varphi_\theta(\varepsilon_n). \quad (3.1.4)$$

Le premier modèle (3.1.2) a notamment été étudié par Tsybakov *et al.* [16] pour un bruit gaussien. Gamboa *et al.* [30] puis Vimond [66] se sont quant à eux intéressés au second modèle (3.1.3). Ici, on suppose que tous les paramètres des modèles sont déterministes. Le cas de paramètres aléatoires a été étudié par Trigano *et al.* [63] ou encore par Castillo et Loubes [10]. Le troisième modèle (3.1.4) de déformation de variables aléatoires a été récemment étudié dans [46]. Alors que l'estimation dans ces modèles est très souvent faite en calculant la vraisemblance du modèle quand le bruit est gaussien, ou bien en faisant des approximations de type Fourier pour les fonctions de lien, l'objectif de cette thèse est de proposer des algorithmes stochastiques récurrents pour l'estimation des paramètres qui ne nécessitent pas l'évaluation de la fonction f et qui sont très faciles à implémenter. De plus, on s'intéresse dans un second temps à l'estimation de la fonction de lien f .

Une première chose cruciale à étudier avant d'effectuer l'estimation des paramètres dans ces modèles est l'étude des contraintes d'identifiabilité. Dans les papiers [66] et [67], il semblerait que les conditions d'identifiabilité proposées soient insuffisantes, comme nous le montrerons à la fin de ce chapitre. Il est donc nécessaire de les établir de manière précise. C'est ce qu'on se propose de faire dans la suite.

3.2 Conditions d'identifiabilité

Dans cette partie, nous proposons de détailler les conditions d'identifiabilité pour les modèles (3.1.2) et (3.1.3). Pour les deux modèles, on considère une fonction f de période 1, c'est-à-dire qui vérifie pour tout $x \in \mathbb{R}$, $f(x+1) = f(x)$.

3.2.1 Modèle $Y_n = f(X_n - \theta) + \varepsilon_n$

Pour obtenir les conditions d'identifiabilité du modèle (3.1.2), on suppose qu'il existe f^* de période 1 et $\theta^* \in \mathbb{R}$ tels que, pour tout $x \in \mathbb{R}$,

$$f(x - \theta) = f^*(x - \theta^*). \quad (3.2.1)$$

On détaille la détermination de ces contraintes d'identifiabilité dans deux cas : f est paire et f est quelconque.

1. On suppose ici que f est paire c'est-à-dire qu'elle vérifie pour tout $x \in \mathbb{R}$, $f(x) = f(-x)$. En utilisant la parité de f et de f^* , on déduit alors de (3.2.1) que, pour tout $x \in \mathbb{R}$,

$$\begin{aligned} f(x) &= f^*(x + \theta - \theta^*) \\ &= f^*(-x - \theta + \theta^*) \\ &= f(-x - 2\theta + 2\theta^*) \\ &= f(x + 2(\theta - \theta^*)). \end{aligned}$$

La périodicité de f entraîne donc que

$$2(\theta - \theta^*) \in \mathbb{Z}.$$

Une première condition pour l'identifiabilité du modèle (3.1.2) est donc $|\theta - \theta^*| < 1/2$. Si on prend $|\theta| < 1/4$, on a donc, dans ce cas, $\theta = \theta^*$, et immédiatement, pour tout $x \in \mathbb{R}$, $f(x) = f^*(x)$. Un jeu de contraintes d'identifiabilité naturelles pour le modèle (3.1.2) est donc

$$C_0 = \{f \text{ paire, de période minimale 1 et } |\theta| < 1/4\}. \quad (3.2.2)$$

2. On suppose ici que f est quelconque et on note f_1 et g_1 les deux premiers coefficients de Fourier de la fonction f

$$f_1 = \int_{-1/2}^{1/2} f(x) \cos(2\pi x) dx \quad \text{et} \quad g_1 = \int_{-1/2}^{1/2} f(x) \sin(2\pi x) dx.$$

On déduit alors de (3.2.1) que

$$\begin{aligned} \int_{-1/2}^{1/2} f(x - \theta) \cos(2\pi x) dx &= \int_{-1/2}^{1/2} f^*(x - \theta^*) \cos(2\pi x) dx \\ &\Downarrow \\ \int_{-1/2-\theta}^{1/2-\theta} f(x) \cos(2\pi(x + \theta)) dx &= \int_{-1/2-\theta^*}^{1/2-\theta^*} f^*(x) \cos(2\pi(x + \theta^*)) dx. \end{aligned}$$

Puis, en utilisant l'égalité trigonométrique

$$\cos(2\pi(x + \theta)) = \cos(2\pi x) \cos(2\pi\theta) - \sin(2\pi x) \sin(2\pi\theta)$$

et la 1-périodicité de f , on obtient donc que

$$\cos(2\pi\theta) f_1 - \sin(2\pi\theta) g_1 = \cos(2\pi\theta^*) f_1^* - \sin(2\pi\theta^*) g_1^*. \quad (3.2.3)$$

De même,

$$\begin{aligned} \int_{-1/2}^{1/2} f(x - \theta) \sin(2\pi x) dx &= \int_{-1/2}^{1/2} f^*(x - \theta^*) \sin(2\pi x) dx \\ &\Downarrow \\ \cos(2\pi\theta) g_1 + \sin(2\pi\theta) f_1 &= \cos(2\pi\theta^*) g_1^* + \sin(2\pi\theta^*) f_1^*. \end{aligned} \quad (3.2.4)$$

On aboutit donc au système

$$\begin{cases} \cos(2\pi\theta) f_1 - \sin(2\pi\theta) g_1 = \cos(2\pi\theta^*) f_1^* - \sin(2\pi\theta^*) g_1^*, \\ \cos(2\pi\theta) g_1 + \sin(2\pi\theta) f_1 = \cos(2\pi\theta^*) g_1^* + \sin(2\pi\theta^*) f_1^*. \end{cases} \quad (3.2.5)$$

Si on impose de connaître f_1 et g_1 , alors (3.2.5) devient

$$\begin{cases} \cos(2\pi\theta) f_1 - \sin(2\pi\theta) g_1 = \cos(2\pi\theta^*) f_1 - \sin(2\pi\theta^*) g_1, \\ \cos(2\pi\theta) g_1 + \sin(2\pi\theta) f_1 = \cos(2\pi\theta^*) g_1 + \sin(2\pi\theta^*) f_1. \end{cases} \quad (3.2.6)$$

Dans ce cas, il s'ensuit que

$$\begin{cases} \cos(2\pi\theta) f_1 g_1 - \sin(2\pi\theta) g_1^2 = \cos(2\pi\theta^*) f_1 g_1 - \sin(2\pi\theta^*) g_1^2, \\ \cos(2\pi\theta) f_1 g_1 + \sin(2\pi\theta) f_1^2 = \cos(2\pi\theta^*) f_1 g_1 + \sin(2\pi\theta^*) f_1^2. \end{cases} \quad (3.2.7)$$

ou

$$\begin{cases} \cos(2\pi\theta)f_1^2 - \sin(2\pi\theta)f_1g_1 = \cos(2\pi\theta^*)f_1^2 - \sin(2\pi\theta^*)f_1g_1, \\ \cos(2\pi\theta)g_1^2 + \sin(2\pi\theta)f_1g_1 = \cos(2\pi\theta^*)g_1^2 + \sin(2\pi\theta^*)f_1g_1. \end{cases} \quad (3.2.8)$$

On suppose que soit f_1 , soit g_1 est non nul. Le système (3.2.7) implique que

$$\sin(2\pi\theta) = \sin(2\pi\theta^*)$$

et le système (3.2.8) entraîne que

$$\cos(2\pi\theta) = \cos(2\pi\theta^*).$$

Ainsi, si $|\theta| < 1/2$, on obtient que $\theta = \theta^*$ et donc pour tout $x \in \mathbb{R}$, $f(x) = f^*(x)$. De plus, si f_1 et g_1 sont tous les deux nuls alors on considère le premier f_p ou g_p non nul où f_p ou g_p sont définis, pour $p \geq 1$, par

$$f_p = \int_{-1/2}^{1/2} f(x) \cos(2\pi px) dx \quad \text{et} \quad g_p = \int_{-1/2}^{1/2} f(x) \sin(2\pi px) dx.$$

De la même manière, on aboutit aux égalités $\cos(2p\pi\theta) = \cos(2p\pi\theta^*)$ et $\sin(2p\pi\theta) = \sin(2p\pi\theta^*)$. Ainsi, si $|\theta| < 1/2p$, il s'ensuit que $\theta = \theta^*$. Un jeu de contraintes d'identifiabilité naturelles pour le modèle (3.1.2) est donc

$$\mathcal{C}_1 = \{f \text{ de période minimale } 1 \text{ et il existe } p \geq 1 \text{ tel que } f_p \text{ et } g_p \text{ sont connus, } \min(|f_p|, |g_p|) > 0 \text{ et } |\theta| < 1/2p\}. \quad (3.2.9)$$

3.2.2 Modèle $Y_{i,j} = a_j f(X_i - \theta_j) + v_j + \varepsilon_{i,j}$

Pour le modèle plus général, on pose comme jeu de contraintes d'identifiabilité

$$\mathcal{C}_2 = \{f \text{ de période minimale } 1 \text{ et de carré intégrable et } \int_{-1/2}^{1/2} f(x) dx = 0, \theta_1 = 0, a_1 = 1 \text{ et } \max_{1 \leq j \leq p} |\theta_j| < 1/4\}. \quad (3.2.10)$$

Nous allons montrer que sous les conditions \mathcal{C}_2 , le modèle (3.1.3) est identifiable. Pour cela, on suppose qu'il existe f^* de période 1 et un triplet (θ^*, a^*, v^*) appartenant à \mathcal{C}_2 tels que, pour tout $1 \leq j \leq p$ et pour tout $x \in \mathbb{R}$,

$$a_j f(x - \theta_j) + v_j = a_j^* f^*(x - \theta_j^*) + v_j^*. \quad (3.2.11)$$

En passant à l'intégrale, comme $\int_{-1/2}^{1/2} f(x) dx = 0$ et $\int_{-1/2}^{1/2} f^*(x) dx = 0$, on obtient, pour tout $1 \leq j \leq p$,

$$v_j = v_j^*.$$

La relation (3.2.11) devient donc pour tout $1 \leq j \leq p$ et pour tout $x \in \mathbb{R}$,

$$a_j f(x - \theta_j) = a_j^* f^*(x - \theta_j^*). \quad (3.2.12)$$

De plus, en $j = 1$, $\theta_1 = 0$, $a_1 = 1$ et $\theta_1^* = 0$, $a_1^* = 1$, donc pour tout $x \in \mathbb{R}$,

$$f(x) = f^*(x).$$

Il en découle que pour tout $1 \leq j \leq p$ et pour tout $x \in \mathbb{R}$,

$$a_j f(x - \theta_j) = a_j^* f(x - \theta_j^*). \quad (3.2.13)$$

En passant au carré, en intégrant et en utilisant la périodicité de f et f^* , il s'ensuit que

$$a_j^2 I_2 = (a_j^*)^2 I_2, \quad (3.2.14)$$

où $I_2 = \int_{-1/2}^{1/2} f^2(x) dx$. La fonction f n'étant pas nulle, on a $I_2 \neq 0$ et (3.2.14) devient, pour tout $1 \leq j \leq p$,

$$a_j^2 = (a_j^*)^2. \quad (3.2.15)$$

Deux cas sont alors possibles :

1. $a_j = a_j^*$, dans ce cas, de (3.2.13), on a $f(x - \theta_j) = f(x - \theta_j^*)$. Ainsi, $f(x) = f(x + \theta_j - \theta_j^*)$. Donc $\theta_j - \theta_j^* \in \mathbb{Z}$. Par hypothèse, $|\theta_j - \theta_j^*| < 1/2$ donc nécessairement $\theta_j - \theta_j^* = 0$ c'est-à-dire $\theta_j = \theta_j^*$.
2. $a_j = -a_j^* \neq 0$, dans ce cas, de (3.2.13), on a $f(x - \theta_j) + f(x - \theta_j^*) = 0$. Ainsi, $f(x) = -f(x + \theta_j - \theta_j^*) = f(x + 2(\theta_j - \theta_j^*))$. Donc $2(\theta_j - \theta_j^*) \in \mathbb{Z}$. Par hypothèse, $|\theta_j - \theta_j^*| < 1/2$ donc nécessairement $\theta_j - \theta_j^* = 0$. Ainsi, $\theta_j = \theta_j^*$. D'où $f(x) + f(x) = 0$, c'est-à-dire $f(x) = 0$, ce qui est impossible.

3.2.3 Deux contre exemples

Dans cette partie, nous étudions les conditions d'identifiabilité proposées par Vimond [66] et Wang et Brown [67]. Nous montrons que les conditions proposées ne suffisent pas à garantir l'identifiabilité du modèle $Y_{i,j} = a_j f(X_i - \theta_j) + v_j + \varepsilon_{i,j}$.

L'article de Vimond

Dans son papier [66], Vimond s'intéresse au modèle défini, pour tout $1 \leq i \leq n$ et pour tout $1 \leq j \leq p$, par

$$Y_{i,j} = a_j f(t_i - \theta_j) + v_j + \varepsilon_{i,j}$$

avec les temps d'observation

$$t_i = \frac{2\pi(i-1)}{n}.$$

Nous prenons les notations utilisées dans l'article [66]. La fonction f est supposée de période 2π et continue tandis que la suite $(\varepsilon_{i,j})$ est une suite de variables aléatoires indépendantes et identiquement distribuées selon la loi normale centrée réduite. De plus, on suppose que

$$(\theta, a, v) \in \mathcal{A} = [0, 2\pi]^p \times \mathbb{R}^p \times [-v_{\max}, v_{\max}]^p.$$

Les contraintes d'identifiabilité proposées par Vimond sont

$$\int_0^{2\pi} f(x) dx = 0 \quad \text{et} \quad (\theta, a, v) \in \mathcal{A}_0,$$

où

$$\mathcal{A}_0 = \{(\theta, a, v) \in \mathcal{A}, \theta_1 = 0, \sum_{j=1}^p a_j^2 = p, a_1 > 0\}.$$

Ces conditions d'identifiabilité ne sont alors pas suffisantes. En effet, supposons qu'il existe une fonction f^* de période 2π et continue telle que $\int_0^{2\pi} f^*(x)dx = 0$ et $(\theta^*, a^*, v^*) \in \mathcal{A}_0$ tels que, pour tout $1 \leq j \leq p$ et pour tout $x \in \mathbb{R}$,

$$a_j f(x - \theta_j) + v_j = a_j^* f^*(x - \theta_j^*) + v_j^*. \quad (3.2.16)$$

En passant à l'intégrale dans l'égalité (3.2.16), puisque $\int_0^{2\pi} f(x)dx = 0$ et $\int_0^{2\pi} f^*(x)dx = 0$, on obtient immédiatement que, pour tout $1 \leq j \leq p$,

$$v_j = v_j^*.$$

Ainsi, il découle de l'équation (3.2.16) que

$$a_j f(x - \theta_j) = a_j^* f^*(x - \theta_j^*). \quad (3.2.17)$$

De plus, en $j = 1$, $\theta_1 = 0$, $\theta_1^* = 0$, $a_1 > 0$ et $a_1^* > 0$, donc pour tout $x \in \mathbb{R}$,

$$f^*(x) = \frac{a_1}{a_1^*} f(x). \quad (3.2.18)$$

On déduit alors de (3.2.17) que pour tout $x \in \mathbb{R}$,

$$a_j f(x - \theta_j) = a_j^* \frac{a_1}{a_1^*} f(x - \theta_j^*). \quad (3.2.19)$$

En passant au carré puis à l'intégrale dans (3.2.19), il s'ensuit que

$$a_j^2 I_2 = \left(a_j^* \frac{a_1}{a_1^*} \right)^2 I_2, \quad (3.2.20)$$

où $I_2 = \int_0^{2\pi} f^2(x)dx$. Puisque f n'est pas nulle, on a $I_2 \neq 0$ et donc on déduit de (3.2.20) que

$$a_j^2 (a_1^*)^2 = a_1^2 (a_j^*)^2. \quad (3.2.21)$$

Ainsi, en sommant la relation précédente de 1 à p , la condition $\sum_{j=1}^p a_j^2 = p$ entraîne immédiatement que

$$(a_1^*)^2 p = a_1^2 p, \quad (3.2.22)$$

d'où,

$$(a_1^*)^2 = a_1^2. \quad (3.2.23)$$

Comme $a_1 > 0$ et $a_1^* > 0$, on a donc $a_1 = a_1^*$. L'équation (3.2.21) mène alors à l'égalité, pour tout $1 \leq j \leq p$,

$$(a_j^*)^2 = a_j^2. \quad (3.2.24)$$

Deux cas sont alors possibles :

1. $a_j = a_j^*$. Dans ce cas, on déduit de (3.2.19) que, pour tout $x \in \mathbb{R}$,

$$f(x - \theta_j) = f(x - \theta_j^*). \quad (3.2.25)$$

On peut étudier par exemple la fonction

$$f(x) = \cos(x).$$

Cette fonction est d'intégrale nulle. Si on pose $\theta_j = 0$ et $\theta_j^* = 2\pi$, alors l'égalité (3.2.25) est vraie mais $\theta_j \neq \theta_j^*$. Le modèle n'est donc pas identifiable.

2. $a_j = -a_j^* \neq 0$. Dans ce cas, on déduit de (3.2.19) que pour tout $x \in \mathbb{R}$,

$$f(x - \theta_j) + f(x - \theta_j^*) = 0. \quad (3.2.26)$$

Pour la même fonction f que précédemment, si on pose, pour tout $1 \leq j \leq p$,

$$\theta_j = \frac{\pi}{2} \quad \text{et} \quad \theta_j^* = \frac{3\pi}{2},$$

alors l'égalité (3.2.26) est vraie mais $\theta_j \neq \theta_j^*$. Le modèle n'est donc à nouveau pas identifiable.

Afin d'assurer l'identifiabilité du modèle, on peut supposer en plus des conditions précédentes que, d'une part, pour tout $1 \leq j \leq p$, $a_j \geq 0$ et $\theta_j \in [0, 2\pi[$, ou, d'autre part, que pour tout $1 \leq j \leq p$, $\theta_j \in [0, \pi[$. Dans notre cas, le jeu de contraintes d'identifiabilité semblable à celui de Vimond [66] serait donc

$$\begin{aligned} \mathcal{C}_3 = \{ & f \text{ de période minimale 1 et de carré intégrable et} \\ & \theta_1 = 0, v_1 = 0, a_1 > 0 \text{ et } \sum_{j=1}^p a_j^2 = p, \max_{1 \leq j \leq p} |\theta_j| < 1/4 \}. \end{aligned} \quad (3.2.27)$$

Remarque 3.2.1 On peut également remarquer que l'égalité (3.2.26) n'a jamais lieu si on suppose qu'il n'existe pas de $\tau \in [0, 2\pi]$ telle que, pour tout $x \in \mathbb{R}$,

$$f(x + \tau) = -f(x).$$

Ainsi, dans notre cadre, un autre jeu de conditions d'identifiabilité est celui donné par

$$\begin{aligned} \mathcal{C}'_3 = \{ & f \text{ de période minimale 1 et de carré intégrable,} \\ & \text{et } \nexists \tau \text{ tel que } |\tau| \leq 1 \text{ et pour tout } x \in \mathbb{R}, f(x + \tau) = -f(x), \\ & \theta_1 = 0, v_1 = 0, a_1 > 0 \text{ et } \sum_{j=1}^p a_j^2 = p, \max_{1 \leq j \leq p} |\theta_j| < 1/2 \}. \end{aligned} \quad (3.2.28)$$

En particulier, considérer ce jeu d'hypothèses permet de disposer de l'intervalle le plus large possible pour θ_j . De ce fait, c'est un jeu d'hypothèses très pratique pour les applications numériques.

L'article de Wang et Brown

Dans leur papier [67], Wang et Brown s'intéressent aussi au modèle défini, pour tout $1 \leq i \leq n$ et pour tout $1 \leq j \leq p$, par

$$Y_{i,j} = a_j f(t_i - \theta_j) + v_j + \varepsilon_{i,j},$$

avec les temps d'observation

$$t_i = \frac{i-1}{n}.$$

La fonction f est supposée de période 1 et telle que $\sup_{x \in [0,1]} |f(x)| = 1$. De plus, f est supposée deux fois dérivable, absolument continue et f' est absolument continue. La suite

$\varepsilon_{i,j}$ est un bruit blanc gaussien. Les contraintes d'identifiabilité proposées par Wang et Brown sont données par

$$\int_0^1 f(x)dx = 0, \quad \int_0^1 (f''(x))^2 dx < +\infty, \quad f(0) = f(1) \quad \text{et} \quad f'(0) = f'(1).$$

De plus, il est supposé que pour tout $1 \leq j \leq p$, $a_j > 0$ et $0 \leq \theta_j < 1$. Ces conditions d'identifiabilité ne sont alors pas suffisantes. En effet, supposons qu'il existe f^* et (θ^*, a^*, v^*) vérifiant les conditions précédentes et tels que pour tout $1 \leq j \leq p$ et pour tout $x \in \mathbb{R}$,

$$a_j f(x - \theta_j) + v_j = a_j^* f^*(x - \theta_j^*) + v_j^*. \quad (3.2.29)$$

Comme précédemment, en passant à l'intégrale on obtient immédiatement que, pour tout $1 \leq j \leq p$,

$$v_j = v_j^*.$$

Ainsi, on est ramené à l'équation, pour tout $x \in \mathbb{R}$,

$$a_j f(x - \theta_j) = a_j^* f^*(x - \theta_j^*). \quad (3.2.30)$$

Comme $a_j > 0$ et $a_j^* > 0$, on en déduit que

$$a_j |f(x - \theta_j)| = a_j^* |f^*(x - \theta_j^*)|. \quad (3.2.31)$$

Or, par hypothèse, f et f^* sont 1-périodiques et vérifient

$$\sup_{x \in [0,1]} |f(x)| = 1 \quad \text{et} \quad \sup_{x \in [0,1]} |f^*(x)| = 1.$$

En passant à la borne sup dans l'égalité (3.2.31), on en déduit immédiatement que, pour tout $1 \leq j \leq p$,

$$a_j = a_j^*.$$

Il s'ensuit alors de (3.2.30) que, pour tout $x \in \mathbb{R}$,

$$f(x - \theta_j) = f^*(x - \theta_j^*). \quad (3.2.32)$$

On peut étudier par exemple les fonctions f et f^* définies, pour tout $x \in \mathbb{R}$, par

$$f(x) = \cos(2\pi(x + \theta))$$

et

$$f^*(x) = \cos(2\pi(x + \theta^*)),$$

avec $0 \leq \theta, \theta^* < 1$ et $\theta \neq \theta^*$. Les fonctions f et f^* vérifient bien les conditions requises. De plus, si pour tout $1 \leq j \leq p$, $\theta_j^* = \theta^*$ et $\theta_j = \theta$, alors l'égalité (3.2.32) est vraie mais $f \neq f^*$ et $\theta \neq \theta^*$. Le modèle n'est donc pas identifiable.

Afin d'assurer l'identifiabilité du modèle, on peut supposer que $\theta_1 = 0$. Dans ce cas, en prenant $j = 1$ dans (3.2.32), on a alors, pour tout $x \in \mathbb{R}$, $f(x) = f^*(x)$. Ainsi, pour tout $2 \leq j \leq p$, $f(x) = f(x + \theta_j - \theta_j^*)$. Cette égalité implique donc que

$$(\theta_j - \theta_j^*) \in \mathbb{Z}.$$

Et puisque $|\theta_j - \theta_j^*| < 1$, on en déduit alors que $\theta_j = \theta_j^*$ et le modèle est identifiable. Un jeu de contraintes d'identifiabilité adapté est donc

$$\mathcal{C}_4 = \{f \text{ de période minimale } 1, \text{ d'intégrale nulle et} \quad (3.2.33)$$

$$\sup_{x \in [0;1]} |f(x)| = 1, \theta_1 = 0 \text{ et } \min_{1 \leq j \leq p} a_j > 0, \max_{1 \leq j \leq p} |\theta_j| < 1/2\}.$$

Un autre jeu de contraintes possible pour ce type de modèle est aussi

$$\mathcal{C}_5 = \{f \text{ de période minimale } 1 \text{ et de carré intégrable et} \quad (3.2.34)$$

$$\theta_1 = 0, v_1 = 0, a_1 = 1 \text{ et } \min_{1 \leq j \leq p} a_j > 0, \max_{1 \leq j \leq p} |\theta_j| < 1/2\}.$$

Chapitre 4

Le modèle de translation

Ce chapitre est consacré à l'étude du modèle de régression semi-paramétrique (3.1.2) défini, pour tout $n \geq 0$, par

$$Y_n = f(X_n - \theta) + \varepsilon_n \quad (4.0.1)$$

où (X_n) et (ε_n) sont deux suites indépendantes de variables aléatoires indépendantes et de même loi. L'objectif de ce travail est d'estimer le paramètre de translation θ et dans un second temps la fonction de lien f .

4.1 Procédure d'estimation

On fait les hypothèses classiques sur ce type de modèle, voir par exemple [10] ou [16].

- (\mathcal{H}_1) Les temps d'observations (X_n) sont indépendants et de même loi qui admet une densité de probabilité g , ne s'annulant pas sur son support $[-1/2, 1/2]$. De plus, g est continue, deux fois dérivables, à dérivées bornées.
- (\mathcal{H}_2) La fonction de lien f est symétrique, bornée, périodique de période 1.

Comme nous l'avons vu dans le chapitre précédent, l'identifiabilité du modèle (4.0.1) impose de choisir $|\theta| < 1/4$. Pour l'estimation de θ , on met en place un algorithme stochastique de type Robbins-Monro, algorithme qui ne nécessite pas l'évaluation de la fonction de régression f . Ensuite, on utilise l'estimateur préalable de θ pour construire un estimateur récursif à noyau de type Nadaraya-Watson de la fonction f .

4.1.1 Estimation de θ

Afin de construire un algorithme de Robbins-Monro pour l'estimation de θ , on a besoin de trouver une fonction ϕ qui va s'annuler en θ . Plus précisément, soit X une variable aléatoire de même loi que la suite (X_n) , on considère la fonction ϕ définie, pour tout $t \in \mathbb{R}$, par

$$\phi(t) = \mathbb{E}\left[\frac{\sin(2\pi(X-t))}{g(X)} f(X-\theta)\right]. \quad (4.1.1)$$

La périodicité de f entraîne alors que

$$\begin{aligned}\phi(t) &= \int_{-1/2}^{1/2} \sin(2\pi(x-t))f(x-\theta) dx = \int_{-1/2}^{1/2} \sin(2\pi(y+\theta-t))f(y) dy, \\ &= \sin(2\pi(\theta-t)) \int_{-1/2}^{1/2} \cos(2\pi y)f(y) dy + \cos(2\pi(\theta-t)) \int_{-1/2}^{1/2} \sin(2\pi y)f(y) dy.\end{aligned}$$

Par conséquent, la symétrie de f entraîne que

$$\phi(t) = f_1 \sin(2\pi(\theta-t)) \quad (4.1.2)$$

où f_1 est le premier coefficient de Fourier de f

$$f_1 = \int_{-1/2}^{1/2} \cos(2\pi x)f(x) dx.$$

On suppose $f_1 \neq 0$. La fonction ϕ est une fonction continue et bornée et vérifie $\phi(\theta) = 0$. De plus, on peut vérifier facilement que, pour tout $t \in \mathbb{R}$ tel que $|t - \theta| < 1/2$, en particulier si $|t| \leq 1/4$, le produit $(t - \theta)\phi(t)$ est de signe constant. Il est de signe négatif si $f_1 > 0$, et de signe positif si $f_1 < 0$. Ainsi, on peut mettre en place un algorithme de Robbins-Monro pour l'estimation de θ . Soit $K = [-1/4, 1/4]$ et π_K la projection sur le compact K définie, pour tout $x \in \mathbb{R}$, par

$$\pi_K(x) = \begin{cases} x & \text{si } |x| \leq 1/4, \\ 1/4 & \text{si } x \geq 1/4, \\ -1/4 & \text{si } x \leq -1/4. \end{cases}$$

Soit (γ_n) une suite positive déterministe, décroissante vers 0 telle que

$$\sum_{n=1}^{\infty} \gamma_n = +\infty \quad \text{et} \quad \sum_{n=1}^{\infty} \gamma_n^2 < +\infty. \quad (4.1.3)$$

Dans toute la suite, on choisira pour simplifier $\gamma_n = 1/n$. On propose d'estimer θ via l'algorithme de Robbins-Monro projeté

$$\widehat{\theta}_{n+1} = \pi_K(\widehat{\theta}_n + \text{sign}(f_1)\gamma_{n+1}T_{n+1}) \quad (4.1.4)$$

où la valeur initiale $\widehat{\theta}_0 \in K$ et la variable aléatoire T_{n+1} est définie par

$$T_{n+1} = \frac{\sin(2\pi(X_{n+1} - \widehat{\theta}_n))}{g(X_{n+1})} Y_{n+1}. \quad (4.1.5)$$

Remarque 4.1.1 Si $f_1 = 0$, alors on considère le premier entier $p \geq 1$, tel que $f_p \neq 0$ où f_p est défini par

$$f_p = \int_{-1/2}^{1/2} \cos(2\pi p x)f(x) dx.$$

La fonction ϕ définie par (4.1.2) est alors égale à

$$\phi(t) = f_p \sin(2\pi p(\theta - t)). \quad (4.1.6)$$

et il faut supposer que $|\theta| < 1/4p$ et choisir $K = [-1/4p, 1/4p]$.

Remarque 4.1.2 *A la vue de (4.1.4), l'algorithme nécessite la connaissance du signe du coefficient de Fourier f_1 de f . Néanmoins, il est possible de s'en affranchir. Pour ce faire, on considère les deux algorithmes définis, pour tout $n \geq 1$, par*

$$\widehat{\theta}_{n+1}^+ = \pi_K \left(\widehat{\theta}_n^+ + \gamma_{n+1} T_{n+1}^+ \right).$$

et

$$\widehat{\theta}_{n+1}^- = \pi_K \left(\widehat{\theta}_n^- - \gamma_{n+1} T_{n+1}^- \right).$$

où les suites (T_{n+1}^+) et (T_{n+1}^-) sont telles que

$$T_{n+1}^+ = \frac{\sin(2\pi(X_{n+1} - \widehat{\theta}_n^+))}{g(X_{n+1})} Y_{n+1}$$

et

$$T_{n+1}^- = \frac{\sin(2\pi(X_{n+1} - \widehat{\theta}_n^-))}{g(X_{n+1})} Y_{n+1}.$$

Il est alors possible de montrer que la seule valeur d'adhérence de $(\widehat{\theta}_n^+)$ et de $(\widehat{\theta}_n^-)$ située à l'intérieur de K est θ , les autres valeurs d'adhérence étant localisées sur les bords du compact K , soit $\{-1/4; 1/4\}$. Donc, $\widehat{\theta}_n^-$ ou $\widehat{\theta}_n^+$ converge nécessairement presque sûrement vers $-1/4$ ou $1/4$. Ainsi, $\widehat{\theta}_n^+$ ou $\widehat{\theta}_n^-$ va converger presque sûrement vers θ . Pour avoir plus de détails, on renvoie au Lemme 3 page 403 de l'article de Brandière et Duflot [9].

L'avantage de l'algorithme (4.1.4) est qu'il ne nécessite pas de connaissance préalable sur la fonction f . En particulier, f n'a besoin d'aucune hypothèse de régularité. De plus, il est très facile à mettre en place. *A contrario*, l'inconvénient de l'algorithme est qu'il nécessite la connaissance de la densité de probabilité g associée à (X_n) . Néanmoins, ce n'est pas très gênant puisqu'il est souvent fait l'hypothèse que la suite (X_n) est une suite de variables aléatoires indépendantes et de même loi uniforme sur $[-1/2; 1/2]$ et ainsi la fonction de densité g est alors constante, égale à 1 sur son support.

4.1.2 Estimation de f

Ensuite, nous nous intéressons à l'estimation de la fonction de lien f . Pour ce faire, nous avons besoin d'ajouter une hypothèse de régularité un peu plus forte sur la fonction f qui est la suivante.

(\mathcal{H}_3) La fonction de régression f est une fonction lipschitzienne.

La fonction f étant symétrique, on propose d'estimer f par l'estimateur symétrique récursif de Nadaraya-Watson défini, pour tout $n \geq 1$ et pour tout $|x| \leq 1/2$, par

$$\widehat{f}_n(x) = \frac{\sum_{k=1}^n (W_k(x) + W_k(-x)) Y_k}{\sum_{k=1}^n (W_k(x) + W_k(-x))} \quad (4.1.7)$$

où

$$W_n(x) = \frac{1}{h_n} K \left(\frac{X_n - \widehat{\theta}_{n-1} - x}{h_n} \right).$$

La fenêtre de lissage (h_n) est choisie égale à $h_n = 1/n^\alpha$ avec $\alpha \in]0; 1[$. De plus, on suppose que le noyau K est une fonction positive symétrique, bornée à support compact, deux fois dérivable à dérivées bornées et satisfaisant

$$\int_{\mathbb{R}} K(x) dx = 1 \quad \text{et} \quad \int_{\mathbb{R}} K^2(x) dx = \nu^2.$$

4.2 Résultats de convergence paramétrique

Théorème 4.2.1 *Supposons (\mathcal{H}_1) et (\mathcal{H}_2) et que $|\theta| < 1/4$. Alors, $\widehat{\theta}_n$ converge presque sûrement vers θ . De plus, le nombre de fois que la variable aléatoire $\widehat{\theta}_n + \text{sign}(f_1)\gamma_{n+1}T_{n+1}$ sort de K est fini presque sûrement.*

Afin d'établir la normalité asymptotique de la suite $(\widehat{\theta}_n)$, on a besoin d'introduire une nouvelle fonction auxiliaire φ définie, pour tout $t \in \mathbb{R}$, par

$$\begin{aligned} \varphi(t) &= \mathbb{E}\left[\frac{\sin^2(2\pi(X-t))}{g^2(X)}(f^2(X-\theta) + \sigma^2)\right], \\ &= \int_{-1/2}^{1/2} \frac{\sin^2(2\pi(x-t))}{g(x)}(f^2(x-\theta) + \sigma^2) dx. \end{aligned} \quad (4.2.1)$$

Si $4\pi|f_1| > 1$, on note

$$\xi^2(\theta) = \frac{\varphi(\theta)}{4\pi|f_1| - 1}. \quad (4.2.2)$$

Théorème 4.2.2 *Supposons (\mathcal{H}_1) et (\mathcal{H}_2) et que $|\theta| < 1/4$. Si (ε_n) a un moment d'ordre > 2 fini et si $4\pi|f_1| > 1$, alors on a la normalité asymptotique*

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \xi^2(\theta)). \quad (4.2.3)$$

Remarque 4.2.1 *Il est facile de voir que $\phi'(t) = -2\pi f_1 \cos(2\pi(\theta - t))$. En particulier, $\phi'(\theta) = -2\pi f_1$. D'une part, si le premier coefficient de Fourier f_1 de f est connu, il est possible de modifier l'algorithme (4.1.4) pour obtenir un estimateur $\widehat{\theta}_n$ asymptotiquement efficace. Plus précisément, si l'on remplace le pas $\gamma_n = 1/n$ dans (4.1.4) par $\gamma_n = \gamma/n$ avec*

$$\gamma = \frac{1}{2\pi|f_1|},$$

alors, d'après Fabian [25], $\widehat{\theta}_n$ est un estimateur asymptotiquement efficace de θ

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\varphi(\theta)}{4\pi^2 f_1^2}\right). \quad (4.2.4)$$

D'autre part, si f_1 est inconnu, la même procédure nous permet de construire un estimateur asymptotiquement efficace $\widehat{\theta}_n$ de θ en remplaçant f_1 par son estimateur naturel

$$\widehat{f}_{1,n} = \frac{1}{n} \sum_{k=1}^n \frac{Y_k \cos(2\pi(X_k - \widehat{\theta}_{k-1}))}{g(X_k)}.$$

Remarque 4.2.2 Dans le cas particulier où $4\pi|f_1| = 1$, il est possible de montrer [23] que

$$\sqrt{\frac{n}{\log(n)}}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \varphi(\theta)).$$

On peut également établir des résultats asymptotiques si $0 < 4\pi|f_1| < 1$. Cependant, dans toute la suite, on se place sous l'hypothèse que $4\pi|f_1| > 1$.

Théorème 4.2.3 Supposons (\mathcal{H}_1) et (\mathcal{H}_2) et que $|\theta| < 1/4$. Si (ε_n) a un moment d'ordre > 2 fini et si $4\pi|f_1| > 1$, alors on a la loi du log-itéré

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left(\frac{n}{2 \log \log n} \right)^{1/2} (\widehat{\theta}_n - \theta) &= - \liminf_{n \rightarrow \infty} \left(\frac{n}{2 \log \log n} \right)^{1/2} (\widehat{\theta}_n - \theta) \\ &= \xi(\theta) \quad p.s. \end{aligned} \quad (4.2.5)$$

En particulier,

$$\limsup_{n \rightarrow \infty} \left(\frac{n}{2 \log \log n} \right) (\widehat{\theta}_n - \theta)^2 = \xi^2(\theta) \quad p.s. \quad (4.2.6)$$

De plus, on a la loi forte quadratique

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{k=1}^n (\widehat{\theta}_k - \theta)^2 = \xi^2(\theta) \quad p.s. \quad (4.2.7)$$

Remarque 4.2.3 Il est également possible de se passer de l'hypothèse de parité sur f . D'après l'étude des conditions d'identifiabilité du Chapitre 3, cela implique alors de supposer connus les coefficients de Fourier f_1 et g_1 de f , donnés par

$$f_1 = \int_{-1/2}^{1/2} \cos(2\pi x) f(x) dx \quad \text{et} \quad g_1 = \int_{-1/2}^{1/2} \sin(2\pi x) f(x) dx,$$

et de supposer qu'on a soit $f_1 \neq 0$ soit $g_1 \neq 0$. On remplace alors la fonction ϕ définie par (4.1.1), par la fonction

$$\begin{aligned} \Phi(t) &= f_1 \mathbb{E} \left[\frac{\sin(2\pi(X-t))}{g(X)} f(X-\theta) \right] - g_1 \mathbb{E} \left[\frac{\cos(2\pi(X-t))}{g(X)} f(X-\theta) \right], \\ &= (f_1^2 + g_1^2) \sin(2\pi(\theta-t)). \end{aligned}$$

Le Théorème 4.2.1 est encore vrai pour l'algorithme de Robbins-Monro projeté

$$\widehat{\theta}_{n+1} = \pi_K \left(\widehat{\theta}_n + \gamma_{n+1} T_{n+1} \right)$$

où la valeur initiale $\widehat{\theta}_0 \in K$ et où la variable aléatoire T_{n+1} est définie par

$$T_{n+1} = \frac{f_1 \sin(2\pi(X_{n+1} - \widehat{\theta}_n))}{g(X_{n+1})} Y_{n+1} - \frac{g_1 \cos(2\pi(X_{n+1} - \widehat{\theta}_n))}{g(X_{n+1})} Y_{n+1}.$$

Ensuite, afin d'obtenir la normalité asymptotique, on remplace la fonction φ définie par (4.2.1), par la fonction

$$\begin{aligned} \Psi(t) &= \mathbb{E} \left[\frac{(f_1 \sin(2\pi(X-t)) - g_1 \cos(2\pi(X-t)))^2}{g^2(X)} (f^2(X-\theta) + \sigma^2) \right], \\ &= \int_{-1/2}^{1/2} \frac{(f_1 \sin(2\pi(x-t)) - g_1 \cos(2\pi(x-t)))^2}{g(x)} (f^2(x-\theta) + \sigma^2) dx. \end{aligned}$$

Les Théorèmes 4.2.2 et 4.2.3 sont alors de nouveau valables dès que $4\pi(f_1^2 + g_1^2) > 1$ avec

$$\xi^2(\theta) = \frac{\Psi(\theta)}{4\pi(f_1^2 + g_1^2) - 1}.$$

Dans toute la suite, on se place sous l'hypothèse de parité de f , cette hypothèse ne nécessitant aucune connaissance a priori des coefficients de Fourier de f .

4.3 Résultats de convergence non paramétrique

Théorème 4.3.1 *Supposons (\mathcal{H}_1) , (\mathcal{H}_2) , (\mathcal{H}_3) et que $|\theta| < 1/4$. Si la suite (ε_n) a un moment d'ordre > 2 fini, alors, pour tout $x \in \mathbb{R}$ tel que $|x| \leq 1/2$,*

$$\lim_{n \rightarrow \infty} \widehat{f}_n(x) = f(x) \quad p.s. \quad (4.3.1)$$

On établit aussi la normalité asymptotique de l'estimateur $\widehat{f}_n(x)$.

Théorème 4.3.2 *Supposons (\mathcal{H}_1) , (\mathcal{H}_2) , (\mathcal{H}_3) et que $|\theta| < 1/4$. Si la suite (ε_n) a un moment fini d'ordre > 2 , et que la fenêtre de lissage (h_n) est telle que $h_n = 1/n^\alpha$ avec $\alpha > 1/3$, alors on a, pour tout $x \in \mathbb{R}$ tels que $|x| \leq 1/2$ et $x \neq 0$,*

$$\sqrt{nh_n}(\widehat{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 \nu^2}{(1 + \alpha)(g(\theta + x) + g(\theta - x))}\right). \quad (4.3.2)$$

De plus, si $x = 0$,

$$\sqrt{nh_n}(\widehat{f}_n(0) - f(0)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 \nu^2}{(1 + \alpha)g(\theta)}\right). \quad (4.3.3)$$

Remarque 4.3.1 *Si on regarde le théorème précédent, on constate qu'en faisant $x = 0$ dans l'équation (4.3.2), on obtient une variance 2 fois plus petite que celle obtenue dans l'équation (4.3.3). Ceci s'explique par le fait que nous avons choisi un estimateur \widehat{f}_n pair pour estimer la fonction paire f . Ainsi, si $x \neq 0$, l'estimateur de $f(x)$ est également un estimateur de $f(-x)$, il est donc naturel que la variance asymptotique soit deux fois plus petite en $x \neq 0$ qu'en $x = 0$.*

4.4 Preuves des résultats paramétriques

4.4.1 Preuve du Théorème 4.2.1.

Sans perdre de généralité, on suppose que $f_1 > 0$. On note \mathcal{F}_n la σ -algèbre des événements s'étant produits avant l'instant n , $\mathcal{F}_n = \sigma(X_0, \varepsilon_0, \dots, X_n, \varepsilon_n)$. Tout d'abord, calculons les deux premiers moments conditionnels de la variable T_n définie par (4.1.5). D'après (4.0.1), nous avons

$$\begin{aligned} \mathbb{E}[T_{n+1} | \mathcal{F}_n] &= \mathbb{E}\left[\frac{\sin(2\pi(X_{n+1} - \widehat{\theta}_n))Y_{n+1}}{g(X_{n+1})} \middle| \mathcal{F}_n\right], \\ &= \mathbb{E}\left[\frac{\sin(2\pi(X_{n+1} - \widehat{\theta}_n))(f(X_{n+1} - \theta) + \varepsilon_{n+1})}{g(X_{n+1})} \middle| \mathcal{F}_n\right]. \end{aligned}$$

D'une part, puisque (X_n) est une suite de variables aléatoires de même loi que X , on a

$$\mathbb{E}\left[\frac{\sin(2\pi(X_{n+1} - \widehat{\theta}_n))f(X_{n+1} - \theta)}{g(X_{n+1})} \middle| \mathcal{F}_n\right] = \phi(\widehat{\theta}_n) \quad \text{p.s.} \quad (4.4.1)$$

où ϕ est la fonction définie par (4.1.2). D'autre part, les suites (X_n) et (ε_n) étant indépendantes et (ε_n) étant une suite de variables aléatoires indépendantes et centrées, on a

$$\mathbb{E}\left[\frac{\sin(2\pi(X_{n+1} - \widehat{\theta}_n))\varepsilon_{n+1}}{g(X_{n+1})} \middle| \mathcal{F}_n\right] = \mathbb{E}\left[\frac{\sin(2\pi(X - \widehat{\theta}_n))}{g(X)}\right]\mathbb{E}[\varepsilon_{n+1}] = 0.$$

Ainsi, (4.4.1) entraîne que

$$\mathbb{E}[T_{n+1} | \mathcal{F}_n] = \phi(\widehat{\theta}_n) \quad \text{p.s.} \quad (4.4.2)$$

De plus,

$$\begin{aligned} T_{n+1}^2 &= \frac{\sin^2(2\pi(X_{n+1} - \widehat{\theta}_n))Y_{n+1}^2}{g^2(X_{n+1})}, \\ &= \frac{\sin^2(2\pi(X_{n+1} - \widehat{\theta}_n))(f^2(X_{n+1} - \theta) + 2\varepsilon_{n+1}f(X_{n+1} - \theta) + \varepsilon_{n+1}^2)}{g^2(X_{n+1})}. \end{aligned}$$

Ainsi, puisque ε_{n+1} est centrée et indépendante de X_{n+1} et $\mathbb{E}[\varepsilon_{n+1}^2 | \mathcal{F}_n] = \mathbb{E}[\varepsilon_{n+1}^2] = \sigma^2$, on obtient que

$$\mathbb{E}[T_{n+1}^2 | \mathcal{F}_n] = \mathbb{E}\left[\frac{\sin^2(2\pi(X - \widehat{\theta}_n))}{g^2(X)}(f^2(X - \theta) + \sigma^2)\right] = \varphi(\widehat{\theta}_n) \quad (4.4.3)$$

où φ est donnée par (4.2.1). La fonction f étant bornée et la densité g ne s'annulant pas sur son support $[-1/2, 1/2]$, on déduit de (4.4.3) qu'il existe $M > 0$ telle que

$$\sup_{n \geq 0} \mathbb{E}[T_{n+1}^2 | \mathcal{F}_n] \leq M \quad \text{p.s.} \quad (4.4.4)$$

On pose, pour tout $n \geq 0$, $V_n = (\widehat{\theta}_n - \theta)^2$. On a clairement que

$$\begin{aligned} V_{n+1} &= (\widehat{\theta}_{n+1} - \theta)^2, \\ &= (\pi_K(\widehat{\theta}_n + \gamma_{n+1}T_{n+1}) - \theta)^2, \\ &= (\pi_K(\widehat{\theta}_n + \gamma_{n+1}T_{n+1}) - \pi_K(\theta))^2 \end{aligned}$$

puisque θ appartient à K . De plus, π_K est une fonction lipschitzienne de constante de Lipschitz égale à 1, donc

$$\begin{aligned} V_{n+1} &\leq (\widehat{\theta}_n + \gamma_{n+1}T_{n+1} - \theta)^2, \\ &\leq V_n + \gamma_{n+1}^2 T_{n+1}^2 + 2\gamma_{n+1}T_{n+1}(\widehat{\theta}_n - \theta). \end{aligned}$$

Ainsi, on déduit des équations (4.4.2) et (4.4.4) qu'on a la relation de récurrence

$$\begin{aligned} \mathbb{E}[V_{n+1} | \mathcal{F}_n] &\leq V_n + \gamma_{n+1}^2 \mathbb{E}[T_{n+1}^2 | \mathcal{F}_n] + 2\gamma_{n+1}(\widehat{\theta}_n - \theta)\mathbb{E}[T_{n+1} | \mathcal{F}_n], \\ &\leq V_n + \gamma_{n+1}^2 M + 2\gamma_{n+1}(\widehat{\theta}_n - \theta)\phi(\widehat{\theta}_n) \quad \text{p.s.} \end{aligned} \quad (4.4.5)$$

De plus, comme $\widehat{\theta}_n \in K$, on a $|\widehat{\theta}_n| < 1/4$ et donc $|\widehat{\theta}_n - \theta| < 1/2$. On a donc, presque sûrement, $(\widehat{\theta}_n - \theta)\phi(\widehat{\theta}_n) < 0$. On déduit alors de (4.4.5) et du lemme 1.3.1 de Robbins-Siegmund, que la suite (V_n) converge presque sûrement vers une variable aléatoire V finie presque sûrement et que

$$\sum_{n=1}^{\infty} \gamma_{n+1}(\theta - \widehat{\theta}_n)\phi(\widehat{\theta}_n) < +\infty \quad \text{p.s.} \quad (4.4.6)$$

Montrons que $V = 0$ p.s. Supposons le contraire, c'est-à-dire que $V \neq 0$ p.s. Cela signifie qu'on peut trouver $0 < a < b < 1/2$ tels que, pour n assez grand, l'événement $\{a < |\widehat{\theta}_n - \theta| < b\}$ ne soit pas négligeable. Cependant, sur cet anneau, on peut trouver une constante $c > 0$ telle que $(\theta - \widehat{\theta}_n)\phi(\widehat{\theta}_n) > c$. On déduit alors de (4.4.6) que

$$\sum_{n=1}^{\infty} \gamma_n < +\infty.$$

Ceci est en contradiction avec (4.1.3). Par conséquent, on en déduit que $V = 0$ p.s. et donc que la suite $(\widehat{\theta}_n)$ converge presque sûrement vers θ .

Pour conclure la preuve du Théorème 4.2.1, il reste à montrer que le nombre de fois que la suite $\widehat{\theta}_n + \gamma_{n+1}T_{n+1}$ sort de K est fini presque sûrement. Pour tout $n \geq 1$, notons

$$N_n = \sum_{k=0}^{n-1} \mathbf{1}_{\{|\widehat{\theta}_k + \gamma_{k+1}T_{k+1}| > 1/4\}}.$$

La suite (N_n) est une suite croissante. Par contradiction, supposons que N_n converge p.s. vers $+\infty$. Alors, on peut trouver une sous-suite (n_k) telle que (N_{n_k}) soit strictement croissante. Par conséquent, pour tout $n_k > 0$,

$$|\widehat{\theta}_{n_k} + \gamma_{n_k+1}T_{n_k+1}| > \frac{1}{4} \quad \text{p.s.}$$

ce qui implique immédiatement que $|\widehat{\theta}_{n_k+1}| = 1/4$ p.s. Ainsi,

$$\lim_{n_k \rightarrow \infty} |\widehat{\theta}_{n_k}| = |\theta| = \frac{1}{4} \quad \text{p.s.}$$

ce qui est une contradiction avec le fait que $|\theta| < 1/4$. Finalement, (N_n) converge presque sûrement vers une variable aléatoire finie p.s., ce qui termine la preuve du Théorème 4.2.1. \square

4.4.2 Preuve du Théorème 4.2.2.

Comme précédemment, on suppose que $f_1 > 0$. Afin d'obtenir le Théorème 4.2.2, nous allons appliquer le Théorème 2.1 du livre de Kushner et Yin [40] page 330. Tout d'abord, comme $\gamma_n = 1/n$, l'hypothèse requise sur la suite de pas décroissante est vérifiée. De plus, on sait que $\widehat{\theta}_n$ converge presque sûrement vers θ . Par conséquent, toutes les hypothèses locales du Théorème 2.1 de [40] sont satisfaites. De plus, on déduit de (4.4.2) que $\mathbb{E}[T_{n+1}|\mathcal{F}_n] = \phi(\widehat{\theta}_n)$ p.s. et que la fonction ϕ est continûment dérivable puisque $\phi(t) = f_1 \sin(2\pi(\theta - t))$. Ainsi, $\phi(\theta) = 0$ et $\phi'(\theta) = -2\pi f_1$ et, par hypothèse, $4\pi f_1 > 1$. On a également, d'après (4.4.3), que

$$\mathbb{E}[T_{n+1}^2|\mathcal{F}_n] = \varphi(\widehat{\theta}_n) \quad \text{p.s.}$$

ce qui entraîne

$$\lim_{n \rightarrow \infty} \mathbb{E}[T_{n+1}^2 | \mathcal{F}_n] = \varphi(\theta) \quad \text{p.s.}$$

Par conséquent, si on est capable de montrer que la suite (W_n) définie par, pour tout $n \geq 1$, par

$$W_n = \frac{(\widehat{\theta}_n - \theta)^2}{\gamma_n}$$

est tendue, alors on pourra déduire du Théorème 2.1 de [40] que

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \xi^2(\theta))$$

où

$$\xi^2(\theta) = \varphi(\theta) \int_0^{+\infty} \exp((1 - 4\pi f_1)t) dt = \frac{\varphi(\theta)}{4\pi f_1 - 1}.$$

Ainsi, il reste à prouver la tension de la suite (W_n) . Tout d'abord, on déduit de (4.4.5) qu'il existe une constante $M > 0$ telle que, pour tout $n \geq 1$,

$$\mathbb{E}[W_{n+1} | \mathcal{F}_n] \leq (1 + \gamma_n)W_n + \gamma_{n+1}M + 2(\widehat{\theta}_n - \theta)\phi(\widehat{\theta}_n). \quad (4.4.7)$$

De plus, pour tout $x \in \mathbb{R}$, on a $\phi(x) = 2\pi f_1(\theta - x) + f_1(\theta - x)v(x)$ où

$$v(x) = \frac{\sin(2\pi(\theta - x)) - 2\pi(\theta - x)}{(\theta - x)}.$$

La fonction v étant continue, il est possible de choisir $0 < \varepsilon < 1/2$ tel que, si $|x - \theta| < \varepsilon$,

$$\frac{q}{2f_1} < v(x) < 0. \quad (4.4.8)$$

On déduit également de (4.4.7) que, pour tout $n \geq 1$,

$$\mathbb{E}[W_{n+1} | \mathcal{F}_n] \leq W_n + 2\gamma_n W_n (q - f_1 v(\widehat{\theta}_n)) + \gamma_n M \quad (4.4.9)$$

avec $2q = 1 - 4\pi f_1$. Par hypothèse, on a $q < 0$. Ensuite, On note A_n et B_n les ensembles définis par $A_n = \{|\widehat{\theta}_n - \theta| \leq \varepsilon\}$ et

$$B_n = \bigcap_{k=m}^n A_k$$

avec $1 \leq m \leq n$. On déduit alors de (4.4.8) que

$$0 < -f_1 v(\widehat{\theta}_n) \mathbf{I}_{B_n} < -\left(\frac{q}{2}\right) \mathbf{I}_{B_n}. \quad (4.4.10)$$

Ainsi, les équations (4.4.9) et (4.4.10) entraînent que, pour tout $n \geq m$,

$$\begin{aligned} \mathbb{E}[W_{n+1} \mathbf{I}_{B_n} | \mathcal{F}_n] &\leq W_n \mathbf{I}_{B_n} + 2\gamma_n W_n \mathbf{I}_{B_n} \left(q - \frac{q}{2}\right) + \gamma_n M, \\ &\leq W_n \mathbf{I}_{B_n} (1 + q\gamma_n) + \gamma_n M. \end{aligned} \quad (4.4.11)$$

Puisque $B_{n+1} = B_n \cap A_{n+1}$, $B_{n+1} \subset B_n$, en passant à l'espérance des deux côtés de l'équation (4.4.11), on a que, pour tout $n \geq m$,

$$\mathbb{E}[W_{n+1} \mathbf{I}_{B_{n+1}}] \leq (1 + q\gamma_n) \mathbb{E}[W_n \mathbf{I}_{B_n}] + \gamma_n M. \quad (4.4.12)$$

Notons aussi $\alpha_n = \mathbb{E}[W_n I_{B_n}]$. L'équation (4.4.12) entraîne que, pour $n \geq m$,

$$\alpha_{n+1} \leq \beta_n \alpha_m + M \beta_n \sum_{k=m}^n \frac{\gamma_k}{\beta_k} \quad \text{où} \quad \beta_n = \prod_{k=m}^n (1 + q\gamma_k). \quad (4.4.13)$$

De plus, comme $\gamma_n = 1/n$, il est facile de voir que $\beta_n = O(n^q)$ et donc que

$$\sum_{k=1}^n \frac{\gamma_k}{\beta_k} = O(n^{-q}).$$

Par conséquent, (4.4.13) mène immédiatement à

$$\sup_{n \geq m} \alpha_n < +\infty. \quad (4.4.14)$$

On peut maintenant montrer la tension de la suite (W_n) . En effet, nous avons établi dans le Théorème 4.2.1 que $\widehat{\theta}_n$ converge vers θ p.s. Ainsi, si C_n est l'ensemble défini par

$$C_n = \bigcup_{k \geq n} \overline{A}_k,$$

alors $\mathbb{P}(C_n)$ converge vers 0 quand n tend vers $+\infty$. De plus, pour $n \geq m$, $\overline{B}_n \subset C_m$, ce qui implique que, si m, n tendent vers $+\infty$, $\mathbb{P}(\overline{B}_n)$ tend vers 0. Alors, pour tout $\xi, K > 0$ et pour tout $n \geq m$ avec m assez grand,

$$\begin{aligned} \mathbb{P}(W_n > K) &\leq \mathbb{P}(W_n I_{B_n} > K/2) + \mathbb{P}(W_n I_{\overline{B}_n} > K/2), \\ &\leq \frac{2}{K} \mathbb{E}[W_n I_{B_n}] + \mathbb{P}(\overline{B}_n). \end{aligned} \quad (4.4.15)$$

On déduit alors de (4.4.14) qu'on peut trouver K dépendant de ξ tel que le premier terme à droite de (4.4.15) soit plus petit que $\xi/2$. C'est aussi le cas pour le deuxième terme puisque $\mathbb{P}(\overline{B}_n)$ tend vers 0. Finalement, pour tout $\xi > 0$, il existe $K > 0$ tel que pour m assez grand,

$$\sup_{n \geq m} \mathbb{P}(W_n > K) < \xi.$$

Ainsi, la suite (W_n) est tendue, ce qui termine la preuve du Théorème 4.2.2. \square

4.4.3 Preuve du Théorème 4.2.3.

Comme le nombre de fois que la variable aléatoire $\widehat{\theta}_n + \gamma_{n+1} T_{n+1}$ sort de K est fini presque sûrement, la suite $(\widehat{\theta}_n)$ a le même comportement asymptotique presque sûr que l'algorithme de Robbins-Monro standard. Par conséquent, on déduit la loi du log itéré (4.2.5) du Théorème 1.3.2, et la loi forte quadratique (4.2.7) du Théorème 1.3.3. \square

4.5 Preuves des résultats non paramétriques

4.5.1 Preuve du Théorème 4.3.1.

Afin d'établir la convergence ponctuelle presque sûre donnée par le Théorème 4.3.1, on note pour tout $x \in \mathbb{R}$,

$$\widehat{h}_n(x) = \frac{1}{n} \sum_{k=1}^n W_k(x) Y_k \quad \text{et} \quad \widehat{g}_n(x) = \frac{1}{n} \sum_{k=1}^n W_k(x).$$

On obtient de (4.0.1) la décomposition

$$n\widehat{h}_n(x) = M_n(x) + P_n(x) + Q_n(x) + n\widehat{g}_n(x)f(x), \quad (4.5.1)$$

$$n\widehat{g}_n(x) = N_n(x) + R_n(x) + ng(\theta + x) \quad (4.5.2)$$

où

$$M_n(x) = \sum_{k=1}^n W_k(x)\varepsilon_k, \quad (4.5.3)$$

$$N_n(x) = \sum_{k=1}^n W_k(x) - \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}], \quad (4.5.4)$$

et

$$P_n(x) = \sum_{k=1}^n W_k(x)(f(X_k - \widehat{\theta}_{k-1}) - f(x)), \quad (4.5.5)$$

$$Q_n(x) = \sum_{k=1}^n W_k(x)(f(X_k - \theta) - f(X_k - \widehat{\theta}_{k-1})), \quad (4.5.6)$$

$$R_n(x) = \sum_{k=1}^n (\mathbb{E}[W_k(x)|\mathcal{F}_{k-1}] - g(\theta + x)). \quad (4.5.7)$$

D'une part,

$$\mathbb{E}[W_n(x)|\mathcal{F}_{n-1}] = \int_{\mathbb{R}} \frac{1}{h_n} K\left(\frac{x_n - \widehat{\theta}_{n-1} - x}{h_n}\right) g(x_n) dx_n.$$

Après le changement de variables $z = h_n^{-1}(x_n - \widehat{\theta}_{n-1} - x)$, comme la densité g est continue, deux fois dérivable, à dérivées bornées, un développement de Taylor entraîne que

$$\begin{aligned} \mathbb{E}[W_n(x)|\mathcal{F}_{n-1}] &= \int_{\mathbb{R}} K(z)g(\widehat{\theta}_{n-1} + x + h_n z) dz, \quad (4.5.8) \\ &= \int_{\mathbb{R}} K(z) \left(g(\widehat{\theta}_{n-1} + x) + h_n z g'(\widehat{\theta}_{n-1} + x) \right. \\ &\quad \left. + \frac{h_n^2 z^2}{2} g''(\widehat{\theta}_{n-1} + x + h_n z \xi) \right) dz, \\ &= g(\widehat{\theta}_{n-1} + x) + \frac{h_n^2}{2} \int_{\mathbb{R}} z^2 K(z) g''(\widehat{\theta}_{n-1} + x + h_n z \xi) dz \end{aligned}$$

où $0 < \xi < 1$. Par conséquent, pour tout $n \geq 1$,

$$|\mathbb{E}[W_n(x)|\mathcal{F}_{n-1}] - g(\widehat{\theta}_{n-1} + x)| \leq M_g \tau^2 h_n^2 \quad \text{p.s.} \quad (4.5.9)$$

où $M_g = \sup_{x \in \mathbb{R}} |g''(x)|$ et

$$\tau^2 = \frac{1}{2} \int_{\mathbb{R}} x^2 K(x) dx.$$

La continuité de la fonction g ainsi que le fait que $\widehat{\theta}_n$ converge vers θ p.s. mène alors à

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}] = g(\theta + x) \quad \text{p.s.} \quad (4.5.10)$$

On déduit alors immédiatement que, pour tout $x \in \mathbb{R}$,

$$R_n(x) = o(n) \quad \text{p.s.} \quad (4.5.11)$$

D'autre part, $(N_n(x))$ est une martingale de carré intégrable dont le crochet est donné par

$$\begin{aligned} \langle N(x) \rangle_n &= \sum_{k=1}^n \mathbb{E}[(N_k(x) - N_{k-1}(x))^2 | \mathcal{F}_{k-1}], \\ &= \sum_{k=1}^n \mathbb{E}[W_k^2(x) | \mathcal{F}_{k-1}] - \mathbb{E}^2[W_k(x) | \mathcal{F}_{k-1}]. \end{aligned}$$

Les mêmes calculs que dans (4.5.8) entraînent que

$$\begin{aligned} \mathbb{E}[W_n^2(x) | \mathcal{F}_{n-1}] &= \frac{1}{h_n} \int_{\mathbb{R}} K^2(z) g(\widehat{\theta}_{n-1} + x + h_n z) dz, \\ &= \frac{\nu^2}{h_n} g(\widehat{\theta}_{n-1} + x) + \frac{h_n}{2} \int_{\mathbb{R}} z^2 K^2(z) g''(\widehat{\theta}_{n-1} + x + h_n z \xi) dz \end{aligned}$$

où $0 < \xi < 1$. Ainsi,

$$|\mathbb{E}[W_n^2(x) | \mathcal{F}_{n-1}] - \frac{\nu^2}{h_n} g(\widehat{\theta}_{n-1} + x)| \leq M_g \mu^2 h_n \quad \text{p.s.} \quad (4.5.12)$$

avec

$$\nu^2 = \int_{\mathbb{R}} K^2(x) dx \quad \text{et} \quad \mu^2 = \frac{1}{2} \int_{\mathbb{R}} x^2 K^2(x) dx.$$

Finalement, puisque

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1+\alpha}} \sum_{k=1}^n h_k^{-1} = \frac{1}{1+\alpha}$$

on déduit de (4.5.9), (4.5.12), du lemme 1.3.2 de Toeplitz et de la convergence presque sûre de $g(\widehat{\theta}_n + x)$ vers $g(\theta + x)$ que

$$\lim_{n \rightarrow \infty} \frac{\langle N(x) \rangle_n}{n^{1+\alpha}} = \frac{\nu^2 g(\theta + x)}{1 + \alpha} \quad \text{p.s.} \quad (4.5.13)$$

Par conséquent, il découle de la loi forte des grands nombres pour les martingales, donnée par exemple au Théorème 1.3.15 page 20 de [23], que pour tout $\gamma > 0$, $(N_n(x))^2 = o(n^{1+\alpha} (\log n)^{1+\gamma})$ p.s. ce qui assure que, pour tout $x \in \mathbb{R}$,

$$N_n(x) = o(n) \quad \text{p.s.} \quad (4.5.14)$$

Ainsi, on déduit de (4.5.2), (4.5.11) et (4.5.14) que, pour tout $x \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \widehat{g}_n(x) = g(\theta + x) \quad \text{p.s.} \quad (4.5.15)$$

De plus, le noyau K étant à support compact, on peut trouver une constante $A > 0$ telle que K s'annule en dehors de $[-A, A]$. Ainsi, pour tout $n \geq 1$ et pour tout $x \in \mathbb{R}$,

$$W_n(x) = \frac{1}{h_n} K\left(\frac{X_n - \widehat{\theta}_{n-1} - x}{h_n}\right) \mathbb{I}_{\{|X_n - \widehat{\theta}_{n-1} - x| \leq Ah_n\}}.$$

La fonction f étant une fonction lipschitzienne, il existe une constante $C_f > 0$ telle que, pour tout $n \geq 1$,

$$|f(X_n - \widehat{\theta}_{n-1}) - f(x)| \leq C_f |X_n - \widehat{\theta}_{n-1} - x|.$$

Par conséquent, (4.5.5) entraîne que, pour tout $x \in \mathbb{R}$,

$$\begin{aligned} |P_n(x)| &\leq C_f \sum_{k=1}^n W_k(x) |X_k - \widehat{\theta}_{k-1} - x|, \\ &\leq AC_f \sum_{k=1}^n h_k W_k(x). \end{aligned} \quad (4.5.16)$$

Ainsi, les convergences (4.5.10), (4.5.14) et (4.5.16) nous permettent de déduire que, pour tout $x \in \mathbb{R}$,

$$P_n(x) = o(n) \quad \text{p.s.} \quad (4.5.17)$$

Ensuite, on obtient par (4.5.6) que, pour tout $x \in \mathbb{R}$,

$$|Q_n(x)| \leq C_f \sum_{k=1}^n W_k(x) |\widehat{\theta}_{k-1} - \theta|. \quad (4.5.18)$$

Ainsi, l'inégalité de Cauchy-Schwarz implique que

$$Q_n^2(x) \leq C_f^2 \sum_{k=1}^n W_k^2(x) \sum_{k=1}^n |\widehat{\theta}_{k-1} - \theta|^2. \quad (4.5.19)$$

On peut décomposer la première somme à droite de (4.5.19) en la somme de deux termes,

$$\sum_{k=1}^n W_k^2(x) = I_n(x) + J_n(x)$$

où

$$\begin{aligned} I_n(x) &= \sum_{k=1}^n W_k^2(x) - \mathbb{E}[W_k^2(x) | \mathcal{F}_{k-1}], \\ J_n(x) &= \sum_{k=1}^n \mathbb{E}[W_k^2(x) | \mathcal{F}_{k-1}]. \end{aligned}$$

En suivant exactement le même raisonnement que pour établir (4.5.14), il n'est pas difficile de voir que

$$I_n(x) = o(n^{1+\alpha}) \quad \text{p.s.}$$

On déduit également de (4.5.13) que

$$J_n(x) = O(n^{1+\alpha}) \quad \text{p.s.}$$

Par conséquent, on obtient que, pour tout $x \in \mathbb{R}$,

$$\sum_{k=1}^n W_k^2(x) = O(n^{1+\alpha}) \quad \text{p.s.} \quad (4.5.20)$$

Ainsi, on déduit de la loi forte quadratique donnée par (4.2.7) et de (4.5.19) et (4.5.20) que $Q_n^2(x) = O(n^{1+\alpha} \log n)$ p.s. ce qui implique que, pour tout $x \in \mathbb{R}$,

$$Q_n(x) = o(n) \quad \text{p.s.} \quad (4.5.21)$$

Maintenant, il ne reste plus qu'à étudier le comportement asymptotique de $(M_n(x))$ définie par (4.5.3). Les suites (X_n) et (ε_n) sont deux suites indépendantes de variables indépendantes et de même loi, donc il est immédiat que $(M_n(x))$ est une martingale de carré intégrable dont le crochet est donné par

$$\begin{aligned} \langle M(x) \rangle_n &= \sum_{k=1}^n \mathbb{E}[(M_k(x) - M_{k-1}(x))^2 | \mathcal{F}_{k-1}], \\ &= \sigma^2 \sum_{k=1}^n \mathbb{E}[W_k^2(x) | \mathcal{F}_{k-1}]. \end{aligned}$$

Alors, il s'ensuit de (4.5.13) que

$$\lim_{n \rightarrow \infty} \frac{\langle M(x) \rangle_n}{n^{1+\alpha}} = \frac{\sigma^2 \nu^2 g(\theta + x)}{1 + \alpha} \quad \text{p.s.} \quad (4.5.22)$$

Par conséquent, par la loi forte des grands nombres pour les martingales, on obtient que pour tout $\gamma > 0$, $(M_n(x))^2 = o(n^{1+\alpha} (\log n)^{1+\gamma})$ p.s. ce qui entraîne que

$$M_n(x) = o(n) \quad \text{p.s.} \quad (4.5.23)$$

Finalement, on déduit de (4.5.1) et (4.5.15) ainsi que de (4.5.17), (4.5.21) et (4.5.23) que, pour tout $x \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \widehat{h}_n(x) = f(x)g(\theta + x) \quad \text{p.s.} \quad (4.5.24)$$

On tire alors de l'égalité

$$\widehat{f}_n(x) = \frac{\widehat{h}_n(x) + \widehat{h}_n(-x)}{\widehat{g}_n(x) + \widehat{g}_n(-x)} \quad (4.5.25)$$

et de la parité de la fonction f que, pour tout $x \in \mathbb{R}$ tel que $|x| \leq 1/2$,

$$\lim_{n \rightarrow \infty} \widehat{f}_n(x) = f(x) \quad \text{p.s.} \quad (4.5.26)$$

4.5.2 Preuve du Théorème 4.3.2.

On s'intéresse maintenant à la normalité asymptotique de \widehat{f}_n . En utilisant (4.5.1), (4.5.2) et (4.5.25), on a la décomposition, pour tout $x \in \mathbb{R}$,

$$\widehat{f}_n(x) - f(x) = \frac{\mathcal{M}_n(x) + \mathcal{P}_n(x) + \mathcal{Q}_n(x)}{n\mathcal{G}_n(x)} \quad (4.5.27)$$

où $\mathcal{G}_n(x) = \widehat{g}_n(x) + \widehat{g}_n(-x)$ et

$$\begin{aligned} \mathcal{M}_n(x) &= M_n(x) + M_n(-x), \\ \mathcal{P}_n(x) &= P_n(x) + P_n(-x), \\ \mathcal{Q}_n(x) &= Q_n(x) + Q_n(-x), \end{aligned}$$

où $M_n(x)$, $P_n(x)$ et $Q_n(x)$ sont donnés respectivement par (4.5.3), (4.5.5) et (4.5.6). On a déjà vu, grâce à (4.5.15), que pour tout $x \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \mathcal{G}_n(x) = g(\theta + x) + g(\theta - x) \quad \text{p.s.} \quad (4.5.28)$$

Afin d'établir la normalité asymptotique, il est nécessaire d'être plus précis sur les vitesses de convergence données par (4.5.17) et (4.5.21). D'abord, on déduit de (4.5.16) que, pour tout $x \in \mathbb{R}$,

$$|P_n(x)| \leq AC_f(L_n(x) + \Lambda_n(x)) \quad (4.5.29)$$

où

$$\begin{aligned} L_n(x) &= \sum_{k=1}^n h_k(W_k(x) - \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}]), \\ \Lambda_n(x) &= \sum_{k=1}^n h_k \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}]. \end{aligned}$$

D'une part, l'équation (4.5.9) entraîne que

$$\Lambda_n(x) = O\left(\sum_{k=1}^n h_k\right) = O(n^{1-\alpha}) \quad \text{p.s.} \quad (4.5.30)$$

D'autre part, $(L_n(x))$ est une martingale de carré intégrable dont le crochet est donné par

$$\langle L(x) \rangle_n = \sum_{k=1}^n h_k^2 (\mathbb{E}[W_k^2(x)|\mathcal{F}_{k-1}] - \mathbb{E}^2[W_k(x)|\mathcal{F}_{k-1}]).$$

On déduit alors de (4.5.9), (4.5.12) et du Lemme 1.3.2 de Toeplitz que

$$\lim_{n \rightarrow \infty} \frac{\langle L(x) \rangle_n}{n^{1-\alpha}} = \frac{\nu^2 g(\theta + x)}{1 - \alpha} \quad \text{p.s.} \quad (4.5.31)$$

Par conséquent, il découle de la loi forte des grands nombres pour les martingales, que pour tout $\gamma > 0$, $(L_n(x))^2 = o(n^{1-\alpha}(\log n)^{1+\gamma})$ p.s. ce qui implique immédiatement que $(L_n(x))^2 = o(n^{1+\alpha})$ p.s. Aussi, les équations (4.5.29) et (4.5.30) nous permettent de dire que, si $\alpha > 1/3$, alors

$$(P_n(x))^2 = O(n^{2-2\alpha}) + o(n^{1+\alpha}) = o(n^{1+\alpha}) \quad \text{p.s.}$$

On a donc immédiatement que

$$(\mathcal{P}_n(x))^2 = o(n^{1+\alpha}) \quad \text{p.s.} \quad (4.5.32)$$

En suivant la preuve de (4.5.32), on déduit de (4.5.18) que, pour tout $x \in \mathbb{R}$,

$$|Q_n(x)| \leq C_f(S_n(x) + \Sigma_n(x)) \quad (4.5.33)$$

où

$$\begin{aligned} S_n(x) &= \sum_{k=1}^n \ell_k(W_k(x) - \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}]), \\ \Sigma_n(x) &= \sum_{k=1}^n \ell_k \mathbb{E}[W_k(x)|\mathcal{F}_{k-1}] \end{aligned}$$

où $\ell_n = |\widehat{\theta}_{n-1} - \theta|$. Il s'ensuit alors de (4.5.9), de l'inégalité de Cauchy-Schwarz et de la loi forte quadratique (4.2.7), que

$$\Sigma_n(x) = O\left(\sum_{k=1}^n \ell_k\right) = O(\sqrt{n \log n}) \quad \text{p.s.} \quad (4.5.34)$$

De plus, on déduit de (4.5.12) que $(S_n(x))$ est une martingale de carré intégrable dont le crochet est donné par

$$\langle S(x) \rangle_n = O(n^\alpha \log n) \quad \text{p.s.}$$

Par conséquent, on obtient à nouveau de la loi forte des grands nombres pour les martingales que, pour tout $\gamma > 0$, $(S_n(x))^2 = o(n^\alpha (\log n)^{2+\gamma})$ p.s. Ainsi $(S_n(x))^2 = o(n^{1+\alpha})$ p.s. et on déduit de (4.5.33) et (4.5.34) que

$$(Q_n(x))^2 = O(n \log n) + o(n^{1+\alpha}) = o(n^{1+\alpha}) \quad \text{p.s.}$$

Ceci implique alors que

$$(Q_n(x))^2 = o(n^{1+\alpha}) \quad \text{p.s.} \quad (4.5.35)$$

Il reste donc à établir la normalité asymptotique du terme $\mathcal{M}_n(x)$. On a déjà vu que $(M_n(x))$ est une martingale de carré intégrable. Ainsi, $(\mathcal{M}_n(x))$ est aussi une martingale de carré intégrable dont le crochet est donné par

$$\langle \mathcal{M}(x) \rangle_n = \sigma^2 \sum_{k=1}^n \mathbb{E}[(W_k(x) + W_k(-x))^2 | \mathcal{F}_{k-1}].$$

Il est donc nécessaire d'étudier le terme $\mathbb{E}[W_n(x)W_n(-x) | \mathcal{F}_{n-1}]$. Les mêmes calculs que dans (4.5.8) impliquent que

$$\begin{aligned} \mathbb{E}[W_n(x)W_n(-x) | \mathcal{F}_{n-1}] &= \frac{1}{h_n} \int_{\mathbb{R}} K(z)K(z + 2h_n^{-1}x)g(\widehat{\theta}_{n-1} + x + h_n z) dz, \\ &= \frac{1}{h_n} g(\widehat{\theta}_{n-1} + x)I_n(x) + g'(\widehat{\theta}_{n-1} + x)J_n(x) \\ &\quad + \frac{h_n}{2} \int_{\mathbb{R}} z^2 K(z)K(z + 2h_n^{-1}x)g''(\widehat{\theta}_{n-1} + x + h_n z \xi) dz \end{aligned}$$

avec $0 < \xi < 1$. Par conséquent, on a

$$\left| \mathbb{E}[W_n(x)W_n(-x) | \mathcal{F}_{n-1}] - \frac{1}{h_n} g(\widehat{\theta}_{n-1} + x)I_n(x) - g'(\widehat{\theta}_{n-1} + x)J_n(x) \right| \leq M_g H_n(x) h_n \quad \text{p.s.}$$

où

$$\begin{aligned} I_n(x) &= \int_{\mathbb{R}} K(z)K(z + 2h_n^{-1}x) dz, \\ J_n(x) &= \int_{\mathbb{R}} z K(z)K(z + 2h_n^{-1}x) dz, \\ H_n(x) &= \int_{\mathbb{R}} z^2 K(z)K(z + 2h_n^{-1}x) dz. \end{aligned}$$

Cependant, le noyau K étant à support compact, on a pour tout $x \in \mathbb{R}$ tel que $x \neq 0$,

$$\lim_{n \rightarrow \infty} K(z + 2h_n^{-1}x) = 0.$$

Le théorème de convergence dominée entraîne alors que les trois intégrales $I_n(x)$, $J_n(x)$, et $H_n(x)$ tendent vers 0 quand n tend vers l'infini, ce qui implique que, pour tout $x \in \mathbb{R}$ tel que $x \neq 0$,

$$\sum_{k=1}^n \mathbb{E}[W_k(x)W_k(-x)|\mathcal{F}_{k-1}] = o\left(\sum_{k=1}^n h_k^{-1}\right) = o(n^{1+\alpha}) \quad \text{p.s.} \quad (4.5.36)$$

Ainsi, les équations (4.5.22) et (4.5.36) entraînent que pour tout $x \in \mathbb{R}$ avec $x \neq 0$,

$$\lim_{n \rightarrow \infty} \frac{\langle \mathcal{M}(x) \rangle_n}{n^{1+\alpha}} = \frac{\sigma^2 \nu^2}{1+\alpha} (g(\theta+x) + g(\theta-x)) \quad \text{p.s.} \quad (4.5.37)$$

Si $x = 0$, il s'ensuit immédiatement de (4.5.22) que

$$\lim_{n \rightarrow \infty} \frac{\langle \mathcal{M}(0) \rangle_n}{n^{1+\alpha}} = \frac{4\sigma^2 \nu^2 g(\theta)}{1+\alpha} \quad \text{p.s.} \quad (4.5.38)$$

De plus, il n'est pas difficile de voir que la condition de Lindeberg est satisfaite. En effet, comme la suite (ε_n) a un moment d'ordre $a > 2$ fini, si on note $\Delta \mathcal{M}_n(x) = \mathcal{M}_n(x) - \mathcal{M}_{n-1}(x)$, on a

$$\mathbb{E}[|\Delta \mathcal{M}_n(x)|^a | \mathcal{F}_{n-1}] = \mathbb{E}[|\varepsilon_n|^a] \mathbb{E}[|W_n(x) - W_n(-x)|^a | \mathcal{F}_{n-1}].$$

Ainsi,

$$\mathbb{E}[|\Delta \mathcal{M}_n(x)|^a | \mathcal{F}_{n-1}] \leq 2^{a-1} \mathbb{E}[|\varepsilon_n|^a] \mathbb{E}[W_n^a(x) + W_n^a(-x) | \mathcal{F}_{n-1}].$$

Cependant, les mêmes calculs que dans (4.5.8) mènent à

$$\sum_{k=1}^n \mathbb{E}[W_k^a(x) | \mathcal{F}_{k-1}] = O\left(\sum_{k=1}^n h_k^{1-a}\right) = O(n^{1+\alpha(a-1)}) \quad \text{p.s.} \quad (4.5.39)$$

On a donc, pour tout $\varepsilon > 0$,

$$\frac{1}{n^{1+\alpha}} \sum_{k=1}^n \mathbb{E}[(\Delta \mathcal{M}_k(x))^2 \mathbb{I}_{|\Delta \mathcal{M}_k(x)| \geq \varepsilon \sqrt{n^{1+\alpha}}} | \mathcal{F}_{k-1}] \leq \frac{1}{\varepsilon^{a-2} n^b} \sum_{k=1}^n \mathbb{E}[|\Delta \mathcal{M}_k(x)|^a | \mathcal{F}_{k-1}]$$

où $b = a(1+\alpha)/2$. Par conséquent, on déduit de (4.5.39) que, pour tout $\varepsilon > 0$,

$$\frac{1}{n^{1+\alpha}} \sum_{k=1}^n \mathbb{E}[(\Delta \mathcal{M}_k(x))^2 \mathbb{I}_{|\Delta \mathcal{M}_k(x)| \geq \varepsilon \sqrt{n^{1+\alpha}}} | \mathcal{F}_{k-1}] = O(n^c) \quad \text{p.s.}$$

où $c = (2-a)(1-\alpha)/2$. Comme $c < 0$, la condition de Lindeberg est satisfaite. On tire alors du théorème central limite pour les martingales, donné par exemple par le Corollaire 2.1.10 de [23], que pour tout $x \in \mathbb{R}$ tel que $x \neq 0$,

$$\frac{\mathcal{M}_n(x)}{\sqrt{n^{1+\alpha}}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 \nu^2}{1+\alpha} (g(\theta+x) + g(\theta-x))\right) \quad (4.5.40)$$

tandis que, pour $x = 0$,

$$\frac{\mathcal{M}_n(0)}{\sqrt{n^{1+\alpha}}} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{4\sigma^2 \nu^2}{1+\alpha} g(\theta)\right). \quad (4.5.41)$$

Finalement, les équations (4.5.27) et (4.5.28) ainsi que (4.5.32), (4.5.35), (4.5.40), (4.5.41) et le lemme de Slutsky nous permettent de conclure que, pour tout $x \in \mathbb{R}$ tel que $|x| \leq 1/2$ et $x \neq 0$,

$$\sqrt{nh_n}(\widehat{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 \nu^2}{(1+\alpha)(g(\theta+x) + g(\theta-x))}\right)$$

tandis que, pour $x = 0$,

$$\sqrt{nh_n}(\widehat{f}_n(0) - f(0)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2 \nu^2}{(1+\alpha)g(\theta)}\right)$$

ce qui termine la preuve du Théorème 4.3.2. \square

4.6 Extension au cas où la densité g est inconnue

Nous avons vu que l'estimateur défini par $|\tilde{\theta}_0| < 1/4$ et, pour tout $n \geq 0$,

$$\tilde{\theta}_{n+1} = \pi_K\left(\tilde{\theta}_n + \text{sign}(f_1)\gamma_{n+1}T_{n+1}\right)$$

où

$$T_{n+1} = \frac{\sin(2\pi(X_{n+1} - \tilde{\theta}_n))}{g(X_{n+1})} Y_{n+1}$$

converge presque sûrement vers θ . Cependant, cet estimateur requiert la connaissance de la densité de probabilité g associée à la suite (X_n) . On propose ici de généraliser cet estimateur au cas où g est inconnue en remplaçant g dans l'expression de T_{n+1} par un estimateur de g . Plus précisément, on considère l'algorithme défini par $|\hat{\theta}_0| < 1/4$, et pour tout $n \geq 0$,

$$\hat{\theta}_{n+1} = \pi_K\left(\hat{\theta}_n + \text{sign}(f_1)\gamma_{n+1}\widehat{T}_{n+1}\right) \quad (4.6.1)$$

où

$$\widehat{T}_{n+1} = \frac{\sin(2\pi(X_{n+1} - \hat{\theta}_n))}{\widehat{g}_n(X_{n+1})} Y_{n+1}, \quad (4.6.2)$$

où \widehat{g}_n est l'estimateur récursif de Parzen-Rosenblatt de g défini, pour tout $x \in [-1/2; 1/2]$ et pour tout $n \geq 0$, par

$$\widehat{g}_n(x) = \frac{1}{n} \sum_{k=1}^n \frac{1}{h_k} K\left(\frac{X_k - x}{h_k}\right). \quad (4.6.3)$$

où K est une fonction symétrique, positive et bornée et telle que

$$\int_{-\infty}^{+\infty} K(x)dx = 1, \quad \int_{-\infty}^{+\infty} K^2(x)dx = \mu^2 < +\infty, \quad \frac{1}{2} \int_{-\infty}^{+\infty} x^2 K(x)dx = \nu^2 < +\infty.$$

On choisit $h_n = 1/n^\alpha$ avec $0 < \alpha < 1$. On fait également le choix de $\gamma_n = 1/n$. On a alors le théorème suivant.

Théorème 4.6.1 *Si K est une fonction lipschitzienne, alors la suite $(\hat{\theta}_n)$ définie par (4.6.1) converge presque sûrement vers θ .*

Remarque 4.6.1 *L'algorithme définissant la suite $(\hat{\theta}_n)$ a l'avantage de ne pas nécessiter la connaissance de g . Cependant, nous perdons la récursivité que l'on avait avec l'algorithme définissant $(\tilde{\theta}_n)$. L'implémentation de $(\hat{\theta}_n)$ sera donc plus longue que celle de $(\tilde{\theta}_n)$.*

4.6.1 Preuve du Théorème 4.6.1

Sans perte de généralité, on suppose que $f_1 > 0$. On note \mathcal{F}_n la tribu naturelle $\mathcal{F}_n = \sigma(X_0, Y_0, \dots, X_n, Y_n)$. Nous commençons par calculer les deux premiers moments conditionnels de \widehat{T}_{n+1} . D'une part,

$$\mathbb{E}[\widehat{T}_{n+1}|\mathcal{F}_n] = \mathbb{E}[T_{n+1}|\mathcal{F}_n] + \mathbb{E}[(\widehat{T}_{n+1} - T_{n+1})|\mathcal{F}_n]. \quad (4.6.4)$$

D'après (4.4.2), on sait que

$$\mathbb{E}[T_{n+1}|\mathcal{F}_n] = \phi(\widehat{\theta}_n) \quad \text{p.s.}$$

Ainsi, on déduit de (4.6.4) que

$$\mathbb{E}[\widehat{T}_{n+1}|\mathcal{F}_n] = \phi(\widehat{\theta}_n) + \mathbb{E}[(\widehat{T}_{n+1} - T_{n+1})|\mathcal{F}_n] \quad \text{p.s.} \quad (4.6.5)$$

D'autre part,

$$\begin{aligned} \mathbb{E}[\widehat{T}_{n+1}^2|\mathcal{F}_n] &= \mathbb{E}[(\widehat{T}_{n+1} - T_{n+1} + T_{n+1})^2|\mathcal{F}_n] \\ &\leq 2\mathbb{E}[T_{n+1}^2|\mathcal{F}_n] + 2\mathbb{E}[(\widehat{T}_{n+1} - T_{n+1})^2|\mathcal{F}_n] \end{aligned} \quad (4.6.6)$$

D'après (4.4.4), on sait aussi qu'il existe $M > 0$ telle que

$$\sup_{n \geq 0} \mathbb{E}[T_{n+1}^2|\mathcal{F}_n] \leq M \quad \text{p.s.}$$

On déduit donc de (4.6.6) que

$$\mathbb{E}[\widehat{T}_{n+1}^2|\mathcal{F}_n] \leq 2M + 2\mathbb{E}[(\widehat{T}_{n+1} - T_{n+1})^2|\mathcal{F}_n]. \quad (4.6.7)$$

Pour tout $n \geq 0$, on pose $V_n = (\widehat{\theta}_n - \theta)^2$. La projection π_K étant lipschitzienne de constante de Lipschitz plus petite que 1, on a alors la relation

$$\mathbb{E}[V_{n+1}|\mathcal{F}_n] \leq V_n + \gamma_{n+1}^2 \mathbb{E}[\widehat{T}_{n+1}^2|\mathcal{F}_n] + 2\gamma_{n+1}(\widehat{\theta}_n - \theta) \mathbb{E}[\widehat{T}_{n+1}|\mathcal{F}_n] \quad (4.6.8)$$

En utilisant (4.6.4) et (4.6.6), il s'ensuit de (4.6.8) que

$$\mathbb{E}[V_{n+1}|\mathcal{F}_n] \leq V_n + 2\gamma_{n+1}^2(M + P_n) + 2\gamma_{n+1}(\widehat{\theta}_n - \theta)\phi(\widehat{\theta}_n) + 2\gamma_{n+1}Q_n, \quad (4.6.9)$$

avec

$$P_n = \mathbb{E}[(\widehat{T}_{n+1} - T_{n+1})^2|\mathcal{F}_n]$$

et

$$Q_n = \mathbb{E}[(\widehat{T}_{n+1} - T_{n+1})|\mathcal{F}_n].$$

Afin d'utiliser le lemme 1.3.1 de Robbins-Siegmund pour conclure au fait que V_n converge p.s. vers une variable finie p.s., on a donc besoin d'avoir

$$\sum_{n=0}^{+\infty} \gamma_{n+1}^2 P_n < +\infty \quad \text{p.s.} \quad (4.6.10)$$

et

$$\sum_{n=0}^{+\infty} \gamma_{n+1} Q_n < +\infty \quad \text{p.s.} \quad (4.6.11)$$

Tout d'abord,

$$\begin{aligned} \widehat{T}_{n+1} - T_{n+1} &= \sin(2\pi(X_{n+1} - \widehat{\theta}_n)) Y_{n+1} \left(\frac{1}{\widehat{g}_n(X_{n+1})} - \frac{1}{g(X_{n+1})} \right) \\ &= \frac{\sin(2\pi(X_{n+1} - \widehat{\theta}_n)) Y_{n+1}}{g(X_{n+1}) \widehat{g}_n(X_{n+1})} (g(X_{n+1}) - \widehat{g}_n(X_{n+1})). \end{aligned}$$

Comme g ne s'annule pas sur son support, que f est bornée et que le bruit ε_{n+1} est indépendant de \mathcal{F}_n et a un moment fini d'ordre 2, on a immédiatement l'existence d'une constante $C > 0$ telle que

$$\begin{aligned} P_n &= \mathbb{E} \left[\frac{\sin^2(2\pi(X_{n+1} - \widehat{\theta}_n)) Y_{n+1}^2}{g^2(X_{n+1}) \widehat{g}_n^2(X_{n+1})} (g(X_{n+1}) - \widehat{g}_n(X_{n+1}))^2 \middle| \mathcal{F}_n \right] \\ &\leq C \mathbb{E} \left[\frac{(g(X_{n+1}) - \widehat{g}_n(X_{n+1}))^2}{\widehat{g}_n^2(X_{n+1})} \middle| \mathcal{F}_n \right], \end{aligned} \quad (4.6.12)$$

et l'existence d'une constante $D > 0$ telle que

$$\begin{aligned} Q_n &\leq \mathbb{E} \left[\frac{|\sin(2\pi(X_{n+1} - \widehat{\theta}_n)) Y_{n+1}|}{g(X_{n+1}) \widehat{g}_n(X_{n+1})} |g(X_{n+1}) - \widehat{g}_n(X_{n+1})| \middle| \mathcal{F}_n \right] \\ &\leq D \mathbb{E} \left[\frac{|g(X_{n+1}) - \widehat{g}_n(X_{n+1})|}{\widehat{g}_n(X_{n+1})} \middle| \mathcal{F}_n \right]. \end{aligned} \quad (4.6.13)$$

D'une part,

$$\mathbb{E} \left[\frac{(g(X_{n+1}) - \widehat{g}_n(X_{n+1}))^2}{\widehat{g}_n^2(X_{n+1})} \middle| \mathcal{F}_n \right] = \int_{-1/2}^{1/2} (g(x) - \widehat{g}_n(x))^2 \frac{g(x)}{\widehat{g}_n^2(x)} dx \quad (4.6.14)$$

et d'autre part,

$$\mathbb{E} \left[\frac{|g(X_{n+1}) - \widehat{g}_n(X_{n+1})|}{\widehat{g}_n(X_{n+1})} \middle| \mathcal{F}_n \right] = \int_{-1/2}^{1/2} |g(x) - \widehat{g}_n(x)| \frac{g(x)}{\widehat{g}_n(x)} dx. \quad (4.6.15)$$

De plus, pour tout $x \in [-1/2; 1/2]$, on a la décomposition

$$\begin{aligned} \widehat{g}_n(x) - g(x) &= \widehat{g}_n(x) - \mathbb{E}[\widehat{g}_n(x)] + \mathbb{E}[\widehat{g}_n(x)] - g(x) \\ &= \frac{M_n(x)}{n} + \frac{R_n(x)}{n} \end{aligned} \quad (4.6.16)$$

où, pour tout $n \geq 1$ et pour tout $x \in [-1/2; 1/2]$,

$$M_n(x) = \sum_{k=1}^n K_{h_k}(X_k - x) - \mathbb{E}[K_{h_k}(X_k - x)], \quad (4.6.17)$$

et

$$R_n(x) = \sum_{k=1}^n \mathbb{E}[K_{h_k}(X_k - x)] - g(x), \quad (4.6.18)$$

avec, pour tout $x \in \mathbb{R}$,

$$K_{h_k}(x) = \frac{1}{h_k} K\left(\frac{x}{h_k}\right).$$

Tout d'abord, $(M_n(x))$ est une martingale de carré intégrable dont le crochet est donné par

$$\begin{aligned} \langle M(x) \rangle_n &= \sum_{k=1}^n \mathbb{E}[K_{h_k}(X_k - x)^2] - \mathbb{E}[K_{h_k}(X_k - x)]^2 \\ &\leq \sum_{k=1}^n \mathbb{E}[K_{h_k}(X_k - x)^2]. \end{aligned}$$

Cependant, on a pour tout $1 \leq k \leq n$,

$$\mathbb{E}[K_{h_k}(X_k - x)^2] \leq \frac{1}{h_k} \int_{\mathbb{R}} K^2(y) g(x + h_k y) dy.$$

Ainsi, comme g est bornée, on en déduit que, pour tout $x \in [-1/2; 1/2]$,

$$\langle M(x) \rangle_n \leq \|g\|_{\infty} \mu^2 \sum_{k=1}^n \frac{1}{h_k} \leq \|g\|_{\infty} \mu^2 \frac{n}{h_n} \quad \text{p.s.} \quad (4.6.19)$$

De plus, si on note, pour tout $n \geq 1$, $\Delta M_n(x) = M_n(x) - M_{n-1}(x)$, on a

$$|\Delta M_n(x)| \leq \frac{2}{h_n} \|K\|_{\infty}. \quad (4.6.20)$$

En particulier, avec $h_n = 1/n^\alpha$, on déduit de (4.6.19) et de (4.6.20) qu'il existe deux constantes a et b telles que

$$\langle M(0) \rangle_n \leq a n^{1+\alpha} \quad \text{et} \quad |\Delta M_n(0)| \leq b n^\alpha. \quad (4.6.21)$$

De plus, comme le noyau K est borné et lipschitzien, pour tout $\delta \in]0; 1[$, il existe une constante C_δ telle que, pour tout $x, y \in \mathbb{R}$,

$$|K(x) - K(y)| \leq C_\delta |x - y|^\delta. \quad (4.6.22)$$

Ainsi, pour tout $x, y \in [-1/2; 1/2]$, on a

$$|\Delta M_n(x) - \Delta M_n(y)| \leq 2C_\delta |x - y|^\delta n^{\alpha(1+\delta)}.$$

De plus, pour tout $x, y \in [-1/2; 1/2]$,

$$\begin{aligned} \langle M(x) - M(y) \rangle_n &\leq \sum_{k=1}^n \mathbb{E}[(K_{h_k}(X_k - x) - K_{h_k}(X_k - y))^2], \\ &\leq \sum_{k=1}^n k^{2\alpha} \int_{\mathbb{R}} (K(k^\alpha(u - x)) - K(k^\alpha(u - y)))^2 g(u) du. \end{aligned}$$

Grâce au changement de variables $t = k^\alpha (u - x)$, on obtient donc que

$$\langle M(x) - M(y) \rangle_n \leq \|g\|_\infty \sum_{k=1}^n k^\alpha \int_{\mathbb{R}} (K(t) - K(t + k^\alpha(x - y)))^2 dt. \quad (4.6.23)$$

De plus, comme K est une densité, il s'ensuit de (4.6.22) que, pour tout $1 \leq k \leq n$,

$$\int_{\mathbb{R}} (K(t) - K(t + k^\alpha(x - y)))^2 dt \leq 2C_{2\delta} |x - y|^{2\delta} k^{2\alpha\delta}.$$

Alors, on déduit de (4.6.23) que pour tout $x, y \in [-1/2; 1/2]$,

$$\langle M(x) - M(y) \rangle_n \leq 2C_{2\delta} |x - y|^{2\delta} n^{1+\alpha+2\alpha\delta}.$$

Puisque δ peut être choisi aussi petit qu'on veut, les quatre conditions du Théorème 6.4.34 page 220 de [23] sont satisfaites. La martingale $(M_n(x))$ vérifie donc la loi des grands nombres uniforme dilatée, et plus particulièrement, pour tout $(1 + \alpha)/2 < \beta < 1$,

$$\sup_{|x| \leq 1/2} |M_n(x)| = o(n^\beta) \quad \text{p.s.} \quad (4.6.24)$$

Il reste à majorer convenablement le terme de reste $R_n(x)$. On a pour tout $x \in [-1/2; 1/2]$ et pour tout $1 \leq k \leq n$,

$$\begin{aligned} |\mathbb{E}[K_{h_k}(X_k - x)] - g(x)| &\leq \int_{\mathbb{R}} K(y) |g(x + h_k y) - g(x)| dy \\ &\leq h_k^2 \|g''\|_\infty \nu^2. \end{aligned}$$

Ainsi,

$$\sup_{|x| \leq 1/2} \frac{|R_n(x)|}{n} = \mathcal{O}\left(\frac{1}{n} \sum_{k=1}^n h_k^2\right) = \mathcal{O}(h_n^2) \quad \text{p.s.} \quad (4.6.25)$$

On déduit alors de (4.6.16), (4.6.24) et (4.6.25) que

$$\sup_{|x| \leq 1/2} |\widehat{g}_n(x) - g(x)| = \mathcal{O}(n^{-2\alpha} + n^{\beta-1}) \quad \text{p.s.} \quad (4.6.26)$$

Ainsi, il découle de (4.6.14), (4.6.15) que, d'une part,

$$\mathbb{E}\left[\frac{(g(X_{n+1}) - \widehat{g}_n(X_{n+1}))^2}{\widehat{g}_n^2(X_{n+1})} \middle| \mathcal{F}_n\right] = \mathcal{O}\left((n^{-4\alpha} + n^{2(\beta-1)}) \int_{-1/2}^{1/2} \frac{g(x)}{\widehat{g}_n^2(x)} dx\right) \quad \text{p.s.} \quad (4.6.27)$$

et d'autre part,

$$\mathbb{E}\left[\frac{|g(X_{n+1}) - \widehat{g}_n(X_{n+1})|}{\widehat{g}_n(X_{n+1})} \middle| \mathcal{F}_n\right] = \mathcal{O}\left((n^{-2\alpha} + n^{\beta-1}) \int_{-1/2}^{1/2} \frac{g(x)}{\widehat{g}_n(x)} dx\right) \quad \text{p.s.} \quad (4.6.28)$$

ce qui implique, grâce à (4.6.12) et (4.6.13) que

$$P_n = \mathcal{O}\left((n^{-4\alpha} + n^{2(\beta-1)}) \int_{-1/2}^{1/2} \frac{g(x)}{\widehat{g}_n^2(x)} dx\right) \quad \text{p.s.} \quad (4.6.29)$$

et

$$Q_n = \mathcal{O}\left((n^{-2\alpha} + n^{\beta-1}) \int_{-1/2}^{1/2} \frac{g(x)}{\widehat{g}_n(x)} dx\right) \quad \text{p.s.} \quad (4.6.30)$$

Enfin, on sait que, pour tout $x \in [-1/2; 1/2]$,

$$\lim_{n \rightarrow +\infty} \widehat{g}_n(x) = g(x) \quad \text{p.s.}$$

Comme g et \widehat{g}_n sont définies sur le compact $[-1/2; 1/2]$, ceci entraîne que

$$P_n = \mathcal{O}\left(n^{-4\alpha} + n^{2(\beta-1)}\right) \quad \text{p.s.} \quad (4.6.31)$$

et

$$Q_n = \mathcal{O}\left(n^{-2\alpha} + n^{\beta-1}\right) \quad \text{p.s.} \quad (4.6.32)$$

On en déduit alors immédiatement que, pour tout $0 < \alpha < 1$,

$$\sum_{n=1}^{+\infty} \gamma_{n+1}^2 P_n < +\infty \quad \text{p.s.}$$

et

$$\sum_{n=1}^{+\infty} \gamma_{n+1} Q_n < +\infty \quad \text{p.s.}$$

La fin de la preuve du Théorème 4.6.1 suit alors exactement la fin de la preuve du Théorème 4.2.1. \square

Remarque 4.6.2 *Une perspective de recherche immédiate est d'établir la normalité asymptotique de cette suite. Je travaille actuellement sur le sujet.*

Chapitre 5

Modèle de déformation à forme commune

Ce chapitre est consacré à l'étude du modèle de déformation à forme commune (3.1.3) défini, pour tout $1 \leq j \leq p$ et pour tout $i \geq 1$, par

$$Y_{i,j} = a_j f(X_i - \theta_j) + v_j + \varepsilon_{i,j} \quad (5.0.1)$$

où pour tout $1 \leq j \leq p$, (X_i) et $(\varepsilon_{i,j})$ sont deux suites indépendantes de variables aléatoires indépendantes et de même loi. Dans ce travail, on étend les résultats précédemment établis dans le chapitre 4 en estimant, pour tout $p \geq 1$ fixé, le paramètre de moyenne v , le paramètre de translation θ et le paramètre d'échelle a , donnés respectivement par

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_p \end{pmatrix}, \quad \theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_p \end{pmatrix}, \quad a = \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix}. \quad (5.0.2)$$

L'apport essentiel de ce chapitre en contraste avec le chapitre précédent est que l'on est maintenant capable d'estimer récursivement un paramètre d'échelle a , crucial dans l'étude de certaines données comme par exemple les données de variations de températures. A cet effet, on pourra consulter l'article de Vimond [66] et son application aux données de températures où la connaissance du signe de a est très importante car elle permet de dire dans quel hémisphère se situe la ville dont on observe les variations climatiques.

Remarque 5.0.3 *Le modèle (5.0.1) est une version simplifiée du modèle plus général défini, pour tout $1 \leq j \leq p$ et pour tout $i \geq 1$, par*

$$Y_{i,j} = a_j f(X_{i,j} - \theta_j) + v_j + \varepsilon_{i,j}$$

où les variables $X_{i,j}$ sont indépendantes et de même loi. Néanmoins, tous les résultats qui suivent sont aussi valides pour ce modèle.

5.1 Procédure d'estimation

Afin d'estimer les paramètres v , θ et a , on suppose que le bruit $(\varepsilon_{i,j})$ est une suite de variables aléatoires indépendantes de moyenne nulle et de variances $\mathbb{E}[\varepsilon_{i,j}^2] = \sigma_j^2$. De plus,

comme dans le chapitre 4, nous faisons les hypothèses classiques suivantes.

- (\mathcal{H}_1) Les temps d'observations (X_i) sont indépendants et de même loi qui admet une densité de probabilité g , ne s'annulant pas sur son support $[-1/2; 1/2]$.
De plus, g est continue, deux fois dérivables, à dérivées bornées.
- (\mathcal{H}_2) La fonction f est symétrique, bornée, périodique de période 1.

De plus, comme nous avons remarqué dans le chapitre 3, ces hypothèses ne suffisent pas à garantir l'identifiabilité du modèle (5.0.1). C'est pourquoi nous rajoutons aux hypothèses (\mathcal{H}_1) et (\mathcal{H}_2) les deux hypothèses suivantes.

- (\mathcal{H}_3) $\int_{-1/2}^{1/2} f(x)dx = 0$,
- (\mathcal{H}_4) $a_1 = 1$, $\theta_1 = 0$ et $\max_{1 \leq j \leq p} |\theta_j| < 1/4$.

Ces hypothèses correspondent à l'ensemble \mathcal{C}_2 défini par (3.2.10) au chapitre 3. L'hypothèse (\mathcal{H}_3) nous permet de définir v de manière unique tandis que la contrainte sur θ de l'hypothèse (\mathcal{H}_4) nous permet de définir θ de manière unique. De plus, $a_1 = 1$, $\theta_1 = 0$ signifie que la première courbe $j = 1$ est choisie comme courbe de référence. Ces contraintes d'identifiabilité sont bien adaptées à notre cadre de travail. Comme signalé au chapitre 3, on peut remplacer (\mathcal{H}_3) et (\mathcal{H}_4) par (\mathcal{H}'_3) et (\mathcal{H}'_4) définies par

- (\mathcal{H}'_3) $\int_{-1/2}^{1/2} f(x)dx = 0$ et $\sup_{x \in [0;1]} |f(x)| = 1$,
- (\mathcal{H}'_4) $\theta_1 = 0$, $\min_{1 \leq j \leq p} a_j > 0$ et $\max_{1 \leq j \leq p} |\theta_j| < 1/2$.

Une alternative est aussi de remplacer (\mathcal{H}_4) par (\mathcal{H}''_4) ou (\mathcal{H}'''_4) qui sont données par

- (\mathcal{H}''_4) $a_1 = 1$, $\theta_1 = 0$, $\min_{1 \leq j \leq p} a_j > 0$, $\max_{1 \leq j \leq p} |\theta_j| < 1/2$,
- (\mathcal{H}'''_4) $a_1 > 0$, $\theta_1 = 0$, $\sum_{j=1}^p a_j^2 = p$ et $\max_{1 \leq j \leq p} |\theta_j| < 1/4$.

Sous ces hypothèses, on peut facilement définir un estimateur fortement consistant de f_1 , le premier coefficient de Fourier de f

$$f_1 = \int_{-1/2}^{1/2} \cos(2\pi x) f(x) dx.$$

Cet estimateur $\widehat{f}_{1,n}$ est donné par

$$\widehat{f}_{1,n} = \frac{1}{n} \sum_{i=1}^n \frac{\cos(2\pi X_i)}{g(X_i)} Y_{i,1}. \quad (5.1.1)$$

Dans toute la suite, X sera une variable aléatoire de même loi que (X_i) tandis que, pour tout $1 \leq j \leq p$, Y_j et ε_j seront des variables aléatoires de même loi que $(Y_{i,j})$ et $(\varepsilon_{i,j})$. Enfin, on note

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_p \end{pmatrix}, \quad (5.1.2)$$

où

$$Y_j = a_j f(X - \theta_j) + v_j + \varepsilon_j.$$

5.1.1 Estimation de v

L'estimation du paramètre de moyenne v se fait aisément via (\mathcal{H}_2) et (\mathcal{H}_3) . En effet, nous avons, pour tout $1 \leq j \leq p$,

$$\mathbb{E} \left[\frac{Y_{i,j}}{g(X_i)} \right] = v_j.$$

Ainsi, un estimateur naturel \widehat{v}_n de v est donné, pour tout $1 \leq j \leq p$, par

$$\widehat{v}_{n,j} = \frac{1}{n} \sum_{i=1}^n \frac{Y_{i,j}}{g(X_i)}. \quad (5.1.3)$$

5.1.2 Estimation de θ

Comme dans le chapitre 4, on propose d'utiliser la fonction ϕ définie, pour tout $t \in \mathbb{R}^p$, par

$$\phi(t) = \mathbb{E} \left[D(X, t) \begin{pmatrix} a_1 f(X - \theta_1) \\ \vdots \\ a_p f(X - \theta_p) \end{pmatrix} \right] \quad (5.1.4)$$

où $D(X, t)$ est la matrice diagonale de taille p définie par

$$D(X, t) = \frac{1}{g(X)} \text{diag} \left(\sin(2\pi(X - t_1)), \dots, \sin(2\pi(X - t_p)) \right). \quad (5.1.5)$$

En utilisant la symétrie de f , les mêmes calculs que dans le chapitre 4 mènent, pour tout $1 \leq j \leq p$, à

$$\mathbb{E} \left[\frac{\sin(2\pi(X - t_j))}{g(X)} a_j f(X - \theta_j) \right] = a_j f_1 \sin(2\pi(\theta_j - t_j)) \quad (5.1.6)$$

Par conséquent, pour tout $t \in \mathbb{R}^p$,

$$\phi(t) = f_1 \begin{pmatrix} a_1 \sin(2\pi(\theta_1 - t_1)) \\ \vdots \\ a_p \sin(2\pi(\theta_p - t_p)) \end{pmatrix}. \quad (5.1.7)$$

Sans perdre de généralité, on suppose que $f_1 \neq 0$. Pour estimer le paramètre θ , nous allons mettre en place la même procédure de Robbins-Monro que dans le chapitre 4 pour chaque coordonnée de θ . Plus précisément, pour tout $1 \leq j \leq p$, en notant $\phi_j(t) = a_j f_1 \sin(2\pi(\theta_j - t_j))$, alors si $|t_j - \theta_j| < 1/2$, $(t_j - \theta_j)\phi_j(t)$ est de signe négatif si $\text{sign}(a_j f_1) > 0$ et de signe positif sinon. De plus, en notant $K = [-1/4; 1/4]$, on définit la projection π_K sur K , pour tout $x \in \mathbb{R}$, par

$$\pi_K(x) = \begin{cases} x & \text{if } |x| \leq 1/4, \\ 1/4 & \text{if } x > 1/4, \\ -1/4 & \text{if } x < -1/4. \end{cases}$$

Soit (γ_n) une suite déterministe positive décroissante vers 0 et telle que

$$\sum_{n=1}^{\infty} \gamma_n = +\infty \quad \text{et} \quad \sum_{n=1}^{\infty} \gamma_n^2 < +\infty. \quad (5.1.8)$$

On choisira pour simplifier $\gamma_n = 1/n$. Alors, pour tout $1 \leq j \leq p$, nous allons estimer θ_j par la suite $(\widehat{\theta}_{n,j})$ définie, pour tout $n \geq 1$, par

$$\widehat{\theta}_{n+1,j} = \pi_K \left(\widehat{\theta}_{n,j} + \text{sign}(a_j f_1) \gamma_{n+1} T_{n+1,j} \right) \quad (5.1.9)$$

où la valeur initiale $\widehat{\theta}_0 \in K^p$ et le vecteur aléatoire T_{n+1} est donné par

$$T_{n+1} = D(X_{n+1}, \widehat{\theta}_n) \begin{pmatrix} Y_{n+1,1} \\ \vdots \\ Y_{n+1,p} \end{pmatrix}. \quad (5.1.10)$$

5.1.3 Estimation de a

Pour l'estimation de a , on introduit une autre fonction auxiliaire ψ définie, pour tout $t \in \mathbb{R}^p$, par

$$\psi(t) = \mathbb{E} \left[C(X, t) \begin{pmatrix} a_1 f(X - \theta_1) \\ \vdots \\ a_p f(X - \theta_p) \end{pmatrix} \right] \quad (5.1.11)$$

où $C(X, t)$ est la matrice diagonale d'ordre p , donnée par

$$C(X, t) = \frac{1}{g(X)} \text{diag} \left(\cos(2\pi(X - t_1)), \dots, \cos(2\pi(X - t_p)) \right). \quad (5.1.12)$$

Comme pour (5.1.7), en utilisant la périodicité et la symétrie de f , nous avons après calculs,

$$\psi(t) = f_1 \begin{pmatrix} a_1 \cos(2\pi(\theta_1 - t_1)) \\ \vdots \\ a_p \cos(2\pi(\theta_p - t_p)) \end{pmatrix}. \quad (5.1.13)$$

Si on suppose que f_1 est connu, nous allons donc choisir pour estimateur de a , la suite (\widehat{a}_n) définie, pour $n \geq 1$ et pour tout $1 \leq j \leq p$, par

$$\widehat{a}_{n,j} = \frac{1}{n f_1} \sum_{i=1}^n \frac{\cos(2\pi(X_i - \widehat{\theta}_{i-1,j}))}{g(X_i)} Y_{i,j}. \quad (5.1.14)$$

Si f_1 est inconnu alors on remplace f_1 par son estimateur $\widehat{f}_{1,n}$ donné par (5.1.1) pour obtenir comme estimateur de a , la suite (\widetilde{a}_n) définie, pour $n \geq 1$ et pour tout $1 \leq j \leq p$, par

$$\widetilde{a}_{n,j} = \frac{1}{n \widehat{f}_{1,n}} \sum_{i=1}^n \frac{\cos(2\pi(X_i - \widehat{\theta}_{i-1,j}))}{g(X_i)} Y_{i,j}. \quad (5.1.15)$$

5.1.4 Estimation de f

On s'intéresse maintenant à l'estimation de la fonction de régression f . Tout d'abord, comme dans le chapitre 4, on ajoute l'hypothèse de régularité suivante à f .

(\mathcal{H}_5) La fonction de régression f est lipschitzienne.

Pour des raisons de clarté, on suppose que f_1 est connu (voir la Remarque 5.1.1 ci-dessous). Pour l'estimation de f , on considère alors l'estimateur pondéré de Nadaraya-Watson défini, pour tout $n \geq 1$ et pour tout $|x| \leq 1/2$, par

$$\widehat{f}_n(x) = \sum_{j=1}^p \omega_j(x) \widehat{f}_{n,j}(x), \quad (5.1.16)$$

où, pour tout $1 \leq j \leq p$,

$$\omega_j(x) = \omega_j(-x), \quad \omega_j(x) \geq 0 \quad \text{et} \quad \sum_{j=1}^p \omega_j(x) = 1, \quad (5.1.17)$$

$$\widehat{f}_{n,j}(x) = \frac{1}{\widehat{a}_{n,j}} \frac{\sum_{i=1}^n (W_{i,j}(x) + W_{i,j}(-x)) (Y_{i,j} - \widehat{v}_{i-1,j})}{\sum_{i=1}^n (W_{i,j}(x) + W_{i,j}(-x))}, \quad (5.1.18)$$

et

$$W_{n,j}(x) = \frac{1}{h_n} K\left(\frac{X_n - \widehat{\theta}_{n-1,j} - x}{h_n}\right). \quad (5.1.19)$$

On choisit une fenêtre de lissage $h_n = 1/n^\alpha$ avec $\alpha \in]0; 1[$. De plus, on suppose que le noyau K est une fonction positive, symétrique et bornée à support compact, et satisfaisant

$$\int_{\mathbb{R}} K(x) dx = 1 \quad \text{et} \quad \int_{\mathbb{R}} K^2(x) dx = \nu^2.$$

Remarque 5.1.1 Si f_1 est inconnu, il est nécessaire de remplacer $\widehat{f}_n(x)$ défini par (5.1.16) par $\widetilde{f}_n(x)$ donné par

$$\widetilde{f}_n(x) = \sum_{j=1}^p \omega_j(x) \widetilde{f}_{n,j}(x), \quad (5.1.20)$$

où $\widetilde{f}_{n,j}(x)$ est défini, pour tout $x \in [-1/2; 1/2]$ et pour tout $n \geq 1$, par

$$\widetilde{f}_{n,j}(x) = \frac{1}{\widetilde{a}_{n,j}} \frac{\sum_{i=1}^n (W_{i,j}(x) + W_{i,j}(-x)) (Y_{i,j} - \widehat{v}_{i-1,j})}{\sum_{i=1}^n (W_{i,j}(x) + W_{i,j}(-x))}.$$

et $\widetilde{a}_{n,j}$ est donné par (5.1.15).

5.2 Résultats de convergence paramétrique

5.2.1 Estimation de v

Théorème 5.2.1 Supposons les hypothèses (\mathcal{H}_1) à (\mathcal{H}_4) . Alors, on a la convergence presque sûre

$$\lim_{n \rightarrow +\infty} \widehat{v}_n = v \quad \text{p.s.} \quad (5.2.1)$$

et la normalité asymptotique

$$\sqrt{n} (\widehat{v}_n - v) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, \Gamma(v)), \quad (5.2.2)$$

où $\Gamma(v)$ est la matrice de covariance donnée par

$$\Gamma(v) = \text{Cov}\left(\frac{Y}{g(X)}\right).$$

De plus, on a la forte quadratique

$$\lim_{n \rightarrow +\infty} \frac{1}{\log(n)} \sum_{i=1}^n (\widehat{v}_i - v) (\widehat{v}_i - v)^T = \Gamma(v) \quad \text{p.s.} \quad (5.2.3)$$

5.2.2 Estimation de θ

Théorème 5.2.2 *Supposons les hypothèses (\mathcal{H}_1) à (\mathcal{H}_4) . Alors, $\widehat{\theta}_n$ converge presque sûrement vers θ . De plus, pour tout $1 \leq j \leq p$, le nombre de fois que la variable aléatoire $\widehat{\theta}_{n,j} + \gamma_{n+1} \text{sign}(a_j f_1) T_{n+1,j}$ sort de K est fini presque sûrement.*

Afin d'établir la normalité asymptotique de $\widehat{\theta}_n$, nous avons besoin d'introduire la fonction auxiliaire φ définie, pour tout $t \in \mathbb{R}^p$, par

$$\varphi(t) = \mathbb{E}[V(t)V(t)^T] \quad (5.2.4)$$

où $V(t)$ est donné par

$$V(t) = \text{diag}\left(\text{sign}(a_1 f_1), \dots, \text{sign}(a_p f_1)\right) D(X, t) Y.$$

Dès que $4\pi|f_1| \min_{1 \leq j \leq p} |a_j| > 1$, on note pour tout $1 \leq k, l \leq p$,

$$\Sigma(\theta)_{k,l} = \frac{\varphi(\theta)_{k,l}}{2\pi(|a_k| + |a_l|)|f_1| - 1}.$$

Théorème 5.2.3 *Supposons les hypothèses (\mathcal{H}_1) à (\mathcal{H}_4) . De plus, supposons que $(\varepsilon_{i,j})$ a un moment fini d'ordre > 2 et que*

$$4\pi|f_1| \min_{1 \leq j \leq p} |a_j| > 1.$$

Alors, nous avons la normalité asymptotique

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, \Sigma(\theta)). \quad (5.2.5)$$

Dans le théorème qui suit, on note e_j le j -ème vecteur de la base canonique de \mathbb{R}^p .

Théorème 5.2.4 *Supposons les hypothèses (\mathcal{H}_1) à (\mathcal{H}_4) . De plus, supposons que $(\varepsilon_{i,j})$ a un moment d'ordre > 2 fini et que*

$$4\pi|f_1| \min_{1 \leq j \leq p} |a_j| > 1.$$

Alors, on a la loi du log-itéré vectorielle donnée, pour tout $1 \leq j \leq p$, par

$$\begin{aligned} \limsup_{n \rightarrow \infty} \left(\frac{n}{2 \log \log n} \right)^{1/2} e_j^T (\widehat{\theta}_n - \theta) &= - \liminf_{n \rightarrow \infty} \left(\frac{n}{2 \log \log n} \right)^{1/2} e_j^T (\widehat{\theta}_n - \theta) \\ &= \sqrt{e_j^T \Sigma(\theta) e_j} \quad \text{p.s.} \end{aligned} \quad (5.2.6)$$

En particulier,

$$\limsup_{n \rightarrow \infty} \left(\frac{n}{2 \log \log n} \right) \|\widehat{\theta}_n - \theta\|^2 \leq \text{Tr}(\Sigma(\theta)) \quad \text{p.s.} \quad (5.2.7)$$

De plus, on a la loi forte quadratique vectorielle

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{i=1}^n (\widehat{\theta}_i - \theta)(\widehat{\theta}_i - \theta)^T = \Sigma(\theta) \quad \text{p.s.} \quad (5.2.8)$$

Remarque 5.2.1 De manière similaire à la Remarque 4.2.2, dans le cas particulier où

$$4\pi |f_1| \min_{1 \leq j \leq p} |a_j| = 1,$$

il est aussi possible de montrer, en utilisant le Théorème 2.2.12 page 52 de [23], que

$$\sqrt{\frac{n}{\log(n)}} (\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \varphi(\theta)).$$

On peut également avoir des résultats asymptotiques si

$$0 < 4\pi |f_1| \min_{1 \leq j \leq p} |a_j| < 1.$$

Plus précisément, sous cette hypothèse, en notant $\tau = 2\pi |f_1| \min_{1 \leq j \leq p} |a_j|$, on a

$$n^\tau (\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} Z,$$

où Z est un vecteur aléatoire fini presque-sûrement.

Remarque 5.2.2 La différentielle de ϕ est donnée par

$$D\phi(t) = -2\pi f_1 \text{diag}\left(a_1 \cos(2\pi(\theta_1 - t_1)), \dots, a_p \cos(2\pi(\theta_p - t_p))\right).$$

Par conséquent,

$$D\phi(\theta) = -2\pi f_1 \text{diag}(a_1, \dots, a_p).$$

D'une part, si le coefficient de Fourier f_1 de f et le vecteur a sont connus avec toutes les composantes de a non nulles, il est possible d'obtenir un estimateur $\widehat{\theta}_n$ de θ asymptotiquement efficace, via une petite modification de (5.1.9). Plus précisément, pour tout $1 \leq j \leq p$, il est nécessaire de remplacer le pas $\gamma_n = 1/n$ dans (5.1.9) par $\gamma_n = \gamma_j/n$ où

$$\gamma_j = \frac{1}{2\pi a_j f_1}.$$

On déduit alors de la partie 10.2.2 du livre de Kushner et Yin [40] page 331 que $\widehat{\theta}_n$ est un estimateur de θ asymptotiquement efficace avec

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \ell(\theta)). \quad (5.2.9)$$

où, pour tout $1 \leq k, l \leq p$,

$$\ell(\theta)_{k,l} = \frac{\varphi(\theta)_{k,l}}{4\pi^2 |f_1|^2 |a_k a_l|},$$

avec $\varphi(\theta)$ donnée par (5.2.4). D'autre part, si f_1 et a sont inconnus, il est aussi possible d'obtenir un estimateur $\widehat{\theta}_n$ de θ asymptotiquement efficace en remplaçant f_1 par son estimateur $\widehat{f}_{1,n}$ donné par (5.1.1) et en remplaçant a_j par son estimateur $\widetilde{a}_{n,j}$ défini par (5.1.15).

Remarque 5.2.3 Comme dans le Chapitre 4, on peut également s'affranchir de l'hypothèse de parité de f . Pour ce faire, on a besoin de connaître les coefficients de Fourier de f

$$f_1 = \int_{-1/2}^{1/2} \cos(2\pi x) f(x) dx \quad \text{et} \quad g_1 = \int_{-1/2}^{1/2} \sin(2\pi x) f(x) dx.$$

D'une part, si $f_1 \neq 0$ ou $g_1 \neq 0$, alors on remplace dans (5.1.4) la matrice diagonale $D(X, t)$ définie par (5.1.5), par

$$\Delta(X, t) = \frac{1}{g(X)} \text{diag}(\delta(X, t_1), \dots, \delta(X, t_p)), \quad (5.2.10)$$

où, pour tout $1 \leq j \leq p$,

$$\delta(X, t_j) = f_1 \sin(2\pi(X - t_j)) - g_1 \cos(2\pi(X - t_j)). \quad (5.2.11)$$

Ainsi, le Théorème 5.2.2 est encore vrai pour l'algorithme de Robbins-Monro projeté défini, pour tout $1 \leq j \leq p$, par

$$\widehat{\theta}_{n+1,j} = \pi_K(\widehat{\theta}_{n,j} + \text{sign}(a_j) \gamma_{n+1} T_{n+1,j}),$$

où la valeur initiale $\widehat{\theta}_0 \in K^p$ et où le vecteur aléatoire T_{n+1} est donné par

$$T_{n+1} = \Delta(X_{n+1}, \widehat{\theta}_n) \begin{pmatrix} Y_{n+1,1} \\ \vdots \\ Y_{n+1,p} \end{pmatrix}.$$

D'autre part, on remplace également la deuxième fonction auxiliaire φ définie par (5.2.4), par

$$\Psi(t) = \mathbb{E}[W(t)W(t)^T],$$

où $W(t)$ est donnée par

$$W(t) = \text{diag}(\text{sign}(a_1 f_1), \dots, \text{sign}(a_p f_1)) \Delta(X, t) Y.$$

Ainsi, dès que $4\pi(f_1^2 + g_1^2) \min_{1 \leq j \leq p} |a_j| > 1$, les Théorèmes 5.2.3 et 5.2.4 sont encore valables avec $\Sigma(\theta)$ donnée pour tout $1 \leq k, l \leq p$, par

$$\Sigma(\theta)_{k,l} = \frac{\Psi(\theta)_{k,l}}{2\pi(f_1^2 + g_1^2)(|a_k| + |a_l|) - 1}.$$

Si on ne connaît pas f_1 et g_1 , il suffit de remplacer leurs expressions dans (5.2.11) par leurs estimateurs respectifs $\widehat{f}_{1,n}$ donné par (5.1.1) et $\widehat{g}_{1,n}$ donné par

$$\widehat{g}_{1,n} = \frac{1}{n} \sum_{i=1}^n \frac{\sin(2\pi X_i)}{g(X_i)} Y_{i,1}. \quad (5.2.12)$$

5.2.3 Estimation de a

Afin d'établir les résultats de convergence de \widehat{a}_n et \widetilde{a}_n , on a besoin d'introduire la matrice carrée M_p , définie par

$$M_p = I_p - ae_1^T. \quad (5.2.13)$$

où I_p est la matrice identité d'ordre p et e_1 est le premier vecteur de la base canonique de \mathbb{R}^p . On a alors le résultat suivant.

Théorème 5.2.5 *Supposons les hypothèses (\mathcal{H}_1) à (\mathcal{H}_4) . Alors, on a les convergences presque sûres*

$$\lim_{n \rightarrow +\infty} \widehat{a}_n = a \quad \text{p.s.} \quad (5.2.14)$$

et

$$\lim_{n \rightarrow +\infty} \widetilde{a}_n = a \quad \text{p.s.} \quad (5.2.15)$$

ainsi que les normalités asymptotiques

$$\sqrt{n}(\widehat{a}_n - a) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, \Gamma(a)), \quad (5.2.16)$$

et

$$\sqrt{n}(\widetilde{a}_n - a) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, M_p \Gamma(a) M_p^T), \quad (5.2.17)$$

où $\Gamma(a)$ est la matrice de covariance définie par

$$\Gamma(a) = \frac{1}{f_1^2} \text{Cov}(C(X, \theta)Y). \quad (5.2.18)$$

De plus, nous avons les lois fortes quadratiques

$$\lim_{n \rightarrow +\infty} \frac{1}{\log(n)} \sum_{i=1}^n (\widehat{a}_i - a)(\widehat{a}_i - a)^T = \Gamma(a) \quad \text{p.s.} \quad (5.2.19)$$

et

$$\lim_{n \rightarrow +\infty} \frac{1}{\log(n)} \sum_{i=1}^n (\widetilde{a}_i - a)(\widetilde{a}_i - a)^T = M_p \Gamma(a) M_p^T \quad \text{p.s.} \quad (5.2.20)$$

5.3 Résultats de convergence non paramétrique

Théorème 5.3.1 *Supposons les hypothèses (\mathcal{H}_1) à (\mathcal{H}_5) . Supposons de plus que $(\varepsilon_{i,j})$ a un moment d'ordre > 2 fini. Alors, pour tout $x \in [-1/2; 1/2]$, on a*

$$\lim_{n \rightarrow \infty} \widehat{f}_n(x) = f(x) \quad \text{p.s.} \quad (5.3.1)$$

Théorème 5.3.2 *Supposons les hypothèses (\mathcal{H}_1) à (\mathcal{H}_5) . Supposons de plus que $(\varepsilon_{i,j})$ a un moment d'ordre > 2 fini. Si la fenêtre de lissage (h_n) est telle que $h_n = 1/n^\alpha$ avec $\alpha > 1/3$, alors on a, pour tout $x \in [-1/2; 1/2]$ avec $x \neq 0$, la normalité asymptotique ponctuelle*

$$\sqrt{nh_n}(\widehat{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\nu^2}{1+\alpha} \sum_{j=1}^p \frac{\sigma_j^2 \omega_j^2(x)}{a_j^2 (g(\theta_j + x) + g(\theta_j - x))}\right). \quad (5.3.2)$$

De plus, si $x = 0$, alors

$$\sqrt{nh_n}(\widehat{f}_n(0) - f(0)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\nu^2}{1+\alpha} \sum_{j=1}^p \frac{\sigma_j^2 \omega_j^2(0)}{a_j^2 g(\theta_j)}\right). \quad (5.3.3)$$

Remarque 5.3.1 *D'après la Remarque 5.1.1, si f_1 est inconnu, on remplace $\widehat{f}_n(x)$ défini par (5.1.16) par $\widetilde{f}_n(x)$ donné par (5.1.20). Les Théorèmes 5.3.1 et 5.3.2 sont alors à nouveau valables pour $\widetilde{f}_n(x)$ avec les mêmes variances asymptotiques.*

Remarque 5.3.2 *Le choix des poids $\omega_j(x)$ peut être important. Intuitivement, les variances asymptotiques données par (5.3.2) et (5.3.3) sont minimales si pour $1 \leq j \leq p$, $\omega_j(x)$ est inversement proportionnelle à la variance du bruit σ_j^2 . Plus précisément, le Théorème des extrema liés nous donne les valeurs des poids $\omega_j(x)$ pour lesquels les variances asymptotiques des équations (5.3.2) et (5.3.3) sont minimales sous la contrainte (5.1.17). Ces poids sont donnés, pour tout $1 \leq j \leq p$ et pour tout $x \in [-1/2; 1/2]$, par*

$$\omega_j(x) = \frac{m_j(x)}{\sum_{k=1}^p m_k(x)}$$

où

$$m_j(x) = \frac{a_j^2 (g(\theta_j + x) + g(\theta_j - x))}{\sigma_j^2}.$$

Par conséquent, si l'on suppose que les variances σ_j^2 sont connues et si l'on utilise les poids donnés, pour tout $1 \leq j \leq p$ et pour tout $x \in [-1/2; 1/2]$, par

$$\widehat{\omega}_j(x) = \frac{\widehat{m}_j(x)}{\sum_{k=1}^p \widehat{m}_k(x)}$$

où

$$\widehat{m}_j(x) = \frac{\widehat{a}_{n,j}^2 (g(\widehat{\theta}_{n,j} + x) + g(\widehat{\theta}_{n,j} - x))}{\sigma_j^2},$$

les variances asymptotiques des équations (5.3.2) et (5.3.3) sont respectivement données, pour $x \neq 0$, par

$$\frac{\nu^2}{1 + \alpha} \left(\sum_{j=1}^p m_j(x) \right)^{-1},$$

et pour $x = 0$, par

$$\frac{\nu^2}{1 + \alpha} \left(\sum_{j=1}^p \frac{m_j(0)}{2} \right)^{-1}.$$

5.4 Preuves des résultats paramétriques

5.4.1 Preuve du Théorème 5.2.1.

Les convergences (5.2.1) et (5.2.2) s'obtiennent immédiatement en utilisant la loi forte des grands nombres et le théorème central limite pour les martingales dont les accroissements sont indépendants. De plus, on déduit (5.2.3) du Théorème 2.1 du papier de Chaabane et Maaouia [11]. \square

5.4.2 Preuve du Théorème 5.2.2.

Le résultat se déduit immédiatement du Théorème 4.2.1. \square

5.4.3 Preuve du Théorème 5.2.3.

La preuve qui suit est une version vectorielle de la preuve du Théorème 4.2.2. Nous allons appliquer le Théorème 2.1 page 330 de Kushner et Yin [40]. Tout d'abord, $\gamma_n = 1/n$, donc les conditions sur le pas sont satisfaites. De plus, on sait que $\widehat{\theta}_n$ converge presque sûrement vers θ . Par conséquent, les hypothèses locales du Théorème 2.1 de Kushner et Yin [40] sont vérifiées. On déduit de (5.1.10) que

$$\mathbb{E}[T_{n+1} | \mathcal{F}_n] = \phi(\widehat{\theta}_n) \quad \text{p.s.}$$

et la fonction ϕ est continûment différentiable. Ainsi, $\phi(\theta) = 0$ et $D\phi(\theta)$ est la matrice diagonale définie par

$$D\phi(\theta) = -2\pi f_1 \text{diag}(a_1, \dots, a_p).$$

De plus, la condition $4\pi|f_1| \min_{1 \leq j \leq p} |a_j| > 1$ implique immédiatement que la matrice

$$D\phi(\theta) + \frac{1}{2}I_p$$

est une matrice définie négative. On a aussi que, pour tout $1 \leq k, l \leq p$,

$$\mathbb{E}[\text{sign}(a_k f_1) T_{n+1,k} \text{sign}(a_l f_1) T_{n+1,l} | \mathcal{F}_n] = \varphi(\widehat{\theta}_n)_{k,l} \quad \text{p.s.}$$

ce qui mène à

$$\lim_{n \rightarrow \infty} \mathbb{E}[\text{sign}(a_k f_1) T_{n+1,k} \text{sign}(a_l f_1) T_{n+1,l} | \mathcal{F}_n] = \varphi(\theta)_{k,l} \quad \text{p.s.}$$

Par conséquent, il suffit d'établir la tension de la suite (W_n) définie par

$$W_n = \frac{\|\widehat{\theta}_n - \theta\|^2}{\gamma_n},$$

afin d'appliquer le Théorème 2.1 de [40] et d'obtenir la normalité asymptotique

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, \Sigma(\theta))$$

où, pour tout $1 \leq k, l \leq p$,

$$\Sigma(\theta)_{k,l} = \varphi(\theta)_{k,l} \int_0^{+\infty} \exp\left(\left(1 - 2\pi|f_1|(|a_k| + |a_l|)\right)t\right) dt = \frac{\varphi(\theta)_{k,l}}{2\pi|f_1|(|a_k| + |a_l|) - 1}.$$

Nous procédons donc à la preuve de la tension de la suite (W_n) . Soit (V_n) la suite définie, pour tout $n \geq 1$, par

$$V_n = \|\widehat{\theta}_n - \theta\|^2, \quad (5.4.1)$$

et soit (T'_n) la suite de vecteurs aléatoires de \mathbb{R}^p définie, pour tout $n \geq 1$ et pour tout $1 \leq j \leq p$, par

$$T'_{n,j} = \text{sign}(a_j f_1) T_{n,j}. \quad (5.4.2)$$

Alors, on a clairement que

$$\begin{aligned} V_{n+1} &= \|\widehat{\theta}_{n+1} - \theta\|^2 \\ &= \|\pi_{K^p}(\widehat{\theta}_n + \gamma_{n+1} T'_{n+1}) - \theta\|^2 \\ &= \|\pi_{K^p}(\widehat{\theta}_n + \gamma_{n+1} T'_{n+1}) - \pi_{K^p}(\theta)\|^2 \\ &\leq \|\widehat{\theta}_n + \gamma_{n+1} T'_{n+1} - \theta\|^2 \end{aligned}$$

puisque $\pi_{K^p} = (\pi_K, \dots, \pi_K)^T$ est une fonction Lipschitzienne. Il s'ensuit que

$$V_{n+1} \leq V_n + \gamma_{n+1}^2 \|T'_{n+1}\|^2 + 2\gamma_{n+1} \langle \widehat{\theta}_n - \theta, T'_{n+1} \rangle \quad \text{p.s.}$$

En passant à l'espérance conditionnelle dans l'inégalité précédente, on obtient l'existence d'une constante $M > 0$ telle que

$$\mathbb{E}[V_{n+1} | \mathcal{F}_n] \leq V_n + \gamma_{n+1}^2 M + 2\gamma_{n+1} \langle \widehat{\theta}_n - \theta, \mathbb{E}[T'_{n+1} | \mathcal{F}_n] \rangle \quad \text{p.s.} \quad (5.4.3)$$

De plus, les équations (5.1.10) et (5.4.2) entraînent que

$$\mathbb{E}[T'_{n+1} | \mathcal{F}_n] = S_p(a) \phi(\widehat{\theta}_n), \quad (5.4.4)$$

où

$$S_p(a) = \text{diag}(\text{sign}(a_1 f_1), \dots, \text{sign}(a_p f_1)).$$

Ainsi, on déduit de (5.4.3) et de (5.4.4) que

$$\mathbb{E}[W_{n+1} | \mathcal{F}_n] \leq \frac{V_n}{\gamma_{n+1}} + \gamma_{n+1} M + 2 \langle \widehat{\theta}_n - \theta, S_p(a) \phi(\widehat{\theta}_n) \rangle \quad \text{p.s.} \quad (5.4.5)$$

Un développement de la fonction ϕ nous permet d'écrire la relation

$$\begin{aligned} \langle \widehat{\theta}_n - \theta, S_p(a)\phi(\widehat{\theta}_n) \rangle &= \langle \widehat{\theta}_n - \theta, 2\pi f_1 S_p(a) \text{diag}(a_1, \dots, a_p) (\theta - \widehat{\theta}_n) \rangle \\ &+ f_1 \langle \widehat{\theta}_n - \theta, S_p(a) \text{diag}(a_1, \dots, a_p) \mathcal{V}(\widehat{\theta}_n) (\theta - \widehat{\theta}_n) \rangle \end{aligned} \quad (5.4.6)$$

où, pour tout $t \neq \theta$,

$$\mathcal{V}(t) = \text{diag} \left(\frac{\sin(2\pi(\theta_1 - t_1)) - 2\pi(\theta_1 - t_1)}{\theta_1 - t_1}, \dots, \frac{\sin(2\pi(\theta_p - t_p)) - 2\pi(\theta_p - t_p)}{\theta_p - t_p} \right).$$

De plus, on a l'égalité

$$f_1 S_p(a) \text{diag}(a_1, \dots, a_p) = L(a),$$

où

$$L(a) = \text{diag}(|f_1 a_1|, \dots, |f_1 a_p|).$$

Avec l'équation (5.4.6), on obtient que

$$\langle \widehat{\theta}_n - \theta, S_p(a)\phi(\widehat{\theta}_n) \rangle = -2\pi (\widehat{\theta}_n - \theta)^T L(a) (\widehat{\theta}_n - \theta) - (\widehat{\theta}_n - \theta)^T L(a) \mathcal{V}(\widehat{\theta}_n) (\widehat{\theta}_n - \theta). \quad (5.4.7)$$

Ainsi, (5.4.5) devient

$$\begin{aligned} \mathbb{E}[W_{n+1} | \mathcal{F}_n] &\leq (1 + \gamma_n) W_n + \gamma_n M - 4\pi (\widehat{\theta}_n - \theta)^T L(a) (\widehat{\theta}_n - \theta) \\ &\quad - 2 (\widehat{\theta}_n - \theta)^T L(a) \mathcal{V}(\widehat{\theta}_n) (\widehat{\theta}_n - \theta). \end{aligned} \quad (5.4.8)$$

De plus, comme

$$L(a) \geq \min_{1 \leq j \leq p} |a_j f_1| I_p,$$

on a, d'après (5.4.8),

$$\mathbb{E}[W_{n+1} | \mathcal{F}_n] \leq W_n + 2q\gamma_n W_n + M\gamma_n - 2 (\widehat{\theta}_n - \theta)^T L(a) (\widehat{\theta}_n - \theta), \quad (5.4.9)$$

où

$$2q = 1 - 4\pi |f_1| \min_{1 \leq j \leq p} |a_j|,$$

ce qui signifie, par hypothèse, que $q < 0$. La fonction \mathcal{V} étant continue, on peut trouver $0 < \varepsilon < 1/2$ tel que, si $\|t - \theta\| < \varepsilon$, alors

$$\frac{q}{2|f_1| \min_{1 \leq j \leq p} |a_j|} I_p < \mathcal{V}(t) < 0. \quad (5.4.10)$$

De plus, on note A_n et B_n les ensembles $A_n = \{\|\widehat{\theta}_n - \theta\| \leq \varepsilon\}$ et

$$B_n = \bigcap_{k=m}^n A_k$$

avec $1 \leq m \leq n$. Alors, il s'ensuit de (5.4.10) que

$$0 < -2|f_1| \min_{1 \leq j \leq p} |a_j| \mathcal{V}(\widehat{\theta}_n) I_{B_n} < -\left(\frac{q}{2}\right) I_p I_{B_n}. \quad (5.4.11)$$

Ainsi, on déduit de (5.4.9) et (5.4.11) que, pour tout $n \geq m$,

$$\begin{aligned} \mathbb{E}[W_{n+1}\mathbf{I}_{B_n}|\mathcal{F}_n] &\leq W_n\mathbf{I}_{B_n} + 2\gamma_n W_n q \mathbf{I}_{B_n} - q\gamma_n W_n \mathbf{I}_{B_n} + \gamma_n M, \\ &\leq W_n\mathbf{I}_{B_n}(1 + q\gamma_n) + \gamma_n M. \end{aligned} \quad (5.4.12)$$

De plus, puisque $B_{n+1} = B_n \cap A_{n+1}$ alors $B_{n+1} \subset B_n$, et en passant à l'espérance dans (5.4.12), on obtient que, pour tout $n \geq m$,

$$\mathbb{E}[W_{n+1}\mathbf{I}_{B_{n+1}}] \leq (1 + q\gamma_n)\mathbb{E}[W_n\mathbf{I}_{B_n}] + \gamma_n M. \quad (5.4.13)$$

Finalement, en suivant exactement la même preuve que celle du Théorème 4.2.2, on obtient que pour tout $\xi > 0$, il existe $K > 0$ tel que pour m suffisamment grand,

$$\sup_{n \geq m} \mathbb{P}(W_n > K) < \xi$$

ce qui permet de conclure à la tension de la suite (W_n) et termine la preuve du Théorème 5.2.3. \square

5.4.4 Preuve du Théorème 5.2.4.

La différentielle de ϕ en θ est la matrice diagonale

$$-2\pi f_1 \text{diag}(a_1, \dots, a_p).$$

Les vecteurs propres de cette matrice sont les vecteurs de la base canonique de \mathbb{R}^p . Ainsi, la loi du log-itéré vectorielle donnée par (5.2.6) s'obtient grâce au Théorème 1.3.2 alors que la loi forte quadratique vectorielle donnée par (5.2.8) s'obtient grâce au Théorème 1.3.3. \square

5.4.5 Preuve du Théorème 5.2.5.

On rappelle que, pour tout $1 \leq j \leq p$,

$$\widehat{a}_{n,j} = \frac{1}{nf_1} \sum_{i=1}^n \frac{\cos(2\pi(X_i - \widehat{\theta}_{i-1,j}))}{g(X_i)} Y_{i,j}$$

et

$$\widetilde{a}_{n,j} = \frac{1}{n\widehat{f}_{1,n}} \sum_{i=1}^n \frac{\cos(2\pi(X_i - \widehat{\theta}_{i-1,j}))}{g(X_i)} Y_{i,j}.$$

Ainsi, il est clair que

$$\widehat{a}_n = \frac{1}{nf_1} \sum_{i=1}^n C(X_i, \widehat{\theta}_{i-1}) Y_i,$$

et

$$\widetilde{a}_n = \frac{1}{n\widehat{f}_{1,n}} \sum_{i=1}^n C(X_i, \widehat{\theta}_{i-1}) Y_i,$$

où on a noté

$$Y_i = \begin{pmatrix} Y_{i,1} \\ \vdots \\ Y_{i,p} \end{pmatrix}.$$

Nous avons aussi les décompositions

$$\widehat{a}_n - a = \frac{1}{nf_1} S_n(a) + \frac{1}{nf_1} R_n(a), \quad (5.4.14)$$

et

$$\widetilde{a}_n - a = \frac{1}{n\widehat{f}_{1,n}} (S_n(a) + (f_1 - \widehat{f}_{1,n})a) + \frac{1}{n\widehat{f}_{1,n}} R_n(a), \quad (5.4.15)$$

avec

$$S_n(a) = \sum_{i=1}^n (C(X_i, \theta) Y_i - f_1 a),$$

et

$$R_n(a) = \sum_{i=1}^n (C(X_i, \widehat{\theta}_{i-1}) - C(X_i, \theta)) Y_i.$$

De plus,

$$\begin{aligned} (f_1 - \widehat{f}_{1,n}) a &= \sum_{i=1}^n \left(f_1 - \frac{\cos(2\pi X_i)}{g(X_i)} Y_{i,1} \right) a, \\ &= -e_1^T S_n(a) a \end{aligned}$$

Ainsi, on déduit immédiatement de (5.4.15) que

$$\widetilde{a}_n - a = \frac{1}{n\widehat{f}_{1,n}} M_p S_n(a) + \frac{1}{n\widehat{f}_{1,n}} R_n(a), \quad (5.4.16)$$

où la matrice M_p est donnée par (5.2.13). De plus, pour tout $1 \leq j \leq p$,

$$R_{n,j}(a) = a_j R_{n,j}^1(a) + v_j R_{n,j}^2(a) + R_{n,j}^3(a), \quad (5.4.17)$$

où

$$R_{n,j}^1(a) = \sum_{i=1}^n \frac{\Delta c_{i,j}}{g(X_i)} f(X_i - \theta_j), \quad R_{n,j}^2(a) = \sum_{i=1}^n \frac{\Delta c_{i,j}}{g(X_i)}, \quad R_{n,j}^3(a) = \sum_{i=1}^n \frac{\Delta c_{i,j}}{g(X_i)} \varepsilon_{i,j},$$

et

$$\Delta c_{i,j} = \cos(2\pi(X_i - \widehat{\theta}_{i-1,j})) - \cos(2\pi(X_i - \theta_j)).$$

Tout d'abord, puisque

$$\mathbb{E}[C(X_i, \theta) Y_i | \mathcal{F}_{i-1}] = f_1 a,$$

la suite $(S_n(a))$ est une martingale vectorielle à accroissements indépendants et de carré intégrable. Pour tout $n \geq 1$, son crochet $\langle S(a) \rangle_n$ est donné par

$$\begin{aligned} \langle S(a) \rangle_n &= \sum_{i=1}^n \mathbb{E}[(C(X_i, \theta) Y_i - f_1 a) (C(X_i, \theta) Y_i - f_1 a)^T | \mathcal{F}_{i-1}] \\ &= \sum_{i=1}^n \text{Cov}(C(X_i, \theta) Y_i | \mathcal{F}_{i-1}) \end{aligned}$$

Ainsi, on a

$$\lim_{n \rightarrow +\infty} \frac{\langle S(a) \rangle_n}{n} = \Gamma(a) \quad \text{p.s.}$$

où $\Gamma(a)$ est donnée par (5.2.18). Ensuite, comme

$$\begin{aligned} \mathbb{E}\left[\frac{\Delta c_{i,j}}{g(X_i)}|\mathcal{F}_{i-1}\right] &= \int_{-1/2}^{1/2} \cos(2\pi(x - \widehat{\theta}_{i-1,j})) dx - \int_{-1/2}^{1/2} \cos(2\pi(x - \theta_j)) dx \\ &= 0, \end{aligned}$$

la suite $(R_{n,j}^2(a))$ est une martingale de carré intégrable dont le crochet est donné par

$$\langle R_j^2(a) \rangle_n = \sum_{i=1}^n \mathbb{E}\left[\frac{\Delta c_{i,j}^2}{g^2(X_i)}|\mathcal{F}_{i-1}\right].$$

De plus, comme la fonction cosinus est une fonction lipschitzienne, on a

$$|\Delta c_{i,j}| \leq |\widehat{\theta}_{i-1,j} - \theta_j|.$$

Par conséquent, g ne s'annulant pas sur $[-1/2; 1/2]$, il existe une constante $C > 0$ telle que

$$\mathbb{E}\left[\frac{\Delta c_{i,j}^2}{g^2(X_i)}|\mathcal{F}_{i-1}\right] \leq C (\widehat{\theta}_{i-1,j} - \theta_j)^2. \quad (5.4.18)$$

Ainsi, on déduit de la loi forte quadratique (5.2.8) et de l'inégalité précédente (5.4.18) que

$$\langle R_j^2(a) \rangle_n = \mathcal{O}(\log(n)) \quad \text{p.s.} \quad (5.4.19)$$

De plus, la loi forte des grands nombres pour les martingales, donnée par exemple par le Théorème 1.3.15 de [23], entraîne que, pour tout $1 \leq j \leq p$,

$$R_{n,j}^2(a) = o(\log(n)) \quad \text{p.s.} \quad (5.4.20)$$

La suite $(R_{n,j}^3(a))$ est encore une martingale de carré intégrable dont le crochet est donné par

$$\langle R_j^3(a) \rangle_n = \sigma_j^2 \langle R_j^2(a) \rangle_n.$$

Ainsi, on déduit immédiatement de (5.4.19) que

$$\langle R_j^3(a) \rangle_n = \mathcal{O}(\log(n)) \quad \text{p.s.}, \quad (5.4.21)$$

et de la loi forte des grands nombres pour les martingales que, pour tout $1 \leq j \leq p$,

$$R_{n,j}^3(a) = o(\log(n)) \quad \text{p.s.} \quad (5.4.22)$$

Ensuite, pour tout $1 \leq j \leq p$, le changement de variable $u = x - \theta_j$ entraîne que

$$\begin{aligned} \mathbb{E}\left[\frac{\Delta c_{i,j}}{g(X_i)} f(X_i - \theta_j) |\mathcal{F}_{i-1}\right] &= \int_{-1/2}^{1/2} (\cos(2\pi(x - \widehat{\theta}_{i-1,j})) - \cos(2\pi(x - \theta_j))) f(x - \theta_j) dx \\ &= \int_{-1/2-\theta_j}^{1/2-\theta_j} (\cos(2\pi(u + \theta_j - \widehat{\theta}_{i-1,j})) - \cos(2\pi u)) f(u) du. \end{aligned}$$

L'égalité trigonométrique élémentaire

$$\cos(2\pi(u + \theta_j - \widehat{\theta}_{i-1,j})) = \cos(2\pi u) \cos(2\pi(\theta_j - \widehat{\theta}_{i-1,j})) - \sin(2\pi u) \sin(2\pi(\theta_j - \widehat{\theta}_{i-1,j})),$$

la symétrie et la périodicité de f impliquent alors que

$$\mathbb{E}\left[\frac{\Delta c_{i,j}}{g(X_i)} f(X_i - \theta_j) | \mathcal{F}_{i-1}\right] = f_1(\cos(2\pi(\theta_j - \widehat{\theta}_{i-1,j})) - 1) \quad \text{p.s.}$$

De plus, pour tout $|x| < 1/2$, on a

$$|\cos(2\pi x) - 1| \leq 2\pi^2 x^2,$$

ce qui conduit à

$$\left| \mathbb{E}\left[\frac{\Delta c_{i,j}}{g(X_i)} f(X_i - \theta_j) | \mathcal{F}_{i-1}\right] \right| \leq 2\pi^2 (\theta_j - \widehat{\theta}_{i-1,j})^2 \quad \text{p.s.} \quad (5.4.23)$$

On a également la décomposition

$$R_{n,j}^1(a) = A_{n,j}(a) + B_{n,j}(a),$$

avec

$$A_{n,j}(a) = \sum_{i=1}^n \left(\frac{\Delta c_{i,j}}{g(X_i)} f(X_i - \theta_j) - \mathbb{E}\left[\frac{\Delta c_{i,j}}{g(X_i)} f(X_i - \theta_j) | \mathcal{F}_{i-1}\right] \right),$$

et

$$B_{n,j}(a) = \sum_{i=1}^n \mathbb{E}\left[\frac{\Delta c_{i,j}}{g(X_i)} f(X_i - \theta_j) | \mathcal{F}_{i-1}\right].$$

Il s'ensuit à nouveau de la loi forte quadratique (5.2.8) et de (5.4.23) que, pour tout $1 \leq j \leq p$,

$$B_{n,j}(a) = \mathcal{O}(\log(n)) \quad \text{p.s.} \quad (5.4.24)$$

De plus, pour tout $1 \leq j \leq p$, $(A_{n,j}(a))$ est une martingale de carré intégrable de processus croissant $\langle A_j(a) \rangle_n$ satisfaisant

$$\langle A_j(a) \rangle_n \leq \sum_{i=1}^n \mathbb{E}\left[\frac{\Delta c_{i,j}^2}{g(X_i)^2} f^2(X_i - \theta_j) | \mathcal{F}_{i-1}\right].$$

La fonction f étant bornée, on déduit de (5.2.8) et (5.4.18) que

$$\langle A_j(a) \rangle_n = \mathcal{O}(\log(n)) \quad \text{p.s.}$$

Il découle alors de la loi forte quadratique pour les martingales que, pour tout $1 \leq j \leq p$,

$$A_{n,j}(a) = o(\log(n)) \quad \text{p.s.} \quad (5.4.25)$$

Ainsi, les équations (5.4.20), (5.4.22), (5.4.24) et (5.4.25) impliquent que, pour tout $1 \leq j \leq p$,

$$R_{n,j}(a) = \mathcal{O}(\log(n)) \quad \text{p.s.} \quad (5.4.26)$$

Finalement, on obtient de (5.4.14) que

$$\widehat{a}_n - a = \frac{1}{nf_1} S_n(a) + \mathcal{O}\left(\frac{\log(n)}{n}\right) \quad \text{p.s.} \quad (5.4.27)$$

et de (5.4.16) que

$$\tilde{a}_n - a = \frac{1}{n\widehat{f}_{1,n}} M_p S_n(a) + \mathcal{O}\left(\frac{\log(n)}{n}\right) \quad \text{p.s.} \quad (5.4.28)$$

Par conséquent, comme $\widehat{f}_{1,n}$ converge presque sûrement vers f_1 , (5.2.14) et (5.2.15) se déduisent de la loi forte des grands nombres pour les martingales alors que (5.2.16) et (5.2.17) se déduisent du théorème central limite pour les martingales et du lemme de Slutsky. Enfin, on obtient (5.2.19) et (5.2.20) grâce au Théorème 2.1 de [11]. \square

5.5 Preuves des résultats non paramétriques

5.5.1 Preuve du Théorème 5.3.1.

Pour $x \in [-1/2; 1/2]$, on note $(\check{f}_{n,j}(x))$ la suite définie pour tout $n \geq 1$ et $1 \leq j \leq p$, par

$$\check{f}_{n,j}(x) = \widehat{a}_{n,j} \widehat{f}_{n,j}(x). \quad (5.5.1)$$

On peut écrire la décomposition

$$\check{f}_{n,j}(x) = \check{f}_{n,j}^1(x) + \check{f}_{n,j}^2(x), \quad (5.5.2)$$

où

$$\check{f}_{n,j}^1(x) = \frac{\sum_{i=1}^n (W_{i,j}(x) + W_{i,j}(-x)) (Y_{i,j} - v_j)}{\sum_{i=1}^n (W_{i,j}(x) + W_{i,j}(-x))},$$

et

$$\check{f}_{n,j}^2(x) = \frac{\sum_{i=1}^n (W_{i,j}(x) + W_{i,j}(-x)) (v_j - \widehat{v}_{i-1,j})}{\sum_{i=1}^n (W_{i,j}(x) + W_{i,j}(-x))}.$$

D'une part, il découle du Théorème 4.3.1 que pour tout $x \in [-1/2; 1/2]$,

$$\lim_{n \rightarrow +\infty} \check{f}_{n,j}^1(x) = a_j f(x) \quad \text{p.s.}$$

D'autre part, puisque $\widehat{v}_{i-1,j}$ converge p.s. vers v_j quand i tend vers $+\infty$, le Lemme 1.3.2 de Toeplitz implique que pour tout $x \in [-1/2; 1/2]$,

$$\lim_{n \rightarrow +\infty} \check{f}_{n,j}^2(x) = 0 \quad \text{p.s.}$$

Ainsi, on obtient que

$$\lim_{n \rightarrow \infty} \check{f}_{n,j}(x) = a_j f(x) \quad \text{p.s.} \quad (5.5.3)$$

Par conséquent, comme $\widehat{a}_{n,j}$ converge p.s. vers $a_j \neq 0$ quand n tend vers $+\infty$, il s'ensuit que

$$\lim_{n \rightarrow \infty} \widehat{f}_{n,j}(x) = f(x) \quad \text{p.s.} \quad (5.5.4)$$

Finalement, les équations (5.1.16) et (5.5.4) nous permettent de conclure la preuve du Théorème 5.3.1. \square

5.5.2 Preuve du Théorème 5.3.2.

On s'intéresse maintenant à la normalité asymptotique de \widehat{f}_n . Pour tout $x \in [-1/2; 1/2]$, on a

$$\begin{aligned} \widehat{f}_n(x) - f(x) &= \sum_{j=1}^p \omega_j(x) (\widehat{f}_{n,j}(x) - f(x)) \\ &= \sum_{j=1}^p \omega_j(x) \frac{\mathcal{M}_{n,j}(x) + \mathcal{P}_{n,j}(x) + \mathcal{Q}_{n,j}(x) + \mathcal{R}_{n,j}(x) + \mathcal{S}_{n,j}(x)}{n\mathcal{G}_{n,j}(x)} \end{aligned} \quad (5.5.5)$$

où

$$\begin{aligned} \mathcal{G}_{n,j}(x) &= \widehat{a}_{n,j} (\widehat{g}_{n,j}(x) + \widehat{g}_{n,j}(-x)), \\ \mathcal{M}_{n,j}(x) &= M_{n,j}(x) + M_{n,j}(-x), \\ \mathcal{P}_{n,j}(x) &= P_{n,j}(x) + P_{n,j}(-x), \\ \mathcal{Q}_{n,j}(x) &= Q_{n,j}(x) + Q_{n,j}(-x), \\ \mathcal{R}_{n,j}(x) &= R_{n,j}(x) + R_{n,j}(-x), \\ \mathcal{S}_{n,j}(x) &= S_{n,j}(x) + S_{n,j}(-x), \end{aligned}$$

où $\widehat{g}_{n,j}(x)$, $M_{n,j}(x)$, $P_{n,j}(x)$, $Q_{n,j}(x)$, $R_{n,j}(x)$ et $S_{n,j}(x)$ sont définis par

$$\begin{aligned} \widehat{g}_{n,j}(x) &= \frac{1}{n} \sum_{i=1}^n W_{i,j}(x), \\ M_{n,j}(x) &= \sum_{i=1}^n W_{i,j}(x) \varepsilon_{i,j}, \\ P_{n,j}(x) &= \widehat{a}_{n,j} \sum_{i=1}^n W_{i,j}(x) (f(X_i - \widehat{\theta}_{i-1,j}) - f(x)), \\ Q_{n,j}(x) &= \widehat{a}_{n,j} \sum_{i=1}^n W_{i,j}(x) (f(X_i - \theta_j) - f(X_i - \widehat{\theta}_{i-1,j})), \\ R_{n,j}(x) &= (a_j - \widehat{a}_{n,j}) \sum_{i=1}^n W_{i,j}(x) f(X_i - \theta_j), \\ S_{n,j}(x) &= \sum_{i=1}^n W_{i,j}(x) (v_j - \widehat{v}_{i-1,j}). \end{aligned}$$

Tout d'abord, on déduit de (4.5.28) et de la convergence presque sûre de \widehat{a}_n vers a , que

$$\lim_{n \rightarrow +\infty} \mathcal{G}_{n,j}(x) = a_j (g(\theta_j + x) + g(\theta_j - x)) \quad \text{p.s.} \quad (5.5.6)$$

De plus, on déduit de (4.5.32) et de (4.5.35) que, pour $\alpha > 1/3$,

$$\mathcal{P}_{n,j}^2(x) = o(n^{1+\alpha}) \quad \text{p.s.}, \quad (5.5.7)$$

$$\mathcal{Q}_{n,j}^2(x) = o(n^{1+\alpha}) \quad \text{p.s.} \quad (5.5.8)$$

Ainsi, pour tout $x \in [-1/2; 1/2]$, on trouve que

$$\lim_{n \rightarrow \infty} \sqrt{\frac{h_n}{n}} \sum_{j=1}^p \omega_j(x) \frac{\mathcal{P}_{n,j}(x) + \mathcal{Q}_{n,j}(x)}{\mathcal{G}_{n,j}(x)} = 0 \quad \text{p.s.} \quad (5.5.9)$$

Ensuite, la fonction f étant bornée, on a

$$R_{n,j}(x) = \mathcal{O} \left(|a_j - \widehat{a}_{n,j}| \sum_{i=1}^n W_{i,j}(x) \right) \quad \text{p.s.}$$

Par suite,

$$\sum_{i=1}^n W_{i,j}(x) = \mathcal{O}(n) \quad \text{p.s.,}$$

ce qui assure que

$$R_{n,j}(x) = \mathcal{O}(n|a_j - \widehat{a}_{n,j}|) \quad \text{p.s.} \quad (5.5.10)$$

De plus, on obtient de (5.4.27) que, pour tout $1 \leq j \leq p$,

$$|a_j - \widehat{a}_{n,j}| = \mathcal{O} \left(\sqrt{\frac{\log(n)}{n}} \right) \quad \text{p.s.}$$

qui, grâce à (5.5.10), nous permet de dire que

$$R_{n,j}(x) = \mathcal{O} \left(\sqrt{n \log(n)} \right) \quad \text{p.s.} \quad (5.5.11)$$

Par conséquent,

$$\mathcal{R}_{n,j}^2(x) = o(n^{1+\alpha}) \quad \text{p.s.} \quad (5.5.12)$$

Puis, on a l'inégalité suivante

$$|S_{n,j}(x)| \leq \Lambda_{n,j}(x) + \Sigma_{n,j}(x) \quad (5.5.13)$$

où

$$\Lambda_{n,j}(x) = \sum_{i=1}^n \mathcal{L}_{i,j} (W_{i,j}(x) - \mathbb{E}[W_{i,j}(x)|\mathcal{F}_{i-1}])$$

et

$$\Sigma_{n,j}(x) = \sum_{i=1}^n \mathcal{L}_{i,j} \mathbb{E}[W_{i,j}(x)|\mathcal{F}_{i-1}]$$

avec $\mathcal{L}_{i,j} = |v_j - \widehat{v}_{i-1,j}|$. On déduit alors de (4.5.34), de l'inégalité de Cauchy-Schwarz et de la loi forte quadratique donnée par (5.2.3) que

$$\Sigma_{n,j}(x) = \mathcal{O} \left(\sqrt{n} \left(\sum_{i=1}^n \mathcal{L}_{i,j}^2 \right)^{1/2} \right) = \mathcal{O} \left(\sqrt{n \log(n)} \right) \quad \text{p.s.} \quad (5.5.14)$$

De plus, la suite $(\Lambda_{n,j}(x))$ est une martingale dont le processus croissant est donné par

$$\langle \Lambda_j(x) \rangle_n = \mathcal{O} \left(\sum_{i=1}^n \mathcal{L}_{i,j}^2 \mathbb{E}[W_{i,j}^2(x)|\mathcal{F}_{i-1}] \right) \quad \text{p.s.}$$

Ainsi, il découle à nouveau de la loi forte quadratique (5.2.3) que

$$\langle \Lambda(x) \rangle_{n,j} = \mathcal{O} \left(n^\alpha \sum_{i=1}^n \mathcal{L}_{i,j}^2 \right) = \mathcal{O} (n^\alpha \log(n)) \quad \text{p.s.} \quad (5.5.15)$$

ce qui, grâce à la loi forte des grands nombres pour les martingales, nous permet de dire que, pour tout $\gamma > 0$, on a

$$\Lambda_{n,j}^2(x) = o(n^\alpha \log(n)^{2+\gamma}) \quad \text{p.s.} \quad (5.5.16)$$

Alors,

$$\begin{aligned} \mathcal{S}_{n,j}^2(x) &= o(n^\alpha \log(n)^{2+\gamma}) + \mathcal{O}(n \log(n)) \\ &= o(n^{1+\alpha}) \quad \text{p.s.} \end{aligned} \quad (5.5.17)$$

A présent, on étudie le comportement asymptotique du terme dominant $\mathcal{M}_{n,j}(x)$. Pour tout $x \in [-1/2; 1/2]$ et tout $1 \leq j \leq p$, la suite $(\mathcal{M}_{n,j}(x))$ est une martingale de carré intégrable dont le crochet vaut

$$\langle \mathcal{M}_j(x) \rangle_n = \sigma_j^2 \sum_{i=1}^n \mathbb{E} \left[(W_{i,j}(x) + W_{i,j}(-x))^2 \mid \mathcal{F}_{i-1} \right].$$

On déduit alors de (4.5.37) que, pour $x \neq 0$,

$$\lim_{n \rightarrow \infty} \frac{\langle \mathcal{M}_j(x) \rangle_n}{n^{1+\alpha}} = \frac{\sigma_j^2 \nu^2}{1+\alpha} (g(\theta_j + x) + g(\theta_j - x)). \quad (5.5.18)$$

On obtient aussi de (4.5.38) que, pour $x = 0$,

$$\lim_{n \rightarrow \infty} \frac{\langle \mathcal{M}_j(0) \rangle_n}{n^{1+\alpha}} = 4 \frac{\sigma_j^2 \nu^2}{1+\alpha} g(\theta_j). \quad (5.5.19)$$

De plus, comme $(\varepsilon_{i,j})$ a un moment d'ordre > 2 fini, l'équation (4.5.39) nous permet de vérifier que la condition de Lindeberg est satisfaite pour $(\mathcal{M}_{n,j}(x))$. On peut donc conclure par le théorème central limite pour les martingales que, pour tout $x \in [-1/2; 1/2]$ et $x \neq 0$,

$$\frac{\mathcal{M}_{n,j}(x)}{\sqrt{n^{1+\alpha}}} \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\sigma_j^2 \nu^2}{1+\alpha} (g(\theta_j + x) + g(\theta_j - x)) \right), \quad (5.5.20)$$

tandis que, pour $x = 0$,

$$\frac{\mathcal{M}_{n,j}(0)}{\sqrt{n^{1+\alpha}}} \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, 4 \frac{\sigma_j^2 \nu^2}{1+\alpha} g(\theta_j) \right). \quad (5.5.21)$$

Finalement, il s'ensuit de (5.5.20) et (5.5.21), de l'indépendance de $\varepsilon_{i,1}, \dots, \varepsilon_{i,p}$ et du lemme de Slutsky que, pour tout $x \in [-1/2; 1/2]$ tel que $x \neq 0$,

$$\frac{1}{\sqrt{n^{1+\alpha}}} \sum_{j=1}^p \omega_j(x) \frac{\mathcal{M}_{n,j}(x)}{\mathcal{G}_{n,j}(x)} \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\nu^2}{1+\alpha} \sum_{j=1}^p \frac{\sigma_j^2 \omega_j^2(x)}{a_j^2 (g(\theta_j + x) + g(\theta_j - x))} \right), \quad (5.5.22)$$

alors que, pour $x = 0$,

$$\frac{1}{\sqrt{n^{1+\alpha}}} \sum_{j=1}^p \omega_j(0) \frac{\mathcal{M}_{n,j}(0)}{\mathcal{G}_{n,j}(0)} \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\nu^2}{1+\alpha} \sum_{j=1}^p \frac{\sigma_j^2 \omega_j^2(0)}{a_j^2 g(\theta_j)}\right). \quad (5.5.23)$$

Finalement, les équations (5.5.9), (5.5.12), (5.5.17) ainsi que les deux convergences précédentes (5.5.22) et (5.5.23) et le lemme de Slutsky nous permettent de conclure que, pour tout $x \in [-1/2; 1/2]$ avec $x \neq 0$,

$$\sqrt{nh_n}(\widehat{f}_n(x) - f(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\nu^2}{1+\alpha} \sum_{j=1}^p \frac{\sigma_j^2 \omega_j^2(x)}{a_j^2 (g(\theta_j + x) + g(\theta_j - x))}\right),$$

tandis que, pour $x = 0$,

$$\sqrt{nh_n}(\widehat{f}_n(0) - f(0)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\nu^2}{1+\alpha} \sum_{j=1}^p \frac{\sigma_j^2 \omega_j^2(0)}{a_j^2 g(\theta_j)}\right).$$

Ceci achève la preuve du Théorème 5.3.2. □

Chapitre 6

Déformation de variables aléatoires

Ce chapitre est consacré à l'étude du modèle de déformation de variables aléatoires (3.1.4), défini, pour tout $n \geq 0$, par

$$X_n = \varphi_\theta(\varepsilon_n). \quad (6.0.1)$$

Plus précisément, on suppose que l'on dispose d'une suite (ε_n) de variables aléatoires indépendantes et de même loi ainsi qu'une déformation φ_θ , dépendant d'un paramètre inconnu θ , produisant la suite d'observations (X_n) donnée par (6.0.1). Il est important de noter ici que (ε_n) est observable. On considère que, pour tout $t \in \mathbb{R}$, la déformation φ_t est connue et inversible et que la fonction $t \rightarrow \varphi_t$ n'est pas inversible (voir la Remarque 6.0.1 ci-dessous). Seul le paramètre θ est inconnu. Ce problème peut être vu comme un problème déterministe puisque si on dispose d'une suite déterministe ε_n et d'une suite x_n provenant du modèle

$$x_n = \varphi_\theta(\varepsilon_n)$$

alors, une méthode de gradient déterministe permet d'approcher θ en considérant, à x fixé, le gradient de la fonction

$$M_x(t) = (x - \varphi_t^{-1} \circ \varphi_\theta(x))^2.$$

Plus précisément, dans un intervalle où la fonction M_x est strictement concave, l'algorithme déterministe

$$\theta_{n+1} = \theta_n + \gamma_{n+1} \nabla M_{x_n}(\theta_n) \quad (6.0.2)$$

avec $\nabla M_{x_n}(\theta_n) = -2\partial\varphi_{\theta_n}^{-1}(x_n)(\varepsilon_n - \varphi_{\theta_n}(x_n))$, va converger vers θ quand le nombre d'itérations n va tendre vers l'infini. Cependant, nous allons voir que considérer (ε_n) comme une suite de variables aléatoires va nous permettre de caractériser de manière naturelle la fonction de contraste à utiliser. Plus précisément, on considère la variable aléatoire

$$Z_n(t) = \varphi_t^{-1}(X_n). \quad (6.0.3)$$

Alors, un critère naturel pour estimer θ est maintenant de minimiser en t la moyenne quadratique de la différence entre $Z_n(t)$ et ε_n

$$M(t) = \mathbb{E}[(Z_n(t) - \varepsilon_n)^2]. \quad (6.0.4)$$

On note ε et X des variables aléatoires de même loi que (ε_n) et (X_n) , respectivement. D'une part, on peut réécrire la variable aléatoire $Z_n(t)$ définie par (6.0.3) de la façon suivante

$$Z_n(t) = \varphi_t^{-1}(X_n) = \varphi_t^{-1}(\varphi_\theta(\varepsilon_n)). \quad (6.0.5)$$

On note aussi $Z(t)$ une variable aléatoire qui a la même loi que $Z_n(t)$. D'autre part, pour estimer θ , on se ramène donc au problème de minimiser en t le contraste défini par

$$M(t) = \mathbb{E}[(Z(t) - \varepsilon)^2]. \quad (6.0.6)$$

En notant F^{-1} la fonction quantile de ε , on peut donc écrire que

$$M(t) = \mathbb{E}[(\varphi_t^{-1}(\varphi_\theta(\varepsilon)) - \varepsilon)^2] = \int_0^1 (\varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) - F^{-1}(x))^2 dx.$$

Ainsi, si on suppose, que pour tout t , φ_t est croissante, alors la fonction quantile de $Z(t)$ vaut $F_{Z(t)}^{-1} = \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}$ et donc

$$M(t) = \int_0^1 (F_{Z(t)}^{-1}(x) - F^{-1}(x))^2 dx. \quad (6.0.7)$$

Cette quantité correspond à la distance de Wasserstein entre la loi de $Z(t)$ et de celle de ε , définie et étudiée dans [15] dans un cadre général. Utiliser la distance de Wasserstein pour aligner des lois est une approche naturelle puisque la distance de Wasserstein correspond au coût de transport entre deux lois de probabilité. Ainsi, dans notre cadre, considérer l'écart quadratique moyen entre les variables aléatoires ε_n et X_n est équivalent à étudier la distance de Wasserstein entre leurs lois. Finalement, considérer le modèle aléatoire nous permet d'avoir une interprétation probabiliste du contraste à minimiser, en terme de distance de Wasserstein.

Dans ce chapitre, nous proposons une méthode d'estimation récursive de θ de type gradient basé sur la distance de Wasserstein. On dérive alors de cet estimateur un estimateur de gradient stochastique en ajoutant une perturbation aléatoire. Puis, si la loi ε possède une densité de probabilité f , en utilisant l'estimation préalable de θ , nous construisons un estimateur de f plus robuste que l'estimateur naturel de Parzen-Rosenblatt.

Remarque 6.0.1 *Si, pour tout $x \in \mathbb{R}$, la fonction $t \rightarrow \varphi_t(x)$ est inversible, alors une estimation exacte du paramètre θ est possible. Nous excluons ce cas. Un exemple de fonction non inversible est la fonction de Box-Cox présentée au chapitre 7.*

6.1 Procédure d'estimation

Dans un premier temps, on s'intéresse donc à l'estimation de θ dans le modèle (6.0.1). Comme $M(\theta) = 0$ et que la fonction M est positive, il est clair que M admet un minimum global en θ , ce qui nous permet d'avoir une caractérisation du paramètre θ . Formellement, on suppose que $\theta \in \Theta$ où Θ est un intervalle \mathbb{R} . Nous avons également besoin d'ajouter des hypothèses de régularité sur les fonctions de déformation φ_t . Plus précisément, pour tout $t \in \Theta$, on suppose que

- (\mathcal{H}_1) φ_t est inversible, croissante de I_1 vers I_2 , des intervalles de \mathbb{R} .
- (\mathcal{H}_2) Pour tout $x \in I_2$, $\varphi_t^{-1}(x)$ est de classe \mathcal{C}^1 par rapport à $t \in \Theta$.
- (\mathcal{H}_3) $\varphi_t^{-1} \circ \varphi_\theta \in L^2(\varepsilon)$.
- (\mathcal{H}_4) Pour tout compact B de Θ , $\mathbb{E}\left[\sup_{t \in B} |\partial \varphi_t^{-1} \circ \varphi_\theta(\varepsilon)|^4\right] < +\infty$.

Dans toute la suite, pour tout $x \in I_2$, on note $\partial \varphi_t^{-1}(x)$ la dérivée de $\varphi_t^{-1}(x)$ par rapport à t . De plus, on déduit immédiatement de l'hypothèse (\mathcal{H}_1) que la fonction de répartition de X est $F_X = F \circ \varphi_\theta^{-1}$ alors que celle de $Z(t)$ est $F \circ \varphi_\theta^{-1} \circ \varphi_t$. Nous avons alors le lemme suivant.

Lemme 6.1.1 *Supposons les hypothèses (\mathcal{H}_1) à (\mathcal{H}_4). Alors M est de classe $\mathcal{C}^1(\Theta)$.*

Preuve du Lemme 6.1.1

Tout d'abord, (\mathcal{H}_4) implique en particulier que, pour tout compact B de Θ ,

$$\mathbb{E}\left[\sup_{t \in B} |\partial \varphi_t^{-1} \circ \varphi_\theta(\varepsilon)|^2\right] < +\infty.$$

De plus, en rappelant que F^{-1} est la fonction quantile associé à ε , nous avons

$$\mathbb{E}\left[\sup_{t \in B} |\partial \varphi_t^{-1} \circ \varphi_\theta(\varepsilon)|^2\right] = \int_0^1 \sup_{t \in B} |\partial \varphi_t^{-1} \circ \varphi_\theta(F^{-1}(x))|^2 dx < +\infty. \quad (6.1.1)$$

On déduit aussi de (\mathcal{H}_2) que, pour tout $x \in I_2$,

$$\partial \left[(F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x))^2 \right] = -2\partial \varphi_t^{-1}(\varphi_\theta \circ F^{-1}(x)) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)) \quad (6.1.2)$$

est une fonction continue en t . De plus, si B est un compact contenant θ , alors il découle de (\mathcal{H}_2) et du théorème des valeurs intermédiaires qu'il existe une constante $C_B > 0$ telle que

$$\sup_{t \in B} |F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta(F^{-1}(x))| \leq C_B \sup_{t \in B} |\partial \varphi_t^{-1} \circ \varphi_\theta(F^{-1}(x))|. \quad (6.1.3)$$

Ainsi, on déduit de (6.1.2) et de l'inégalité précédente que

$$\sup_{t \in B} \left| \partial \left[(F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x))^2 \right] \right| \leq 2C_B \sup_{t \in B} |\partial \varphi_t^{-1} \circ \varphi_\theta(F^{-1}(x))|^2$$

ce qui, par (6.1.1), implique que

$$\sup_{t \in B} \left| \partial \left[(F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x))^2 \right] \right|$$

est intégrable. Finalement, M est de classe $\mathcal{C}^1(\Theta)$ et pour tout $t \in \Theta$,

$$M'(t) = \int_0^1 -2\partial \varphi_t^{-1}(\varphi_\theta \circ F^{-1}(x)) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)) dx.$$

□

On déduit immédiatement du lemme 6.1.1 que la dérivée M' de M a l'expression suivante, pour tout $t \in \Theta$,

$$\begin{aligned} M'(t) &= -2 \int_0^1 \partial \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)) dx \\ &= -2\mathbb{E}[\partial \varphi_t^{-1}(X) (\varepsilon - \varphi_t^{-1}(X))]. \end{aligned} \quad (6.1.4)$$

Il est alors clair que $M'(\theta) = 0$. Ainsi, on peut supposer qu'il existe $\{a, b\} \in \Theta^2$ tels que $a < b$ et $\theta \in]a; b[\subset \Theta$ tel que, pour tout $t \in [a; b]$,

$$(\mathcal{H}_5) \quad (t - \theta)M'(t) > 0.$$

Afin que la fonction M soit un peu plus régulière, nous avons aussi besoin d'ajouter une hypothèse de régularité supplémentaire sur les fonctions de déformation. Cette hypothèse est la suivante.

$$(\mathcal{H}_6) \quad \text{Pour tout } x \in I_2, \varphi_t^{-1}(x) \text{ est deux fois dérivable en } t \in \Theta, \text{ et} \\ \text{pour tout compact } B \text{ de } \Theta, \mathbb{E} \left[\sup_{t \in B} |\partial^2 \varphi_t^{-1} \circ \varphi_\theta(\varepsilon)|^2 \right] < +\infty.$$

Lemme 6.1.2 *Supposons les hypothèses (\mathcal{H}_1) à (\mathcal{H}_6) . Alors, M est de classe $\mathcal{C}^2(\Theta)$.*

Preuve du Lemme 6.1.2

En effet, l'hypothèse (\mathcal{H}_6) implique que la fonction

$$-2\partial \varphi_t^{-1}(\varphi_\theta \circ F^{-1}(x)) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)) \quad (6.1.5)$$

est continûment dérivable en t . De plus, on a

$$\begin{aligned} &\partial [\partial \varphi_t^{-1}(\varphi_\theta \circ F^{-1}(x)) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x))] \\ &= -[\partial \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)]^2 + \partial^2 \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)). \end{aligned}$$

On déduit alors de (6.1.3) que, pour tout compact B contenant t et θ ,

$$\begin{aligned} \sup_{t \in B} |\partial^2 \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x))| \\ \leq C_B \sup_{t \in B} |\partial^2 \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)| \sup_{t \in B} |\partial \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)|. \end{aligned}$$

Ainsi, (\mathcal{H}_6) et (6.1.1) et l'inégalité de Cauchy Schwartz implique que

$$\sup_{t \in B} \partial^2 \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x))$$

est intégrable. On a donc

$$\int_0^1 \sup_{t \in B} |\partial [\partial \varphi_t^{-1}(\varphi_\theta \circ F^{-1}(x)) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x))]| dx < +\infty$$

qui nous permet de conclure que M est de classe $\mathcal{C}^2(\Theta)$ et on a, pour tout $t \in \Theta$,

$$\begin{aligned} M''(t) &= 2 \int_0^1 [\partial\varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)]^2 dx \\ &\quad - 2 \int_0^1 \partial^2\varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)) dx. \end{aligned}$$

□

On déduit alors du lemme 6.1.2 que la dérivée seconde M'' de M est définie, pour tout $t \in \Theta$, par

$$\begin{aligned} M''(t) &= 2 \int_0^1 [\partial\varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)]^2 dx \\ &\quad - 2 \int_0^1 \partial^2\varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x) (F^{-1}(x) - \varphi_t^{-1} \circ \varphi_\theta \circ F^{-1}(x)) dx \end{aligned} \quad (6.1.6)$$

c'est-à-dire

$$M''(t) = 2\mathbb{E}[(\partial\varphi_t^{-1}(X))^2] - 2\mathbb{E}[\partial^2\varphi_t^{-1}(X)(\varepsilon - \varphi_t^{-1}(X))]. \quad (6.1.7)$$

6.1.1 Estimation de θ

A ce stade, nous pouvons mettre en place notre algorithme de Robbins-Monro pour l'estimation de θ . Plus précisément, notons $\pi_{[a;b]}$ la projection sur le compact $[a; b]$ définie, pour tout $x \in [a; b]$, par

$$\pi_{[a;b]}(x) = x\mathbf{I}_{\{a \leq x \leq b\}} + a\mathbf{I}_{\{x \leq a\}} + b\mathbf{I}_{\{x \geq b\}}.$$

Soit (γ_n) une suite de nombres réels positifs, décroissante vers 0, et telle que

$$\sum_{n=1}^{\infty} \gamma_n = +\infty \quad \text{et} \quad \sum_{n=1}^{\infty} \gamma_n^2 < +\infty. \quad (6.1.8)$$

Nous allons estimer θ grâce à l'algorithme de Robbins-Monro projeté, défini, pour tout $n \geq 0$, par

$$\widehat{\theta}_{n+1} = \pi_{[a;b]}(\widehat{\theta}_n - \gamma_{n+1}T_{n+1}) \quad (6.1.9)$$

où la valeur initiale $\widehat{\theta}_0 \in [a; b]$ et la variable aléatoire T_{n+1} est définie par

$$T_{n+1} = -2\partial\varphi_{\widehat{\theta}_n}^{-1}(X_{n+1}) (\varepsilon_{n+1} - \varphi_{\widehat{\theta}_n}^{-1}(X_{n+1})). \quad (6.1.10)$$

Naturellement, cet algorithme est analogue à l'algorithme de gradient déterministe (6.0.2) où la fonction de contraste est la distance de Wasserstein. De plus, nous pouvons modifier l'algorithme (6.1.9) afin d'obtenir un algorithme de gradient stochastique. Plus précisément, cette modification consiste à remplacer l'algorithme (6.1.9) par sa version « excité »

$$\widetilde{\theta}_{n+1} = \pi_{[a;b]}(\widetilde{\theta}_n - \gamma_{n+1}\widetilde{T}_{n+1}) \quad (6.1.11)$$

où la valeur initiale $\widetilde{\theta}_0 \in [a; b]$ et où la variable aléatoire \widetilde{T}_{n+1} est définie par

$$\widetilde{T}_{n+1} = -2\partial\varphi_{\widetilde{\theta}_n}^{-1}(X_{n+1}) (\varepsilon_{n+1} - \varphi_{\widetilde{\theta}_n}^{-1}(X_{n+1})) + V_{n+1}. \quad (6.1.12)$$

L'excitation (V_n) est une suite exogène de variables aléatoires indépendantes et de même loi de moyenne 0 et de variance $\sigma^2 > 0$. Naturellement, comme nous le verrons dans la suite, le théorème central limite pour $(\widehat{\theta}_n)$ sera dégénéré, alors que l'ajout de la perturbation V_{n+1} permettra d'obtenir un théorème central limite non dégénéré pour l'algorithme excité $(\widetilde{\theta}_n)$.

6.1.2 Estimation de la loi de ε

Dans un second temps, on s'intéresse à l'estimation de la loi de ε . Pour cela, on suppose que la variable aléatoire ε possède une densité f et on s'intéresse à l'estimation non paramétrique de f . Une façon naturelle d'estimer f est de considérer l'estimateur de Parzen-Rosenblatt récursif défini, pour tout $n \geq 1$ et pour tout $x \in I_1$, par

$$\widetilde{f}_n(x) = \frac{1}{n} \sum_{k=1}^n \frac{1}{h_k} K\left(\frac{\varepsilon_k - x}{h_k}\right) \quad (6.1.13)$$

où K est une fonction noyau symétrique, positive et bornée et telle que

$$\int_{\mathbb{R}} K(x) dx = 1, \quad \text{et} \quad \int_{\mathbb{R}} x^2 K(x) dx < +\infty.$$

De plus, dans notre cas, on sait que $Z_n(\theta) = \varepsilon_n$. L'idée est alors d'utiliser l'estimation préalable de θ pour construire un autre estimateur de Parzen-Rosenblatt de f . A cette fin, nous allons étudier l'estimateur récursif de Parzen-Rosenblatt défini, pour tout $n \geq 1$ et pour tout $x \in I_2$, par

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{k=1}^n \frac{1}{h_k} K\left(\frac{Z_k(\widehat{\theta}_{k-1}) - x}{h_k}\right) \quad (6.1.14)$$

où l'estimateur $\widehat{\theta}_{k-1}$ est donné par (6.1.9) et où $h_n = 1/n^\alpha$ avec $0 < \alpha < 1$. Afin d'étudier le comportement asymptotique de \widehat{f}_n , nous avons besoin d'ajouter d'autres hypothèses aux hypothèses (\mathcal{H}_1) à (\mathcal{H}_6) . Plus précisément, si on note ∂ l'opérateur dérivé par rapport à t et d celui par rapport à x , on a besoin d'ajouter les hypothèses suivantes sur f et sur les fonctions de déformation φ_t .

- (\mathcal{H}'_1) f est bornée et de classe $\mathcal{C}^2(I_1)$ et ses dérivées sont bornées.
- (\mathcal{H}'_2) Pour tout $t \in \Theta$, φ_t est de classe $\mathcal{C}^3(I_1)$.
- (\mathcal{H}'_3) φ_{θ}^{-1} est de classe $\mathcal{C}^3(I_2)$ et ses dérivées sont bornées.
- (\mathcal{H}'_4) $d\varphi, d^2\varphi, d^3\varphi$ sont bornées.

De plus, nous aurons besoin d'une hypothèse un peu plus forte sur la régularité des fonctions de déformation φ .

- (\mathcal{H}'_5) φ est de classe $\mathcal{C}^2(\Theta \times I_1)$ et $\partial\varphi_t(x), \partial d\varphi_t(x)$ sont bornées en t .

6.2 Résultats de convergence paramétrique

Théorème 6.2.1 *On suppose les hypothèses (\mathcal{H}_1) à (\mathcal{H}_5) avec $\theta \in]a; b[$ où $a < b$. Alors, $\widehat{\theta}_n$ converge presque sûrement vers θ .*

Dans toute la suite, on fait le choix d'un pas $\gamma_n = 1/n$.

Théorème 6.2.2 *On suppose les hypothèses (\mathcal{H}_1) à (\mathcal{H}_6) avec $\theta \in]a; b[$ où $a < b$. Si $M''(\theta) > 1/2$ et s'il existe $\alpha > 4$ tel que, pour tout compact B de Θ ,*

$$\mathbb{E} \left[\sup_{t \in B} |\partial \varphi_t^{-1} \circ \varphi_\theta(\varepsilon)|^\alpha \right] < +\infty$$

alors, on a la normalité asymptotique dégénérée

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \delta_0. \quad (6.2.1)$$

Cependant, pour la suite $(\widetilde{\theta}_n)$ définie par (6.1.11), si l'excitation (V_n) a un moment d'ordre > 2 fini, on a la normalité asymptotique

$$\sqrt{n}(\widetilde{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\sigma^2}{2M''(\theta) - 1} \right). \quad (6.2.2)$$

De plus, si pour tout $t \in [a; b]$,

$$(\mathcal{H}_7) \quad M''(t) \geq 1/2,$$

alors pour tout $n \geq 0$,

$$\mathbb{E} \left[(\widehat{\theta}_n - \theta)^2 \right] \leq (\widehat{\theta}_0 - \theta)^2 \frac{\exp(C_1 \pi^2/6)}{n+1} \quad (6.2.3)$$

où la constante C_1 est donnée par

$$C_1 = 4\mathbb{E} \left[\sup_{t \in [a; b]} |\partial \varphi_t^{-1} \circ \varphi_\theta(\varepsilon)|^4 \right]. \quad (6.2.4)$$

Remarque 6.2.1 *Il est immédiat, par (6.1.6), que*

$$M''(\theta) = 2 \int_0^1 [\partial \varphi_\theta^{-1} \circ \varphi_\theta \circ F^{-1}(x)]^2 dx = 2\mathbb{E} \left[(\partial \varphi_\theta^{-1}(X))^2 \right].$$

Ainsi, l'inégalité $M''(\theta) > 0$ a lieu d'une manière générale. De plus, si on remplace M par λM avec $\lambda > 0$, alors les résultats ne changent pas. Ainsi, la condition $M''(t) \geq 1/2$ n'est pas restrictive.

6.3 Résultats de convergence non paramétrique

Théorème 6.3.1 *On suppose les hypothèses (\mathcal{H}_1) à (\mathcal{H}_5) avec $\theta \in]a; b[$ où $a < b$. On suppose de plus les hypothèses (\mathcal{H}'_1) à (\mathcal{H}'_4) . Alors, pour tout $x \in I_1$, on a*

$$\lim_{n \rightarrow +\infty} \widehat{f}_n(x) = f(x) \quad \text{p.s.} \quad (6.3.1)$$

Le deuxième résultat concerne la convergence en moyenne quadratique de $\widehat{f}_n(x)$ vers $f(x)$.

Théorème 6.3.2 *On suppose les hypothèses (\mathcal{H}_1) à (\mathcal{H}_7) avec $\theta \in]a; b[$ où $a < b$. De plus, on suppose les hypothèses (\mathcal{H}'_1) à (\mathcal{H}'_5) . Alors, pour tout $x \in I_1$, on a*

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[\left| \widehat{f}_n(x) - f(x) \right|^2 \right] = 0. \quad (6.3.2)$$

6.4 Preuves des convergences paramétriques

6.4.1 Preuve du Théorème 6.2.1.

La preuve du Théorème 6.2.1 suit essentiellement les mêmes lignes que celle du Théorème 4.2.1. On note \mathcal{F}_n la σ -algèbre des événements s'étant produits avant l'instant n , $\mathcal{F}_n = \sigma(\varepsilon_0, \dots, \varepsilon_n)$. On commence par calculer les deux premiers moments conditionnels de la variable aléatoire T_n donnée par (6.1.10). D'une part, on a

$$\begin{aligned} \mathbb{E}[T_{n+1} | \mathcal{F}_n] &= -2\mathbb{E} \left[\partial \varphi_{\widehat{\theta}_n}^{-1}(X_{n+1}) \left(\varepsilon_{n+1} - \varphi_{\widehat{\theta}_n}^{-1}(X_{n+1}) \right) | \mathcal{F}_n \right], \\ &= -2\mathbb{E} \left[\partial \varphi_{\widehat{\theta}_n}^{-1} \circ \varphi_{\theta}(\varepsilon_{n+1}) \left(\varepsilon_{n+1} - \varphi_{\widehat{\theta}_n}^{-1} \circ \varphi_{\theta}(\varepsilon_{n+1}) \right) | \mathcal{F}_n \right]. \end{aligned}$$

Ainsi, comme (ε_n) est une suite de variables aléatoires indépendantes et $\widehat{\theta}_n \in \mathcal{F}_n$, on déduit immédiatement de (6.1.4) que

$$\begin{aligned} &-2\mathbb{E} \left[\partial \varphi_{\widehat{\theta}_n}^{-1} \circ \varphi_{\theta}(\varepsilon_{n+1}) \left(\varepsilon_{n+1} - \varphi_{\widehat{\theta}_n}^{-1} \circ \varphi_{\theta}(\varepsilon_{n+1}) \right) | \mathcal{F}_n \right] \\ &= -2 \int_0^1 \partial \varphi_{\widehat{\theta}_n}^{-1} \circ \varphi_{\theta} \circ F^{-1}(x) \left(F^{-1}(x) - \varphi_{\widehat{\theta}_n}^{-1} \circ \varphi_{\theta} \circ F^{-1}(x) \right) dx \\ &= M'(\widehat{\theta}_n) \quad \text{p.s.} \end{aligned}$$

ce qui entraîne que

$$\mathbb{E}[T_{n+1} | \mathcal{F}_n] = M'(\widehat{\theta}_n) \quad \text{p.s.} \quad (6.4.1)$$

D'autre part, on a

$$\begin{aligned} \mathbb{E}[T_{n+1}^2 | \mathcal{F}_n] &= 4\mathbb{E} \left[\partial \varphi_{\widehat{\theta}_n}^{-1}(X_{n+1})^2 \left(\varepsilon_{n+1} - \varphi_{\widehat{\theta}_n}^{-1}(X_{n+1}) \right)^2 | \mathcal{F}_n \right], \\ &= 4\mathbb{E} \left[\partial \varphi_{\widehat{\theta}_n}^{-1}(X_{n+1})^2 \left(\varphi_{\theta}^{-1}(X_{n+1}) - \varphi_{\widehat{\theta}_n}^{-1}(X_{n+1}) \right)^2 | \mathcal{F}_n \right] / \end{aligned} \quad (6.4.2)$$

Le théorème des accroissements finis assure que

$$\left| \varphi_{\theta}^{-1}(X_{n+1}) - \varphi_{\widehat{\theta}_n}^{-1}(X_{n+1}) \right| \leq \sup_{t \in [a; b]} |\partial \varphi_t^{-1}(X_{n+1})| \times |\widehat{\theta}_n - \theta|. \quad (6.4.3)$$

Par conséquent, les équations (6.4.2) et (6.4.3) impliquent que

$$\mathbb{E}[T_{n+1}^2 | \mathcal{F}_n] \leq 4(\widehat{\theta}_n - \theta)^2 \mathbb{E}\left[\sup_{t \in [a; b]} |\partial \varphi_t^{-1}(X)|^4\right]. \quad (6.4.4)$$

Ainsi, on en déduit qu'il existe une constante C_1 donnée par (6.2.4) et telle que

$$\mathbb{E}[T_{n+1}^2 | \mathcal{F}_n] \leq C_1 (\widehat{\theta}_n - \theta)^2 \quad \text{p.s.} \quad (6.4.5)$$

Maintenant, pour tout $n \geq 0$, on note $V_n = (\widehat{\theta}_n - \theta)^2$. On a clairement que

$$\begin{aligned} V_{n+1} &= (\widehat{\theta}_{n+1} - \theta)^2, \\ &= \left(\pi_{[a; b]}(\widehat{\theta}_n - \gamma_{n+1} T_{n+1}) - \theta\right)^2, \\ &= \left(\pi_{[a; b]}(\widehat{\theta}_n - \gamma_{n+1} T_{n+1}) - \pi_{[a; b]}(\theta)\right)^2 \end{aligned}$$

puisque $\theta \in]a; b[$. La fonction $\pi_{[a; b]}$ étant une fonction lipschitzienne de constante de Lipschitz 1, on obtient que

$$\begin{aligned} V_{n+1} &\leq (\widehat{\theta}_n - \gamma_{n+1} T_{n+1} - \theta)^2, \\ &\leq V_n + \gamma_{n+1}^2 T_{n+1}^2 - 2\gamma_{n+1} T_{n+1} (\widehat{\theta}_n - \theta). \end{aligned}$$

Ainsi, les équations (6.4.1) et (6.4.5) impliquent que

$$\mathbb{E}[V_{n+1} | \mathcal{F}_n] \leq V_n (1 + C_1 \gamma_{n+1}^2) - 2\gamma_{n+1} (\widehat{\theta}_n - \theta) M'(\widehat{\theta}_n) \quad \text{p.s.} \quad (6.4.6)$$

De plus, $\widehat{\theta}_n \in [a; b]$, donc d'après (\mathcal{H}_5) on a $(\widehat{\theta}_n - \theta) M'(\widehat{\theta}_n) > 0$. Ainsi, on déduit de (6.4.6) et du Lemme 1.3.1 de Robbins-Siegmund que la suite (V_n) converge p.s. vers une variable aléatoire V finie p.s. et

$$\sum_{n=1}^{\infty} \gamma_{n+1} (\widehat{\theta}_n - \theta) M'(\widehat{\theta}_n) < +\infty \quad \text{p.s.} \quad (6.4.7)$$

Supposons par l'absurde que $V \neq 0$ p.s. Alors, on peut trouver deux constantes c et d telles que

$$0 < c < d < 2 \max(|a|, |b|),$$

et pour n assez grand, l'événement $\{c < |\widehat{\theta}_n - \theta| < d\}$ n'est pas négligeable. Cependant, sur cet anneau, on peut trouver une constante $e > 0$ telle que $(\widehat{\theta}_n - \theta) M'(\widehat{\theta}_n) \geq e$. Ainsi, (6.4.7) implique que

$$\sum_{n=1}^{\infty} \gamma_n < +\infty,$$

ce qui contredit (6.1.8). Par conséquent, on obtient que $V = 0$ p.s., et donc que $\widehat{\theta}_n$ converge p.s. vers θ . \square

6.4.2 Preuve du Théorème 6.2.2.

On va suivre la preuve du Théorème 4.2.2. Plus précisément, on veut appliquer le Théorème 2.1 de Kushner et Yin [40] page 330. Tout d'abord, comme $\gamma_n = 1/n$, les conditions sur le pas sont satisfaites. De plus, on a vu que $\widehat{\theta}_n$ converge p.s. vers θ . Par conséquent, toutes les hypothèses locales du Théorème 2.1 de [40] sont vérifiées. De plus, il découle de (6.4.1) que $\mathbb{E}[T_{n+1}|\mathcal{F}_n] = M'(\widehat{\theta}_n)$ p.s. et la fonction M est deux fois continûment dérivable. On a $M(\theta) = 0$, $M'(\theta) = 0$ et $M''(\theta) > 1/2$. On déduit alors de (6.4.5) et de la convergence presque sûre de $\widehat{\theta}_n$ vers θ que

$$\lim_{n \rightarrow \infty} \mathbb{E}[T_{n+1}^2|\mathcal{F}_n] = 0 \quad \text{p.s.}$$

Enfin, le Théorème 4.1 de [40] page 341 assure que la suite (W_n) définie par

$$W_n = \sqrt{n}(\widehat{\theta}_n - \theta)$$

est tendue. Ainsi, on déduit du Théorème 2.1 de [40] que

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \delta_0.$$

Des calculs analogues à ceux effectués dans la preuve du Théorème 4.2.2 nous permettent également de conclure que

$$\sqrt{n}(\widetilde{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2}{2M''(\theta) - 1}\right).$$

De plus, en passant à l'espérance dans (6.4.6), on a, pour tout $n \geq 0$,

$$v_{n+1} \leq v_n(1 + C_1\gamma_{n+1}^2) - 2\gamma_{n+1}\mathbb{E}[(\widehat{\theta}_n - \theta)M'(\widehat{\theta}_n)] \quad (6.4.8)$$

avec

$$v_n = \mathbb{E}[(\widehat{\theta}_n - \theta)^2].$$

Comme $M'(\theta) = 0$, on a

$$M'(\widehat{\theta}_n) = (\widehat{\theta}_n - \theta) \int_0^1 M''(\theta + x(\widehat{\theta}_n - \theta))dx \quad \text{p.s.} \quad (6.4.9)$$

Par conséquent, on déduit de (6.4.8) et (6.4.9) que

$$v_{n+1} \leq v_n(1 + C_1\gamma_{n+1}^2) - 2\gamma_{n+1}\mathbb{E}\left[(\widehat{\theta}_n - \theta)^2 \int_0^1 M''(\theta + x(\widehat{\theta}_n - \theta))dx\right]. \quad (6.4.10)$$

Finalement, puisque $\theta \in]a; b[$ et $\widehat{\theta}_n \in [a; b]$, $\theta + x(\widehat{\theta}_n - \theta) \in [a; b]$ pour tout $x \in [0; 1]$. Aussi, d'après \mathcal{H}_7 , pour tout $t \in [a; b]$, $M''(t) \geq 1/2$ et on peut donc écrire que

$$\int_0^1 M''(\theta + x(\widehat{\theta}_n - \theta))dx \geq 1/2.$$

On déduit alors de (6.4.10) que, pour tout $n \geq 0$,

$$v_{n+1} \leq v_n(1 + C_1\gamma_{n+1}^2 - \gamma_{n+1}). \quad (6.4.11)$$

De plus, pour $x \in \mathbb{R}$, on a l'inégalité de convexité

$$1 - x \leq \exp(-x).$$

On en déduit donc que

$$v_{n+1} \leq v_n \exp(C_1 \gamma_{n+1}^2 - \gamma_{n+1}). \quad (6.4.12)$$

Par une récurrence immédiate dans (6.4.12), il advient que

$$\begin{aligned} v_n &\leq v_0 \prod_{k=1}^n \exp(C_1 \gamma_k^2 - \gamma_k), \\ &\leq v_0 \exp\left(C_1 \sum_{k=1}^n \gamma_k^2 - \sum_{k=1}^n \gamma_k\right), \\ &\leq v_0 \exp\left(C_1 \sum_{k=1}^{+\infty} \gamma_k^2 - \sum_{k=1}^n \gamma_k\right). \end{aligned} \quad (6.4.13)$$

Finalement, comme $\gamma_k = 1/k$, il s'ensuit de (6.4.13) et du fait que

$$\sum_{k=1}^{+\infty} \gamma_k^2 = \frac{\pi^2}{6}$$

et

$$\sum_{k=1}^n \gamma_k \geq \log(n+1)$$

que, pour tout $n \geq 0$,

$$v_n \leq v_0 \frac{\exp(C_1 \pi^2/6)}{n+1}$$

ce qui achève la preuve du Théorème 6.2.2. \square

6.5 Preuves des convergences non paramétriques

On rappelle que f est la densité de ε et on note f_t la densité de la variable $Z(t)$. Comme la distribution de $Z(t)$ est $F \circ \varphi_\theta^{-1} \circ \varphi_t$, on a, pour tout $x \in I_1$,

$$f_t(x) = f(\varphi_\theta^{-1} \circ \varphi_t(x)) d[\varphi_\theta^{-1} \circ \varphi_t](x).$$

En particulier, $f_\theta = f$. Avant de s'attaquer aux preuves des deux théorèmes de convergence de \widehat{f}_n , on établit des résultats de régularité sur $f_t(x)$ qui seront utiles dans la suite. Tout d'abord, on a

$$\begin{aligned} f_t(x) &= f(\varphi_\theta^{-1} \circ \varphi_t(x)) d[\varphi_\theta^{-1} \circ \varphi_t](x) \\ &= f(\varphi_\theta^{-1} \circ \varphi_t(x)) d\varphi_t(x) d[\varphi_\theta^{-1}](\varphi_t(x)). \end{aligned}$$

Les hypothèses (\mathcal{H}'_1) , (\mathcal{H}'_2) et (\mathcal{H}'_3) impliquent que f_t est deux fois continûment dérivable en x . De plus, pour tout $x \in I_1$,

$$df_t(x) = f(\varphi_\theta^{-1} \circ \varphi_t(x)) d^2[\varphi_\theta^{-1} \circ \varphi_t](x) + f'(\varphi_\theta^{-1} \circ \varphi_t(x)) (d[\varphi_\theta^{-1} \circ \varphi_t](x))^2$$

et

$$\begin{aligned} d^2 f_t(x) &= f(\varphi_\theta^{-1} \circ \varphi_t(x)) d^3[\varphi_\theta^{-1} \circ \varphi_t](x) \\ &\quad + 3f'(\varphi_\theta^{-1} \circ \varphi_t(x)) d[\varphi_\theta^{-1} \circ \varphi_t](x) d^2[\varphi_\theta^{-1} \circ \varphi_t](x) \\ &\quad + f''(\varphi_\theta^{-1} \circ \varphi_t(x)) (d[\varphi_\theta^{-1} \circ \varphi_t](x))^3. \end{aligned}$$

On déduit alors des hypothèses (\mathcal{H}'_1) à (\mathcal{H}'_4) que $f_t(x)$, $df_t(x)$ et $d^2 f_t(x)$ sont bornées sur $\Theta \times I_1$. De plus, l'hypothèse (\mathcal{H}'_5) assure que $f_t(x)$ est aussi continûment dérivable en (t, x) et on a pour tout $t \in \Theta$ et pour tout $x \in I_1$,

$$\partial f_t(x) = f(\varphi_\theta^{-1} \circ \varphi_t(x)) \partial d[\varphi_\theta^{-1} \circ \varphi_t](x) + f'(\varphi_\theta^{-1} \circ \varphi_t(x)) d[\varphi_\theta^{-1} \circ \varphi_t](x) \partial[\varphi_\theta^{-1} \circ \varphi_t](x),$$

où

$$\partial[\varphi_\theta^{-1} \circ \varphi_t](x) = \partial \varphi_t(x) d[\varphi_\theta^{-1}](\varphi_t(x)),$$

et

$$\partial d[\varphi_\theta^{-1} \circ \varphi_t](x) = \partial d \varphi_t(x) d[\varphi_\theta^{-1}](\varphi_t(x)) + \partial \varphi_t(x) d \varphi_t(x) d^2[\varphi_\theta^{-1}](\varphi_t(x)).$$

Ainsi, sous (\mathcal{H}'_4) et (\mathcal{H}'_5)

$$\sup_{t \in \Theta} |\partial f_t(x)| < +\infty. \quad (6.5.1)$$

6.5.1 Preuve du Théorème 6.3.1

On note, pour tout $x \in I_1$,

$$W_n(x) = \frac{1}{h_n} K\left(\frac{Z_n(\hat{\theta}_{n-1}) - x}{h_n}\right).$$

Pour tout $x \in I_1$, on a la décomposition

$$n \widehat{f}_n(x) = M_n(x) + N_n(x),$$

avec

$$M_n(x) = \sum_{k=1}^n \mathbb{E}[W_k(x) | \mathcal{F}_{k-1}] \quad (6.5.2)$$

et

$$N_n(x) = \sum_{k=1}^n (W_k(x) - \mathbb{E}[W_k(x) | \mathcal{F}_{k-1}]). \quad (6.5.3)$$

D'une part, $f_{\hat{\theta}_{n-1}}$ est la densité de $Z_n(\hat{\theta}_{n-1})$. Il en découle que, pour tout $1 \leq k \leq n$,

$$\begin{aligned} \mathbb{E}[W_k(x) | \mathcal{F}_{k-1}] &= \int_{\mathbb{R}} \frac{1}{h_k} K\left(\frac{u-x}{h_k}\right) f_{\hat{\theta}_{k-1}}(u) du \\ &= \int_{\mathbb{R}} K(v) f_{\hat{\theta}_{k-1}}(x + h_k v) dv. \end{aligned}$$

Par suite, pour tout $1 \leq k \leq n$, on a

$$\mathbb{E}[W_k(x) | \mathcal{F}_{k-1}] - f_{\hat{\theta}_{k-1}}(x) = \int_{\mathbb{R}} (f_{\hat{\theta}_{k-1}}(x + v h_k) - f_{\hat{\theta}_{k-1}}(x)) K(v) dv.$$

De plus, on a vu que f_t est deux fois continûment dérivable. Ainsi, pour tout $t \in \Theta$, il existe un réel $z_k = x + vh_k y$, avec $0 < y < 1$, tel que

$$f_t(x + vh_k) - f_t(x) = vh_k df_t(x) + \frac{(vh_k)^2}{2} d^2 f_t(z_k). \quad (6.5.4)$$

En utilisant la symétrie de K et les remarques préliminaires sur $d^2 f_t$, on obtient alors que

$$\int_{\mathbb{R}} (f_t(x + vh_k) - f_t(x)) K(v) dv = \int_{\mathbb{R}} \frac{(vh_k)^2}{2} d^2 f_t(z_k) K(v) dv$$

ce qui implique que

$$\sup_{t \in \Theta} \left| \int_{\mathbb{R}} (f_t(x + vh_k) - f_t(x)) K(v) dv \right| \leq \frac{h_k^2}{2} \sup_{t \in \Theta, z \in I_1} |d^2 f_t(z)| \int_{\mathbb{R}} v^2 K(v) dv.$$

Par conséquent, il existe $C_2 > 0$ telle que, pour tout $1 \leq k \leq n$,

$$\left| \mathbb{E}[W_k(x) | \mathcal{F}_{k-1}] - f_{\widehat{\theta}_{k-1}}(x) \right| \leq C_2 h_k^2. \quad (6.5.5)$$

De plus, puisque f_t est continue en t , et $\widehat{\theta}_n$ converge p.s. vers θ , on a, pour tout $x \in I_1$,

$$\lim_{n \rightarrow +\infty} f_{\widehat{\theta}_{n-1}}(x) = f(x) \quad \text{p.s.} \quad (6.5.6)$$

Le lemme de Césaro et l'équation (6.5.5) entraînent alors que

$$\lim_{n \rightarrow +\infty} \frac{1}{n} M_n(x) = f(x) \quad \text{p.s.} \quad (6.5.7)$$

D'autre part, K étant borné, $(N_n(x))$ est une martingale de carré intégrable dont le crochet est donné par

$$\langle N(x) \rangle_n = \sum_{k=1}^n \mathbb{E}[W_k^2(x) | \mathcal{F}_{k-1}] - \mathbb{E}^2[W_k(x) | \mathcal{F}_{k-1}].$$

De plus, on a, pour tout $1 \leq k \leq n$,

$$\mathbb{E}[W_k^2(x) | \mathcal{F}_{k-1}] = \frac{1}{h_k} \int K^2(v) f_{\widehat{\theta}_{k-1}}(x + h_k v) dv.$$

Cependant, l'équation (6.5.4), la régularité de $f_t(x)$ et la symétrie de K entraînent que

$$\sup_{t \in \Theta} \left| \int_{\mathbb{R}} \frac{1}{h_k} (f_t(x + vh_k) - f_t(x)) K^2(v) dv \right| \leq \frac{h_k}{2} \sup_{t \in \Theta, z \in I_1} |d^2 f_t(z)| \int_{\mathbb{R}} v^2 K^2(v) dv.$$

Par conséquent, il existe $C_3 > 0$ telle que, pour tout $1 \leq k \leq n$,

$$\left| \mathbb{E}[W_k^2(x) | \mathcal{F}_{k-1}] - \frac{\nu^2}{h_k} f_{\widehat{\theta}_{k-1}}(x) \right| \leq C_3 h_k \quad (6.5.8)$$

où $\nu^2 = \int_{\mathbb{R}} K^2(u) du$. Il s'ensuit également de (6.5.6) et du Lemme 1.3.2 de Toeplitz que

$$\lim_{n \rightarrow \infty} \frac{1}{\sum_{i=1}^n h_k^{-1}} \sum_{i=1}^n \frac{1}{h_k} f_{\widehat{\theta}_{k-1}}(x) = f(x) \quad \text{p.s.}$$

On déduit alors de l'équivalence élémentaire

$$\sum_{k=1}^n \frac{1}{h_k} \sim \frac{n^{1+\alpha}}{\alpha+1}$$

que

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1+\alpha}} \sum_{k=1}^n \frac{\nu^2}{h_k} f_{\hat{\theta}_{k-1}}(x) = \frac{\nu^2}{\alpha+1} f(x) \quad \text{p.s.}$$

Finalement, (6.5.8) mène à

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1+\alpha}} \sum_{k=1}^n \mathbb{E} [W_k^2(x) | \mathcal{F}_{k-1}] = \frac{\nu^2}{\alpha+1} f(x) \quad \text{p.s.} \quad (6.5.9)$$

D'autre part, les équations (6.5.5) et (6.5.6) ainsi que le lemme de Cesaro entraînent que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E}^2 [W_k(x) | \mathcal{F}_{k-1}] = f^2(x) \quad \text{p.s.} \quad (6.5.10)$$

On peut conclure de (6.5.9) et (6.5.10) que

$$\lim_{n \rightarrow \infty} \frac{\langle N(x) \rangle_n}{n^{1+\alpha}} = \frac{\nu^2}{\alpha+1} f(x) \quad \text{p.s.}$$

La loi des grands nombres pour les martingales entraîne que, pour tout $\gamma > 0$, $(N_n(x))^2 = o(n^{1+\alpha} (\log(n))^{1+\gamma})$ p.s. et donc que, pour tout $x \in I_1$,

$$\lim_{n \rightarrow +\infty} \frac{1}{n} N_n(x) = 0 \quad \text{p.s.} \quad (6.5.11)$$

Finalement, en combinant (6.5.7) et (6.5.11), on obtient que, pour tout $x \in I_1$,

$$\lim_{n \rightarrow +\infty} \widehat{f}_n(x) = f(x) \quad \text{p.s.} \quad (6.5.12)$$

ce qui termine la preuve du Théorème 6.3.1. \square

6.5.2 Preuve du Théorème 6.3.2

Notre objectif est de montrer que pour tout $x \in I_1$,

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[|\widehat{f}_n(x) - f(x)|^2 \right] = 0.$$

On fait la décomposition classique biais-variance pour obtenir

$$\mathbb{E} \left[|\widehat{f}_n(x) - f(x)|^2 \right] = B_n(x) + V_n(x) \quad (6.5.13)$$

avec

$$B_n(x) = |\mathbb{E} [\widehat{f}_n(x)] - f(x)|^2 \quad (6.5.14)$$

et

$$V_n(x) = \mathbb{E} \left[|\widehat{f}_n(x) - \mathbb{E} [\widehat{f}_n(x)]|^2 \right]. \quad (6.5.15)$$

Tout d'abord, on peut écrire que

$$\begin{aligned}\mathbb{E}[\widehat{f}_n(x)] - f(x) &= \frac{1}{n} \sum_{k=1}^n \mathbb{E}[W_k(x) - f(x)], \\ &= \frac{1}{n} \sum_{k=1}^n \mathbb{E}[\mathbb{E}[W_k(x)|\mathcal{F}_{k-1}] - f(x)].\end{aligned}$$

De plus, l'équation (6.5.5) entraîne que

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[\left| \mathbb{E}[W_n(x)|\mathcal{F}_{n-1}] - f_{\widehat{\theta}_{n-1}}(x) \right| \right] = 0. \quad (6.5.16)$$

Il s'ensuit aussi de la bornitude de $f_{\widehat{\theta}_{n-1}}(x)$, de (6.5.6) et du théorème de convergence dominée, que

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[\left| f_{\widehat{\theta}_{n-1}}(x) - f(x) \right| \right] = 0. \quad (6.5.17)$$

Ainsi, on déduit de (6.5.16) et de (6.5.17) que

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left[\left| \mathbb{E}[W_n(x)|\mathcal{F}_{n-1}] - f(x) \right| \right] = 0.$$

Le lemme de Césaro entraîne alors que

$$\lim_{n \rightarrow +\infty} \left| \mathbb{E}[\widehat{f}_n(x)] - f(x) \right| = 0$$

ce qui implique que

$$\lim_{n \rightarrow +\infty} B_n(x) = 0. \quad (6.5.18)$$

Ensuite, on s'intéresse au terme de variance $V_n(x)$. Pour tout $1 \leq k \leq n$ et pour tout $x \in I_1$, on note $U_k(x)$

$$U_k(x) = W_k(x) - \mathbb{E}[W_k(x)]. \quad (6.5.19)$$

On a alors la décomposition

$$V_n(x) = \frac{1}{n^2} \sum_{k=1}^n \mathbb{E}[U_k(x)^2] + \frac{2}{n^2} \sum_{k=1, i < l}^n \mathbb{E}[U_k(x)U_l(x)]. \quad (6.5.20)$$

Remarquons que, si $k < l$, on a

$$\mathbb{E}[U_k(x)U_l(x)|\mathcal{F}_{l-1}] = U_k(x)\mathbb{E}[U_l(x)|\mathcal{F}_{l-1}].$$

De plus, (6.5.5) implique que

$$\left| \mathbb{E}[U_l(x)|\mathcal{F}_{l-1}] - f_{\widehat{\theta}_{l-1}}(x) + \mathbb{E}[f_{\widehat{\theta}_{l-1}}(x)] \right| \leq 2C_2 h_l^2.$$

Ainsi, on obtient que

$$-2C_2 h_l^2 |U_k(x)| \leq \mathbb{E}[U_k(x)U_l(x)|\mathcal{F}_{l-1}] - U_k(x)f_{\widehat{\theta}_{l-1}}(x) + U_k(x)\mathbb{E}[f_{\widehat{\theta}_{l-1}}(x)] \leq 2C_2 h_l^2 |U_k(x)|.$$

En passant à l'espérance dans la précédente inégalité, on a alors

$$-2C_2 h_l^2 \mathbb{E}[|U_k(x)|] \leq \mathbb{E}[U_k(x)U_l(x)] - \mathbb{E}[U_k(x)f_{\widehat{\theta}_{l-1}}(x)] + \mathbb{E}[U_k(x)]\mathbb{E}[f_{\widehat{\theta}_{l-1}}(x)] \leq 2C_2 h_l^2 \mathbb{E}[|U_k(x)|].$$

Finalement, on obtient que

$$|\mathbb{E}[U_k(x)U_l(x)]| \leq \left| \mathbb{E}\left[U_k(x)f_{\widehat{\theta}_{l-1}}(x)\right] - \mathbb{E}[U_k(x)]\mathbb{E}\left[f_{\widehat{\theta}_{l-1}}(x)\right] \right| + 2C_2h_l^2\mathbb{E}[|U_k(x)|]. \quad (6.5.21)$$

De plus, on a la décomposition

$$\begin{aligned} \mathbb{E}\left[U_k(x)f_{\widehat{\theta}_{l-1}}(x)\right] - \mathbb{E}[U_k(x)]\mathbb{E}\left[f_{\widehat{\theta}_{l-1}}(x)\right] &= \mathbb{E}\left[U_k(x)\left(f_{\widehat{\theta}_{l-1}}(x) - f(x)\right)\right] \\ &+ \left(f(x) - \mathbb{E}\left[f_{\widehat{\theta}_{l-1}}(x)\right]\right)\mathbb{E}[U_k(x)]. \end{aligned} \quad (6.5.22)$$

Par conséquent, (6.5.21) et (6.5.22) ainsi que l'inégalité de Cauchy-Schwarz nous permettent de dire que

$$\mathbb{E}[|U_k(x)U_l(x)|] \leq 2\sqrt{\mathbb{E}[U_k(x)^2]}\left(\sqrt{\mathbb{E}\left[\left(f_{\widehat{\theta}_{l-1}}(x) - f(x)\right)^2\right]} + C_2h_l^2\right). \quad (6.5.23)$$

De plus, par définition (6.5.19) de $U_k(x)$, on a

$$\mathbb{E}[U_k^2(x)] \leq \mathbb{E}[W_k^2(x)].$$

Ainsi, grâce à (6.5.8),

$$\mathbb{E}[U_k^2(x)] \leq \frac{\nu^2}{h_k}\mathbb{E}\left[f_{\widehat{\theta}_{k-1}}(x)\right] + C_3h_k. \quad (6.5.24)$$

A partir de maintenant, on note C toute constante indépendante de n . D'une part, rappelons que (6.2.3) implique que pour tout $n \geq 0$,

$$\mathbb{E}\left[|\widehat{\theta}_n - \theta|^2\right] \leq \frac{C}{n}. \quad (6.5.25)$$

D'autre part, l'inégalité des accroissements finis entraîne que, pour tout $x \in I_1$,

$$|f_t(x) - f(x)| \leq \sup_{t \in \Theta} |\partial f_t(x)| |t - \theta|.$$

Il découle alors de (6.5.1) et (6.5.25) que

$$\sqrt{\mathbb{E}\left[|f_{\widehat{\theta}_{n-1}}(x) - f(x)|^2\right]} \leq \frac{C}{\sqrt{n}}. \quad (6.5.26)$$

Ainsi, on déduit des équations (6.5.23), (6.5.24) et (6.5.26) que

$$\mathbb{E}[|U_k(x)U_l(x)|] \leq 2\left(\sqrt{\frac{\nu^2}{h_k}\mathbb{E}\left[f_{\widehat{\theta}_{k-1}}(x)\right]} + C_3h_k\right)\left(\frac{C}{\sqrt{l}} + C_2h_l^2\right). \quad (6.5.27)$$

Finalement, puisque $f_t(x)$ est bornée, on a

$$\mathbb{E}[|U_k(x)U_l(x)|] \leq C\left(\frac{1}{\sqrt{lh_k}} + \frac{h_l^2}{\sqrt{h_k}}\right). \quad (6.5.28)$$

Cependant, si $h_n = 1/n^\alpha$, alors

$$\sum_{k=1, k < l}^n \frac{1}{\sqrt{lh_k}} = \sum_{l=2}^n \frac{1}{l^{1/2}} \sum_{k=1}^{l-1} k^{\alpha/2} \leq \sum_{l=2}^n \frac{l^{\alpha/2+1}}{l^{1/2}} \leq n^{\frac{3+\alpha}{2}}$$

et

$$\sum_{k=1, k < l}^n \frac{h_l^2}{\sqrt{h_k}} = \sum_{l=2}^n h_l^2 \sum_{k=1}^{l-1} k^{\frac{\alpha}{2}} \leq \sum_{l=2}^n \frac{l^{\frac{\alpha}{2}+1}}{l^{2\alpha}} \leq n^{2-3\frac{\alpha}{4}}.$$

Par conséquent, on déduit des deux inégalités précédentes et de (6.5.28) que

$$\frac{1}{n^2} \sum_{k=1, k < l}^n \mathbb{E}[|U_k(x)U_l(x)|] \leq C \left(n^{-\frac{1+\alpha}{2}} + n^{-3\frac{\alpha}{4}} \right) \quad (6.5.29)$$

terme qui tend vers 0 si n tend vers $+\infty$, car $0 < \alpha < 1$. De plus, l'équation (6.5.24) et la bornitude de f_t entraînent que

$$\frac{1}{n^2} \sum_{k=1}^n \mathbb{E}[U_k^2(x)] \leq C \frac{1}{n^2} \sum_{k=1}^n \frac{1}{h_k} \leq C \frac{n^{\alpha+1}}{n^2} \leq C n^{-1+\alpha} \quad (6.5.30)$$

terme qui tend aussi vers 0 si n tend vers $+\infty$, car $\alpha < 1$. Les équations (6.5.20), (6.5.29) et (6.5.30) nous permettent donc de conclure que, pour tout $x \in I_1$,

$$\lim_{n \rightarrow +\infty} V_n(x) = 0. \quad (6.5.31)$$

Ainsi, la combinaison de (6.5.13), (6.5.18) et (6.5.31) achèvent la preuve du Théorème 6.3.2. \square

6.6 Extension au cas où φ_θ n'est pas inversible

Nous nous sommes placés précédemment dans le cadre d'une fonction φ_t strictement croissante et nous avons vu que la fonction M que l'on minimisait pour estimer θ était construite sur la relation donnée, pour tout $x \in I_1$ et pour tout $t \in [a, b]$, par

$$\varphi_t^{-1} \circ \varphi_t(x) = x, \quad (6.6.1)$$

où φ_t^{-1} représente l'inverse de φ_t . Maintenant, supposons que φ_t est croissante mais pas strictement croissante. La fonction φ_t n'est plus inversible mais on peut définir son inverse généralisée. Celle-ci est donnée, pour tout $y \in I_2$, par

$$\varphi_t^{-1}(y) = \inf\{x \in I_1 : \varphi_t(x) \geq y\}. \quad (6.6.2)$$

Dans ce cas, pour tout $x \in I_1$,

$$\varphi_t^{-1} \circ \varphi_t(x) \leq x.$$

La relation (6.6.1) est donc mise en défaut et on ne peut plus utiliser le critère précédent pour estimer θ . En revanche, pour tout $y \in I_2$, on a toujours

$$\varphi_t \circ \varphi_t^{-1}(y) = y.$$

Ainsi, on peut considérer la variable aléatoire $\tilde{Z}_n(t)$ définie, pour tout $n \geq 0$ et pour tout $t \in [a, b]$, par

$$\tilde{Z}_n(t) = \varphi_t(\varepsilon_n). \quad (6.6.3)$$

En notant $\tilde{Z}(t)$ une variable aléatoire de même loi que $\tilde{Z}_n(t)$ et X une variable aléatoire de même loi que X_n , nous allons estimer θ en minimisant la fonction \tilde{M} définie, pour tout $t \in [a, b]$, par

$$\begin{aligned} \tilde{M}(t) &= \mathbb{E}[(\tilde{Z}(t) - X)^2], \\ &= \int_0^1 (\varphi_t^{-1} \circ F^{-1}(x) - \varphi_\theta^{-1} \circ F^{-1}(x))^2 dx. \end{aligned}$$

Par des manipulations simples sur les fonctions quantiles, on a, pour tout $x \in [0, 1]$,

$$\varphi_t^{-1} \circ F^{-1}(x) = F_{\tilde{Z}(t)}^{-1}(x).$$

Finalement, on aboutit à la fonction

$$\tilde{M}(t) = \int_0^1 (F_{\tilde{Z}(t)}^{-1}(x) - F_X^{-1}(x))^2 dx.$$

qui correspond à la distance de Wasserstein entre les lois de $\tilde{Z}(t)$ et de X . Sans rentrer dans les détails techniques, en utilisant exactement les mêmes arguments que précédemment, on peut mettre en oeuvre un algorithme de gradient basé sur la fonction \tilde{M} pour l'estimation de θ , et ainsi généraliser notre procédure d'estimation à des fonctions φ_t non nécessairement inversibles.

Chapitre 7

Simulations

7.1 Le modèle de translation

7.1.1 Données simulées

Convergence numérique des estimateurs

Dans un premier temps, on considère des données simulées suivant le premier modèle semi-paramétrique de translation défini, pour tout $n \geq 0$, par

$$Y_n = f(X_n - \theta) + \varepsilon_n \quad (7.1.1)$$

où $\theta = 1/10$ et la fonction périodique f est donnée, pour $p \geq 1$ et pour tout $x \in \mathbb{R}$, par

$$f(x) = \sum_{k=1}^p \cos(2k\pi x).$$

On a choisi $p = 8$ et (X_n) et (ε_n) deux suites indépendantes de variables aléatoires indépendantes de loi respectives $\mathcal{U}[-1/2, 1/2]$ et $\mathcal{N}(0, 1)$. Les données simulées se trouvent à gauche de la Figure 7.1.

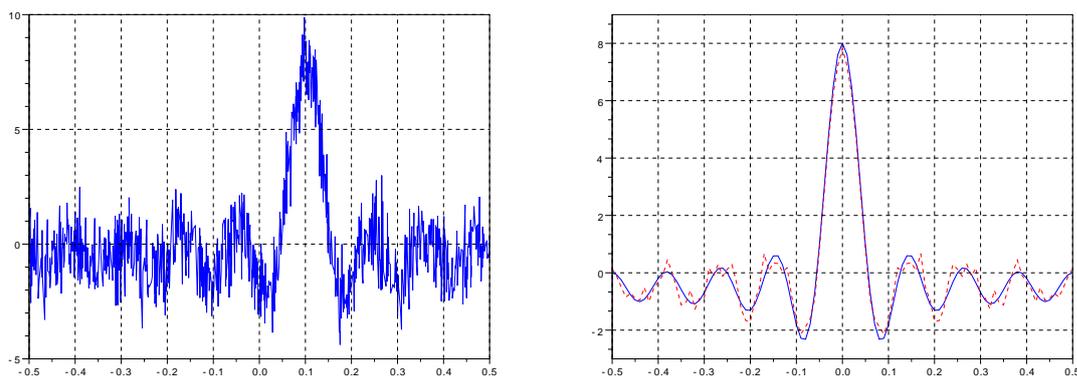


FIGURE 7.1 – Données simulées et convergence presque sûre

Pour l'estimation du paramètre θ , on met en place notre procédure de Robbins-Monro avec $n = 1000$ itérations. On obtient pour estimateur $\widehat{\theta}_n = 0.1014$ ce qui montre le bon comportement asymptotique de l'estimateur $\widehat{\theta}_n$ comparé à la vraie valeur $\theta = 1/10$. De plus, en utilisant (4.2.3), on peut obtenir des intervalles de confiance pour θ . Plus précisément, ils sont donnés, pour tout $n \geq 1$, par

$$I_n(\theta) = \left[\widehat{\theta}_n - q_\beta \frac{\widehat{\xi}_n(\theta)}{\sqrt{n}}, \widehat{\theta}_n + q_\beta \frac{\widehat{\xi}_n(\theta)}{\sqrt{n}} \right],$$

où q_β est le quantile d'ordre $0 < \beta < 1$ de la loi $\mathcal{N}(0,1)$ et où $\widehat{\xi}_n(\theta)$ est un estimateur consistant de $\xi(\theta)$ donné par (4.2.2). Dans notre cas, nous n'avons pas besoin d'estimer $\xi^2(\theta)$ puisque $f_1 = 1/2$ et

$$\xi^2(\theta) = \frac{7}{8(2\pi - 1)}.$$

De plus, pour $n = 1000$ et pour un risque $\beta = 5\%$, l'intervalle de confiance est précisément $I_n(\theta) = [0.0762; 0.1266]$. L'intervalle $I_n(\theta)$ est donc de longueur petite, on peut donc conclure que notre procédure d'estimation est satisfaisante sur ce jeu de données. On a tracé en rouge sur la gauche de la Figure 7.2 les intervalles de confiance $I_n(\theta)$, pour $n = 1, \dots, 1000$.

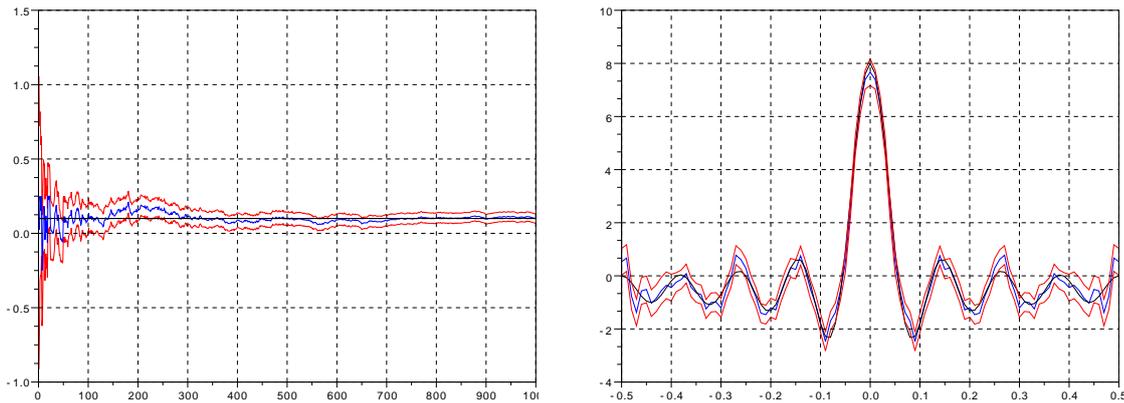


FIGURE 7.2 – Intervalles de confiance pour θ et f

En ce qui concerne l'estimation de la fonction de régression f , on choisit le noyau K uniforme sur l'intervalle $[-1,1]$, et on prend pour fenêtre de lissage $h_n = 1/n^\alpha$ with $\alpha = 9/10$. On déduit des convergences (4.3.2) et (4.3.3) que pour $n = 1000$ et pour tout $x \in [-1/2, 1/2]$, un intervalle de confiance de $f(x)$ est donné par

$$J_n(x) = \left[\widehat{f}_n(x) - q_\beta \frac{\widehat{v}_n(x, \widehat{\theta}_n)}{\sqrt{nh_n}}, \widehat{f}_n(x) + q_\beta \frac{\widehat{v}_n(x, \widehat{\theta}_n)}{\sqrt{nh_n}} \right],$$

où q_β est le quantile d'ordre $0 < \beta < 1$ de la loi $\mathcal{N}(0,1)$ tandis que $\widehat{v}_n^2(x, \widehat{\theta}_n)$ est un estimateur consistant de la variance asymptotique $v^2(x, \theta)$ du Théorème 4.3.2. Dans notre cas, $v^2 = 1/2$ et

$$v^2(x, \theta) = \begin{cases} \frac{5}{19} & \text{si } -1/2 \leq x < -2/5 \text{ ou } 2/5 < x \leq 1/2, \\ \frac{5}{38} & \text{si } -2/5 \leq x \leq 2/5 \text{ et } x \neq 0, \\ \frac{5}{19} & \text{si } x = 0. \end{cases}$$

On a tracé en rouge sur la droite de la Figure 7.2 les intervalles de confiance $J_n(x)$, pour tout $x \in [-1/2, 1/2]$. D'une part, les simulations montrent que la plus grande longueur des intervalles de confiance $J_n(x)$ est atteinte en $x = -0.47$ et $x = 0.47$ et que cette longueur est précisément égale à 1.0066. D'autre part, la plus petite longueur des intervalles de confiance $J_n(x)$ est atteinte en $x = -0.04$ et $x = 0.04$ et est égale à 0.7118. Le fait qu'il y ait deux valeurs de x pour la plus grande et la plus petite valeur des intervalles de confiance n'est pas très étonnant. Ceci est dû au fait que l'estimateur \widehat{f}_n est une fonction paire. On peut donc conclure, sur ce premier jeu de données, que les performances numériques de l'estimateur de Robbins-Monro $\widehat{\theta}_n$ de θ ainsi que celui de Nadaraya-Watson \widehat{f}_n de f sont très satisfaisantes.

Enfin, nous avons comparé les performances de l'estimateur de θ quand la densité de probabilité g associée à la suite (X_n) est connue et quand elle est inconnue. On a simulé des données suivant le modèle

$$Y_n = f(X_n - \theta) + \varepsilon_n$$

où $\theta = 0.05$ et où la fonction f est la même que précédemment. Nous avons simulé les variables (X_n) suivant la loi $\mathcal{N}(0,1)$ tronquée sur $[-1/2; 1/2]$ et le bruit (ε_n) suivant la loi $\mathcal{N}(0,1)$. On compare alors la convergence des estimateurs $\widehat{\theta}_n$ et $\widetilde{\theta}_n$ définis respectivement par

$$\widehat{\theta}_{n+1} = \pi_K(\widehat{\theta}_n + \gamma_{n+1} T_{n+1})$$

et

$$\widetilde{\theta}_{n+1} = \pi_K(\widetilde{\theta}_n + \gamma_{n+1} \widetilde{T}_{n+1})$$

où

$$T_{n+1} = \frac{\sin(2\pi(X_{n+1} - \widehat{\theta}_n))}{g(X_{n+1})} \quad \text{et} \quad \widetilde{T}_{n+1} = \frac{\sin(2\pi(X_{n+1} - \widetilde{\theta}_n))}{\widetilde{g}_n(X_{n+1})}$$

avec

$$\widetilde{g}_n(X_{n+1}) = \frac{1}{n} \sum_{k=1}^n \frac{1}{h_k} K\left(\frac{X_k - X_{n+1}}{h_k}\right),$$

où $h_n = 1/n^\alpha$ avec $\alpha = 1/5$ et où K est le noyau gaussien. L'algorithme définissant $\widetilde{\theta}_n$ a l'avantage d'être implémentable quand la densité g est inconnue. En revanche, cet algorithme n'est plus récursif et il est donc plus long à implémenter que celui définissant $\widehat{\theta}_n$. On a représenté en rouge sur la Figure 7.3 le comportement de l'algorithme $\widehat{\theta}_n$ et en bleu celui de l'algorithme $\widetilde{\theta}_n$ jusqu'à $n = 500$ itérations puis jusqu'à $n = 1000$ itérations et $n = 5000$ itérations. On peut remarquer, que même si l'estimateur $\widehat{\theta}_n$ semble être un peu plus proche de la vraie valeur de θ que l'estimateur $\widetilde{\theta}_n$, leur comportement est vraiment similaire. Ainsi, la convergence de $\widetilde{\theta}_n$ est très satisfaisante, et même si la suite n'est plus récursive, elle ne nécessite aucune connaissance de g et donc l'utilisation de $\widetilde{\theta}_n$ constitue une réelle amélioration par rapport à $\widehat{\theta}_n$.

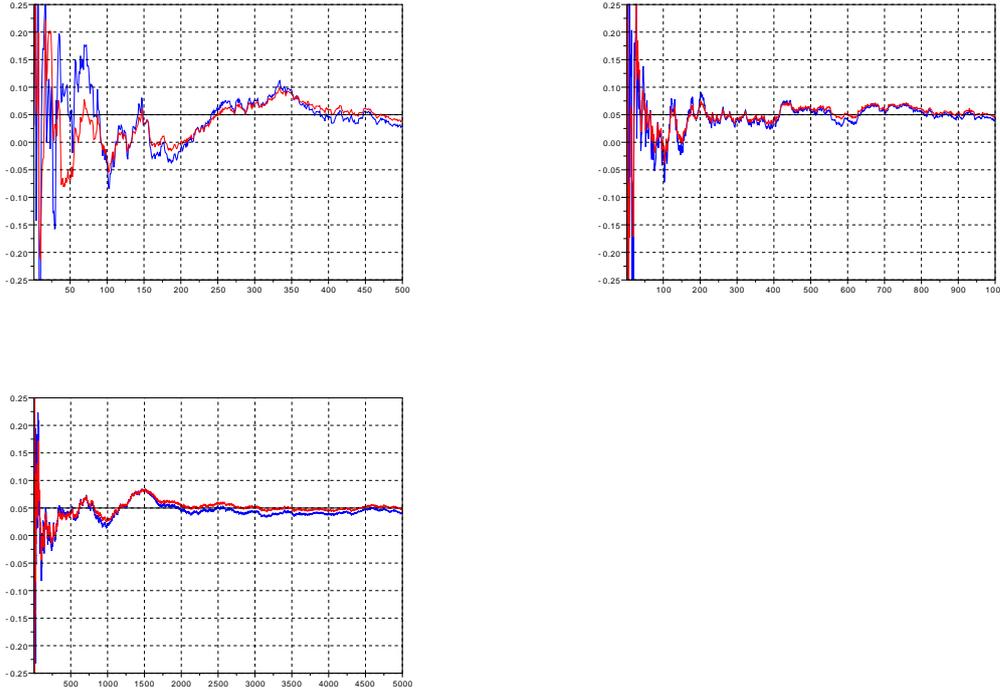


FIGURE 7.3 – Convergence de $\hat{\theta}_n$ et $\tilde{\theta}_n$ pour $n = 500$, $n = 1000$ et $n = 5000$.

Détection de décalage

Dans un deuxième temps, on considère 30 courbes simulées suivant le modèle (7.1.1)

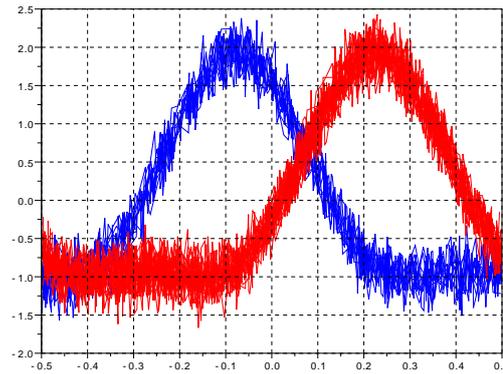
$$Y_n = f(X_n - \theta) + \varepsilon_n$$

où on choisit $\theta = -1/5$ pour les 10 premières courbes et $\theta' = 1/10$ pour les 20 dernières. La fonction f est définie, pour tout $x \in [-1/2, 1/2]$, par

$$f(x) = \cos(2\pi x) + \sin(2\pi x) + \cos(2\pi x) \sin(2\pi x).$$

Notre but est de proposer une procédure statistique permettant de détecter l'écart entre les 10 courbes correspondant à $\theta = -1/5$ et les 20 correspondant à $\theta' = 1/10$. En d'autres termes, on veut voir si la valeur $\Delta = \theta' - \theta$ est éloignée de 0. On a simulé (X_n) et (ε_n) comme deux suites indépendantes de variables aléatoires indépendantes de loi uniforme sur $[-1/2, 1/2]$ et $\mathcal{N}(0, 1/5)$. Chaque courbe est tracée avec $n = 200$ points et est représentée sur la Figure 7.4.

La procédure statistique est la suivante. Tout d'abord, on estime la première valeur $\theta = -1/5$ à partir des 10 premières courbes. On met en place notre algorithme de Robbins-Monro avec $n = 200$ itérations pour le premier estimateur $\hat{\theta}_n$ de θ évalué sur la première courbe, puis avec $n = 400$ itérations pour le second estimateur $\hat{\theta}_n$ de θ calculé avec les deux premières courbes, et ainsi de suite, jusqu'au calcul du dernier estimateur $\hat{\theta}_n$ de θ avec $n = 2000$. Ainsi, on obtient $-0,1950$ pour la moyenne des 10 premiers estimateurs $\hat{\theta}_n$ de θ . On poursuit ensuite la même procédure avec toutes les autres courbes. La valeur du

FIGURE 7.4 – Données simulées avec deux valeurs différentes θ et θ'

onzième estimateur correspondant à $n = 2\,200$ est 0,0986. Cette valeur est significativement différente des 10 premières. Elle correspond en fait à la première courbe simulée avec $\theta' = 1/10$. De plus, on obtient 0,0998 pour la moyenne sur les 20 derniers estimateurs $\widehat{\theta}_n$ de θ' . Finalement, cette procédure nous permet de détecter un changement de paramètre de la valeur $\theta = -1/5$ à $\theta' = 1/10$ puisque $\widehat{\Delta}_n = 0,0998 + 0,1950 = 0,2948$. Dans le but d'avoir des estimations $\widehat{\theta}_n$ encore plus précises, on peut remplacer le pas $\gamma_n = 1/n$ dans (4.1.4) par $\gamma_n = 1/n^a$ avec $1/2 < a < 1$, comme on l'a signalé dans la section 1.3.5.

7.1.2 Données ECG

Nous nous sommes intéressés au recalage de signaux d'électrocardiogramme (abrégé en ECG). Un signal ECG est un enregistrement de l'activité électrique du cœur sur une période de temps donnée. On enregistre cet activité grâce à des électrodes placées sur la surface extérieure de la peau. Un signal ECG standard se décompose en 3 parties : une onde P suivie d'une complexe QRS puis d'une onde T. La Figure 7.5 ci-dessous représente un tel signal.

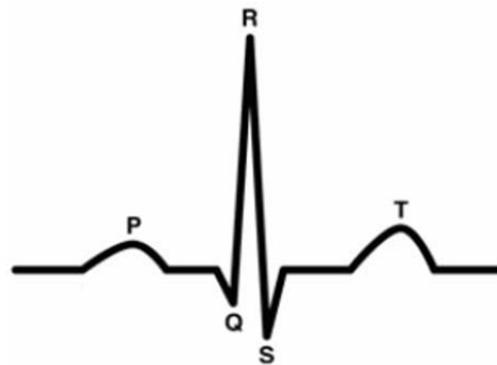


FIGURE 7.5 – Signal ECG normal

L'analyse statistique de données ECG est assez courante. On peut par exemple consulter

les travaux de Trigano *et al.* [63] ou de Bigot [5]. On renvoie également au livre de Clifford *et al.* [13] pour un panorama des méthodes statistiques pour l'analyse de données ECG. La base de données MIT-BIH rassemble un grand nombre de signaux ECG provenant de patients ayant une activité cardiaque normale ou anormale. Sur cette base de données, chaque enregistrement correspond à un signal ECG mesuré sur une heure.

L'activité cardiaque anormale la plus courante est celle appelée fibrillation auriculaire. Un indicateur de fibrillation auriculaire est notamment l'irrégularité d'intervalles RR sur l'ECG. Notre objectif est de proposer une procédure statistique pour détecter ces irrégularités sur un signal ECG. Nous allons nous appuyer sur ce que nous avons fait sur les données simulées pour détecter un décalage de courbes. Nous avons choisi l'enregistrement 04015 dans la « Atrial Fibrillation (AF) database » provenant de la base de données MIT-BIH. Le signal d'origine est représenté à gauche de la Figure 7.6. Nous voyons que le signal a deux parties bien distinctes. Nous découpons la première partie du signal de la manière suivante. On considère que le cycle d'intérêt, c'est-à-dire le cycle PQRST, est à peu près le même à chaque battement cardiaque. La segmentation de la première partie du signal ECG est faite en détectant le maximum de chaque complexe QRS puis en centrant ces segments autour du maximum. C'est ce qui est en bleu sur la droite de la Figure 7.6. Il est important d'avoir des segments de même taille pour assurer l'hypothèse de périodicité de f . En découplant le dernier complexe QRS, un autre complexe apparaît, dû à la deuxième partie du signal. C'est le signal en rouge sur la droite de la Figure 7.6. Une procédure générale de segmentation de signaux ECG bien adaptée est celle proposée par Gasser et Kneip [32].

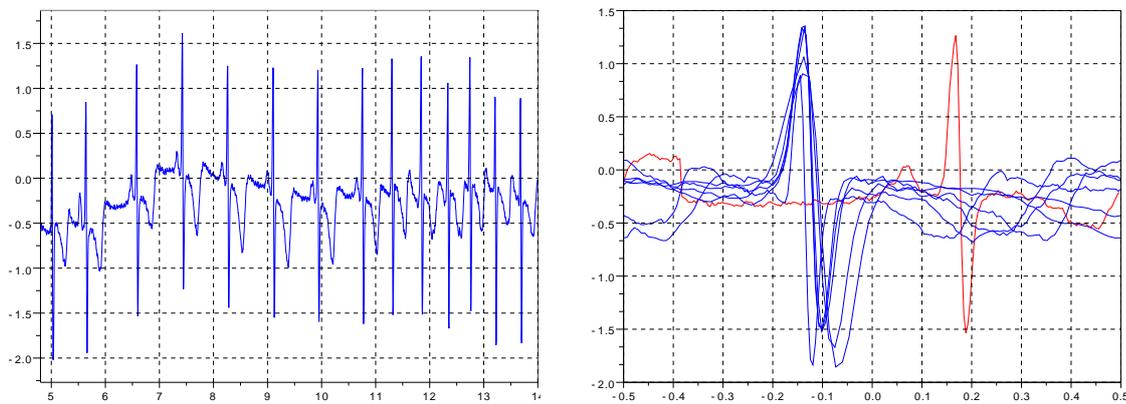


FIGURE 7.6 – Signal d'origine

La taille de l'échantillon est de 1 238 données. On suppose que le modèle de translation

$$Y_n = f(X_n - \theta) + \varepsilon_n$$

est bien adapté aux données, et que les (X_n) sont distribuées uniformément sur $[-1/2, 1/2]$. La fonction f n'est clairement pas symétrique. Cependant, nous avons vu dans la Remarque 4.2.3 que notre procédure de Robbins-Monro fonctionne encore pour une fonction de régression quelconque. Néanmoins, on a alors besoin de connaître les deux premiers

coefficients de Fourier de f , f_1 et g_1 . A la vue du signal, nous allons donc trouver deux paramètres de translation différents θ et θ' . La première valeur θ est associée à la première partie du signal, alors que la seconde θ' correspond à la deuxième partie. La différence entre les deux paramètres $\Delta = \theta' - \theta$ va donc expliquer le décalage que l'on peut observer sur la Figure 7.6. En effet, une valeur de Δ loin de 0 peut s'interpréter comme la détection d'irrégularité des intervalles RR, ce qui confirmera le diagnostic de la présence d'une fibrillation auriculaire. Sur cet enregistrement, notre procédure de Robbins-Monro sur $n = 800$ itérations mène à $\widehat{\theta}_n = -0,1734$ pour θ et le dernier $\widehat{\theta}_n = 0,1278$ pour θ' avec $n = 1238$ itérations. On obtient donc $\widehat{\Delta}_n = -0,17 - 0,1278 = -0,2978$, valeur qui explique le décalage qu'on voit sur le Figure 7.6.

7.2 Modèle à forme commune

7.2.1 Données simulées

On s'intéresse maintenant au modèle de déformation à forme commune défini, pour tout $1 \leq i \leq n$ et pour tout $1 \leq j \leq p$,

$$Y_{i,j} = a_j f(X_i - \theta_j) + v_j + \varepsilon_{i,j} \quad (7.2.1)$$

où $1 \leq j \leq p$ et $1 \leq i \leq n$ avec $p = 5$ et $n = 2000$. On prend également pour paramètre de hauteur $v = (0,1/3, -1,2, -9/10)^T$, pour paramètre de translation $\theta = (0,1/5, -1/20, -1/7,1/6)^T$ et pour paramètre d'échelle $a = (1, -4,3, -5/2, -2)^T$. De plus, le bruit $(\varepsilon_{i,j})$ est une suite de variables aléatoires indépendantes et de même loi $\mathcal{N}(0,1)$. Les temps d'observation aléatoires (X_i) sont simulés selon la loi uniforme sur $[-1/2; 1/2]$ et on choisit pour fonction de régression la fonction f définie, pour tout $x \in [-1/2; 1/2]$, par

$$f(x) = \sum_{k=1}^5 \cos(2k\pi x).$$

En particulier, le premier coefficient de Fourier de f vaut $f_1 = 1/2$. Nous avons représenté ces données sur la Figure 7.7.

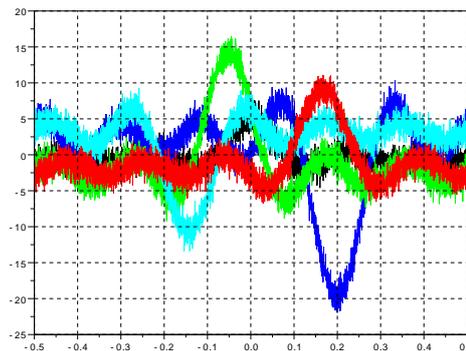


FIGURE 7.7 – Données simulées

Les résultats pour l'estimation des paramètres v , θ et a par \widehat{v}_n , $\widehat{\theta}_n$ et \widehat{a}_n sont représentés sur la Figure 7.8. On a tracé les vraies valeurs en abscisse et leur estimées en ordonnée. On voit sur la Figure 7.8 que les paramètres et leur estimés sont très proches, ce qui montrent le bon comportement numérique de notre procédure d'estimation sur ce jeu de données.

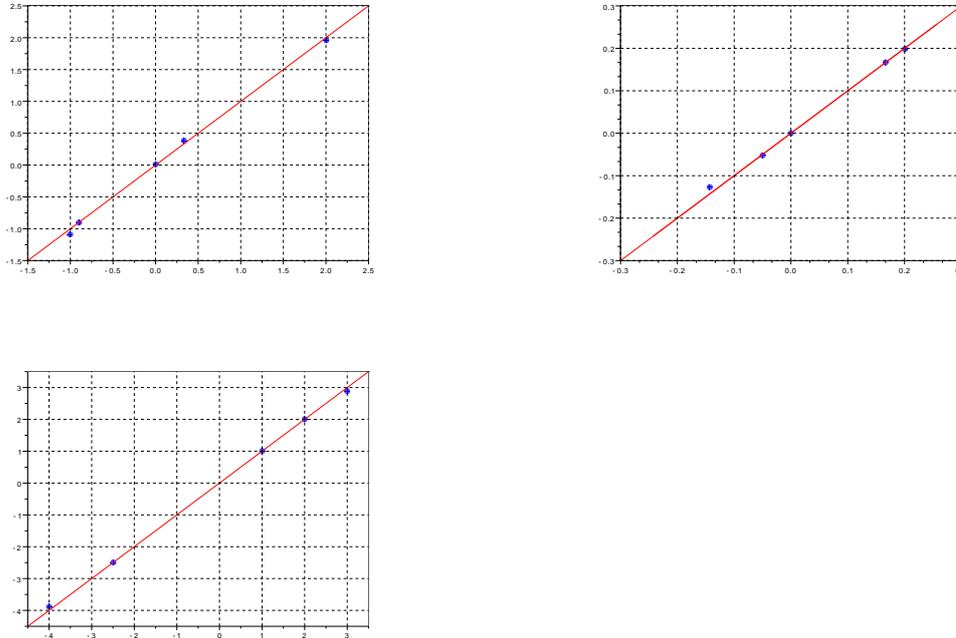
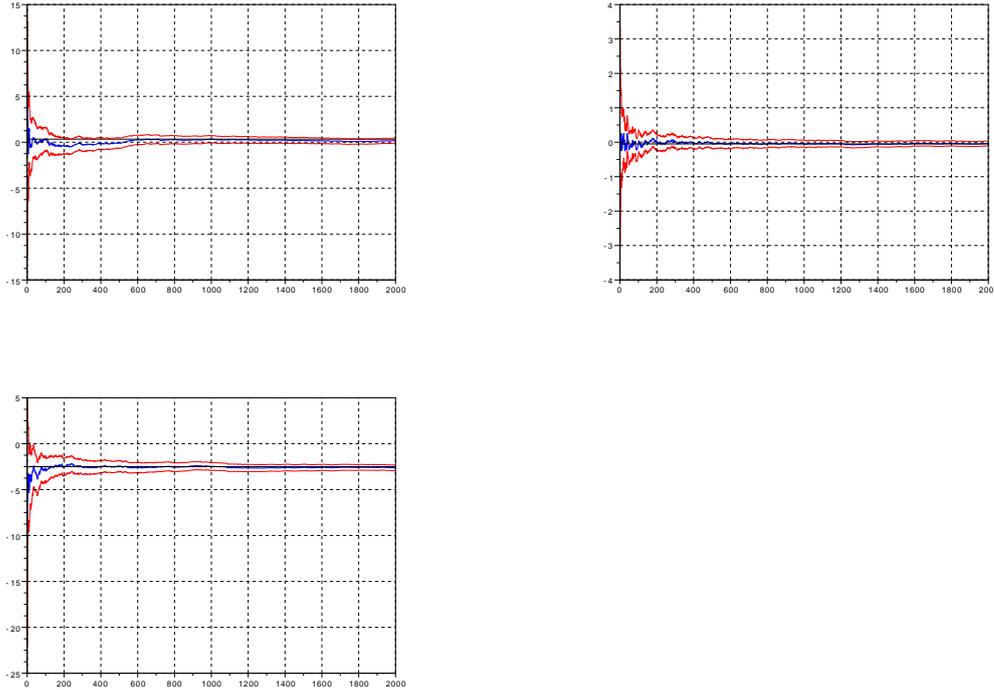


FIGURE 7.8 – Estimation de v , θ et a

De plus, en utilisant les résultats de convergence (5.2.2), (5.2.5) et (5.2.16), on peut obtenir des régions de confiance pour les paramètres v , θ et a . Par exemple, pour $n = 2000$, si on note respectivement $I_n(v_2)$, $I_n(\theta_3)$ et $I_n(a_4)$ les intervalles de confiance associés aux paramètres v_2 , θ_3 et a_4 , on a précisément que, pour un risque 5%, ces intervalles valent

$$\begin{aligned} I_n(v_2) &= [-0.1160; 0.4452], \\ I_n(\theta_3) &= [-0.1211; 0.0226], \\ I_n(a_4) &= [-2.7225; -2.1224]. \end{aligned}$$

Ces intervalles sont de longueurs respectives 0.5612, 0.1437 et 0.6001, valeurs relativement petites, ce qui confirme les bonnes performances numériques de nos estimateurs sur ce jeu de données. Nous avons représenté ces intervalles de confiance sur la Figure 7.9, pour un nombre d'itérations n variant de 1 à 2000.

FIGURE 7.9 – Intervalles de confiance de v_2 , θ_3 et a_4

Ensuite, en ce qui concerne l'estimation de la fonction de régression f , nous avons choisi le paramètre $\alpha = 9/10$ pour la fenêtre de lissage (h_n). De plus, le noyau K considéré est le noyau uniforme sur $[-1; 1]$. On a aussi choisi, pour tout $1 \leq j \leq p$, $\omega_j(x) = 1/p$. Le résultat de l'estimation de f par \widehat{f}_n est représenté sur la gauche de la Figure 7.10, alors que le résultat de l'estimation de f par $\widehat{f}_{n,1}$ est représenté sur la partie droite. De plus, on déduit des résultats de convergence (5.3.2) et (5.3.3) que, pour $n = 2000$ et pour tout $x \in [-1/2; 1/2]$, un intervalle de confiance de $f(x)$ est

$$K_n(x) = \left[\widehat{f}_n(x) - q_\beta \frac{\widehat{w}_n(x, \widehat{\theta}_n)}{\sqrt{nh_n}}, \widehat{f}_n(x) + q_\beta \frac{\widehat{w}_n(x, \widehat{\theta}_n)}{\sqrt{nh_n}} \right]$$

alors qu'on déduit des convergences (4.3.2) et (4.3.3) que, pour $n = 2000$ et pour tout $x \in [-1/2; 1/2]$, un intervalle de confiance de $f(x)$ est donné par

$$J_n(x) = \left[\widehat{f}_{n,1}(x) - q_\beta \frac{\widehat{v}_n(x, \widehat{\theta}_{n,1})}{\sqrt{nh_n}}, \widehat{f}_{n,1}(x) + q_\beta \frac{\widehat{v}_n(x, \widehat{\theta}_{n,1})}{\sqrt{nh_n}} \right],$$

où q_β représente le quantile d'ordre $0 < \beta < 1$ de la loi $\mathcal{N}(0,1)$ et où $\widehat{w}_n^2(x, \widehat{\theta}_n)$ et $\widehat{v}_n^2(x, \widehat{\theta}_{n,1})$ sont respectivement des estimateurs consistants des variances asymptotiques $w^2(x, \theta)$ données dans le Théorème 5.3.2 et $v^2(x, \theta_1)$ donnée dans le Théorème 4.3.2. Dans notre cas

particulier, $\nu^2 = 1/2$, et un calcul numérique donne

$$w^2(x, \theta) = \begin{cases} 0.0114 & \text{si } -1/2 \leq x \leq -23/50 \text{ et } 23/50 \leq x \leq 1/2, \\ 0.0108 & \text{si } -23/50 < x \leq -9/25 \text{ et } 9/25 \leq x < 23/50, \\ 0.0099 & \text{si } -9/25 < x \leq -17/50 \text{ et } 17/50 \leq x < 9/25, \\ 0.0086 & \text{si } -17/50 < x \leq -31/100 \text{ et } 31/100 \leq x < 17/50, \\ 0.0083 & \text{si } -31/100 < x < 0 \text{ et } 0 < x < 31/100, \\ 0.0166 & \text{si } x = 0, \end{cases}$$

et

$$v^2(x, \theta_1) = \begin{cases} 5/38 & \text{si } x \neq 0, \\ 5/19 & \text{si } x = 0. \end{cases}$$

Grosso modo, pour tout $x \in [-1/2; 1/2]$, la variance asymptotique $v^2(x, \theta_1)$ obtenue à partir de $\widehat{f}_{n,1}$ est dix fois plus grande que la variance asymptotique $w^2(x, \theta)$ obtenue à partir de \widehat{f}_n . De plus, dans ce cas, la variance optimale obtenue dans la Remarque 5.3.2 est, pour tout $x \in [-1/2; 1/2]$, de l'ordre de 10^{-3} ce qui est encore dix fois plus petit que $w^2(x, \theta)$. Les intervalles de confiance $K_n(x)$ et $J_n(x)$ sont représentés en rouge sur la Figure 7.11. De plus, on peut observer sur la Figure 7.10 que l'estimateur \widehat{f}_n est une meilleure approximation de f que $\widehat{f}_{n,1}$. Plus précisément, l'estimation de f par \widehat{f}_n est meilleure que l'estimation par $\widehat{f}_{n,1}$ puisque les longueurs des intervalles de confiance $K_n(x)$ sont plus petites que celles des intervalles $J_n(x)$ comme nous pouvons le voir sur la Figure 7.11. Ceci est une conséquence de l'ordre des variances $v^2(x, \theta_1)$ et $w^2(x, \theta)$. Dans notre cas, le plus grand intervalle de confiance $K_n(x)$ est de longueur 0.3460 (pour $x = 0$) et cette longueur est presque trois fois plus petite que la longueur du plus petit intervalle de confiance $J_n(x)$, celui-ci étant de longueur 0.9723 (pour $x = -0.09$ et $x = 0.09$). En particulier, cela justifie le choix de prendre comme estimateur de f une version pondérée de l'estimateur de Nadaraya-Watson.

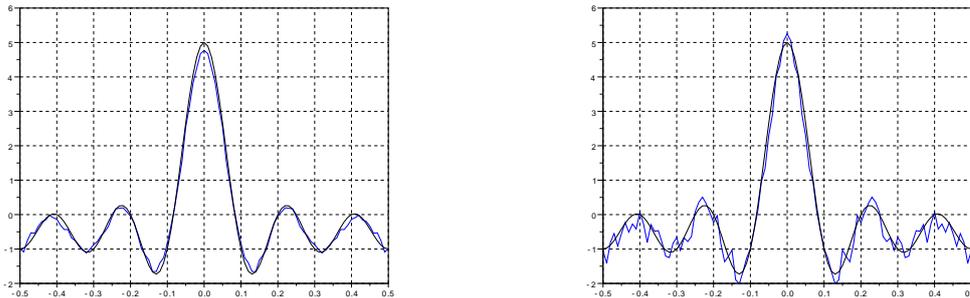
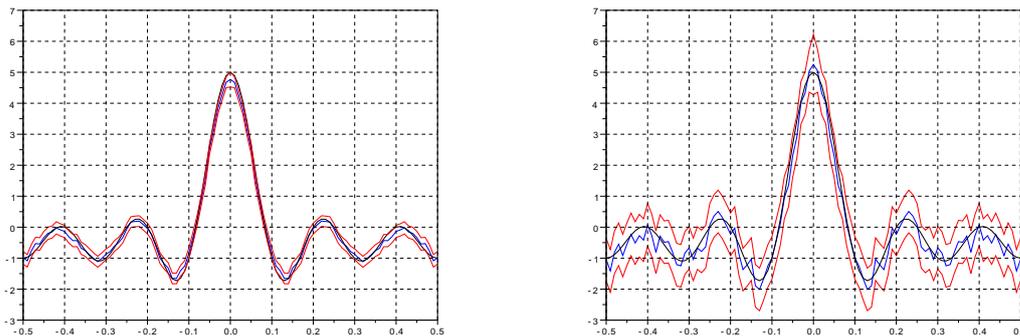


FIGURE 7.10 – Estimation de f par \widehat{f}_n et $\widehat{f}_{n,1}$

FIGURE 7.11 – Intervalles de confiance pour f

7.2.2 Données ECG

Dans la première partie sur les données ECG, on cherchait à repérer un décalage dans un signal ECG irrégulier. On s'intéresse maintenant à l'estimation du comportement moyen d'un signal ECG, ou formellement à l'estimation de f dans le modèle de déformation à forme commune (7.2.1). On a représenté à gauche de la Figure 7.12 un signal ECG d'un patient sain et à droite un signal ECG d'un patient qui présente une anomalie cardiaque.

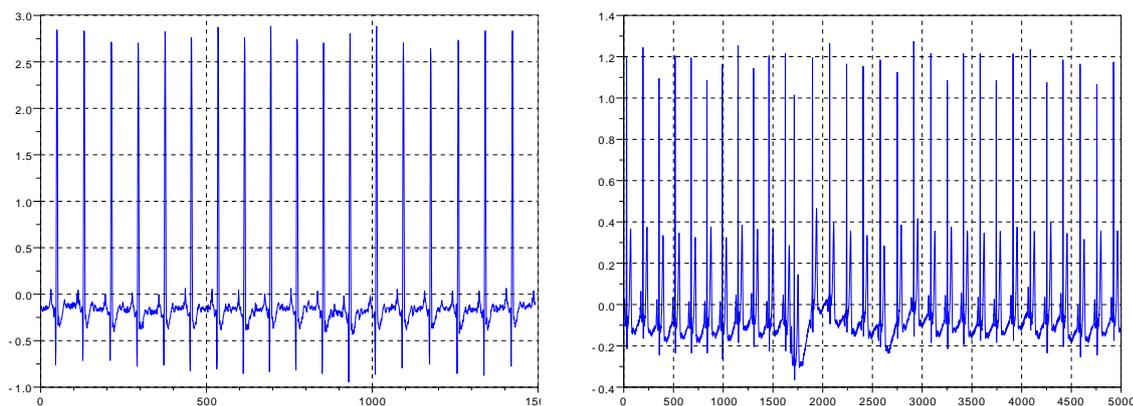


FIGURE 7.12 – ECG d'un patient sain et d'un patient malade

Comme pour la première partie sur les données ECG, on considère que le cycle d'intérêt, c'est-à-dire le cycle PQRST, est à peu près le même à chaque battement cardiaque. Après segmentation des deux signaux ECG, on observe des signaux de même longueur et chacun contient un unique cycle PQRST. On obtient alors, pour le coeur sain, $p = 18$ segments, chacun étant de taille $n = 83$, et pour le coeur malade, $p = 15$ et $n = 91$. Notre objectif est d'estimer la forme commune f de chaque ECG, correspondant à la fonction f du modèle de déformation (7.2.1). Tout d'abord, on va commencer par montrer que l'estimation de f pour le coeur sain nous permet de dire que notre modèle est bien adapté au problème de modéliser un signal ECG. En effet, pour le coeur sain, une bonne approximation du

cycle cardiaque est de faire la moyenne des p signaux différents, puisque les différents paramètres v , θ et a du modèle (7.2.1) sont triviaux. Le résultat pour l'estimation de la forme commune de l'ECG pour le coeur sain par notre procédure est représenté à droite de la Figure 7.13, alors qu'on a représenté à gauche le signal ECG d'origine. La comparaison entre les deux Figures nous permet de dire que notre procédure d'estimation est bien adaptée à la modélisation de signaux ECG.

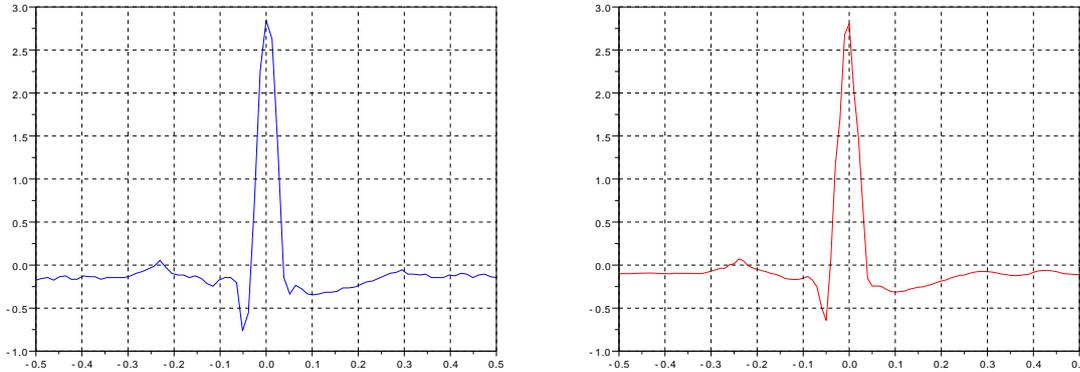


FIGURE 7.13 – Un signal et sa reconstruction

Ensuite, on veut estimer la forme commune pour le coeur malade dont le signal ECG est représenté à droite de la Figure 7.12. On peut voir que l'activité électrique est plus irrégulière que pour un coeur sain et qu'une simple moyenne risque donc de ne pas correspondre à une bonne modélisation du comportement moyen. Ainsi, on suppose que le modèle de déformation à forme commune (7.2.1)

$$Y_{i,j} = a_j f(X_i - \theta_j) + v_j + \varepsilon_{i,j},$$

est bien adapté aux données. La procédure de segmentation nous permet d'avoir une fonction f 1-périodique. De plus, f n'est pas paire, mais nous avons déjà vu que notre procédure d'estimation est encore valable pour une fonction f quelconque. Les paramètres a , θ et v correspondent à la déformation due à l'arythmie liée à la forme commune f que nous voulons estimer. Pour être plus précis, nous devons choisir une courbe de référence parmi les p courbes, c'est-à-dire un $1 \leq j^* \leq p$ tel que $a_{j^*} = 1$, $\theta_{j^*} = 0$ et $v_{j^*} = 0$. Pour ce choix, on considère un critère de variance résiduelle. Plus précisément, on commence par considérer le modèle (7.2.1) avec pour référence la première courbe soit $a_1 = 1$, $\theta_1 = 0$ and $v_1 = 0$. A partir de ce modèle, on applique notre procédure d'estimation pour estimer a , θ et v et la fonction f par \hat{a}_n , $\hat{\theta}_n$, \hat{v}_n et \hat{f}_n , respectivement. A partir de ces estimées, on calcule le vecteur $\hat{\sigma}_{n,1}^2$ dont la j -ème coordonnée $(\hat{\sigma}_{n,1}^2)_j$ est définie par

$$(\hat{\sigma}_{n,1}^2)_j = \frac{1}{n} \sum_{i=1}^n (Y_{i,j} - \hat{a}_{n,j} \hat{f}_n(X_i - \hat{\theta}_{n,j}) - \hat{v}_{n,j})^2.$$

On recommence ensuite la procédure d'estimation en changeant la courbe de référence. Finalement, on obtient donc p vecteurs de taille p qui sont $\hat{\sigma}_{n,1}^2, \dots, \hat{\sigma}_{n,p}^2$. La courbe de

référence que l'on choisit pour modéliser le signal ECG est alors celle réalisant

$$j^* = \arg \min_{1 \leq j \leq p} \|\widehat{\sigma}_{n,j}^2\|_1$$

où $\|\cdot\|_1$ correspond à la norme l^1 . Ainsi, on modélise le signal ECG par le modèle de déformation à forme commune

$$Y_{i,j} = a_j f(X_i - \theta_j) + v_j + \varepsilon_{i,j}$$

avec $a_{j^*} = 1$, $\theta_{j^*} = 0$, $v_{j^*} = 0$. Sur notre jeu de données, l'implémentation de cette méthode entraîne que $j^* = 3$ et $\|\widehat{\sigma}_{n,j^*}^2\|_1 = 0.6095$. Le résultat pour l'estimation de la forme commune f est représenté sur la Figure 7.14. À droite de la Figure 7.14 nous pouvons comparer l'estimation de f par notre estimateur \widehat{f}_n avec une simple moyenne du signal. Nous pouvons en particulier observer que notre estimation est meilleure puisqu'elle permet de mieux estimer l'onde P.

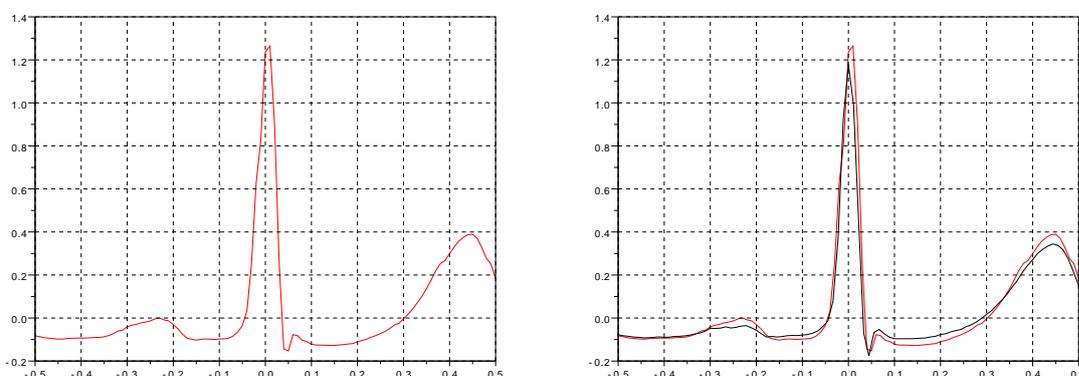


FIGURE 7.14 – Estimation de f par \widehat{f}_n (en rouge) et par la moyenne du signal (en noir)

7.3 Déformation de variables aléatoires

On considère maintenant le modèle de déformation de variables aléatoires défini, pour tout $n \geq 0$, par

$$X_n = \varphi_\theta(\varepsilon_n). \quad (7.3.1)$$

On illustre numériquement les propriétés asymptotiques de l'estimateur $\widehat{\theta}_n$ défini par (6.1.9) et sa version excitée $\widetilde{\theta}_n$ définie par (6.1.11). Les transformations φ_θ qui sont inversibles en θ n'ont pas d'intérêt puisque, dans ce cas, il est possible d'exprimer θ en fonction des variables $X_0, \dots, X_n, \varepsilon_0, \dots, \varepsilon_n$. En revanche, quand φ_θ n'est pas inversible en θ , notre procédure a tout son sens et elle est utile pour estimer θ . Parmi les nombreuses transformations intéressantes, on choisit de se focaliser sur l'une d'entre elles utilisée en Économétrie. Plus précisément, on regarde les performances de notre estimateur pour la

transformation de Box-Cox φ_t qui est définie, pour tout $t \in \mathbb{R}$ et pour tout $x \in \mathbb{R}_*^+$, par

$$\varphi_t(x) = \begin{cases} \frac{x^t - 1}{t} & \text{si } t \neq 0 \\ \log(x) & \text{si } t = 0 \end{cases} \quad (7.3.2)$$

Pour les simulations, on suppose que $\theta > 0$ et, plus particulièrement, on suppose que $\theta \in]a; b[$ avec $a = 1/10$ et $b = 2$. La transformation φ_t est alors inversible sur $]1; +\infty[$ et à valeurs dans \mathbb{R}_*^+ . De plus, la transformations inverse φ_t^{-1} de φ_t est donnée pour tout $x \in \mathbb{R}_*^+$, par

$$\varphi_t^{-1}(x) = (1 + tx)^{1/t}. \quad (7.3.3)$$

Il est alors clair que, pour tout $t \in [a; b]$, $\varphi_t^{-1}(x)$ est continûment dérivable en t et que sa dérivée est donnée, pour tout $x \in \mathbb{R}_*^+$, par

$$\partial \varphi_t^{-1}(x) = \frac{1}{t} \left(\frac{x}{1+tx} - \frac{1}{t} \log(1+tx) \right) (1+tx)^{1/t}. \quad (7.3.4)$$

On note M la fonction de Wasserstein définie par (6.0.7) associée à φ_t . On illustre la procédure d'estimation pour $\theta = 1$. On a tracé la fonction M sur la Figure 7.15 ci-dessous.

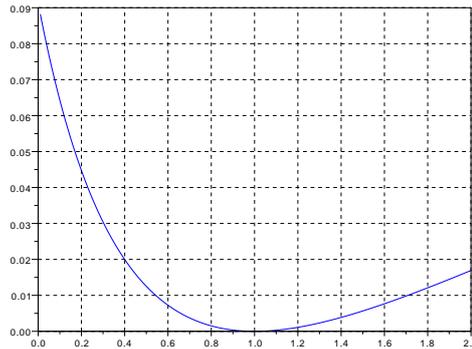


FIGURE 7.15 – Distance de Wasserstein pour la transformation de Box-Cox

On peut voir en particulier que θ est effectivement un minimum global de M . Pour l'estimation de θ , on choisit de générer des variables (ε_n) selon la loi uniforme sur $[1; 2]$. On simule alors (X_n) suivant le modèle de déformation (7.3.1)

$$X_n = \varphi_\theta(\varepsilon_n).$$

Pour un pas $\gamma_n = 1/n$, on simule la suite $(\widehat{\theta}_n)$ suivant (6.1.9). Plus précisément,

$$\widehat{\theta}_{n+1} = \pi_{[a;b]}(\widehat{\theta}_n - \gamma_n T_{n+1})$$

où

$$T_{n+1} = -2\partial \varphi_{\widehat{\theta}_n}^{-1}(X_{n+1}) \left(\varepsilon_{n+1} - \varphi_{\widehat{\theta}_n}^{-1}(X_{n+1}) \right),$$

où $\varphi_{\tilde{\theta}_n}^{-1}$ est donnée par (7.3.3) et où $\partial\varphi_{\tilde{\theta}_n}^{-1}$ est donnée par (7.3.4). On simule aussi la suite $(\tilde{\theta}_n)$ suivant (6.1.11). Plus précisément,

$$\tilde{\theta}_{n+1} = \pi_{[a;b]}(\tilde{\theta}_n - \gamma_n \tilde{T}_{n+1})$$

où

$$\tilde{T}_{n+1} = -2\partial\varphi_{\tilde{\theta}_n}^{-1}(X_{n+1})\left(\varepsilon_{n+1} - \varphi_{\tilde{\theta}_n}^{-1}(X_{n+1})\right) + V_{n+1},$$

où la suite (V_n) est une suite exogène de variables aléatoires indépendantes et de même loi $\mathcal{N}(0,1/2)$. Les valeurs de $\hat{\theta}_n$ et de $\tilde{\theta}_n$ sont calculées jusqu'à $n = 2000$. On a représenté à gauche de la Figure 7.16 la convergence de la suite $(\hat{\theta}_n)$ et à droite, la convergence de la suite $(\tilde{\theta}_n)$. En particulier, on obtient que $|\hat{\theta}_{2000} - \theta| = 0.0018$ et $|\tilde{\theta}_{2000} - \theta| = 0.0036$ ce qui montre les bonnes performances numériques des deux estimateurs.

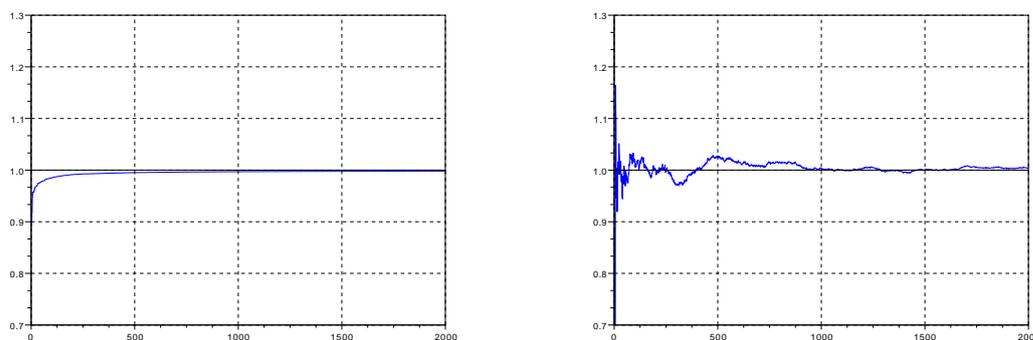


FIGURE 7.16 – Convergence de $\hat{\theta}_n$ et de $\tilde{\theta}_n$.

De plus, sur la Figure 7.17, on a représenté la normalité asymptotique (6.2.2) de l'estimateur excité $(\tilde{\theta}_n)$. Pour ce faire, on a réalisé 200 réalisations de la variable $\sqrt{2000}(\tilde{\theta}_{2000} - \theta)$. Comme prévu, on obtient bien un comportement gaussien non dégénéré.

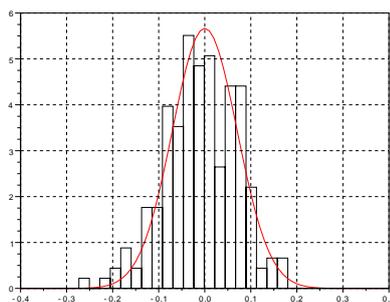


FIGURE 7.17 – Normalité asymptotique de $\sqrt{n}(\tilde{\theta}_n - \theta)$.

Conclusion et perspectives

Dans cette thèse, on a étudié trois modèles de déformation semi-paramétriques. Pour chaque modèle, on a proposé des estimateurs récursifs des paramètres et de la fonction de lien et on a établi la convergence presque sûre de nos estimateurs ainsi que leur normalité asymptotique. Pour le premier modèle de translation, on a proposé un algorithme stochastique récursif double pour l'estimation du paramètre de translation et de la fonction de lien. L'avantage de cet algorithme est qu'il est très facile à implémenter. Ensuite, on a généralisé notre procédure d'estimation à un modèle vectoriel de déformation à forme commune. L'apport principal de cette procédure d'estimation par rapport à celle sur le premier modèle de translation est que l'on est capable d'estimer un paramètre d'échelle. Enfin, on a étudié le modèle de déformation de variables aléatoires dans le cadre où la déformation est connue à un paramètre réel près. L'intérêt d'étudier ce modèle aléatoire est double. D'une part, on a une caractérisation de la fonction de contraste à minimiser en terme de distance de Wasserstein, distance naturelle dans l'étude de distance entre lois de probabilité. D'autre part, on peut dériver de l'estimateur du paramètre un estimateur de gradient stochastique. On a également vu le bon comportement numérique de nos estimateurs à la fois sur des données simulées et des données réelles.

Voici quelques perspectives de recherche dans la continuité de ce travail de thèse.

Etude semi-paramétrique d'un certain processus autorégressif fonctionnel

Une première généralisation « naturelle » du modèle de translation semi-paramétrique (4.0.1) serait d'étudier l'estimation de θ et de f dans le processus autorégressif fonctionnel périodique défini, pour tout $n \geq 0$, par

$$X_{n+1} = f(X_n - \theta) + \varepsilon_{n+1}. \quad (7.3.5)$$

Il existe déjà une littérature assez fournie sur l'étude du modèle non paramétrique

$$X_{n+1} = f(X_n) + \varepsilon_{n+1}.$$

Parmi les premiers à s'être intéressés à ce problème, on peut notamment citer Doukhan et Ghindès [22]. De plus, Senoussi [62] a proposé un estimateur à noyau récursif de f et a établi la convergence presque-sûre ainsi qu'un théorème central limite pour cet estimateur. Duflo [23] a également étudié ce processus sous des hypothèses de mélange. Delyon et Juditsky se sont quant à eux intéressés au cadre multidimensionnel dans [21]. Enfin, Hilgert et Portier [37] ont étudié le cas où le processus X_n n'est pas directement observé. Pour le modèle (7.3.5), l'idée serait de proposer un estimateur récursif pour θ qui ne dépende pas d'une

évaluation préalable de la fonction de lien f . Si $\widehat{\theta}_n$ est un tel estimateur de θ , un estimateur naturel de f serait alors l'estimateur à noyau défini, pour tout $n \geq 1$, par

$$\widehat{f}_n(x) = \frac{\sum_{k=1}^n \frac{1}{h_k} K\left(\frac{X_k - x - \widehat{\theta}_{k-1}}{h_k}\right) X_{k+1}}{\sum_{k=1}^n \frac{1}{h_k} K\left(\frac{X_k - x - \widehat{\theta}_{k-1}}{h_k}\right)}.$$

Je travaille actuellement sur le sujet.

Extension des chapitres 4 et 5 à un paramètre θ aléatoire

Une extension des chapitres 4 et 5 serait d'étudier le modèle défini, pour tout $1 \leq i \leq n$ et pour tout $1 \leq j \leq p$, par

$$Y_{i,j} = f(X_i - \theta_j) + \varepsilon_{i,j}, \quad (7.3.6)$$

où les θ_j sont des paramètres de translation indépendants et de même loi de densité φ . Ce modèle apparaît souvent en biologie ou économétrie, où les issues d'une expérience dépendent d'une variable aléatoire θ qui modélise le cas où les données prennent en compte la variabilité de chaque individu j : chaque sujet j peut réagir de manière différente par rapport à un comportement moyen. Estimer φ nous permettrait alors de comprendre ce comportement moyen. Ce type de modèle a déjà été étudié par Trigano *et al.* [63] ou encore Castillo et Loubès [10].

Adaptation du chapitre 6 à des observations bruitées

Dans le chapitre 6, nous avons étudié le modèle de déformation de variables aléatoires défini, pour tout $n \geq 0$, par

$$X_n = \varphi_\theta(\varepsilon_n). \quad (7.3.7)$$

On supposait que l'on observait les couples (X_n, ε_n) . Nous avons développé une méthode de gradient basé sur la distance de Wasserstein entre ε_n et $\varphi_t^{-1}(X_n)$ quand φ_t est inversible, et nous avons vu que l'on pouvait adapter notre méthode au cas plus général où φ_t n'est plus inversible, en considérant cette fois la distance de Wasserstein entre X_n et $\varphi_t(\varepsilon_n)$. Une première extension naturelle de ce travail serait de modifier notre critère pour prendre en compte des observations bruitées de X_n et de ε_n . Par exemple, on pourrait étudier l'estimation de θ dans les modèles

$$X_n = \varphi_\theta(\varepsilon_n + V_n)$$

ou

$$X_n = \varphi_\theta(\varepsilon_n) + V_n$$

où la suite (V_n) est une suite de variables aléatoires centrées indépendantes et de même loi. Une adaptation de la méthode de gradient stochastique donnée par les équations (6.1.11) et (6.1.12) est très certainement possible pour l'estimation de θ dans ces deux modèles, ce qui permettrait aussi de donner un réel sens statistique aux équations (6.1.11) et (6.1.12).

Bibliographie

- [1] Michel BENAÏM : A dynamical system approach to stochastic approximations. *SIAM J. Control Optim.*, 34(2):437–472, 1996.
- [2] Albert BENVENISTE, Pierre PRIOURET et Michel MÉTIVIER : *Adaptive algorithms and stochastic approximations*. Springer-Verlag New York, Inc., New York, NY, USA, 1990.
- [3] B. BERCU et D. CHAFAÏ : *Modélisation stochastique et simulation*. Mathématiques appliquées pour le Master/SMIAI. Dunod, Paris, 2007.
- [4] B. BERCU et P. FRAYSSE : A robbins-monro procedure for estimation in semiparametric regression models. *Ann. Statist.*, 40:666–693, 2012.
- [5] Jérémie BIGOT : Frechet means of curves for signal averaging and application to ecg data analysis. *soumis*, 2011.
- [6] David BLONDIN : *Lois limites uniformes et estimation non-paramétrique de la régression*. Thèse de doctorat, Université Paris 6, 2004.
- [7] Julius R. BLUM : Approximation methods which converge with probability one. *Ann. Math. Statistics*, 25:382–386, 1954.
- [8] Julius R. BLUM : Multidimensional stochastic approximation methods. *Ann. Math. Statistics*, 25:737–744, 1954.
- [9] Odile BRANDIÈRE et Marie DUFLO : Les algorithmes stochastiques contournent-ils les pièges? *Ann. Inst. H. Poincaré Probab. Statist.*, 32(3):395–427, 1996.
- [10] I. CASTILLO et J.-M. LOUBES : Estimation of the distribution of random shifts deformation. *Math. Methods Statist.* 18, 1:21–42, 2009.
- [11] F. CHAABANE et F. MAAOUIA : Théorèmes limites avec poids pour les martingales vectorielles. *ESAIM PS*, 4:137–189, 2000.
- [12] Han Fu CHEN et Yun Min ZHU : Stochastic approximation procedures with randomly varying truncations. *Sci. Sinica Ser. A*, 29(9):914–926, 1986.
- [13] G. D. CLIFFORD, F. AZUAJE et P. MCSHARRY : *Advanced methods and tools for ECG data analysis*. Artech House, Boston, London, 2006.
- [14] Gérard COLLOMB : Quelques propriétés de la méthode du noyau pour l'estimation non paramétrique de la régression en un point fixé. *C. R. Acad. Sci. Paris Sér. A-B*, 285(4):A289–A292, 1977.
- [15] Juan Antonio CUESTA et Carlos MATRÁN : Notes on the Wasserstein metric in Hilbert spaces. *Ann. Probab.*, 17(3):1264–1276, 1989.

- [16] A. S. DALALYAN, G. K. GOLUBEV et A. B. TSYBAKOV : Penalized maximum likelihood and semiparametric second-order efficiency. *Ann. Statist.* 34, 1:169–201, 2006.
- [17] Paul DEHEUVELS : Sur l'estimation séquentielle de la densité. *C. R. Acad. Sci. Paris Sér. A-B*, 276:A1119–A1121, 1973.
- [18] Paul DEHEUVELS : Sur une famille d'estimateurs de la densité d'une variable aléatoire. *C. R. Acad. Sci. Paris Sér. A-B*, 276:A1013–A1015, 1973.
- [19] Paul DEHEUVELS : Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité. *C. R. Acad. Sci. Paris Sér. A*, 278:1217–1220, 1974.
- [20] Bernard DELYON : General results on the convergence of stochastic algorithms. *IEEE Trans. Automat. Control*, 41(9):1245–1255, 1996.
- [21] Bernard DELYON et Anatoli JUDITSKY : On minimax identification of nonparametric autoregressive models. *Probab. Theory Related Fields*, 116(1):21–39, 2000.
- [22] Paul DOUKHAN et Marcel GHINDÈS : Estimations dans le processus " $X_{n+1} = f(X_n) + \varepsilon_n''$ ". *C. R. Acad. Sci. Paris Sér. A-B*, 291(1):A61–A64, 1980.
- [23] M. DUFLO : *Random iterative models*, volume 34 de *Applications of Mathematics*. Springer-Verlag, Berlin, 1997.
- [24] Marie DUFLO : *Algorithmes stochastiques*, volume 23 de *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 1996.
- [25] V. FABIAN : Asymptotically efficient stochastic approximation; the rm case. *Ann. Statist.*, 1:486–495, 1973.
- [26] Jean-Claude FORT et Gilles PAGÈS : Convergence of stochastic algorithms : from the Kushner-Clark theorem to the Lyapounov functional method. *Adv. in Appl. Probab.*, 28(4):1072–1094, 1996.
- [27] Lescornel H. FRAYSSE, P. et J.-M. LOUBÈS : A robbins monro procedure for the estimation of parametric deformation between random variables. *soumis*, 2013.
- [28] Philippe FRAYSSE : Recursive estimation for a class of models of deformation. *soumis*, 2012.
- [29] N. FRIKHA et S. MENOZZI : Concentration bounds for stochastic approximations. *Elec. com. in proba.*, 17:1–15, 2012.
- [30] F. GAMBOA, J.-M. LOUBES et E. MAZA : Semi-parametric estimation of shifts. *Electron. J. Stat.*, 1:616–640, 2007.
- [31] V. F. GAPOŠKIN et T. P. KRASULINA : The law of the iterated logarithm in stochastic approximation processes. *Teor. Veroyatnost. i Primenen.*, 19:879–886, 1974.
- [32] Theo GASSER et Alois KNEIP : Searching for structure in curve sample. *Journal of the American Statistical Association*.
- [33] P. HALL et C. C. HEYDE : *Martingale limit theory and its application*. Academic Press Inc. New York, 1980.
- [34] Peter HALL : Limit theorems for stochastic measures of the accuracy of density estimators. *Stochastic Process. Appl.*, 13(1):11–25, 1982.
- [35] Peter HALL : Central limit theorem for integrated square error of multivariate non-parametric density estimators. *J. Multivariate Anal.*, 14(1):1–16, 1984.

- [36] Wolfgang HÄRDLE : *Applied nonparametric regression*, volume 19 de *Econometric Society Monographs*. Cambridge University Press, Cambridge, 1990.
- [37] Nadine HILGERT et Bruno PORTIER : Estimation non paramétrique dans un modèle autorégressif fonctionnel non directement observé. *C. R. Acad. Sci. Paris Sér. I Math.*, 327(6):597–600, 1998.
- [38] Gopinath KALLIANPUR : A note on the Robbins-Monro stochastic approximation method. *Ann. Math. Statistics*, 25:386–388, 1954.
- [39] J. KIEFER et J. WOLFOWITZ : Stochastic estimation of the maximum of a regression function. *Ann. Math. Statistics*, 23:462–466, 1952.
- [40] H. J. KUSHNER et G. G. YIN : *Stochastic approximation and recursive algorithms and applications*, volume 35 de *Applications of Mathematics*. Springer-Verlag, New York, 2003.
- [41] Harold Joseph KUSHNER et Dean S. CLARK : *Stochastic approximation methods for constrained and unconstrained systems*. Applied mathematical sciences. Springer, New York, 1978. Includes index.
- [42] W. H. LAWTON, E. A. SYLVESTRE et M. S. MAGGIO : Self modeling nonlinear regression. *Technometrics*, 14:513–532, 1972.
- [43] Alain LE BRETON et Alexander NOVIKOV : Some results about averaging in stochastic approximation. *Metrika*, 42(3-4):153–171, 1995. Second International Conference on Mathematical Statistics (Smolenice Castle, 1994).
- [44] Jérôme LELONG : Almost sure convergence for randomly truncated stochastic algorithms under verifiable conditions. *Statist. Probab. Lett.*, 78(16):2632–2636, 2008.
- [45] Jérôme LELONG : *Etude asymptotique des algorithmes stochastiques et calcul du prix des options Parisiennes*. Thèse de doctorat, Ecole Nationale des Ponts et Chaussées, 2007.
- [46] Hélène LESCORNEL et Jean-Michel LOUBES : Estimation of deformations between distributions by minimal Wasserstein distance.
- [47] Lennart LJUNG : Analysis of recursive stochastic algorithms. *IEEE Trans. Automatic Control*, AC-22(4):551–575, 1977.
- [48] Abdelkader MOKKADEM, Mariane PELLETIER et Yousri SLAOUI : Revisiting Révész's stochastic approximation method for the estimation of a regression function. *ALEA Lat. Am. J. Probab. Math. Stat.*, 6:63–114, 2009.
- [49] È. A. NADARAJA : On a regression estimate. *Teor. Veroyatnost. i Primenen.*, 9:157–159, 1964.
- [50] È. A. NADARAYA : *Nonparametric estimation of probability densities and regression curves*, volume 20 de *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1989. Translated from the Russian by Samuel Kotz.
- [51] M. PELLETIER : On the almost sure asymptotic behaviour of stochastic algorithms. *Stochastic Process. Appl.* 78, 2:217–244, 1998.
- [52] Mariane PELLETIER : Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing. *Ann. Appl. Probab.*, 8(1):10–44, 1998.

- [53] Mariane PELLETIER : Asymptotic almost sure efficiency of averaged stochastic algorithms. *SIAM J. Control Optim.*, 39(1):49–72 (electronic), 2000.
- [54] B. T. POLYAK et A. B. JUDITSKY : Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.*, 30(4):838–855, 1992.
- [55] P. RÉVÉSZ : How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Math. Operationsforsch. Statist. Ser. Statist.*, 8(1):119–126, 1977.
- [56] H. ROBBINS et S. MONRO : A stochastic approximation method. *Ann. Math. Statistics*, 22:400–407, 1951.
- [57] H. ROBBINS et D. SIEGMUND : A convergence theorem for non negative almost supermartingales and some applications. *Optimizing methods in stat.*, pages 233–257, 1971.
- [58] David RUPPERT : Almost sure approximations to the Robbins-Monro and Kiefer-Wolfowitz processes with dependent noise. *Ann. Probab.*, 10(1):178–187, 1982.
- [59] RUPPERT, D. : Efficient estimators from a slowly convergent Robbins-Monro process. Rapport technique 781, School of Operations Research and Industrial Engineering, Cornell University, 1985.
- [60] Jerome SACKS : Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statist.*, 29:373–405, 1958.
- [61] Eugene F. SCHUSTER : Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Statist.*, 43:84–88, 1972.
- [62] R. SENOSSI : Uniform iterated logarithm laws for martingales and their application to functional estimation in controlled Markov chains. *Stochastic Process. Appl.*, 89(2):193–211, 2000.
- [63] T. TRIGANO, U. ISSERLES et Y. RITOV : Semiparametric curve alignment and shift density estimation for biological data. *IEEE Trans. Signal Processing*, 59:1970–1984, 2011.
- [64] Alexandre B. TSYBAKOV : *Introduction à l'estimation non-paramétrique*, volume 41 de *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 2004.
- [65] Alexandre B. TSYBAKOV : *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.
- [66] Myriam VIMOND : Efficient estimation for a subclass of shape invariant models. *Ann. Statist.*, 38(3):1885–1912, 2010.
- [67] Yuedong WANG, Chunlei KE et Morton B. BROWN : Shape-invariant modeling of circadian rhythms with random effects and smoothing spline ANOVA decompositions. *Biometrics*, 59(4):804–812, 2003.
- [68] Geoffrey S. WATSON : Smooth regression analysis. *Sankhyā Ser. A*, 26:359–372, 1964.
- [69] Charles T. WOLVERTON et Terry J. WAGNER : Asymptotically optimal discriminant functions for pattern classification. *IEEE Trans. Information Theory*, IT-15:258–265, 1969.

- [70] Hajime YAMATO : Sequential estimation of a continuous probability density function and mode. *Bull. Math. Statist.*, 14:1–12 ; correction, *ibid.* 15 (1972), 133, 1970/71.