

N° d'ordre :38/2022.C/MT

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université des Sciences et de la Technologie Houari Boumediène

Faculté de Mathématiques



THÈSE DE DOCTORAT

Présentée pour l'obtention du Diplôme de Doctorat

En : MATHÉMATIQUES

Spécialité : Mathématiques Financières et Actuariat

Par : RIH Soumia

Sujet

**Sur l'inférence dans les modèles de données
faiblement dépendantes**

Soutenue publiquement, le 10/05/22 à 13h30, devant le jury composé de :

Mme. Zohra GUESSOUM	Professeure	à L'U.S.T.H.B/ Alger	Présidente
Mr. Abdelkader TATACHAK	Professeur	à L'U.S.T.H.B/ Alger	Directeur de Thèse
Mr. Khaled KHALDI	Professeur	à L'U.M.B.B/ Boumerdès	Examineur
Mme. Ourida SADKI	Professeure	à L'U.S.T.H.B/ Alger	Examinatrice
Mme. Hanya KHERCHI MEDJDEN	Professeure	à L'E.N.S.S.E.A/ Alger	Examinatrice
Mme. Hassiba BENSERADJ	MCB	à L'U.M.B.B/ Boumerdès	Invitée
Mme. Farida HAMRANI	MCB	à L'U.S.T.H.B/ Alger	Invitée

Je dédie cette thèse

*à ma très chère mère, ce rêve n'aurait pas pu se réaliser sans elle,
à l'esprit de mon père.*

Résumé

Le sujet de cette thèse concerne l'inférence statistique dans les modèles conditionnels via l'étude de l'estimation non paramétrique à noyau du mode conditionnel, sous la notion généralisée de dépendance faible établie par [Doukhan and Louhichi \(1999\)](#). Nous étendons le travail effectué par [Hwang and Shin \(2016\)](#) au cas conditionnel et multivarié.

Dans un premier temps, nous rappelons quelques résultats établis sur le comportement asymptotique de certains estimateurs à noyau de densité, mode, de densité conditionnelle et mode conditionnel, dans les cas de données indépendantes et mélangeantes.

Dans un second temps, nous donnons nos résultats en considérant des variables faiblement dépendantes. Nous établissons la convergence et la normalité asymptotique des estimateurs à noyau de la fonction mode conditionnelle et de la densité conditionnelle. Nous obtenons des intervalles de confiance asymptotiques. Une étude de simulation est réalisée sur des modèles faiblement dépendants pour valider nos résultats théoriques. Des prévisions sur des données réelles sont également menées pour montrer l'intérêt de notre étude.

Mots et phrases clés : Faible dépendance, Convergence forte, Estimation non paramétrique, Mode conditionnel, Normalité asymptotique.

Abstract

The topic of this thesis concerns statistical inference in conditional models via the study of non parametric kernel estimation of the conditional mode, under the generalized notion of weak dependence established by [Doukhan and Louhichi \(1999\)](#). We extend the work done by [Hwang and Shin \(2016\)](#) to the conditional and multivariate case.

In the first step, we recall some established results on the asymptotic behavior of some kernel density, mode, conditional density and conditional mode estimators, in the case of independent and mixing data.

In a second step, we give our results considering weakly dependent variables. We establish the convergence and the asymptotic normality of the kernel estimators of the conditional mode function and of the conditional density. We obtain asymptotic confidence intervals. A simulation study is performed on weakly dependent models to validate our theoretical results. Some forecasting on real data is also conducted to show the interest of our study.

Key words and phrases : Weak dependence, Asymptotic normality, Conditional mode, Non-parametric estimation, Strong convergence.

Articles et communications

1. Rih, S., Tatachak, A., "Kernel conditional density and mode estimation for ψ -weakly dependent observations".
<https://doi.org/10.1080/03610926.2021.1944216>
2. Rih, S., Tatachak, A., " Kernel conditional mode estimation under ψ -weak dependence", *TAM-TAM'2019 Conference*, 23-27 Février 2019, Tlemcen, Alger.
https://tamtam2019-univ-tlemcen.weebly.com/uploads/9/6/8/6/96869444/recueil_des_r%C3%A9sum%C3%A9s_de_la_neuvi%C3%A8me_%C3%A9dition_du_tamtam_18-19..pdf
3. Rih, S., Tatachak, A., "Asymptotic normality of a kernel conditional mode function estimator for ψ -weakly dependent data", *colloque Modélisation Stochastique et Stochastique (MSS'2019)*, 24-26 Novembre 2019, Université des sciences et de technologie Houari Boumedién, Alger.
https://hal.archives-ouvertes.fr/hal-02593238/file/MSS%202019_Proceedings.pdf
4. Rih, S., Tatachak, A., " Kernel conditional mode estimation for ψ -weakly dependent observations and some simulations", *JSLMSTD*, 26 Février 2022, Université des sciences et de technologie Houari Boumedién, Alger.

Remerciements

Le moment est venu pour moi d'exprimer ma profonde gratitude à tous ceux qui m'ont aidée et encouragée dans les moments difficiles dans le but d'obtenir le diplôme de Docteur de l'Université des Sciences et de la Technologie Houari Boumediène.

Tout d'abord, je tiens à remercier très vivement mon directeur de thèse Monsieur Abdelkader Tatachak. Ses qualités humaines, son aide et son soutien constant m'ont permis de mener à bien cette thèse. Qu'il trouve ici l'expression de ma profonde gratitude.

Ma gratitude va également aux Professeures Zohra Guessoum et Ourida Sadki. Je tiens également à vous remercier d'avoir toujours été à l'écoute de nos présentations au sein du groupe de travail. Vos réponses à nos questions, les conseils et les remarques que vous nous avez apportés nous ont permis de mieux travailler. Je suis très reconnaissante que vous fassiez partie de mon jury.

Je remercie également les Professeurs Khaled KHALDI et Hanya KHERCHI MEDJDEN d'avoir accepté d'être examinateurs et de bien vouloir participer au jury de cette thèse.

Je suis très heureuse que Mme Hassiba BENSERADJ et Mme Farida HAMRANI soient membres invités de ma thèse. J'ai eu la chance de travailler avec elles et je les remercie sincèrement de leurs encouragements, conseils et aides.

Je tiens également à remercier tous les doctorants et jeunes docteurs du groupe de travail GTDI pour les échanges fructueux lors des séminaires et discussions et avec qui j'ai partagé mes angoisses, mes rêves et des moments inoubliables.

Je tiens à remercier ma très nombreuse famille : mes sœurs, frères, mes nièces et neveux. Je les remercie tous de m'avoir supportée et encouragée dans les moments de doute. Avec une attention particulière à ma mère, toujours présente à mes côtés pour soutenir mes ambitions et m'aider à accomplir mes rêves. Tout ce que je pourrais dire ne sera jamais suffisant pour exprimer tout ce que je lui dois et l'importance qu'elle a eue pour faire de moi ce que je suis aujourd'hui.

Il y a aussi une autre personne à qui je dois exprimer ma reconnaissance. Une personne qui croyait en moi plus que moi-même, qui m'a encouragée et aidée tout au long de mes études, ma sœur Meriem à qui je souhaite tout le bonheur de la vie, merci infiniment.

Je tiens à remercier mon oncle Abdelaziz qui m'a poussée à poursuivre mes études de doctorat. Je lui dédie cette thèse.

Je remercie du fond du cœur tous mes amis et chers collègues pour leur soutien moral.

Avant de clôturer mes remerciements, je tiens à adresser un immense merci à ma chère amie Saloua pour son soutien constant. Je ne trouve pas les mots justes et sincères pour exprimer ma sympathie et mes sentiments, tu es pour moi une sœur et une amie sur qui je peux compter. Je te dédie ce travail en hommage à notre amitié et à la mémoire de tous les moments que nous avons passés ensemble, et je te souhaite une vie pleine de bonheur et de succès.

Enfin, j'adresse mes remerciements à tous ceux qui m'ont aidée de près ou de loin à la réalisation de ce travail.

Abreviations et Notations

- i.i.d : indépendantes et identiquement distribuées.
- v.a : variable aléatoire.
- p.s : presque sûrement.
- \mathbb{L}^p : la convergence en moyenne d'ordre p .
- $\xrightarrow{\mathcal{D}}$: converge en distribution.
- $\mathcal{N}(0, 1)$: la loi normale centrée réduite.
- $:=$: égalité par définition.
- $=:$: égalité par notation.
- $\mathbb{1}_A$: fonction indicatrice de l'ensemble A .
- $\|\cdot\|_\infty$: désigne la norme infinie.
- resp : respectivement.
- Q_p : quantile conditionnel d'ordre p .
- $\mathcal{B}(p)$: la loi de Bernoulli de paramètre p .

Table des figures

2.1	Estimation par la méthode du noyau d'un échantillon de taille 500 issue d'une loi normale pour différentes valeurs de la fenêtre.	30
3.1	Les courbes de $\Theta_n(\cdot)$ par rapport à $\Theta(\cdot)$ pour $n = 50, 100$ et 500 de gauche à droite. .	46
3.2	Les courbes de $f_n(\cdot 0)$ par rapport à $f(\cdot 0)$ pour $n = 50, 100$ et 500 de gauche à droite.	46
3.3	La densité de $\bar{\Theta}(n)$ par rapport à $\mathcal{N}(0, 1)$ et QQ-plot pour $n=50$	47
3.4	La densité de $\bar{\Theta}(n)$ par rapport à $\mathcal{N}(0, 1)$ et QQ-plot pour $n=500$	47
3.5	Les intervalles de confiance de niveau asymptotique de 95% pour $n = 50$ et $n = 500$. .	48
3.6	Le mode conditionnel théorique $\Theta(\cdot, \cdot)$	49
3.7	Surface du mode conditionnel $\Theta_n(\cdot, \cdot)$ pour $n = 50, 100$ et 500 de gauche à droite . . .	49
3.8	La densité de $\bar{\bar{\Theta}}(n)$ par rapport à $\mathcal{N}(0, 1)$ pour $n = 50$ et $n = 500$ de gauche à droite .	49
4.1	L'évolution quotidienne du taux de change USD/EURO (01/01/2015 - 01/01/2020). .	65
4.2	Rendements logarithmiques quotidiens des taux de change USD/EURO	66
4.3	Test de normalité sur les rendements	67
4.4	Représentation du corrélogramme	67
4.5	Estimation du modèle ARCH	68
4.6	Prédiction de rendement par le mode conditionnel et la régression	69
4.7	L'estimateur de la densité conditionnelle $f_n(y x)$	70

Liste des tableaux

3.1	Valeurs du MSE, GMSE pour $f_n(\cdot 0)$ et $\Theta_n(\cdot)$ (resp) avec (h_{Kglob}, h_{Hglob})	46
3.2	Valeurs du GMSE pour $\Theta_n(\cdot, \cdot)$ avec (h_{Kglob}, h_{Hglob})	48

Table des matières

<i>Dédicaces</i>	3
<i>Résumé</i>	4
<i>Abstract</i>	5
<i>Articles et communications</i>	6
<i>Remerciements</i>	8
<i>Abreviations et Notations</i>	9
<i>Liste des figures</i>	9
<i>Liste des tableaux</i>	10
<i>Table des matières</i>	13
1 Introduction générale	14
1.1 Généralités	14
1.2 Contribution de la thèse	17
1.3 Le plan de la thèse	18
1.4 Dépendance faible	18
1.4.1 Le mélange	19
1.4.2 L'association	19
1.4.2.1 Association Positive	20
1.4.2.2 Association Négative	20
1.4.2.3 Quasi-Association (QA)	21
1.4.3 Non mélangeant et non associé	22
1.4.4 Dépendance faible au sens de Doukhan and Louhichi (1999)	22
1.4.4.1 Propriétés de base	23
1.4.4.2 Exemples de processus faiblement dépendants	23
1.4.4.3 Inégalités exponentielles, le théorème central limite de Lindeberg dépendant	25

2	Estimation à noyau de la fonction densité et du mode	28
2.1	Introduction	28
2.2	Estimateur à noyau de la densité	28
2.2.1	Optimalité de l'estimateur de la densité	29
2.2.2	Quelques méthodes de sélection du paramètre de lissage	31
2.2.2.1	Approximation gaussienne (Silverman)	31
2.2.2.2	Méthode de la validation croisée	31
2.2.2.3	Méthode Plug in	32
2.3	Estimateur à noyau du mode simple	33
2.3.1	L'estimateur de Parzen	33
2.3.2	Généralisation de l'estimation du mode de Parzen.	35
2.4	Estimateur à noyau du mode conditionnel	37
2.4.1	Cas des données indépendantes	38
2.4.2	Cas des données mélangées	38
3	Propriétés asymptotiques de l'estimateur du mode conditionnel pour les données	
	ψ-faiblement dépendantes	40
3.1	Préliminaires	41
3.2	Hypothèses	41
3.3	Consistance	43
3.3.1	La vitesse de convergence presque sûre pour $f_n(\cdot x)$	43
3.3.2	La vitesse de convergence presque sûre pour $\Theta_n(\cdot)$	43
3.4	Normalité asymptotique et intervalles de confiance	43
3.4.1	Normalité asymptotique de $\Theta_n(\cdot)$	44
3.4.2	Intervalles de confiance pour $\Theta(\cdot)$	44
3.5	Etude de simulation	45
3.5.1	Cas unidimensionnel	45
3.5.1.1	Consistance	45
3.5.1.2	Normalité asymptotique	46
3.5.2	Cas bidimensionnel	48
3.6	Résultats auxiliaires et preuves	50
4	Application sur des données réelles	64
4.1	Introduction	64
4.2	Les données	64
4.2.1	Description de la base de données	64
4.2.2	Etude de la stationnarité	65
4.2.3	Les statistiques descriptives	67

4.2.3.1	Hétéroscédasticité conditionnelle	68
4.3	Prévision	69
	Conclusion et perspectives	73
	Bibliographie	80

Introduction générale

1.1 Généralités

La statistique inférentielle consiste à déduire d'un échantillon prélevés dans une population, les caractéristiques de cette population. L'une des principales branches de la statistique inférentielle est la théorie de l'estimation. Dans la littérature, on pourra distinguer deux types principaux d'estimation : l'estimation paramétrique et l'estimation non paramétrique. Le problème de l'estimation non paramétrique consiste à estimer à partir des observations une fonction inconnue, appartenant à une certaine classe fonctionnelle. Rappelons que l'approche non paramétrique contrairement à l'estimation paramétrique est définie indépendamment de la distribution ou la loi de l'échantillon d'observations. Plus particulièrement, on parle de méthode d'estimation non paramétrique lorsque celle-ci ne se ramène pas à l'estimation d'un nombre fini de paramètres réels associés à la loi de l'échantillon ce qui correspond d'ailleurs souvent à la réalité.

En statistique, étudier les liens entre deux variables aléatoires dépendantes est un problème de régression. Ce problème, peut-être modélisé comme suit : nous avons deux variables aléatoires dépendantes X (variable explicative ou co-variable) et Y (variable d'intérêt ou réponse), la prévision de Y sachant X se fait à travers X par une fonction r , dite fonction de régression. Plus précisément, on veut trouver une bonne approximation de Y selon un critère donné par la fonction $r(X)$. Alors le problème devient la minimisation de la fonction de risque suivante :

$$err(r) = \mathbb{E}(l(Y - r(X))),$$

où l est une fonction de coût. Si cette fonction de coût est convexe et admet un minimum unique, alors le meilleur prédicteur $r^*(.)$ conditionnel de Y sachant X de $r(.)$ est

$$r^*(x) = \arg \min_{q \in \mathbb{R}} \mathbb{E}(l(Y - q) | X = x), \quad x \in \mathbb{R}^d. \tag{1.1}$$

Par conséquent, on constate qu'en statistique non-paramétrique, les outils naturels de prévision sont les modèles relatifs à la distribution conditionnelle de ces variables aléatoires. Parmi ces modèles,

le plus utilisé est le modèle de régression classique. Cependant, il est bien connu que cette dernière approche n'est pas efficace dans certaines situations. C'est le cas, par exemple, de la fonction de densité de probabilité (pdf) conditionnelle suivante de Y étant donné $X = x$.

$$f(y|x) = \frac{1}{2\sqrt{2\pi}} \left(\exp \left\{ -\frac{1}{2} (y-x)^2 \right\} + \exp \left\{ -\frac{1}{2} (y+x)^2 \right\} \right),$$

pour laquelle la fonction de régression classique disparaît partout (voir [Ould-Saïd \(1997\)](#), Remarque 7, p. 234) et donc, il n'est pas raisonnable d'utiliser $\mathbb{E}(Y|X = x)$ pour la prédiction. D'autres situations où le mode conditionnel fournit une alternative de prédiction de Y pertinente sont celles où la distribution visée est asymétrique ou multimodale (voir [Collomb et al. \(1986\)](#), pp. 228-229). Notez également que la méthode de noyau pour l'estimation de $f(y|x)$ peut être exprimée de manière équivalente dans un cadre de régression. Afin de s'en convaincre, il suffit d'observer que si $F(y|x)$ est la fonction de distribution cumulative conditionnelle de Y compte tenu de $X = x$, alors, lorsque $h \rightarrow 0$, on obtient les résultats suivants

$$\mathbb{E}(\mathbf{1}_{|Y-y|\leq h}|X = x) = F(y+h|x) - F(y-h|x) \approx 2hf(y|x).$$

Par définition, le mode conditionnel est la valeur qui maximise la densité conditionnelle. On note par $\Theta(\cdot)$ ce mode :

$$\Theta(x) = \arg \max_{y \in \mathbb{R}} f(y|x).$$

En considérant la fonction de coût non convexe définie par :

$$l(z) = \begin{cases} 0 & \text{si } z = 0, \\ 1 & \text{sinon} \end{cases}$$

nous pouvons montrer que la solution de (1.1) est la fonction mode conditionnel.

En effet,

$$\begin{aligned} \mathbb{E}(l(Y - q)|X = x) &= \int_{-\infty}^{+\infty} \mathbf{1}_{\{y \neq q\}} f(y|x) dy \\ &= \int_{]-\infty, q[} f(y|x) dy + \int_{]q, +\infty[} f(y|x) dy \\ &= \sum_{n \geq 1} \left(\int_{-\infty}^{q - \frac{1}{n}} f(y|x) dy + \int_{q + \frac{1}{n}}^{+\infty} f(y|x) dy \right). \end{aligned}$$

En différentiant par rapport à q au point $\Theta(x)$, et égalant à 0 on obtient

$$\frac{\partial \mathbb{E} \{l(Y - q)|X = x\}}{\partial q} \Big|_{\Theta(x)} = \sum_{n \geq 1} \left\{ f \left(\Theta(x) - \frac{1}{n} \mid x \right) - f \left(\Theta(x) + \frac{1}{n} \mid x \right) \right\} = 0.$$

Un développement de Taylor de f à l'ordre 1 au voisinage de $\Theta(x)$ donne

$$\sum_{n \geq 1} -\frac{2}{n} f^{(1)}(\Theta(x)) = 0, \tag{1.2}$$

ce qui implique que

$$f^{(1)}(\Theta(x)) = 0.$$

et cela implique que f possède un extremum en $\Theta(x)$. Il ne reste qu'à montrer que $f^{(2)}(\Theta(x)) < 0$ qui se fait en procédant de la même manière que pour (1.2), en utilisant le fait que :

$$\frac{\partial^2 \mathbb{E} \{l(Y - q) | X = x\}}{\partial q^2} \Big|_{\Theta(x)} > 0.$$

L'estimation du mode conditionnel fut traitée pour la première fois par [Collomb et al. \(1986\)](#). Ces auteurs ont montré la convergence uniforme de l'estimateur à noyau de ce modèle conditionnel lorsque les observations sont ϕ -mélangeantes. [Samanta and Thavaneswaran \(1990\)](#), ont obtenu la même propriété asymptotique pour un estimateur à noyau du mode conditionnel en considérant le cas i.i.d alors que les conditions de convergence dans le cas de données α -mélangeantes ont été établies par [Ould-Saïd \(1993\)](#) et dans le cas de données ergodiques par [Martins Rosa \(1993\)](#) et [Ould-Saïd \(1997\)](#). [Quintela-Del-Rio and Vieu \(1997\)](#) ont considéré le mode conditionnel comme le point qui annule la dérivée du premier ordre de la densité conditionnelle et ils ont construit un estimateur pour ce même modèle en utilisant l'estimateur à noyau de la dérivée de la densité conditionnelle. Nous référons à [Berlinet et al. \(1998\)](#), [Louani and Ould-Saïd \(1999\)](#) pour la convergence en loi de l'estimateur à noyau du mode conditionnel dans le cas α -mélangeant. [Mehra et al. \(2000\)](#) ont établi la loi du logarithme itéré, la convergence uniforme presque sûre sur un ensemble compact, et la normalité asymptotique de l'estimateur du plus proche voisin lissé du mode conditionnel, [Ioannides and Matzner-Løber \(2002\)](#) ont construit un estimateur pour le mode conditionnel, lorsque les observations sont affectées d'erreurs. Dans cet article, les auteurs se sont intéressés à la convergence presque sûre de l'estimateur proposé. Alors que sa normalité asymptotique a été démontrée par les mêmes auteurs dans [Ioannides and Matzner-Løber \(2004\)](#). Comme application, [Rossi \(2004\)](#) a utilisé le mode conditionnel dans le domaine de la haute technologie pour décrire un processus de dépollution biologique. Nous citons également les travaux de [Einbeck and Tutz \(2006\)](#) pour plus d'informations sur l'estimateur du mode conditionnel en tant qu'outil de prévision.

Dans le contexte des données fonctionnelles, [Ferraty et al. \(2005\)](#) ont prouvé, sous certaines conditions de régularité de la densité conditionnelle, la convergence presque complète de l'estimateur à noyau qu'ils ont introduit pour la densité conditionnelle, ainsi que celle du mode conditionnel, et ont établi la vitesse de convergence. Depuis la parution de cet article, une importante littérature s'est développée sur l'estimation de la densité conditionnelle et de ses dérivés, en particulier pour l'utiliser afin d'estimer le mode conditionnel dans le cadre de données fonctionnelles. Nous citons aussi les travaux de [Ezzahrioui and Ould-Saïd \(2008\)](#), [Ezzahrioui \(2010\)](#), [Dabo-Niang and Laksaci \(2007\)](#), [Dabo-Niang et al. \(2012\)](#).

D'autres chercheurs se sont intéressés à l'estimation du mode conditionnel pour des données censurées ou tronquées (voir, par exemple, [Ould-Saïd and Cai \(2005\)](#), [Ould-Saïd and Tatachak \(2007\)](#), ([Khardani et al., 2010, 2011, 2012](#)), [Ould-Saïd and Tatachak \(2011\)](#)).

Il est remarquable que l'hypothèse d'indépendance est souvent trop forte pour être réaliste dans de nombreux problèmes pratiques, c'est le cas lorsque les données sont collectées chronologiquement de manière consécutive. Il est alors naturel de supposer que l'observation courante dépend dans une certaine mesure des précédentes. Ceci nous a motivé à étudier le problème de l'estimation de la fonction de mode conditionnelle pour les v.a ψ -faiblement dépendantes (ψ -faiblement), ce problème n'a pas été traité, à notre connaissance, dans la littérature statistique. Le type de ψ -faiblement dépendant comprend le mélange, l'association, les suites gaussiennes et les décalages de Bernoulli et cette notion a été introduite par [Doukhan and Louhichi \(1999\)](#). Comme exemples d'applications statistiques sous ψ -faible dépendance nous citons, les travaux de [Doukhan and Louhichi \(2001\)](#) sur l'estimation de la densité du noyau, l'inégalité de type Bernstein de [Doukhan and Neumann \(2007\)](#), le théorème central limite de Lindeberg dépendant de [Bardet et al. \(2008\)](#), entre autres. En outre, ([Hwang and Shin, 2014, 2016](#)) ont établi des inégalités exponentielles et étudié l'estimation de mode simple (sans covariables) pour des suites de v.a ψ -faiblement dépendantes, respectivement.

1.2 Contribution de la thèse

La contribution principale de cette thèse est d'étudier les propriétés asymptotiques du mode conditionnel dans le cadre de la statistique non paramétrique sous l'hypothèse de la dépendance faible au sens de [Doukhan and Louhichi \(1999\)](#). Pour une compréhension approfondie de ce type de dépendance et de tous les outils probabilistes qui sont liés à cette nouvelle notion introduite par [Doukhan and Louhichi \(1999\)](#), nous renvoyons à la dernière section de ce chapitre.

Dans notre étude, le mode est estimé par le point qui maximise l'estimateur à noyau de la densité conditionnelle. D'abord, nous établissons avec précision la vitesse de convergence presque sûre sur un compact de \mathbb{R}^d de la densité conditionnelle sous certaines conditions de régularité du noyau, de la fonction de densité conjointe et la densité marginale ainsi que des hypothèses adéquates sur les paramètres de lissage. Les preuves reposent principalement sur des calculs de covariance pour lesquels nous utiliserons principalement l'inégalité exponentielle proposée par [Hwang and Shin \(2014\)](#). Par la suite, nous déduisons la convergence du mode conditionnel.

Nous établissons aussi la normalité asymptotique de l'estimateur du mode conditionnel. Ce résultat est obtenu en utilisant le théorème de [Bardet et al. \(2008\)](#) qui généralise le théorème de Lindeberg en proposant une version qui s'adapte à la fois avec les suites indépendantes et dépendantes. Notez que cette procédure ne nécessite pas autant de calculs que celle de Bernstein pour les grands blocs et les petits blocs. La normalité asymptotique de cet estimateur nous permet également de construire des intervalles de confiance. Des simulations sous Matlab montrent la validité de nos résultats théoriques. En termes d'application, cette contribution a une grande importance en pratique. En effet, il s'agit d'adopter dans le cas des variables faiblement dépendantes un outil de prévision alternatives à la régression classique. Notons que cette méthode de prévision est beaucoup plus robuste que celle de la

régression classique qui est très sensible aux erreurs d'observations telle que la présence d'observations aberrantes, l'hétéroscédasticité des données et l'asymétrie.

De plus, cet estimateur offre la possibilité de déterminer des régions prédictives (voir par exemple [De Gooijer and Gannoun \(2000\)](#)) qui sont beaucoup plus intéressantes que la prévision ponctuelle en pratique. Ce modèle est par ailleurs utilisé en finance ou en assurance pour modéliser les risques de valeurs extrêmes.

1.3 Le plan de la thèse

Cette thèse est présentée en quatre chapitres. Dans le reste de ce premier chapitre introductif, nous donnons la définition de différents types de dépendance faible ainsi que des exemples de processus faiblement dépendants. Nous présentons également certains types d'inégalités exponentielles qui se retrouvent dans la littérature sur les séquences faiblement dépendantes.

Dans le deuxième chapitre, nous rappelons les définitions des estimateurs à noyaux de la fonction de densité ; mode simple et conditionnel, en présentant les différentes propriétés asymptotiques de ces estimateurs. Nous présentons également les méthodes qui existent pour la sélection de la fenêtre optimale qui joue un rôle fondamental pour la qualité de l'estimation à noyau.

Le troisième chapitre constitue la principale contribution de cette thèse. Comme détaillé ci-dessus, cet apport concerne les propriétés asymptotiques de l'estimateur du mode conditionnel dans le cadre de la dépendance faible au sens de [Doukhan and Louhichi \(1999\)](#).

Pour illustrer l'utilité de notre étude, un exemple d'application sur des données réelles et une discussion sont présentés au chapitre 4.

Finalement, nous terminons ce document par une conclusion générale qui résume les principaux résultats obtenus et nous donnons quelques perspectives de recherche.

1.4 Dépendance faible

En statistique, on a deux types de dépendance : la dépendance forte et la dépendance faible. La différence entre les deux catégories est que : en dépendance faible, les théorèmes limite associés aux statistiques de Donsker et de Kolmogorov se produisent avec une normalisation en \sqrt{n} et la limite est un processus plutôt irrégulier tandis qu'en dépendance forte, le processus limite est très régulier et on rencontre diverses normalisations (voir [Priour \(2001\)](#)). Dans notre étude, nous allons considérer uniquement le cas de dépendance faible. Commençons par introduire les différents types de cette dépendance.

1.4.1 Le mélange

La notion de mélange s'exprime en termes de coefficients de mélange entre des tribus engendrées par le passé et le futur de la suite $(X_n)_{n \in \mathbb{Z}}$: pour $p \in \mathbb{Z}$, $\mathcal{F}_{-\infty}^p = \sigma(X_i, i \leq p)$ et $\mathcal{F}_p^{+\infty} = \sigma(X_i, i \geq p)$. Si $(\Omega, \mathcal{A}, \mathbb{P})$ désigne un espace probabilisé, on mesure la dépendance entre deux tribus \mathcal{U} et \mathcal{V} grâce aux coefficients qui suivent :

- $\alpha(\mathcal{U}, \mathcal{V}) := \sup | \mathbb{P}(U \cap V) - \mathbb{P}(U)\mathbb{P}(V) |, U \in \mathcal{U}, V \in \mathcal{V}$ (α -mélange ou mélange fort, introduit par [Rosenblatt \(1956\)](#)).
- $\beta(\mathcal{U}, \mathcal{V}) := \sup \frac{1}{2} \sum_{i \in I} \sum_{j \in J} | \mathbb{P}(U_i \cap V_j) - \mathbb{P}(U_i)\mathbb{P}(V_j) |$, (β -mélange, introduit par [Volkonskii and Rozanov \(1959\)](#)).
- Le sup est pris sur toutes les partitions $(U_i)_{i \in I}, (V_j)_{j \in J}$ de Ω avec $U_i \in \mathcal{U}, V_j \in \mathcal{V}$.
- $\phi(\mathcal{U}, \mathcal{V}) := \sup | \mathbb{P}(V|U) - \mathbb{P}(V), \mathbb{P}(U) > 0, U \in \mathcal{U}, V \in \mathcal{V}$ (ϕ -mélange ou mélange uniforme, introduit par [Ibragimov \(1959\)](#)).
- $\rho(\mathcal{U}, \mathcal{V}) := \sup | \text{Corr}(u, v), u \in \mathbb{L}^2(\mathcal{U}), v \in \mathbb{L}^2(\mathcal{V})$ (ρ -mélange, introduit par [Kolmogorov and Rozanov \(1960\)](#)).

Soit $(X_n)_{n \in \mathbb{Z}}$ une suite de variables aléatoires. On dit que cette suite est α -mélangeante si la suite des coefficients de α -mélange $\alpha(n) = \sup_{p \in \mathbb{Z}} (\mathcal{F}_{-\infty}^p, \mathcal{F}_{p+n}^{+\infty})$ converge vers 0 lorsque n tend vers $+\infty$. Les notions de β -mélange, de ϕ -mélange et de ρ -mélange sont définies de façon similaire.

Parmi toutes ces formes de mélanges, l' α -mélange est la plus faible (voir [Doukhan \(1994\)](#), [Rio \(2000\)](#)) et donc le moins restrictif. Toute suite de v.a β, ϕ, ρ mélangeantes sera alors forcément α -mélangeante. Réciproquement, tout résultat énoncé pour des données α -mélangeantes sera valable pour des données soumises à une autre forme de mélange. Cela explique donc que l'étude de variables mélangeantes est généralement consacrée dans la littérature au cas de données α -mélangeantes.

Comme exemples de processus fortement mélangeants, nous citons :

- Les processus linéaire (sous certaines conditions voir [Gorodestkii \(1977\)](#)).
- Les processus ARMA : [Withers \(1981\)](#) élargi la classe des processus linéaires fortement mélangeants donnée par [Gorodestkii \(1977\)](#).
- Chaîne de Markov : voir [Athreya and Pantula \(1986\)](#).

Pour plus de détails sur les conditions de mélange, nous renvoyons le lecteur aux références suivantes : [Doukhan \(1994\)](#), [Bosq \(1998\)](#), [Rio \(2000\)](#).

1.4.2 L'association

Le concept d'association a été introduit indépendamment par [Esary et al. \(1967\)](#) et [Fortuin et al. \(1971\)](#)(FKG). Leur but était de trouver des applications dans la fiabilité des systèmes et dans la statistique mécanique (à travers les inégalités FKG).

Ce concept est une généralisation de la dépendance positive introduite par [Lehmann \(1966\)](#) (plus précisément la dépendance positive par quadrant, en anglais Positive Quadrant Dependence ou PQD).

Définition 1.1. (*Lehmann (1966)*)

Deux variables aléatoires X et Y sont dites PQD si pour tous nombres réels x et y

$$\mathbb{P}(X \geq x, Y \geq y) - \mathbb{P}(X \geq x)\mathbb{P}(Y \geq y) \geq 0.$$

Définition 1.2. (*Esary et al. (1967)*)

Une suite finie de variables aléatoires X_1, \dots, X_n est dite associée si

$$\text{Cov}(f(X_1, \dots, X_n), g(X_1, \dots, X_n)) \geq 0,$$

pour toutes fonctions croissantes f, g de $\mathbb{R}^n \rightarrow \mathbb{R}$ pour lesquelles la covariance existe. Une suite infinie de variables aléatoires est associée si toute sous suite finie est associée.

Il y a plusieurs exemples sur les processus associés citons : les variables aléatoires gaussiennes positivement corrélées (*Pitt (1982)*), les variables aléatoires binaires, les processus autorégressifs et les processus linéaires (*Doukhan and Louhichi (1999)*).

Notons que cette structure de dépendance a plusieurs type appelées associations : positive (ou faible), négative et quasi-association.

1.4.2.1 Association Positive

Définition 1.3. (*Burton et al. (1986)*) Une suite finie de variables aléatoire X_1, \dots, X_n est dite positivement (faiblement) associées si pour tous sous-ensembles disjoints I et $J \in \{1 \dots n\}$

$$\text{Cov}(f(X_i, i \in I), g(X_j, j \in J)) \geq 0,$$

pour toutes fonctions non décroissantes $f : \mathbb{R}^{|I|} \rightarrow \mathbb{R}$, $g : \mathbb{R}^{|J|} \rightarrow \mathbb{R}$ pour lesquelles la covariance existe.

Une suite infinie de variables aléatoires est positivement associée si toute sous suite finie est positivement associée.

Notons que la définition de l'association positive et l'association sont ressemblantes, mais pas équivalentes. L'association positive est strictement plus faible que l'association. Ainsi, l'association implique l'association positive.

1.4.2.2 Association Négative

L'association négative a été introduite par *Joag-Dev and Proschan (1983)*. ils ont donné de nombreuses propriétés et proposé plusieurs applications en Statistique.

Définition 1.4. (*Joag-Dev and Proschan (1983)*)

Une suite finie de variables aléatoires X_1, \dots, X_n est dite négativement associée (NA) si pour tout sous ensembles disjoints I et J de $\{1, 2, \dots, n\}$

$$\text{Cov}(f(X_i, i \in I), g(X_j, j \in J)) \leq 0,$$

pour toutes fonctions croissantes $f : \mathbb{R}^{|I|} \rightarrow \mathbb{R}$, $g : \mathbb{R}^{|J|} \rightarrow \mathbb{R}$ pour lesquelles la covariance existe. Une suite infinie de variables aléatoires est négativement associée si toute sous suite finie est négativement associée.

Une propriété fondamentale vérifiée par les variables associées et négativement associées est l'équivalence entre la non corrélation et l'indépendance. Cette propriété provient de l'inégalité de covariance suivante :

$$| \text{Cov}(f(X_i, i \in I), g(X_j, j \in J)) | \leq \sum_{i \in I} \sum_{j \in J} \left\| \frac{\partial f}{\partial x_i} \right\|_{\infty} \left\| \frac{\partial g}{\partial x_j} \right\|_{\infty} | \text{Cov}(X_i, X_j) |,$$

pour I et $J \in \{1 \dots n\}$ et f, g deux fonctions réelles à dérivées bornées.

Cette inégalité de covariance a été établie par [Birkel \(1988\)](#), a donné naissance à une autre notion de dépendance faible appelée la "quasi-association".

1.4.2.3 Quasi-Association (QA)

Ce concept a été introduit par [Bulinski and Suquet \(2001\)](#). Il permet d'étudier une classe de variables aléatoires indépendantes par la non corrélation. Rappelons que cette classe contient en plus des variables gaussiennes, des variables associées et négativement associées et c'est un cas particulier de dépendance faible introduit par [Doukhan and Louhichi \(1999\)](#).

Définition 1.5. ([Bulinski and Shabanovich \(1998\)](#))

Soit $X = \{X_n, n \in \mathbb{N}\}$ une famille de v.a réelles telle que $\mathbb{E}(X_j^2) < \infty$ pour tout $j \in \mathbb{N}$. La famille X est dite quasi-associée si, pour tous sous-ensembles finis disjoints I et J de \mathbb{N} et toutes fonctions lipschitziennes f, g définies respectivement de $\mathbb{R}^{|I|} \rightarrow \mathbb{R}$ et $\mathbb{R}^{|J|} \rightarrow \mathbb{R}$, on a

$$| \text{Cov}(f(X_i, i \in I), g(X_j, j \in J)) | \leq \sum_{i \in I} \sum_{j \in J} \text{lip}_i(f) \text{lip}_j(g) | \text{Cov}(X_i, X_j) |. \quad (1.3)$$

Les constantes de lipschitz $\text{Lip}_i(f)$ sont telles que pour $x = (x_i, i \in I)$, $y = (y_i, i \in I)$ dans $\mathbb{R}^{|I|}$

$$|f(x) - f(y)| \leq \sum_{i \in I} \text{Lip}_i(f) |x_i - y_i|$$

et

$$\text{Lip}_i(f) = \sup_{x_i \neq x'_i} \frac{|f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_{|I|}) - f(x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_{|I|})|}{|x_i - x'_i|},$$

le sup étant pris pour $x_1, x_2, \dots, x_{|I|}, x'_i \in \mathbb{R}$.

[Bulinski and Shabanovich \(1998\)](#) ont montré que, toute collection de variables aléatoires positivement ou négativement associées admettant un second moment fini satisfait (1.3). Par conséquent, l'association positive ou l'association négative impliquent la quasi-association.

Pour plus de détails sur le concept d'association nous renvoyons le lecteur aux monographies de [Bulinski and Shashkin \(2007\)](#), [Oliveira \(2012\)](#).

1.4.3 Non mélangeant et non associé

En général, les conditions de mélange sont très difficiles à vérifier et on a même des processus qui ne sont pas mélangeants et non associés. Citons le célèbre exemple d'Andrews (voir [Andrews \(1984\)](#)) : si $(\varepsilon_n)_{n \geq 1}$ est une suite de variables aléatoires i.i.d de marginale $\mathcal{B}(1/2)$, alors la solution stationnaire $(X_n)_{n \geq 0}$ de l'équation

$$X_n = \frac{1}{2}(X_{n-1} + \varepsilon_n); X_0 \text{ indépendante de } (\varepsilon_n)_{n \geq 1},$$

n'est pas fortement mélangeante.

Le décalage de Bernoulli en anglais Bernoulli shift (voir la section suivante pour plus de détails), c'est encore un exemple de processus, non mélangeant. La différence entre deux processus indépendants de ce type constitue un exemple de processus non associé et non-mixte.

1.4.4 Dépendance faible au sens de [Doukhan and Louhichi \(1999\)](#)

[Doukhan and Louhichi \(1999\)](#) ont généralisé le mélange et l'association en introduisant une notion de dépendance faible plus simple couvrant plus de processus en s'appuyant sur le contrôle de la covariance entre le **passé** et le **futur** du processus via des fonctions lipschitziennes g ; h i.e **cov (g ("passé"), h ("futur"))**. Sous certaines conditions sur g et h , cette covariance est assez petite lorsque la distance entre le passé et le futur est suffisamment grande.

Pour définir cette notion de dépendance, nous commençons par introduire certaines classes de fonctions.

Soit $\mathbb{L}^\infty = \bigcup_{n=1}^{\infty} \mathbb{L}^\infty(\mathbb{R}^n)$ l'ensemble des fonctions réelles mesurables et bornées, définies sur \mathbb{R}^n , $n \in \mathbb{N}$.

On rappelle le module de Lipschitz de la fonction g (qui est définie sur \mathbb{R}^n) :

$$Lip(g) = \sup_{x \neq y} \frac{|g(x) - g(y)|}{\|x - y\|_1}, \text{ avec } \|(x_1, \dots, x_n)\|_1 = |x_1| + \dots + |x_n|.$$

Soit $\mathcal{L} = \bigcup_{n=1}^{\infty} \mathcal{L}_n$ avec $\mathcal{L}_n = \left\{ g \in \mathbb{L}^\infty(\mathbb{R}^n), Lip(g) < \infty, \|g\|_\infty := \sup_x |g(x)| \leq 1 \right\}$.

Définition 1.6. ([Doukhan and Louhichi \(1999\)](#))

La suite $(X_n)_{n \in \mathbb{Z}}$ est dite $(\epsilon, \mathcal{L}, \psi)$ -faiblement dépendante (simplement, ψ -faiblement dépendante), s'il existe une suite $\epsilon = (\epsilon_r)_{r \in \mathbb{N}}$ décroissante vers 0 à l'infini et une fonction ψ d'arguments $(g, h, n, m) \in \mathcal{L} \times \mathcal{L} \times \mathbb{N}^2$ telles que, pour tous n -uplets (i_1, \dots, i_n) et m -uplets (j_1, \dots, j_m) avec $i_1 \leq \dots \leq i_n < i_n + r \leq j_1 \leq \dots \leq j_m$,

$$\left| Cov(g(X_{i_1}, \dots, X_{i_n}), h(X_{j_1}, \dots, X_{j_m})) \right| \leq \psi(g, h, n, m)\epsilon_r.$$

pour toutes les fonctions g, h qui sont définies respectivement sur \mathbb{R}^n et \mathbb{R}^m .

- Remarque 1.1.** — Cette dépendance rend explicite l'indépendance asymptotique entre "le passé" et le "futur". Cela signifie que le "passé" est progressivement oublié.
- Dans la définition précédente, r désigne l'écart dans le temps entre le "passé" et le "futur".
 - En pratique, la classe \mathcal{L} sera remplacée par la plus petite classe \mathcal{L}_1 qui définit par $\mathcal{L}_1 = \{g \in \mathcal{L}, \|g\|_\infty := \sup_x |g(x)| \leq 1\}$ (voir [Doukhan and Louhichi \(1999\)](#)).
 - Notez que la suite ϵ dépend à la fonction ψ . Selon le choix de cette fonction on obtient les différents coefficients de dépendance.

Citons maintenant quelques exemples sur les différents coefficients de dépendance :

Pour $g, h \in \mathcal{L}_n \times \mathcal{L}_m$:

- Si $\psi(g, h, n, m) = n\|h\|_\infty \text{Lip}(g) + m\|g\|_\infty \text{Lip}(h)$, alors nous disons que la suite X_n est η -faiblement dépendante.
- Si $\psi(g, h, n, m) = n\|h\|_\infty \text{Lip}(g) + m\|g\|_\infty \text{Lip}(h) + nm \text{Lip}(g) \text{Lip}(h)$, alors nous disons que la suite X_n est λ -faiblement dépendante.
- Si $\psi(g, h, n, m) = m\|g\|_\infty \text{Lip}(h)$, alors nous disons que la suite X_n est λ -faiblement dépendante. Pour la définition des coefficients ζ et κ , on a besoin de la classe de fonctions Lipschitziennes non nécessairement bornées, dans ce cas, on suppose que les variables $(X_n)_{n \in \mathbb{Z}}$ sont \mathbb{L}^1 intégrables.
- Si $\psi(g, h, n, m) = \min(n, m) \text{Lip}(g) \text{Lip}(h)$, alors nous disons que la suite X_n est ζ -faiblement dépendante.
- Si $\psi(g, h, n, m) = nm \text{Lip}(g) \text{Lip}(h)$, alors nous disons que la suite X_n est κ -faiblement dépendante.

1.4.4.1 Propriétés de base

Les variables ψ -faiblement dépendantes ont une propriété d'hérédité, c'est-à-dire les propriétés de faible dépendance restent toujours valables à travers des images par des fonctions Lipschitziennes bornées.

- Supposons que $(X_n)_{n \in \mathbb{Z}}$ est un processus stationnaire à valeurs dans \mathbb{R}^d qui est $(\epsilon_r, \mathcal{L}_1, \psi)$ -faiblement dépendant. Alors, pour toute fonction continue lipschitzienne $G : \mathbb{R}^d \rightarrow \mathbb{R}$ avec $\|G\|_\infty = M \leq 1$ et $\text{Lip}(G) \leq 1$, le processus $Y_n = G(X_n)$ est à valeur réelle, stationnaire, et $\|Y_n\|_\infty \leq M$. De plus, il est aussi $(\epsilon_r, \mathcal{L}_1, \psi)$ -faiblement dépendant.
- Dans le cas plus général où $\text{Lip}(G) > 1$, la dépendance faible tient toujours, mais il suffit de remplacer $\psi(g, h, n, m)$ par $\psi^Y(g, h, n, m) = (\text{Lip}(g) \text{Lip}(G), \text{Lip}(h) \text{Lip}(G), n, m)$.

1.4.4.2 Exemples de processus faiblement dépendants

Voir [Doukhan and Louhichi \(1999\)](#), [Doukhan and Louhichi \(2001\)](#).

a-Les variables fortement mélangeantes :

Si une suite $(X_n)_{n \in \mathbb{Z}}$ est fortement mélangée elle est $(\epsilon, \mathcal{L}, \psi)$ -faiblement dépendante avec $\mathcal{L} = \mathbb{L}^\infty$, $\epsilon_r = \alpha_r$ et $\psi(g, h, n, m) = 4\|g\|_\infty\|h\|_\infty$.

b-Les variables associées :

- Les processus associés (resp les processus gaussiens) sont $(\epsilon, \mathcal{L}_1, \psi)$ -faiblement dépendants avec : $\psi(g, h, n, m) = \min(n, m)Lip(g)Lip(h)$ et $\epsilon_r = \sup_{k \geq r} \sum_{|i-j| \geq k} cov(X_i, X_j)$ (resp $\epsilon_r = \sup_{k \geq r} \sum_{|i-j| \geq k} |cov(X_i, X_j)|$).
- Si $(X_n)_{n \in \mathbb{Z}}$ est une suite stationnaire de variables aléatoires réelles associées et centrées, alors elle est λ_r -faiblement dépendante avec $\lambda_r = O(\sup_{t \geq r} |cov(X_0, X_t)|)$.

c-Le décalage de Bernoulli :

Définition 1.7. Soit $(\xi_i)_{i \in \mathbb{Z}}$ une suite stationnaire de variables aléatoires réelles i.i.d et $R : \mathbb{R}^{\mathbb{Z}} \rightarrow \mathbb{R}$ une fonction mesurable. La suite $(X_i)_{i \in \mathbb{Z}}$ définie par :

$$X_i = R(\xi_i, \xi_{i-1}, \xi_{i-2}, \dots) \quad (1.4)$$

est appelée décalage de Bernoulli (schéma de Bernoulli avec des innovations indépendantes $(\xi_i)_{i \in \mathbb{Z}}$).

Proposition 1.1. Soit $(\delta_k)_{k \in \mathbb{N}}$ une suite positive décroissante vers zéro. Si la suite des innovations indépendantes $(\xi_i)_{i \in \mathbb{Z}}$ vérifie :

$$\sup_{i \in \mathbb{Z}} \mathbb{E} |R(\xi_{i-j}, j \in \mathbb{Z}) - R(\xi_{i-j} \mathbf{1}_{|j| < k}, j \in \mathbb{Z})| \leq \delta_k, \quad (1.5)$$

alors le décalage de Bernoulli est $(\epsilon_r, \mathcal{L}_1, \psi)$ -faiblement dépendant avec $\psi = 4(n+m) \min(Lip(g), Lip(h))$ et $\epsilon_r = \delta_{\frac{r}{2}}$.

L'un des principaux intérêts de ces séquences est qu'elles fournissent des exemples de processus très utilisés en statistiques qui sont faiblement dépendantes, mais ne sont pas mélangées. Nous nous donnons ici quelques exemples de tels processus :

1. **Processus de Markov stationnaire :** Soit $(X_i)_{i \in \mathbb{Z}}$ un processus de Markov stationnaire défini récursivement par

$$X_i = m(X_{i-1}, \xi_i),$$

où $(\xi_i)_{i \in \mathbb{Z}}$ est une suite stationnaire de variables aléatoires réelles i.i.d. Alors la fonction R définie dans (1.4) est :

$$R(x) = m(R(x'), x_0) \text{ avec } x = (x_0, x_1, x_2, \dots) \text{ et } x' = (x_1, x_2, x_3, \dots).$$

2. **Le processus autorégressif d'ordre p (AR(p)) :** est défini par l'équation suivante

$$X_i = c + \varphi_1 X_{i-1} + \varphi_2 X_{i-2} + \dots + \varphi_p X_{i-p} + \xi_i,$$

où $\varphi_1, \dots, \varphi_p$ sont les paramètres du modèle, c est une constante et $(\xi_i)_{i \in \mathbb{Z}}$ est une suite de variables indépendantes.

3. **Le processus bilinéaire stationnaire** : est défini par la relation

$$X_i = \xi_i \left(a + \sum_{j \geq 1} a_j X_{i-j} + b + \sum_{j \geq 1} b_j X_{i-j} \right).$$

avec $(a_j)_{j \in \mathbb{N}}$ et $(b_j)_{j \in \mathbb{N}}$ sont deux suites de nombres réels, $a, b \in \mathbb{R}$, et $(\xi_i)_{i \in \mathbb{Z}}$ est une suite de v.a i.i.d, satisfaisant à $(\mathbb{E}(|\xi_i|^p))^{\frac{1}{p}} < \infty$, $p > 0$.

d-Processus Autorégressif Conditionnellement Hétéroscédastique (ARCH) et GARCH(p,q) (Modèle ARCH Généralisé)

Définition 1.8. Soit $(\xi_i)_{i \in \mathbb{Z}}$ une suite de v.a i.i.d et $(X_i)_{i \in \mathbb{Z}}$ un processus GARCH(p,q) donné par les relations suivantes :

$$X_i = \rho_i \xi_i \text{ avec } \rho_i^2 = a_0 + \sum_{j=1}^p a_j X_{i-j}^2 + \sum_{j=1}^q b_j \rho_{i-j}^2, \quad (1.6)$$

où $(p, q) \in \mathbb{N}^2$, $a_0 > 0$, $a_j \geq 0$ et $b_j \geq 0$ pour $j \geq 1$.

Définition 1.9. Soit $(\xi_i)_{i \in \mathbb{Z}}$ une suite de variables aléatoires i.i.d et $(X_i)_{i \in \mathbb{Z}}$ un processus ARCH(1) défini par :

$$X_i = \rho_i \xi_i \text{ avec } \rho_i^2 = b_0 + \sum_{j=1}^{\infty} b_j X_{i-j}^2, \quad (1.7)$$

où $b_j \geq 0$.

On suppose qu'il existe $m > 2$ tel que $\mathbb{E}(|\xi_0|^m) < \infty$ et que la condition de stationnarité $|\xi_0|_m^2 \sum_{k=1}^{\infty} |b_k| < 1$ est vérifiée.

1. S'il existe $C > 0$ et $\mu \in]0, 1[$ tels que $\forall j \in \mathbb{N}$, $0 \leq b_j \leq C\mu^{-j}$, alors le processus (1.6) est λ_r -faiblement dépendant avec $\lambda_r = O(\exp(-c\sqrt{r}))$, $c > 0$ (voir [Doukhan et al. \(2006\)](#)).
2. S'il existe $C > 0$ et $\nu > 1$ tels que $\forall k \in \mathbb{N}$, $0 \leq b_k \leq Ck^{-\nu}$, alors le processus (1.7) est λ_r -faiblement dépendant avec $\lambda_r = O(r^{-\nu+1})$.

Nous renvoyons à [Dedecker et al. \(2007\)](#) pour une liste plus exhaustive.

1.4.4.3 Inégalités exponentielles, le théorème central limite de Lindeberg dépendant

Des travaux portant sur les inégalités exponentielles dans un modèle de données faiblement dépendantes existent et sont relativement récents, nous citons les inégalités du type Bennett ([Dedecker and Prieur \(2004a\)](#)), Fuk-Nagaev ([Dedecker and Prieur \(2004b\)](#)), inégalités du type exponentiel. Nous citons ici deux inégalités exponentielles parmi les plus employées.

Théorème 1.1. ([Doukhan and Neumann \(2007\)](#))

Soient X_1, \dots, X_n des v.a réelles de moyenne nulle définies sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et $\Psi : \mathbb{N}^2 \rightarrow \mathbb{N}$ définie par l'une des fonctions suivantes :

- (i) $\Psi(u, v) = 2v$,
- (ii) $\Psi(u, v) = u + v$,
- (iii) $\Psi(u, v) = uv$,
- (iv) $\Psi(u, v) = a(u + v) + (1 - a)uv, \forall a \in (0, 1)$.

Soient les constantes $K, M, L_1, L_2 < \infty, \mu, \nu \geq 0$ et $u, v \geq 0$. On suppose une suite non décroissante de coefficients réels $(\Phi(n))_{n \geq 0}$ telle que pour tous u -uplets (s_1, \dots, s_u) et v -uplets (t_1, \dots, t_v) avec $1 \leq s_1 \leq \dots \leq s_u \leq t_1 \leq \dots \leq t_v \leq n$, l'inégalité suivante est vérifiée :

$$|Cov(X_{s_1} \cdots X_{s_u}, X_{t_1} \cdots X_{t_v})| \leq K^2 M^{u+v-2} ((u+v)!)^\nu \Psi(u, v) \Phi(t_1 - s_u),$$

avec

$$\sum_{s=0}^{\infty} (s+1)^k (\Phi(s)) \leq L_1 L_2^k (k!)^\mu, \forall k \geq 0,$$

et

$$\mathbb{E}(|X_t|^k) \leq M^k (k!)^\nu, \forall k \geq 0.$$

Alors, $\forall t \geq 0$,

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq t\right) \leq \exp\left(-\frac{t^2/2}{A_n + B_n^{1/(\mu+\nu+2)} t^{(2\mu+2\nu+3)/(\mu+\nu+2)}}\right),$$

où $A_n \geq \sigma_n^2$ avec $\sigma_n^2 := Var\left(\sum_{i=1}^n X_i\right)$ et $B_n = 2(K \vee M)L_2 \left(\frac{2^{4+\mu+\nu} n K^2 L_1}{A_n} \vee 1\right)$.

Hwang and Shin (2014) ont établi une inégalité exponentielle pour les sommes partielles de variables aléatoires ψ -dépendantes ayant un coefficient de dépendance d'ordre $o(r^{-2})$ et une norme asymptotique finie du processus. Ces conditions sont plus simples que celles du résultat des travaux de Doukhan and Neumann (2007).

Théorème 1.2. (Hwang and Shin (2014), Théorème 2, p. 247)

Soit $(X_i)_{i \in \mathbb{N}}$ une suite stationnaire de v.a faiblement dépendante de moyenne nulle et de coefficient de dépendance ϵ_r . Soit $S_n = \sum_{i=1}^n X_i$ et $\sigma_n^2 = Var(S_n)$. Si $\epsilon_r = o(r^{-2})$ tel que

$$\lim_{n \rightarrow \infty} \|X_i\|_\gamma = \lim_{n \rightarrow \infty} \left(\mathbb{E}(|X_i|^\gamma)\right)^{\frac{1}{\gamma}} < \infty, \text{ avec } \gamma \sim \sqrt{n},$$

alors pour tout $t > 0$ et pour n suffisamment grand, on a

$$\mathbb{P}(|S_n| \geq t) \leq C \log n \exp\left(-\frac{t^2}{A_n + t^\rho B_n}\right),$$

où A_n peut être choisi comme tout nombre supérieur ou égal à σ_n^2 et $B_n = \frac{n^{\frac{3}{4}} \log n}{A_n}$ pour une constante $C > 0$ et pour tout $0 < \rho \leq 1$.

Le résultat suivant représente le théorème central limite de Lindeberg pour les processus dépendants. Il convient de noter que ce résultat a été appliqué aux processus faiblement dépendants afin d'établir la normalité asymptotique de l'estimateur à noyau de la densité pour deux cas particuliers de processus faiblement dépendants.

Théorème 1.3. (*Bardet et al. (2008), Théorème 1, p. 157*)

Supposons que $(X_{n,k})_{k \in \mathbb{N}}$, $n \in \mathbb{N}$, soit un schéma triangulaire de vecteurs aléatoires de moyenne nulle ayant des valeurs dans \mathbb{R}^d . On considère qu'il existe une matrice définie positive Σ telle que

$$\sum_{k=1}^n \text{Cov}(X_{n,k}) \rightarrow \Sigma,$$

et pour chaque $0 < \delta \leq 1$,

$$A_n = \sum_{k=1}^n \mathbb{E} \|X_{n,k}\|^{2+\delta} \rightarrow 0,$$

où $\|\cdot\|$ désigne la norme euclidienne. De plus, on suppose que :

$$T(n) = \sum_{k=2}^n \left| \text{Cov} \left(\exp \left(i \langle t, (X_{n,1} + \dots + X_{n,k-1}) \rangle \right), \exp(i \langle t, X_{n,k} \rangle) \right) \right| \rightarrow 0,$$

où $\langle \cdot, \cdot \rangle$ représente le produit scalaire. Alors

$$S_n = \sum_{k=1}^n X_{n,k} \xrightarrow{\mathcal{D}} \mathcal{N}_d(0_d, \Sigma).$$

Estimation à noyau de la fonction densité et du mode

2.1 Introduction

En statistique, l'estimation par noyau (ou encore méthode de Parzen-Rosenblatt ; en anglais, kernel density estimation) est une méthode non paramétrique d'estimation de la densité de probabilité d'une variable aléatoire. Elle se base sur un échantillon d'une population statistique et permet d'estimer la densité en tout point du support.

Dans ce chapitre, nous présentons un aperçu sur les estimateurs à noyau des fonctions de densité ; des modes et leurs importances. Nous étudions le mode de Parzen et sa généralisation, en donnant quelques résultats théoriques établis dans le cas de variables aléatoires i.i.d et α -mélangeantes. Nous examinons également le comportement asymptotique du mode conditionnel.

2.2 Estimateur à noyau de la densité

L'estimateur à noyau de la densité a été inventé par [Rosenblatt \(1956\)](#) basé sur l'idée de "dériver" la fonction de distribution empirique. Soit un n -échantillon d'une variable aléatoire réelle X de densité f . Soient X_1, X_2, \dots, X_n des v.a i.i.d de densité f par rapport à la mesure de Lebesgue sur \mathbb{R} et de fonction de répartition (f.d.r) $F(x) = \int_{-\infty}^x f(t)dt$. Un estimateur naturel de la f.d.r F est l'estimateur empirique défini pour tout $x \in \mathbb{R}$ par :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}.$$

Pour $h > 0$ assez petit on a :

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

en remplaçant F par son estimateur F_n , nous obtenons l'estimateur suivant de f :

$$\begin{aligned}\hat{f}(x) &= \frac{\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{|x-X_i| \leq h\}}}{2h}, \\ &= \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} \mathbb{1}_{\{|x-X_i| \leq h\}}.\end{aligned}$$

Ce dernier est appelé un estimateur de [Rosenblatt \(1956\)](#). Nous pouvons aussi l'écrire sous cette forme :

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K_0 \left(\frac{x - X_i}{h} \right)$$

avec $K_0(u) = \frac{1}{2} \mathbb{1}_{\{|u| \leq 1\}}$. Cet estimateur est en fait un histogramme à fenêtre mobile. [Parzen \(1962\)](#) a proposé de remplacer K_0 par un noyau plus général :

$$\hat{f}_{PR}(x) = \frac{1}{nh_n} \sum_{i=1}^n K \left(\frac{x - X_i}{h_n} \right), \quad (2.1)$$

c'est l'estimateur à noyau de la densité ou l'estimateur de Parzen-Rosenblatt. La fonction K est dite noyau et h_n est une suite de réels positifs tendant vers 0, appelée fenêtre.

La fonction noyau $K : \mathbb{R} \rightarrow \mathbb{R}$ sera supposée mesurable et satisfaisant certaines hypothèses de base parmi lesquelles les suivantes :

- K est bornée, i.e. $\sup_{u \in \mathbb{R}} |K(u)| < \infty$,
- $\lim_{|u| \rightarrow 1} |u|K(u) = 0$
- $\int_{\mathbb{R}} |K(u)| du < \infty$,
- $\int_{\mathbb{R}} K(u) du = 1$.

L'estimateur \hat{f}_{PR} , de la fonction densité a été largement étudié dans la littérature. [Rosenblatt \(1956\)](#) donna dans son article l'erreur quadratique moyenne relative à l'estimation de la densité dans le cas d'observations univariées i.i.d en considérant un noyau uniforme. [Parzen \(1962\)](#) généralisa ce résultat en considérant une classe très vaste de noyaux et établit aussi la normalité asymptotique. Par la suite, le cas multivarié fut traité par [Cacoullos \(1964\)](#). [Deheuvels \(1974\)](#) a étudié les convergences ponctuelle et uniforme presque sûre.

De nombreux auteurs, parmi lesquels nous citons [Roussas \(1969\)](#), [Tran \(1990\)](#) et [Pham and Tran \(1991\)](#), ont par la suite traité le cas où les variables sont dépendantes. Nous renvoyons à [Bosq \(1998\)](#) pour une présentation détaillées des résultats dans ce contexte.

2.2.1 Optimalité de l'estimateur de la densité

La vitesse de convergence d'un estimateur à noyau est étroitement liée au choix de sa fenêtre. Comme l'indique la figure [2.1](#) : un paramètre trop petit fait apparaître des détails artificiels sur le graphique de l'estimateur. Contrairement, lorsque h_n est trop grand la majorité des caractéristiques

sont effacées. La sélection d'une "bonne" ou d'une fenêtre optimale (courbe rouge) est donc soumise à un compromis entre sur-lissage et sous-lissage, qui peut se traduire par un compromis entre biais et variance : l'estimateur rouge a une variance trop élevée, l'estimateur vert est trop biaisé.

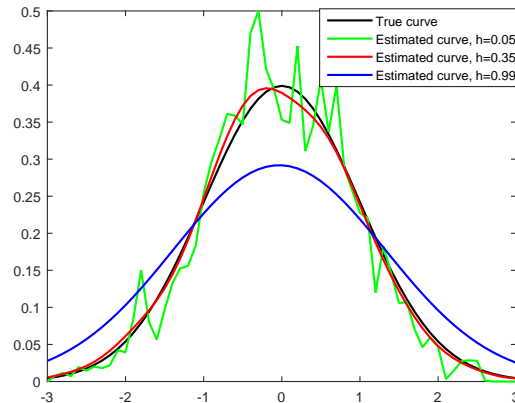


FIGURE 2.1 – Estimation par la méthode du noyau d'un échantillon de taille 500 issue d'une loi normale pour différentes valeurs de la fenêtre.

Le critère d'optimalité le plus couramment utilisé pour sélectionner ce paramètre est le critère de l'erreur quadratique moyenne intégrée :

$$\text{MISE}(\hat{f}_{PR}, h_n) = \mathbb{E} \left(\int (\hat{f}_{PR}(x) - f(x))^2 dx \right)$$

Sous des hypothèses faibles sur f et K , [Rosenblatt \(1956\)](#) et [Parzen \(1962\)](#) ont montré que :

$$\text{MISE}(\hat{f}_{PR}, h_n) = \frac{R(K)}{nh_n} + \frac{1}{4} m_2(K)^2 h_n^4 R(f'') + o \left(\frac{1}{nh_n} + h_n^4 \right),$$

où $R(g) = \int g(x)^2 dx$ pour une fonction g , $m_2(K) = \int x^2 K(x) dx$ et f'' est la seconde dérivée de f . D'où la fenêtre optimale est de la forme :

$$h_{opt} = \frac{R(K)^{1/5}}{m_2(K)^{2/5} R(f'')^{1/5} n^{1/5}}.$$

Ce choix théorique ne peut pas être utilisé en pratique puisqu'il dépend de deux quantités inconnues : la fonction de densité f et sa dérivée seconde f'' , c'est pourquoi une variété de méthodes alternatives basées sur les données ont été développées pour sélectionner la fenêtre optimale. Dans la section suivante, nous présentons les méthodes de sélection les plus couramment utilisées.

2.2.2 Quelques méthodes de sélection du paramètre de lissage

2.2.2.1 Approximation gaussienne (Silverman)

Lorsque les données suivent la loi normale de variance σ^2 , la fenêtre optimale obtenue à partir de celle établie par Parzen (1962) est de la forme,

$$h_{opt} \approx 1.06\sigma_n^2 n^{-\frac{1}{5}},$$

où σ_n^2 est l'estimateur empirique de σ^2 .

2.2.2.2 Méthode de la validation croisée

Cette méthode a été introduite par Rudemo (1982) et Bowman (1984). Pour l'estimateur à noyau \hat{f}_{PR} de la fonction de densité f , l'erreur quadratique intégrée **ISE** est définie par :

$$\mathbf{ISE}(\hat{f}_{PR}, h_n) = \int \hat{f}_{PR}^2(x) dx + \int \hat{f}_{PR}(x) f(x) dx + \int f^2(x) dx.$$

Le paramètre de lissage théorique optimal au sens de **ISE** est donné par

$$h_{opt} = \arg \min_{h_n} \mathbf{ISE}(\hat{f}_{PR}, h_n).$$

Minimiser $\mathbf{ISE}(\hat{f}_{PR}, h_n)$ par rapport à h_n revient à minimiser la quantité $\int \hat{f}_{PR}^2(x) dx + \int \hat{f}_{PR}(x) f(x) dx$. L'idée de la méthode de validation croisée est de construire un estimateur de cette dernière quantité. Cet estimateur, construit à partir des données $(X_i)_{1 \leq i \leq n}$, est donné par le critère de validation croisée suivant :

$$\mathbf{UCV}(\hat{f}_{PR}, h_n) = \int \hat{f}_{PR}^2(x) dx - \frac{2}{n} \sum_{i=1}^n f_{h_n, i}(X_i).$$

où

$$f_{h_n, i}(X_i) = \frac{1}{(n-1)h_n} \sum_{j \neq i, 1 \leq j \leq n} K\left(\frac{x - X_j}{h_n}\right).$$

Cet estimateur est un estimateur sans biais de $\mathbf{MISE}(\hat{f}_{PR}, h_n) - R(f)$.

On a :

$$\begin{aligned} \mathbf{MISE}(\hat{f}_{PR}, h_n) - R(f) &= \mathbb{E} \left(\int \hat{f}_{PR}^2(x) dx - 2 \int \hat{f}_{PR}(x) f(x) dx + \int f^2(x) dx \right) - R(f) \\ &= \mathbb{E} \left(\int \hat{f}_{PR}^2(x) dx \right) - 2 \mathbb{E} \left(\int \hat{f}_{PR}(x) f(x) dx \right). \end{aligned}$$

Montrons que $\int \hat{f}_{PR}^2(x) dx$ et $\frac{1}{n} \sum_{i=1}^n f_{h_n, i}(X_i)$ sont des estimateurs sans biais de $\mathbb{E} \left(\int \hat{f}_{PR}^2(x) dx \right)$ et $\mathbb{E} \left(\int \hat{f}_{PR}(x) f(x) dx \right)$ respectivement. Or $\mathbb{E} \left(\int \hat{f}_{PR}^2(x) dx \right)$ admet l'estimateur sans biais trivial $\int \hat{f}_{PR}^2(x) dx$.

Il reste donc à montrer que $\mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n f_{h_n,i}(X_i) \right) = \mathbb{E} \left(\int \hat{f}_{PR}(x) f(x) dx \right)$.

D'une part, nous avons :

$$\begin{aligned} \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n f_{h_n,i}(X_i) \right) &= \mathbb{E}(f_{h_n,1}(X_1)) \\ &= \mathbb{E} \left(\frac{1}{(n-1)h_n} \sum_{j \neq 1}^n \int K \left(\frac{z - X_j}{h_n} \right) f(z) dz \right) \\ &= \frac{1}{h_n} \int f(x) K \left(\frac{z - x}{h_n} \right) f(z) dz dx. \end{aligned}$$

D'autre part,

$$\begin{aligned} \mathbb{E} \left(\int \hat{f}_{PR} f \right) &= \mathbb{E} \left(\frac{1}{nh_n} \sum_{j=1}^n \int K \left(\frac{z - X_j}{h_n} \right) f(z) dz \right) \\ &= \frac{1}{h_n} \int f(x) K \left(\frac{z - x}{h_n} \right) f(z) dz dx. \end{aligned}$$

Donc $\frac{1}{n} \sum_{i=1}^n f_{h_n,i}(X_i)$ est un estimateur sans biais de $\mathbb{E} \left(\int \hat{f}_{PR}(x) f(x) dx \right)$.

Nous constatons que minimiser le **MISE**, par rapport à h_n est équivalent à minimiser le critère de validation croisée **UCV** par rapport à h_n . La règle de sélection est donnée par

$$h_{opt} = \arg \min_{h_n} \text{UCV}(\hat{f}_{PR}, h_n).$$

2.2.2.3 Méthode Plug in

En remarquant que

$$\int (f'')^2(x) dx = \int f^{(4)} f(x) dx = \mathbb{E} (f^{(4)}(X)),$$

[Sheather and Jones \(1991\)](#) proposent alors d'estimer le terme inconnu $R(f'') = \int (f'')^2(x) dx$ dans l'expression de l'erreur moyenne quadratique intégrée par l'estimateur $S(a)$ suivant :

$$S(a) = \frac{1}{n(n-1)a^5} \sum_{i=1}^n \sum_{j=1}^n K^{(4)} \left(\frac{x_i - x_j}{a} \right),$$

où $K^{(4)}$ désigne la dérivée quatrième du noyau K et où a est un nouveau paramètre de lissage appelé paramètre pilote.

$$h_{SJ} := \arg \min_{h_n} \frac{R(K)}{nh_n} + \frac{1}{4} m_2(K)^2 h_n^4 S(a).$$

Cette fenêtre de lissage vérifie ([Sheather and Jones \(1991\)](#)) :

$$\frac{h_{SJ}}{h_{opt}} = 1 + O_P(n^{-\frac{5}{14}}).$$

2.3 Estimateur à noyau du mode simple

Le mode est l'une des mesures de tendance centrale, qui est utilisée pour trouver les valeurs les plus répétées. Pour une densité de probabilité f , c'est la valeur pour laquelle f admet un maximum (global ou local). Il est important de se référer à la relation entre la moyenne, la médiane et le mode. Ces mesures sont les mesures de localisation les plus utilisées par les enquêteurs, car elles sont faciles à interpréter. On a pour des distributions continues, la relation empirique de [Pearson \(1895\)](#) suivante : $moyenne - mode \simeq 3(moyenne - médiane)$.

Notez que ces trois mesures sont égales lorsque la courbe est unimodale et symétrique.

Exemple 1 : Si X suit une loi gaussienne $X \rightarrow \mathcal{N}(\mu, \sigma^2)$, sa densité f est

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}},$$

alors f admet le mode $\theta = \mu$.

Exemple 2 : Si X suit une loi exponentielle de paramètre $\lambda > 0$, sa densité f est

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

alors f admet le mode $\theta = 0$.

Exemple 3 : Si X suit une loi Gamma de paramètres k et ρ (strictement positifs), $X \sim \Gamma(k, \rho)$, sa fonction de densité de probabilité est donnée par :

$$f(x; k, \rho) = \frac{x^{k-1} e^{-\frac{x}{\rho}}}{\Gamma(k)\rho^k},$$

où $x > 0$ et Γ désigne la fonction Gamma d'Euler. Alors f admet le mode $\theta = (k - 1)\rho$, pour $k \geq 1$. Le problème de l'estimation du mode, qui est souvent une conséquence directe de l'estimation de la densité, a été largement abordé dans la littérature. Il existe plusieurs estimateurs du mode, le plus connu étant l'estimateur de [Parzen \(1962\)](#).

Dans la section suivante, nous présentons cet estimateur et sa généralisation ainsi que certains théorèmes fondamentaux concernant les propriétés asymptotiques de ce dernier pour des données indépendantes.

2.3.1 L'estimateur de Parzen

Soit X_1, X_2, \dots, X_n un échantillon de v.a de même distribution qu'une variable aléatoire X ayant une densité de probabilité f uniformément continue avec un mode unique. L'estimateur du mode de Parzen, noté $\theta_{n,P}$ est défini comme la variable aléatoire qui maximise l'estimateur à noyau \hat{f}_{PR} de la densité f définie dans (2.1). En d'autres termes :

$$\theta_{n,P} = \arg \sup_{x \in \mathbb{R}} \hat{f}_{PR}(x).$$

Rappelons que

$$\hat{f}_{PR}(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right).$$

Théorème 2.1. *Parzen (1962)*

Supposons que K et h_n vérifient les conditions suivantes :

$$\lim_{n \rightarrow +\infty} h_n = 0,$$

$$nh_n^2 \rightarrow \infty$$

$$\sup_{-\infty < t < +\infty} |K(t)| < \infty,$$

$$\int_{-\infty}^{+\infty} |K(t)| dt < \infty,$$

$$\lim_{t \rightarrow +\infty} |tK(t)| = 0,$$

alors pour tout $\varepsilon > 0$, on a

$$\mathbb{P}(|\theta_{n,P} - \theta| < \varepsilon) \rightarrow 1,$$

quand $n \rightarrow +\infty$.

Par la suite, nous donnons le théorème de [Parzen \(1962\)](#) sur la normalité asymptotique du mode. Pour énoncer ce théorème, nous avons besoin des hypothèses suivantes :

H1 f est de classe \mathcal{C}^2 au voisinage de θ , avec $f^2(\theta) < 0$.

H2 $\exists \delta, 0 < \delta < 1$ tel que k vérifie

$$\int_{-\infty}^{+\infty} u^{2+\delta} |k(u)| du < \infty,$$

avec k la transformée de Fourier de K définie par $\int_{-\infty}^{+\infty} k(u) = \exp(-iuy)K(y)dy$.

H3 $\lim_{n \rightarrow \infty} nh_n^6 = \infty, \lim_{n \rightarrow \infty} nh_n^{5+2\delta} = 0$.

H4 La fonction caractéristique ϕ qui est égale à $n^{-1} \sum_{k=1}^n \exp(iuX_k)$ vérifie

$$\int_{-\infty}^{+\infty} u^{2+\delta} |\phi(u)| du < \infty.$$

H5 Supposons que K est de classe \mathcal{C}^1 telle que $\int_{-\infty}^{+\infty} K'^2(y) dy < \infty$.

Théorème 2.2. *Parzen (1962)*

Sous les conditions **H1-H5** on a

$$\mathbb{E} \left(\sup_{-\infty < x < \infty} |\hat{f}_{PR}''(x) - \mathbb{E}(\hat{f}_{PR}''(x))|^2 \right) \rightarrow 0,$$

$$\begin{aligned}\hat{f}_{PR}''(\theta^*) &\xrightarrow{\mathcal{P}} f''(\theta), \\ \sqrt{nh_n^3}(\theta_{n,P} - \theta) &\xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{f(\theta)}{f''(\theta)^2} J\right),\end{aligned}$$

avec $J = \int_{-\infty}^{+\infty} K'^2(u) du = (2\pi)^{-1} \int_{-\infty}^{+\infty} u^2 k^2(u) du$ et $\theta^* \in [\theta, \theta_{n,P}]$.

2.3.2 Généralisation de l'estimation du mode de Parzen.

Samanta (1973) et Konakov (1974) parmi les premiers auteurs qui ont donné des versions multivariées des résultats de Parzen.

Soit $X_j = (X_{1j}, X_{2j}, \dots, X_{dj})$, $j = 1, 2, \dots, n$ une famille de variables aléatoires à valeurs dans \mathbb{R}^d admettant une fonction de distribution commune inconnue F qui est supposée être absolument continue et une fonction de densité f . Soient K_1, K_2, \dots, K_d sont des fonctions de densité de probabilité univariées et h_n une séquence de nombres positifs convergeant vers zéro. L'estimateur f_n de la fonction f est défini comme suit :

$$\begin{aligned}f_n(x) &= \frac{1}{n} \sum_{j=1}^n h_n^{-d} \prod_{i=1}^d K_i\left(\frac{x_i - X_{ij}}{h_n}\right), \quad x = (x_1, \dots, x_d) \\ &= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} h_n^{-d} \prod_{i=1}^d K_i\left(\frac{x_i - u_i}{h_n}\right) dF_n(u),\end{aligned}$$

où $u = (u_1, u_2, \dots, u_d)$, et F_n est l'estimateur de la fonction de répartition définie par

$$F_n(u) = \frac{1}{n} \sum_{j=1}^n \prod_{i=1}^d \mathbb{1}_{\left(\frac{u_i - X_{ij}}{h_n}\right)}.$$

On note par :

$$f^{(j_1, j_2, \dots, j_d)}(x) = \frac{\partial^{j_1 + j_2 + \dots + j_d} f(x)}{\partial(x_1)^{j_1} \partial(x_2)^{j_2} \dots \partial(x_d)^{j_d}},$$

et

$$\begin{aligned}f_n^{(j_1, j_2, \dots, j_d)}(x) &= \frac{\partial^{(j_1 + j_2 + \dots + j_d)} f_n(x)}{\partial(x_1)^{j_1} \partial(x_2)^{j_2} \dots \partial(x_d)^{j_d}} \\ &= \frac{1}{n} \sum_{j=1}^n h_n^{-(d + j_1 + j_2 + \dots + j_d)} \prod_{i=1}^d K_i^{j_i}\left(\frac{x_i - X_{ij}}{h_n}\right),\end{aligned}$$

où K^{j_i} désigne la dérivée d'ordre j_i de K .

Supposons que f soit uniformément continue, alors il existe un vecteur $\theta = (\theta_1, \theta_2, \dots, \theta_d)$ tel que $f(\theta) = \max_{x \in \mathbb{R}^d} f(x)$ dont nous admettons qu'il est unique. Par ailleurs, supposons que pour chaque i , K_i est continu en t et que $K_i(t)$ s'approche de zéro lorsque t tend vers l'infini. Il existe donc une variable aléatoire de dimension d , $\theta_n = (\theta_{1n}, \theta_{2n}, \dots, \theta_{dn})$ telle que

$$f_n(\theta_n) = \max_{x \in \mathbb{R}^d} f_n(x).$$

Si pour chaque i , K_i est deux fois différentiable, alors :

$$f_n^{(1,0,0,\dots,0)}(\theta_n) = f_n^{(0,1,0,\dots,0)}(\theta_n) = \dots = f_n^{(0,0,0,\dots,1)}(\theta_n) = 0.$$

[Samanta \(1973\)](#) a établi sous les conditions décrites ci-dessous la convergence en probabilité et la normalité asymptotique de θ_n .

H6 $f^{(j_1, j_2, \dots, j_d)}(x)$ sont uniformément continues pour $j_1 + j_2 + \dots + j_d \leq 2$.

H7 $K_j^{(i)}$ sont des fonctions à variations bornées pour $j = 1, 2, \dots, d$ et $i = \{0, 1, 2\}$.

H8 $h_n = n^{-\delta}$, $0 < \delta < \frac{1}{2d+4}$.

Théorème 2.3. [Samanta \(1973\)](#)

Sous les conditions **H6-H8**,

$$\theta_n \xrightarrow{\mathcal{P}} \theta.$$

Théorème 2.4. [Samanta \(1973\)](#)

Sous les conditions **H6-H8**, on a

$$\sqrt{nh_n^{d+2}}(\theta_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, C^{-1}\Lambda C^{-1}),$$

où $\Lambda = (\lambda_{rs})$ et $C = (c_{rs})$ avec :

$$c_{rs} = \frac{\partial^2 f(x)}{\partial x_r \partial x_s} \Big|_{x=\theta},$$

$$\lambda_{rs} \leq \begin{cases} f(\theta) \int_{-\infty}^{+\infty} \left(K_r^{(1)}(t)\right)^2 dt \left(\prod_{j=1, j \neq r}^k \int_{-\infty}^{+\infty} K_j^2(t) dt\right) & \text{si } r = s, \\ f(\theta) \int_{-\infty}^{+\infty} K_r(t) K_r^{(1)}(t) dt \int_{-\infty}^{+\infty} K_s(t) K_s^{(1)}(t) dt \left(\prod_{j=1, j \neq r, s}^d \int_{-\infty}^{+\infty} K_j^2(t) dt\right) & \text{si } r \neq s. \end{cases} \quad (2.2)$$

Nous notons que pour $d = 1$, le théorème (2.4) se réduit au théorème de Parzen (2.2).

De nombreux autres auteurs se sont intéressés à l'estimateur de Parzen du mode, notamment : [Nadaraya \(1964\)](#) et [Van Ryzin \(1969\)](#) qui ont établi des résultats de convergence presque sûre en \mathbb{R} et \mathbb{R}^d , respectivement. D'autres études ont affaibli les conditions suffisantes pour la normalité asymptotique (([Eddy, 1980, 1982](#)), ([Romano \(1988\)](#)), ([Grund and Hall \(1995\)](#)) ont étudié la convergence en norme \mathbb{L}^p . [Vieu \(1996\)](#) a obtenu une vitesse de convergence quasi complète de $(\theta_n - \theta)$. [Leclerc and Pierre-Loti-Viaud \(2000\)](#) ont donné un majorant de la vitesse de convergence quasi certaine de θ_n , en utilisant des résultats log itérés. [Abraham et al. \(2004\)](#) ont étudié la normalité asymptotique du mode dans le cas multivarié et [Mokkadem and Pelletier \(2005\)](#) ont étudié sa déviation modérée. [Shi et al. \(2009\)](#) améliorent le taux de convergence de l'estimateur du mode, exprimé en fonction de la fenêtre h_n . D'autres auteurs ont traité le sujet pour des données incomplètes, parmi lesquels nous nous référons à ([Ould-Saïd and Tatachak, 2009b,a, 2011](#)), ([Benrabah et al. \(2015\)](#)), ([Ferrani et al. \(2016\)](#)). Récemment, [Hwang and Shin \(2016\)](#) ont étudié la convergence presque sûre de l'estimateur à noyau du mode pour des données faiblement dépendantes.

2.4 Estimateur à noyau du mode conditionnel

Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathbb{R}^d \times \mathbb{R}$ et $(X_1, Y_1), \dots, (X_n, Y_n)$ n copies de ce couple. On suppose que le couple (X, Y) possède une densité jointe f et que la variable aléatoire X admet une densité marginale v ($v(x) = \int_{\mathbb{R}} f(x, y) dy$). La densité conditionnelle de Y sachant $X = x$ est définie par :

$$f(y|x) = \frac{f(x, y)}{v(x)} \mathbb{1}_{\{v(x) \neq 0\}}.$$

Cette fonction joue un rôle clé dans l'analyse statistique appliquée, et son estimation a donc de nombreuses applications pratiques : en météorologie ([Hyndman et al. \(1996\)](#), [Jeon and Taylor \(2012\)](#)), en économie ([Hall et al. \(2004\)](#)), en actuariat ([Efromovich \(2010\)](#)), en médecine ([Takeuchi et al. \(2009\)](#), [Wen and Wu \(2017\)](#)), dans le domaine de l'énergie nous citons les travaux récents de [Arora and Taylor \(2016\)](#), en astronomie ([Izbicki and Lee \(2016\)](#)). Elle permet aussi d'extraire plusieurs quantités d'intérêt, notamment l'espérance, les modes et les intervalles de prédiction ([Holmes et al. \(2012\)](#))...

Plusieurs méthodes non paramétriques ont été examinées pour estimer la densité conditionnelle : la première méthode d'estimation a été introduite à la fin des années 60 par [Rosenblatt \(1969\)](#). L'idée était de remplacer le numérateur et le dénominateur de ce rapport par des estimateurs à noyau. Nous nous référons à [Youndje \(1993\)](#), [Hyndman et al. \(1996\)](#) et [Laksaci and Yousfate \(2002\)](#) pour une étude de ses propriétés asymptotiques. Une autre démarche a été proposée par [Faugeras \(2009\)](#), il a présenté des résultats asymptotiques pour son estimateur basé sur la copule sous des hypothèses de lissage de la densité marginale de Y et de la fonction copule. Nous citons aussi les estimateurs par projection ([Efromovich \(1999, 2007\)](#)), estimateurs constants par morceaux ([Gyorfi and Kohler \(2007\)](#), [Sart \(2017\)](#)), l'estimateur de [Otneim and Tjøstheim \(2018\)](#)...

Nous nous intéressons à l'estimation non paramétrique de la fonction de densité conditionnelle $f(\cdot|x)$ par la méthode du noyau.

L'estimateur à noyau de la fonction de densité conditionnelle f est donné par

$$f_n(y|x) = \frac{f_n(x, y)}{v_n(x)} \mathbb{1}_{\{v_n(x) \neq 0\}},$$

où

$$f_n(x, y) = \frac{1}{nh_{n,K}^d h_{n,H}} \sum_{i=1}^n K_d \left(\frac{x - X_i}{h_{n,K}} \right) H \left(\frac{y - Y_i}{h_{n,H}} \right),$$

et

$$v_n(x) = \frac{1}{nh_{n,K}^d} \sum_{i=1}^n K_d \left(\frac{x - X_i}{h_{n,K}} \right).$$

K_d et H , sont des noyaux et $h_K = h_{K,n}$ (resp $h_H = h_{H,n}$) est une suite de nombres réels positifs.

L'estimateur à noyau de $\Theta(\cdot)$ est défini comme le maximum de f_n :

$$\Theta_n(x) = \arg \max_{y \in \mathbb{R}} f_n(y|x), \quad x \in \mathbb{R}^d.$$

Dans ce qui suit, nous rappelons quelques résultats de convergence de cet estimateur dans les deux cas d'indépendance et de dépendance.

2.4.1 Cas des données indépendantes

Sous les hypothèses suivantes, [Samanta and Thavaneswaran \(1990\)](#) ont établi la convergence et la normalité asymptotique de l'estimateur à noyau du mode conditionnel dans le cas univarié (i.e $d = 1$) et i.i.d. Notez que dans cet article, les auteurs ont choisi que $h_K = h_H = h_n$ et $K_d = K = H$.

Les hypothèses de [Samanta and Thavaneswaran \(1990\)](#) :

Nous supposons que $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ sont des variables aléatoires i.i.d.

(H1) La fonction de densité de probabilité marginale de X , $v(x)$ est uniformément continue.

(H2) $f^{(i,j)}(x, y) := \frac{\partial^{i+j} f(x,y)}{\partial x_i \partial y_j}$ existe et est bornée pour $1 < i + j < 3$.

(H3) K et ses deux premières dérivées sont des fonctions de variation bornées.

(H4) $\lim_{|u| \rightarrow \infty} |u^2 K^{(i)}(u)| = 0$, $i = 0, 1$.

(H5) $\int_{-\infty}^{+\infty} u^i K(u) du = \begin{cases} 1 & \text{si } i = 0, \\ 0 & \text{si } i = 1, 2 \end{cases}$

(H6) $\int_{-\infty}^{+\infty} |u^3| K(u) du < \infty$.

(H7) $h_n = n^{-\delta}$ avec $\frac{1}{10} < \delta < \frac{1}{8}$.

Théorème 2.1. ([Samanta and Thavaneswaran \(1990\)](#))

Sous les conditions **(H1-H3)** et **(H7)**, si $v(x) > 0$, alors l'estimateur Θ_n converge presque sûrement vers Θ lorsque n tend vers l'infini.

Théorème 2.2. ([Samanta and Thavaneswaran \(1990\)](#))

Sous les conditions **(H1-H7)** et si $v(x) > 0$, on a

$$\frac{nh_n^4}{V(x, \Theta(x))} f^{(0,2)}(x, \Theta(x)) (\Theta_n(x) - \Theta(x)) \rightarrow \mathcal{N}(0, 1),$$

avec $V(x, \Theta(x)) = f(x, \Theta(x)) \int_{\mathbb{R}^2} K(u) K^{(1)}(v)^2 dudv$.

2.4.2 Cas des données mélangées

Théorème 2.3. ([Collomb et al. \(1986\)](#))

Soit $(X_i, Y_i)_{\{1 \leq i \leq n\}}$ un processus à valeurs dans $\mathbb{R}^d \times \mathbb{R}$, strictement stationnaire et ϕ -mélangeant. Sous des hypothèses classiques sur la densité f , les noyaux et si la suite h_n ($h_n = h_K = h_H$) vérifie conjointement avec la suite m_n , la condition

$$\frac{m_n \log n}{nh_n^{d+1}} \rightarrow 0, \quad n \rightarrow \infty,$$

où m_n est une suite d'entiers tendant vers l'infini telle que :

$$\exists A < \infty, \frac{n\phi_{m_n}}{m_n} \leq A, \quad 1 \leq m_n \leq n, \quad n \in \mathbb{N}.$$

Alors,

$$\sup_{x \in S \subset \mathbb{R}^d} \left| \Theta_n(x) - \Theta(x) \right| \rightarrow 0, \quad p.s.$$

Théorème 2.4. (*Louani and Ould-Saïd (1999)*)

Soit $(X_i, Y_i)_{\{1 \leq i \leq n\}}$ un processus à valeurs dans $\mathbb{R}^d \times \mathbb{R}$, strictement stationnaire et α -mélangeant.

Sous des hypothèses standard sur la densité, les deux noyaux K_d , H et si de plus la suite h_n ($h_n = h_K = h_H$) vérifie les conditions suivantes lorsque $n \rightarrow \infty$:

- $nh_n^{d+7} \rightarrow c^2$; où c est une constante positive,
- $m_n h_n^{d+1} \rightarrow 0$ où m_n est une suite d'entiers tendant vers l'infini telle que :

$$1 \leq m_n \leq n \text{ et } \frac{1}{h_n^{(d+1)\delta}} \sum_{l=m_n}^{\infty} (\alpha_l)^\delta, \quad \text{avec } \delta \in (0, 1);$$

- Il existe des suites (p_n) et (q_n) tendant vers l'infini telles que $p_n + q_n \leq n$ et μ_n vérifiant les conditions suivantes, où μ_n , est le plus grand entier positif tel que $\mu_n(p_n + q_n) \leq n$:

1. $\frac{\mu_n q_n}{n} \rightarrow 0$,
2. $p_n h_n \rightarrow \infty$,
3. $\frac{(nh_n^{d+3})^{\frac{1}{2}}}{p_n} \rightarrow \infty$,
4. $\mu_n \alpha_{q_n} \rightarrow 0$,

5. Il existe une constante positive A telle que pour tout $p > 1$:

$$\frac{nh_n^{(d+1)(1-\frac{1}{p})}}{p_n^4} \sum_{i=1}^{p_n} \alpha_i^{\frac{1}{p}} + \frac{n}{p_n^2} \alpha_{p_n}^{\frac{2p_n}{3n}} \leq A,$$

6. $\lim_{n \rightarrow \infty} h_n^{-J} \exp(-Cnh_n^{d+3} p_n^{-2}) = 0$, où C est une constante positive et $J > q + d + 4$ (q est un nombre réel tel que $q > d + 3$).

- $h_n = n^{-\rho}$ avec $\frac{1}{(d+7)} < \rho < \frac{1}{(d+5+\frac{(d+1)}{p})}$ où p est un nombre réel tel que $p > \frac{(d+1)}{2}$.

On a :

$$\sqrt{nh_n^{d+3}} \left(\Theta_n(x) - \Theta(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V(x)),$$

où

$$V(x) := \frac{V(x, \Theta(x))}{\left(f^{(0,2)}(x, \Theta(x))\right)^2} = \frac{f(x, \Theta(x)) \int_{\mathbb{R}^d} \int_{\mathbb{R}} \left(K_d(u) H^{(1)}(v)\right)^2 dudv}{\left(f^{(0,2)}(x, \Theta(x))\right)^2}.$$

Propriétés asymptotiques de l'estimateur du mode conditionnel pour les données ψ -faiblement dépendantes

Dans ce chapitre, nous présentons un article publié dans la revue *Communications in Statistics-Theory and Methods* [Rih and Tatachak \(2021\)](#). Nous établissons la convergence uniforme presque sûre sur un compact avec vitesse et la normalité asymptotique de l'estimateur à noyau du mode conditionnel dans le cadre de ψ -faiblement dépendance. Les principales étapes dans la preuve de nos résultats sont l'utilisation de l'inégalité exponentielle de [Hwang and Shin \(2014\)](#) et le théorème de [Bardet et al. \(2008\)](#) (Linderberg pour des données dépendantes). Nous illustrons nos résultats théoriques par des simulations.

3.1 Préliminaires

Avant d'énoncer les résultats, nous avons besoin de quelques éléments préliminaires. Dans ce qui suit, toutes les relations de limites sont exprimées en $n \rightarrow \infty$. En effet, pour deux fonctions positives $a(n)$ et $b(n)$, on note $a(n) = o(b(n))$ si $\lim a(n)/b(n) = 0$ et $a(n) = O(b(n))$ si $\limsup a(n)/b(n) < \infty$. En outre, l'ensemble I désignera un sous-ensemble compact de \mathbb{R} où $\Theta(x) \in I^\circ$ (I° représente l'intérieur de I), et S désignera un sous-ensemble compact de $\Omega := \{x \in \mathbb{R}^d, \inf_x v(x) > 0\}$. De plus, la lettre C sera utilisée sans distinction comme une constante positive générique.

Soit $Z_i = (X_i, Y_i)_{i \in \mathbb{N}}$ un échantillon d'observations ψ -faiblement dépendantes. Notez que la classe de fonctions ψ que nous allons utiliser est définie par les fonctions suivantes :

$$\begin{aligned} \psi_1(g, h, n, m) &= \min(n, m) \text{Lip}(g) \text{Lip}(h), \\ \psi_2(g, h, n, m) &= 4(n + m) \min(\text{Lip}(g), \text{Lip}(h)). \end{aligned}$$

On rappelle que $f(\cdot|x) = \frac{f(x, \cdot)}{v(x)} \mathbb{1}_{\{v(x) \neq 0\}}$ représente la fonction de densité conditionnelle de Y sachant que $X = x$. Les fonctions $f(\cdot, \cdot)$ et $v(\cdot)$ sont la fonction de densité de probabilité conjointe du couple (X, Y) et la densité marginale de X , respectivement.

Nous rappelons également que si $\Theta(x)$ désigne le mode unique de $f(\cdot|x)$, alors

$$\Theta(x) = \arg \max_{y \in \mathbb{R}} f(y|x).$$

L'estimateur à noyau du mode conditionnel $\Theta(x)$ est défini comme la variable aléatoire $\Theta_n(x)$ qui maximise l'estimateur à noyau $f_n(y|x)$ de $f(y|x)$ (voir [Collomb et al. \(1986\)](#)), soit

$$\Theta_n(x) = \arg \max_{y \in \mathbb{R}} f_n(y|x).$$

Ici

$$f_n(y|x) = \frac{f_n(x, y)}{v_n(x)} \mathbb{1}_{\{v_n(x) \neq 0\}}, \tag{3.1}$$

avec

$$f_n(x, y) = \frac{1}{nh_{n,K}^d h_{n,H}} \sum_{i=1}^n K_d \left(\frac{x - X_i}{h_{n,K}} \right) H \left(\frac{y - Y_i}{h_{n,H}} \right)$$

et

$$v_n(x) = \frac{1}{nh_{n,K}^d} \sum_{i=1}^n K_d \left(\frac{x - X_i}{h_{n,K}} \right).$$

Les fonctions K_d et H sont des noyaux de probabilité définis sur \mathbb{R}^d et \mathbb{R} respectivement, et $h_{n,K} =: h_K(h_{n,H} =: h_H)$ sont des suites positives décroissantes vers zéro, appelées fenêtres.

3.2 Hypothèses

B. Pour tout $d \geq 1$, les fenêtres h_K et h_H satisfont :

- (i) $h_K \rightarrow 0$;

- (ii) $h_H \rightarrow 0$;
- (iii) $\frac{\log n}{nh_K^d h_H} \rightarrow 0$;
- (iv) $h_K = O(n^{-\alpha})$ et $h_H = O(n^{-\beta})$ avec $1 - 3\beta \leq \alpha < 1 - \beta$, $0 < \beta < \frac{1}{3}$;
- (v) $nh_K^d h_H^7 \rightarrow 0$.

M1. La suite $Z_i = (X_i, Y_i)_{1 \leq i \leq n}$ est strictement stationnaire et ψ_ℓ -faiblement dépendante ($\ell = 1$ ou 2);

M2. Le coefficient de dépendance satisfait la condition suivante : $\epsilon_r = o(r^{-2})$ et $U(r) := \sum_{i=r}^{\infty} \epsilon_i = O(r^{-4})$ pour $\ell = 1$ (resp $U(r) = O(r^{-3})$ pour $\ell = 2$).

D1. $f(y | x)$ est deux fois différentiable, uniformément continue et possède une dérivée seconde continue $f^{(2)}(\cdot | x)$ qui ne s'annule pas au voisinage de $\Theta(x)$.

D2. Le mode conditionnel satisfait : pour tout $\varepsilon > 0$ et $\eta(x)$, il existe un $\tau > 0$ tel que

$$\sup_{x \in S} |\Theta(x) - \eta(x)| \geq \varepsilon \Rightarrow \sup |f(\Theta(x)|x) - f(\eta(x)|x)| \geq \tau;$$

D3. La densité conjointe $f(\cdot, \cdot)$ de (X_i, Y_i) est de classe \mathcal{C}^3 , et bornée telle que $\sup_{x,y} |f^{(r,s)}(x, y)| < \infty$ pour $r + s \leq 3$;

D4. La densité conjointe $f_{1,j}(\cdot, \cdot, \cdot, \cdot)$ de $((X_1, Y_1), (X_j, Y_j))$ existe et satisfait :

$$\sup_{(u,v,s,t)} |f_{1,j}(u, v, s, t) - f(u, v)f(s, t)| < \infty;$$

D5. La densité de probabilité $v(\cdot)$ est deux fois continûment différentiable;

D6. La densité conjointe $v_{1,j}(\cdot, \cdot)$ de (X_1, X_j) existe et satisfait : $\sup_{(u,r)} |v_{1,j}(u, r) - v(u)v(r)| < \infty$;

K1. K_d est une densité lipschitzienne, vérifiant $\|K_d\|_\infty \leq 1$;

K2. (i) $\int_{\mathbb{R}^d} s_i K_d(s) ds = 0, i = 1, \dots, d, s = (s_1, s_2, \dots, s_d)^\top$ et $\int_{\mathbb{R}} t H(t) dt = 0$;

(ii) $\int_{\mathbb{R}^d \times \mathbb{R}} |s_1^{i_1} \times \dots \times s_d^{i_d} t^j| K_d(s) H(t) ds dt < \infty$ avec $i_1 + \dots + i_d + j \leq 2$.

K3. Le noyau H est de classe \mathcal{C}^2 et sa dérivée $H^{(2)}$ est lipschitzienne, vérifiant $\|H\|_\infty \leq 1$ et $\|H^{(2)}\|_\infty \leq 1$.

K4. $\int_{\mathbb{R}} |H^{(1)}(t) dt|^{2+\delta} < \infty, 0 < \delta \leq 1$.

Remarque 3.1. (Commentaires sur les hypothèses) Les hypothèses **D1-D6** sont standard dans l'estimation à noyau de la densité conditionnelle et du mode conditionnel. Les hypothèses sur les noyaux ne sont pas très restrictives et celles qui imposent $\|K_d\|_\infty \leq 1, \|H\|_\infty \leq 1$ et $\|H^{(2)}\|_\infty \leq 1$ sont nécessaires pour calculer les covariances sous la ψ -faiblement dépendance (le noyau gaussien satisfait à ces conditions). Les hypothèses **B(i)-B(iii)** sont utilisées pour énoncer le théorème 3.1. Notez que sous **B(iii)**, nous avons $nh_K^d h_H \rightarrow \infty$ qui est une hypothèse classique en estimation

à noyau. Les hypothèses $\mathbf{B}(\mathbf{v})$ et $\mathbf{B}(\mathbf{i}\mathbf{v})$ sont nécessaires pour établir la normalité asymptotique de $\Theta_n(x)$. Enfin, l'hypothèse $\mathbf{M1}$ caractérise le modèle étudié et $\mathbf{M2}$ est utilisée pour prouver les Lemmes 3.2, 3.3 et 3.7.

3.3 Consistance

3.3.1 La vitesse de convergence presque sûre pour $f_n(\cdot|x)$

Théorème 3.1. *Sous les hypothèses $\mathbf{B}(\mathbf{i}\text{-iii})$, $\mathbf{M1}\text{-M2}$, $\mathbf{D3}\text{-D6}$, $\mathbf{K1}$, $\mathbf{K2}(\mathbf{i})$ et $\mathbf{K3}$ et pour n suffisamment grand, on a*

$$\sup_{x \in S} \sup_{y \in I} |f_n(y|x) - f(y|x)| = O \left\{ \sqrt{\frac{\log n}{nh_K^d h_H}} + h_K^2 + h_H^2 \right\}, \text{ p.s.}$$

3.3.2 La vitesse de convergence presque sûre pour $\Theta_n(\cdot)$

Comme application du Théorème 3.1 nous obtenons le taux de convergence presque sûre de $\Theta_n(\cdot)$.

Corollaire 3.1. *Si les hypothèses du Théorème 3.1 et $\mathbf{D2}$ sont satisfaites, alors pour n assez grand nous avons*

$$\sup_{x \in S} |\Theta_n(x) - \Theta(x)| = o(1), \text{ p.s.}$$

De plus, si l'hypothèse $\mathbf{D1}$ est vérifiée, alors pour n assez grand, on a

$$\sup_{x \in S} |\Theta_n(x) - \Theta(x)| = O \left\{ \left(\frac{\log n}{nh_K^d h_H} \right)^{\frac{1}{4}} + h_K + h_H \right\}, \text{ p.s.}$$

3.4 Normalité asymptotique et intervalles de confiance

Nous nous intéressons maintenant à la distribution asymptotique de $\Theta_n(x)$. Pour ce faire, supposons que $f(\cdot|x)$ possède un mode unique $\Theta(x)$. Donc, par l'hypothèse $\mathbf{D1}$ nous avons

$$f^{(1)}(\Theta(x)|x) = 0 \text{ et } f^{(2)}(\Theta(x)|x) < 0,$$

de même, il s'ensuit que

$$f_n^{(1)}(\Theta_n(x)|x) = 0 \text{ et } f_n^{(2)}(\Theta_n(x)|x) < 0.$$

Notez que $f_n^{(j)}(y|x)$ représente l'estimateur à noyau de la dérivée d'ordre j ($j \geq 1$) de la densité conditionnelle $f_n(y|x)$ par rapport à y , défini par :

$$f_n^{(j)}(y|x) = \frac{f_n^{(0,j)}(x, y)}{v_n(x)},$$

et

$$f_n^{(0,j)}(x, y) = \frac{1}{nh_K^d h_H^{j+1}} \sum_{i=1}^n K_d \left(\frac{x - X_i}{h_K} \right) H^{(j)} \left(\frac{y - Y_i}{h_H} \right).$$

Ensuite, un développement de Taylor de $f_n^{(1)}(\cdot|x)$ dans le voisinage de $\Theta(x)$ nous donne

$$0 = f_n^{(1)}(\Theta_n(x)|x) = f_n^{(1)}(\Theta(x)|x) + (\Theta_n(x) - \Theta(x)) f_n^{(2)}(\Theta_n^*(x)|x),$$

où $\Theta_n^*(x)$ est compris entre $\Theta_n(x)$ et $\Theta(x)$. Par conséquent,

$$\Theta_n(x) - \Theta(x) = - \frac{f_n^{(1)}(\Theta(x)|x)}{f_n^{(2)}(\Theta_n^*(x)|x)}.$$

Et, en utilisant (3.1) il résulte que

$$\Theta_n(x) - \Theta(x) = - \frac{f_n^{(0,1)}(x, \Theta(x))}{f_n^{(0,2)}(x, \Theta_n^*(x))}, \quad (3.2)$$

lorsque le dénominateur est non nul.

La distribution asymptotique de $\Theta_n(x)$ est représentée par le résultat suivant.

3.4.1 Normalité asymptotique de $\Theta_n(\cdot)$

Théorème 3.2. *Si les hypothèses **B**, **M1-M2**, **D3-D4** et **K1-K4** sont vérifiées, alors*

$$\sqrt{nh_K^d h_H^3} (\Theta_n(x) - \Theta(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(x)),$$

où

$$\sigma^2(x) := \frac{V(x, \Theta(x))}{\left(f_n^{(0,2)}(x, \Theta(x))\right)^2} := \frac{f(x, \Theta(x)) \int_{\mathbb{R}^d} \int_{\mathbb{R}} \left(K_d(u) H^{(1)}(v)\right)^2 dudv}{\left(f_n^{(0,2)}(x, \Theta(x))\right)^2}.$$

Un estimateur convergent de type plug-in pour la variance $\sigma^2(x)$ est obtenu en utilisant les estimateurs $\Theta_n(x)$ et $V_n(x, \Theta_n(x))$, c'est-à-dire

$$\sigma_n^2(x) := \frac{f_n(x, \Theta_n(x)) \int_{\mathbb{R}^d} \int_{\mathbb{R}} \left(K_d(u) H^{(1)}(v)\right)^2 dudv}{\left(f_n^{(0,2)}(x, \Theta_n(x))\right)^2}.$$

3.4.2 Intervalles de confiance pour $\Theta(\cdot)$

Comme application immédiate du Théorème 3.2, nous obtenons le résultat ci-après.

Corollaire 3.2. *Sous les hypothèses du Théorème 3.2 et pour chaque valeur fixe de $\zeta \in (0, 1)$, l'intervalle de confiance de niveau $(1 - \zeta)$ pour $\Theta(x)$ est :*

$$\Theta_n(x) \pm t_{1-\frac{\zeta}{2}} \frac{\sigma_n(x)}{\sqrt{nh_K^d h_H^3}},$$

où $t_{1-\frac{\zeta}{2}}$ dénote le quantile d'ordre $(1 - \frac{\zeta}{2})$ de la loi normale centrée réduite.

3.5 Etude de simulation

3.5.1 Cas unidimensionnel

Nous avons mené une étude de simulation pour examiner la performance de l'estimateur du mode conditionnel $\Theta_n(x)$ dans le cas d'une covariable unidimensionnelle X (c'est-à-dire $d = 1$). Pour calculer $\Theta_n(x)$, nous avons utilisé un noyau gaussien et les fenêtres optimales globales (glob) qui minimisent l'erreur quadratique moyenne globale (GMSE). Les résultats obtenus illustrent à la fois la consistance et la normalité asymptotique.

3.5.1.1 Consistance

Tout d'abord, nous générons une suite ψ -faiblement dépendante $\{W_\ell; \ell \geq 1\}$ comme suit :

$$W_\ell = \frac{1}{2}W_{\ell-1} + \varepsilon_\ell; W_0 = 0 \text{ et } \varepsilon_\ell \sim \mathcal{N}(0, 1). \quad (3.3)$$

Nous obtenons alors la suite ψ -faiblement dépendante $\{(X_i, Y_i); 1 \leq i \leq n\}$ par

$$X_i = W_i \text{ et } Y_i = W_{i+1}.$$

Notez que selon ce modèle, on a $\Theta(x) = x/2$, puisque la v.a $(Y_i|X_i = x)$ a une distribution $\mathcal{N}(x/2, 1)$. Ensuite, nous calculons l'estimateur $\Theta_n(x)$ et nous évaluons le *GMSE* comme suit :

Étape 1 sous le modèle linéaire présenté dans (3.3), nous simulons $(X_i, Y_i)_{\{1 \leq i \leq n\}}$ pour différentes tailles $n = 50, 100$ et 500 et calculons l'estimateur $\Theta_n(x)$ en utilisant des fenêtre $(h_K, h_H) \in [0.1, 0.9] \times [0.1, 0.9]$.

Étape 2 pour chaque taille n , nous répétons $M = 300$ exécutions de simulation comme décrit à l'étape 1.

Étape 3 nous évaluons le *GMSE* correspondant.

Dans la figure 3.1, pour $x \in [-2, 2]$ nous traçons $\Theta_n(x)$ et $\Theta(x)$ pour $n = 50, 100$ et 500 . Comme nous le voyons, l'estimateur est uniformément performant en x et l'ajustement devient meilleur pour les valeurs élevées de n .

Dans la figure 3.2, pour $y \in [-2, 2]$ nous fixons x de manière arbitraire ($x = 0$) et nous traçons $f(y|0)$ par rapport à $f_n(y|0)$.

Maintenant, afin d'évaluer quantitativement la performance de nos estimateurs, nous calculons le *GMSE* qui représente la moyenne du *MSE* comme indiqué dans le tableau ci-dessous 3.1.

Avant cela, rappelons que pour une fonction g donnée et son estimation $g_{n,h_K,h_H,k}$, la valeur du *GMSE* calculée le long de M essais de Monte Carlo et d'une grille de fenêtres h_K et h_H est définie comme suit

$$GMSE(h_K, h_H) = \frac{1}{Mm} \sum_{k=1}^M \sum_{\ell=1}^m \left(g_{n,h_K,h_H,k}(x_\ell) - g(x_\ell) \right)^2,$$

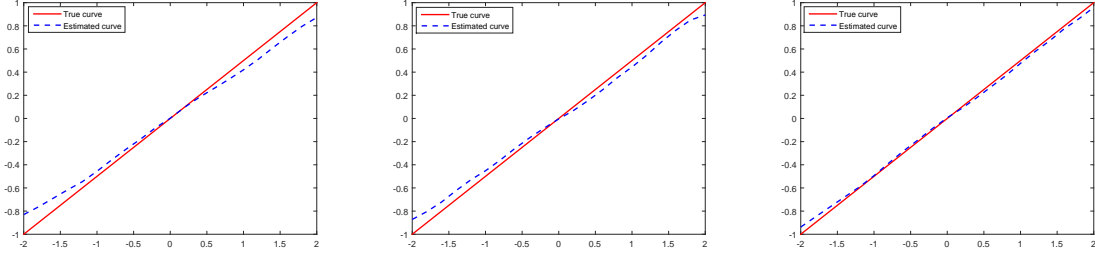


FIGURE 3.1 – Les courbes de $\Theta_n(\cdot)$ par rapport à $\Theta(\cdot)$ pour $n = 50, 100$ et 500 de gauche à droite.

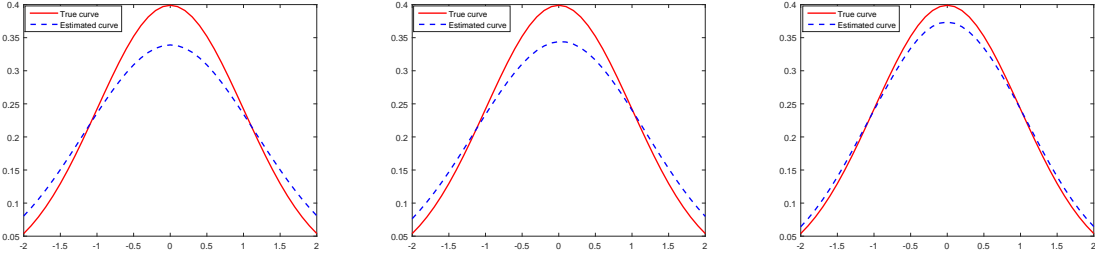


FIGURE 3.2 – Les courbes de $f_n(\cdot|0)$ par rapport à $f(\cdot|0)$ pour $n = 50, 100$ et 500 de gauche à droite.

où m est le nombre de points équidistants x_ℓ appartenant à un ensemble donné et $g_{n,h_K,h_H,k}(x_\ell)$ est la valeur de $g_{n,h_K,h_H,k}(x_\ell)$ calculée à l'itération k .

Les différentes valeurs de $GMSE := \min_{(h_K,h_H) \in [0.1,0.9]^2} GMSE(h_K, h_H)$ et MSE calculées pour $\Theta_n(\cdot)$ et $f_n(\cdot|0)$ respectivement, en utilisant les fenêtres optimales (h_{Kglob}, h_{Hglob}) , sont récapitulées dans le tableau 3.1.

TABLE 3.1 – Valeurs du MSE, GMSE pour $f_n(\cdot|0)$ et $\Theta_n(\cdot)$ (resp) avec (h_{Kglob}, h_{Hglob})

n	$f_n(\cdot 0)$			$\Theta_n(\cdot)$		
	MSE	h_{Kglob}	h_{Hglob}	GMSE	h_{Kglob}	h_{Hglob}
50	0.0030	0.55	0.75	0.2277	0.50	0.80
100	0.0024	0.50	0.65	0.1266	0.40	0.80
500	0.0005	0.20	0.50	0.0146	0.30	0.70

3.5.1.2 Normalité asymptotique

Dans cette partie, sous le modèle (3.3), nous examinerons la normalité asymptotique de notre estimateur dans le cas unidimensionnel et pour des échantillons de tailles finies. Le mode conditionnel est estimé de la même façon que pour la partie consistance. Pour cela, nous allons comparer la forme de la densité normale centrée réduite à celle de l'écart normalisé entre notre estimateur et le mode

conditionnel théorique pour $x = 1/2$, soit

$$\bar{\Theta}(n) := \sqrt{nh_K h_H^3} (\Theta_n(1/2) - \Theta(1/2)) \sigma_n(1/2).$$

La densité de probabilité de cet écart normalisé est estimée par la méthode du noyau.

Pour ce faire, nous tirons $B = 300$ échantillons indépendants $(\bar{\Theta}_1(n), \bar{\Theta}_2(n), \dots, \bar{\Theta}_{300}(n))$ de $\bar{\Theta}(n)$, ($n = 50$ et $n = 500$). Pour calculer la fonction de densité estimée de $\bar{\Theta}_k(n)$, $1 \leq k \leq 300$, nous utilisons la fenêtre de Silverman et le noyau gaussien. Nous traçons les deux densités ensemble, ainsi que les QQ-plot correspondants (Figures 3.3, 3.4). Les graphiques montrent que la distribution $\bar{\Theta}(n)$ correspond assez bien à la distribution normale et que la qualité de l'ajustement s'améliore lorsque la taille de l'échantillon n augmente.

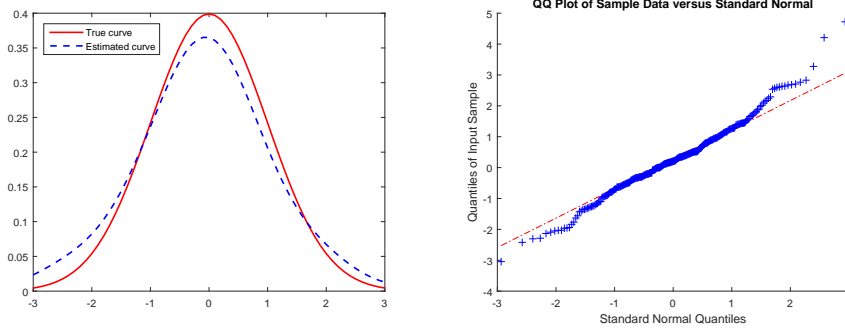


FIGURE 3.3 – La densité de $\bar{\Theta}(n)$ par rapport à $\mathcal{N}(0, 1)$ et QQ-plot pour $n=50$.

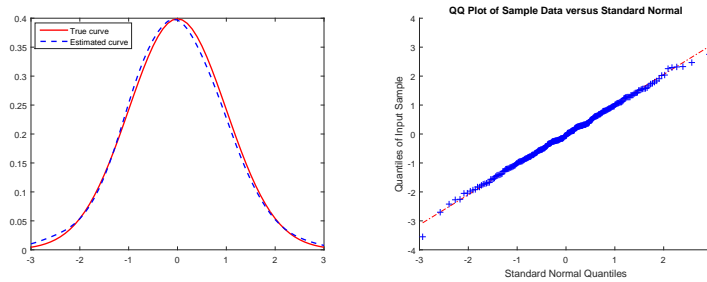


FIGURE 3.4 – La densité de $\bar{\Theta}(n)$ par rapport à $\mathcal{N}(0, 1)$ et QQ-plot pour $n=500$.

Nous nous construisons des intervalles de confiance à 95% pour le mode conditionnel pour différentes valeurs de n (50, 500) et considérons $B = 300$, nous retenons $x = 1/2$.

$$\Theta_n(x) \pm t_{1-\frac{\xi}{2}} \frac{\sigma_n(x)}{\sqrt{nh_K h_H^3}},$$

Dans le même graphique, nous traçons l'estimateur du mode conditionnel, le mode conditionnel théorique ainsi que les bornes inférieure et supérieure de l'intervalle de confiance approximé à 95%. Les graphiques obtenus montrent une diminution de la largeur de l'intervalle lorsque la taille de l'échantillon augmente.

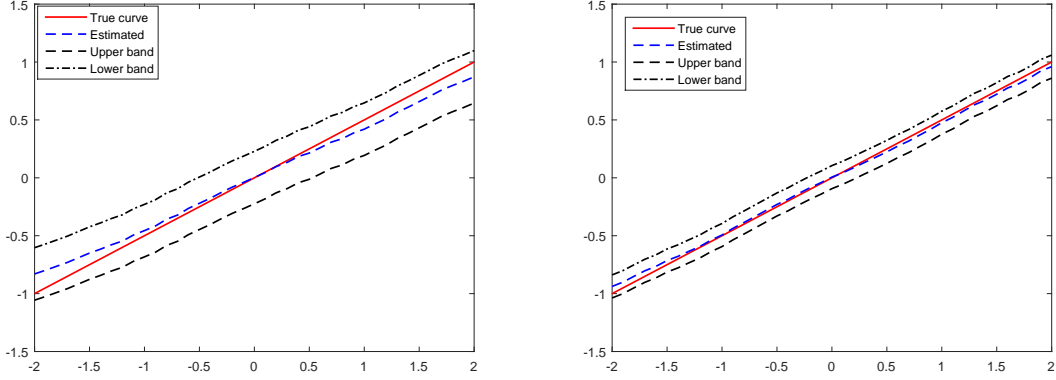


FIGURE 3.5 – Les intervalles de confiance de niveau asymptotique de 95% pour $n = 50$ et $n = 500$.

3.5.2 Cas bidimensionnel

Dans cette section, nous étudions la performance de l'estimateur du mode conditionnel $\Theta_n(x_1, x_2)$ dans le cas d'une covariable bidimensionnelle $X = (X^1, X^2)$. Nous suivons pour cela les mêmes étapes que pour le cas unidimensionnel. Nous générons une suite ψ -faiblement dépendante $\{W_\ell; \ell \geq 2\}$ comme suit :

$$W_\ell = 0.4W_{\ell-1} + 0.5W_{\ell-2} + \varepsilon_\ell; \quad W_1 = 0, W_0 = 0; \quad \varepsilon_\ell \sim \mathcal{N}(0, 1).$$

On obtient alors la suite stationnaire ψ -faiblement dépendante $\left\{ \left((X_i^1, X_i^2), Y_i \right); 1 \leq i \leq n \right\}$ par

$$(X_i^1, X_i^2) = (W_{i+1}, W_{i+2}) \text{ et } Y_i = W_{i+3}.$$

Notez que pour ce modèle, on a $\Theta(x_1, x_2) = 0, 5x_1 + 0, 4x_2$, car la v.a $(Y_i | (X_i^1, X_i^2) = (x_1, x_2))$ possède une distribution $\mathcal{N}(0, 5x_1 + 0, 4x_2, 1)$. Pour calculer l'estimateur $\Theta_n(x_1, x_2)$, nous avons utilisé le noyau normal bivarié (d=2) K_2 défini par : $K_2(x) = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} \exp\left(\frac{1}{2}(X-x)^t \Sigma^{-1}(X-x)\right)$ avec

$$\Sigma = \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}.$$

Les figures suivantes (Figures 3.6, 3.7) présentent la fonction de mode conditionnel théorique et les surfaces estimées pour $n = 50$, $n = 100$, $n = 500$, respectivement. Dans le tableau 3.2, nous avons relevé les valeurs GMSE de $\Theta_n(x_1, x_2)$, $(x_1, x_2) \in [-4, 4] \times [-4, 4]$ pour $n = 50, 100$ et 500 .

TABLE 3.2 – Valeurs du GMSE pour $\Theta_n(\cdot, \cdot)$ avec (h_{Kglob}, h_{Hglob})

n	GMSE	h_{Kglob}	h_{Hglob}
50	0.228	0.45	0.8
100	0.128	0.40	0.75
500	0.0588	0.30	0.75

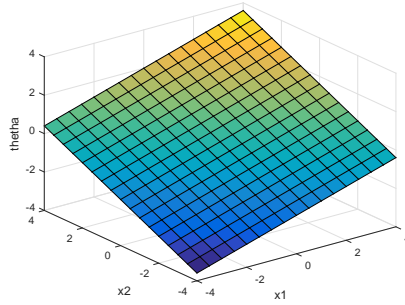


FIGURE 3.6 – Le mode conditionnel théorique $\Theta(.,.)$

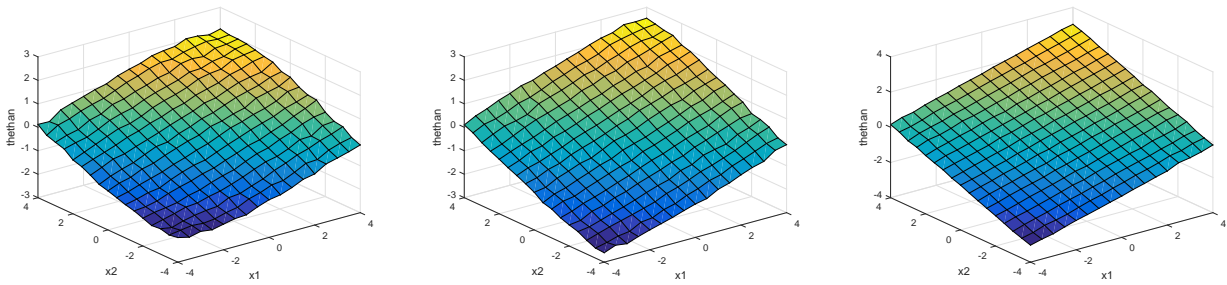


FIGURE 3.7 – Surface du mode conditionnel $\Theta_n(.,.)$ pour $n = 50, 100$ et 500 de gauche à droite

Les résultats indiquent que l'estimateur est très performant dans le cas bidimensionnel. Dans la figure 3.8, nous traçons l'estimation de la densité du noyau de $\bar{\bar{\Theta}}(n)$ par rapport à la fonction de densité $\mathcal{N}(0, 1)$ où

$$\bar{\bar{\Theta}}(n) := \sqrt{nh_K h_H^3} (\Theta_n(0, 0) - \Theta(0, 0)) \sigma_n(0, 0).$$

Les graphiques montrent que, pour des valeurs suffisamment grandes de n , la distribution d'échantillonnage de $\bar{\bar{\Theta}}(n)$ correspond bien à la distribution normale standard.

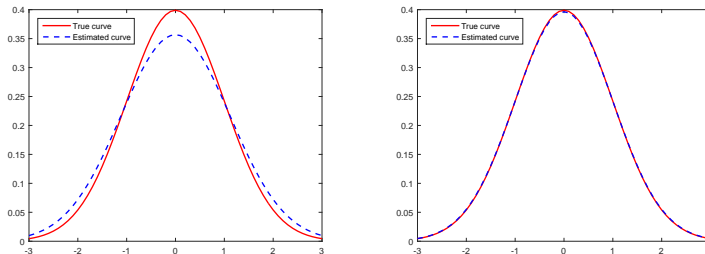


FIGURE 3.8 – La densité de $\bar{\bar{\Theta}}(n)$ par rapport à $\mathcal{N}(0, 1)$ pour $n = 50$ et $n = 500$ de gauche à droite

3.6 Résultats auxiliaires et preuves

Les preuves de nos résultats sont basées sur les Théorèmes 1.2 et 1.3.

Afin de prouver le Théorème 3.1, nous énonçons d'abord les résultats des lemmes 3.1-3.3 présentés ci-après.

Lemme 3.1. *Sous les hypothèses $\mathbf{B(i-ii)}$, $\mathbf{D3}$ et $\mathbf{K1-K2}$, nous avons*

$$\sup_{x \in S} \sup_{y \in I} \left| \mathbb{E}(f_n(x, y)) - f(x, y) \right| = O\left((h_K^2 + h_H^2) \right).$$

Preuve du Lemme 3.1 Le terme du biais ne dépend pas de la structure de dépendances et ainsi, selon les arguments standards et sous les conditions $\mathbf{B(i-ii)}$, $\mathbf{D3}$ et $\mathbf{K1-K2}$, nous avons

$$\sup_{x \in S} \sup_{y \in I} \left| \mathbb{E}(f_n(x, y)) - f(x, y) \right| = O\left((h_K^2 + h_H^2) \right).$$

Lemme 3.2. *Sous les hypothèses $\mathbf{B(i-iii)}$, $\mathbf{M1-M2}$, $\mathbf{D3-D4}$, $\mathbf{K1}$, $\mathbf{K2(i)}$ et $\mathbf{K3}$, on obtient pour n assez grand*

$$\sup_{x \in S} \sup_{y \in I} \left| f_n(x, y) - \mathbb{E}(f_n(x, y)) \right| = O\left(\sqrt{\frac{\log n}{nh_K^d h_H}} \right), \text{ p.s}$$

Preuve du Lemme 3.2

Comme S et I sont des ensembles compacts, S peut être recouvert par un nombre fini p_n de boules $B_k(x_k, a_n)$ centrées en $x_k \in \mathbb{R}^d$ et I est divisé en q_n sous-intervalles I_l de longueurs b_n , centrés en y_l . Pour tous $x \in S$, $y \in I$, il existe des entiers $k \in \{1, \dots, p_n\}$ et $l \in \{1, \dots, q_n\}$ tels que :

$$\max_{1 \leq k \leq p_n} \|x - x_k\| \leq n^{-\frac{1}{2}} h_K^{\frac{d}{2}+1} h_H^{\frac{1}{2}} =: a_n$$

et

$$\max_{1 \leq l \leq q_n} \|y - y_l\| \leq (nh_K^d)^{-\frac{1}{2}} h_H^{\frac{3}{2}} =: b_n,$$

Alors, puisque S et I sont bornés, il existe deux constantes M et M' telles que, $p_n a_n \leq M$, $q_n b_n \leq M'$ et $MM' \leq C$. Par conséquent, nous avons la décomposition suivante

$$\sup_{x \in S} \sup_{y \in I} \left| f_n(x, y) - \mathbb{E}(f_n(x, y)) \right| \leq \mathcal{C}_{1,n} + \mathcal{C}'_{1,n} + \mathcal{C}_{2,n} + \mathcal{C}'_{2,n} + \mathcal{C}_{3,n}, \quad (3.4)$$

sachant que

$$\begin{aligned} \mathcal{C}_{1,n} &:= \max_{1 \leq k \leq p_n} \sup_{x \in B_k} \sup_{y \in I} \left| f_n(x, y) - f_n(x_k, y) \right| \\ \mathcal{C}'_{1,n} &:= \max_{1 \leq k \leq p_n} \sup_{x \in B_k} \sup_{y \in I} \left| \mathbb{E}(f_n(x, y)) - \mathbb{E}(f_n(x_k, y)) \right| \\ \mathcal{C}_{2,n} &:= \max_{1 \leq k \leq p_n} \max_{1 \leq l \leq q_n} \sup_{y \in I_l} \left| f_n(x_k, y_l) - f_n(x_k, y) \right| \end{aligned}$$

$$\mathcal{C}'_{2,n} := \max_{1 \leq k \leq p_n} \max_{1 \leq l \leq q_n} \sup_{y \in I_l} \left| \mathbb{E} \left(f_n(x_k, y_l) \right) - \mathbb{E} \left(f_n(x_k, y) \right) \right|$$

et

$$\mathcal{C}_{3,n} := \max_{1 \leq k \leq p_n} \max_{1 \leq l \leq q_n} \left| f_n(x_k, y_l) - \mathbb{E} \left(f_n(x_k, y_l) \right) \right|.$$

Concernant tout d'abord $\mathcal{C}_{1,n}$ et $\mathcal{C}'_{1,n}$: puisque K_d est lipschitzien et $\|H\|_\infty \leq 1$, on a

$$\begin{aligned} \left| f_n(x, y) - f_n(x_k, y) \right| &\leq \frac{C \|x - x_k\|}{nh_K^{d+1} h_H} \sum_{i=1}^n H \left(\frac{y - Y_i}{h_H} \right) \\ &\leq \frac{C a_n}{h_K^{d+1} h_H} \\ &= O \left(\frac{1}{\sqrt{nh_K^d h_H}} \right). \end{aligned} \quad (3.5)$$

De la même manière, sous **K3** et le Lemme 3.3 nous obtenons les bornes pour $\mathcal{C}_{2,n}$ et $\mathcal{C}'_{2,n}$. En effet,

$$\begin{aligned} \left| f_n(x_k, y_l) - f_n(x_k, y) \right| &\leq \frac{C |y - y_l|}{h_H} \frac{1}{nh_K^d h_H} \sum_{i=1}^n K_d \left(\frac{x_k - X_i}{h_K} \right) \\ &\leq \frac{C b_n}{h_H^2} v_n(x_k) \\ &= O \left(\frac{1}{\sqrt{nh_K^d h_H}} \right). \end{aligned} \quad (3.6)$$

Nous utilisons ensuite une inégalité exponentielle (voir, Théorème 1.2) pour majorer le terme $\mathcal{C}_{3,n}$. Pour ce faire, nous définissons d'abord

$$\Delta_i(x, y) = K_d \left(\frac{x - X_i}{h_K} \right) H \left(\frac{y - Y_i}{h_H} \right) - \mathbb{E} \left(K_d \left(\frac{x - X_i}{h_K} \right) H \left(\frac{y - Y_i}{h_H} \right) \right) =: \Delta_i.$$

Il est clair que

$$f_n(x, y) - \mathbb{E} \left(f_n(x, y) \right) = \frac{1}{nh_K^d h_H} \sum_{i=1}^n \Delta_i.$$

Remarque 3.2. Notez que le processus Δ_i est ψ_ℓ -faiblement dépendant compte tenu de la remarque 7 de [Doukhan and Neumann \(2007\)](#), les hypothèses **K1** et **K3**, et le fait que $Z_i = (X_i, Y_i)$ est ψ_ℓ -faiblement dépendant ($\ell = 1$ ou $\ell = 2$).

Maintenant, on est en mesure d'appliquer le Théorème 1.2, pour cela on évalue d'abord σ_n^2 où $\sigma_n^2 = \text{Var} \left(\sum_{i=1}^n \Delta_i \right)$. Nous avons

$$\begin{aligned} \sigma_n^2 &= n \text{Var}(\Delta_1) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \text{Cov}(\Delta_i, \Delta_j) \\ &=: \mathcal{V} + \mathcal{COV}. \end{aligned}$$

D'une part, on a

$$\begin{aligned} \mathcal{V} &= n \left[\mathbb{E} \left(K_d^2 \left(\frac{x - X_1}{h_K} \right) H^2 \left(\frac{y - Y_1}{h_H} \right) \right) - \mathbb{E}^2 \left(K_d \left(\frac{x - X_1}{h_K} \right) H \left(\frac{y - Y_1}{h_H} \right) \right) \right] \\ &=: n(\mathcal{V}_1 - \mathcal{V}_2). \end{aligned}$$

Donc, sous les conditions **B(i-ii)**, **D3**, **K1**, **K3**, par un changement de variable et un développement de Taylor au voisinage de (x, y) , nous obtenons

$$\begin{aligned} \mathcal{V}_1 &= \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_d^2 \left(\frac{x - u}{h_K} \right) H^2 \left(\frac{y - v}{h_H} \right) f(u, v) dudv \\ &= h_K^d h_H \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_d^2(s) H^2(t) f(x - sh_K, y - th_H) dsdt \\ &= O(h_K^d h_H). \end{aligned}$$

De même, sous **B(i-ii)**, **D3**, **K2(i)** on obtient

$$\mathcal{V}_2 = O(h_K^{2d} h_H^2).$$

Par conséquent,

$$\mathcal{V} = O(nh_K^d h_H). \quad (3.7)$$

D'un autre côté, nous avons

$$\begin{aligned} \mathcal{COV} &= \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n Cov(\Delta_i, \Delta_j) \\ &= \sum_{r=1}^{n-1} (n-r) Cov(\Delta_1, \Delta_{r+1}). \end{aligned}$$

Soit $R = R(n) \leq n-1$, ; $n \rightarrow \infty$, d'où

$$\mathcal{COV} \leq n \sum_{r < R} Cov(\Delta_1, \Delta_{r+1}) + n \sum_{r \geq R} Cov(\Delta_1, \Delta_{r+1}). \quad (3.8)$$

Puis pour $i < j$, remarquez que

$$\begin{aligned} Cov(\Delta_i, \Delta_j) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}} \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_d \left(\frac{x - u}{h_K} \right) H \left(\frac{y - v}{h_H} \right) K_d \left(\frac{x - s}{h_K} \right) H \left(\frac{y - t}{h_H} \right) \\ &\quad \times f_{1,j-i+1}(u, v, s, t) dudv dsdt - \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_d \left(\frac{x - u}{h_K} \right) H \left(\frac{y - v}{h_H} \right) f(u, v) dudv \\ &\quad \times \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_d \left(\frac{x - s}{h_K} \right) H \left(\frac{y - t}{h_H} \right) f(s, t) dsdt. \end{aligned}$$

Alors, par un changement de variables, on obtient

$$\begin{aligned} Cov(\Delta_i, \Delta_j) &= h_K^{2d} h_H^2 \int_{\mathbb{R}^d} \int_{\mathbb{R}} \int_{\mathbb{R}^d} \int_{\mathbb{R}} \left(f_{1,j-i+1}(x - s_1 h_K, y - t_1 h_H, x - s_2 h_K, y - t_2 h_H) \right. \\ &\quad \left. - f(x - s_1 h_K, y - t_1 h_H) f(x - s_2 h_K, y - t_2 h_H) \right) \\ &\quad \times K_d(s_1) H(t_1) K_d(s_2) H(t_2) ds_1 dt_1 ds_2 dt_2. \end{aligned}$$

Ainsi, en utilisant les hypothèses **K1**, **K3** et **D4**, il en découle que

$$Cov(\Delta_i, \Delta_j) = O(h_K^{2d} h_H^2). \quad (3.9)$$

Le second terme du côté droit de (3.8) est borné en utilisant la dépendance faible du processus. Nous avons

$$\begin{aligned} Cov(\Delta_1, \Delta_{r+1}) &= Cov\left(K_d\left(\frac{x - X_1}{h_K}\right) H\left(\frac{y - Y_1}{h_H}\right), K_d\left(\frac{x - X_{r+1}}{h_K}\right) H\left(\frac{y - Y_{r+1}}{h_H}\right)\right) \\ &=: Cov(g(Z_1), g(Z_{r+1})), \end{aligned}$$

avec $\|g\|_\infty \leq \|K_d\|_\infty \|H\|_\infty \leq 1$ (sous les conditions **K1** et **K3**).

De ce fait, l'hypothèse **M1** nous amène à obtenir

$$\left|Cov(g(Z_1), g(Z_{r+1}))\right| \leq \begin{cases} Lip(g)^2 \epsilon_r & \text{pour } \psi_1\text{-faiblement dépendent,} \\ 8Lip(g) \epsilon_r & \text{pour } \psi_2\text{-faiblement dépendent.} \end{cases}$$

En outre, nous avons

$$Lip(g) \leq \left(\|K_d\|_\infty Lip(H) + \|H\|_\infty Lip(K_d)\right) (h_H^{-1} + h_K^{-1}).$$

Alors

$$\left|Cov(g(Z_1), g(Z_{r+1}))\right| \leq \begin{cases} C (h_H^{-1} + h_K^{-1})^2 \epsilon_r & \text{pour } \psi_1\text{-faiblement dépendent,} \\ C (h_H^{-1} + h_K^{-1}) \epsilon_r & \text{pour } \psi_2\text{-faiblement dépendent,} \end{cases}$$

et ensuite

$$\sum_{r \geq R} |Cov(\Delta_1, \Delta_{r+1})| \leq \begin{cases} C (h_H^{-1} + h_K^{-1})^2 U(R) & \text{pour } \psi_1\text{-faiblement dépendent,} \\ C (h_H^{-1} + h_K^{-1}) U(R) & \text{pour } \psi_2\text{-faiblement dépendent.} \end{cases} \quad (3.10)$$

En combinant (3.8), (3.9) et (3.10), on trouve que

$$\mathcal{COV} \leq \begin{cases} Cn \left(Rh_K^{2d} h_H^2 + (h_H^{-1} + h_K^{-1})^2 U(R)\right) & \text{pour } \psi_1\text{-faiblement dépendent,} \\ Cn \left(Rh_K^{2d} h_H^2 + (h_H^{-1} + h_K^{-1}) U(R)\right) & \text{pour } \psi_2\text{-faiblement dépendent.} \end{cases} \quad (3.11)$$

Choisissant maintenant

$$R = \begin{cases} h_K^{-v_1} h_H^{-v_2} & \text{avec } \frac{d+2}{4} < v_1 < d \text{ et } \frac{3}{4} < v_2 < 1 & \text{pour } \psi_1\text{-faiblement dépendent} \\ h_K^{-v_1} h_H^{-v_2} & \text{avec } \frac{d+1}{3} < v_1 < d \text{ et } \frac{2}{3} < v_2 < 1 & \text{pour } \psi_2\text{-faiblement dépendent} \end{cases} \quad (3.12)$$

et sous **M2**, (3.11) est égal à

$$\mathcal{COV} = o(nh_K^d h_H). \quad (3.13)$$

Enfin, en combinant les résultats (3.7) et (3.13), on obtient

$$\sigma_n^2 = O(nh_K^d h_H).$$

Désormais, nous sommes en mesure de déterminer la borne supérieure du terme $\mathcal{C}_{3,n}$. Pour cela, nous employons le Théorème 1.2 et nous écrivons, pour tout $\varepsilon_0 > 0$, ce qui suit

$$\begin{aligned} \mathbb{P} \left(\mathcal{C}_{3,n} \geq \varepsilon_0 \sqrt{\frac{\log n}{nh_K^d h_H}} \right) &= \mathbb{P} \left(\max_{1 \leq k \leq p_n} \max_{1 \leq l \leq q_n} \frac{1}{nh_K^d h_H} \left| \sum_{i=1}^n \Delta_i(x_k, y_l) \right| \geq \varepsilon_0 \sqrt{\frac{\log n}{nh_K^d h_H}} \right) \\ &\leq 2p_n q_n \mathbb{P} \left(\left| \sum_{i=1}^n \Delta_i(x_k, y_l) \right| \geq \varepsilon_0 \sqrt{nh_K^d h_H \log n} \right) \\ &\leq 2Cp_n q_n \log n \exp \left(- \frac{\varepsilon_0^2 nh_K^d h_H \log n}{Cnh_K^d h_H + \frac{n^{\frac{3}{4}} \log n \varepsilon_0 \sqrt{nh_K^d h_H \log n}}{A_n}} \right). \end{aligned}$$

Alors, par un calcul algébrique élémentaire, on arrive à

$$\begin{aligned} \exp \left(- \frac{\varepsilon_0^2 nh_K^d h_H \log n}{Cnh_K^d h_H + \frac{n^{\frac{3}{4}} \log n \varepsilon_0 \sqrt{nh_K^d h_H \log n}}{A_n}} \right) &\leq \exp \left(- \frac{\varepsilon_0^2 \log n}{C + \frac{\varepsilon_0 (\log n)^{\frac{3}{2}}}{n^{\frac{3}{4}} (h_K^d h_H)^{\frac{3}{2}}}} \right) \\ &\leq \exp \left(- \frac{\varepsilon_0^2 \log n}{n^{\frac{3}{4}} \left(\frac{C}{n^{\frac{3}{4}}} + \frac{\varepsilon_0 (\log n)^{\frac{3}{2}}}{(nh_K^d h_H)^{\frac{3}{2}}} \right)} \right). \end{aligned}$$

En conséquence, sous l'hypothèse **B(iii)**

$$\begin{aligned} \mathbb{P} \left(\mathcal{C}_{3,n} \geq \varepsilon_0 \sqrt{\frac{\log n}{nh_K^d h_H}} \right) &\leq Ca_n^{-1} b_n^{-1} n^{-C\varepsilon_0^2} \log n \\ &\leq (n^2 h_K h_H^2)^{-1} n^{3-C\varepsilon_0^2} \log n \\ &= (nh_K h_H)^{-1} (nh_H)^{-1} O(n^{3-C\varepsilon_0^2}) \log n. \end{aligned} \quad (3.14)$$

Ainsi sous **B(i-iii)** et pour un choix convenable de ε_0 , le terme dans (3.14) est bien le terme général d'une série de Bertrand convergente. Par conséquent, grâce aux expressions (3.4), (3.5) et (3.6), il en résulte que

$$\sum_{n \geq 1} \mathbb{P} \left(\sup_{x \in S} \sup_{y \in I} \left| f_n(x, y) - \mathbb{E}(f_n(x, y)) \right| \geq \varepsilon_0 \sqrt{\frac{\log n}{nh_K^d h_H}} \right) \leq \sum_{n \geq 1} \mathbb{P} \left(\mathcal{C}_{3,n} \geq \varepsilon_0 \sqrt{\frac{\log n}{nh_K^d h_H}} \right) < \infty.$$

Enfin, en utilisant le lemme de Borel-Cantelli, nous obtenons

$$\sup_{x \in S} \sup_{y \in I} \left| f_n(x, y) - \mathbb{E}(f_n(x, y)) \right| = O \left(\sqrt{\frac{\log n}{nh_K^d h_H}} \right), \text{ p.s.},$$

ce qui achève la preuve. \square

Lemme 3.3. *Sous les hypothèses **B(i-iii)**, **M1-M2**, **D5-D6**, **K1** et **K2(i)**, pour n suffisamment grand on a*

$$\sup_{x \in S} \left| v_n(x) - v(x) \right| = O \left\{ \max \left(\sqrt{\frac{\log n}{nh_K^d}}, h_K^2 \right) \right\}, \text{ p.s.}$$

Preuve du Lemme 3.3 On a

$$\sup_{x \in S} \left| v_n(x) - v(x) \right| \leq \sup_{x \in S} \left| v_n(x) - \mathbb{E}(v_n(x)) \right| + \sup_{x \in S} \left| \mathbb{E}(v_n(x)) - v(x) \right| =: \mathcal{J}_1 + \mathcal{J}_2.$$

Afin d'obtenir une borne supérieure pour le terme de biais \mathcal{J}_2 , on constate facilement que sous **B(i)**, **D5**, **K1** et **K2(i)** on a,

$$\sup_{x \in S} \left| \mathbb{E}(v_n(x)) - v(x) \right| = O(h_K^2).$$

Pour le terme \mathcal{J}_1 , nous effectuons les mêmes étapes que dans la preuve du Lemme 3.2. Pour cela, pour recouvrir le compact S nous prenons $a_n = n^{-\frac{1}{2}} h_K^{\frac{d}{2}+1}$ et nous remplaçons $\Delta_i(x, y)$ par

$$\Gamma_i(x) = K_d \left(\frac{x - X_i}{h_K} \right) - \mathbb{E} \left(K_d \left(\frac{x - X_i}{h_K} \right) \right).$$

Donc en vertu de **B(iii)**, **D6** et **M1-M2**, nous obtenons

$$\mathcal{J}_1 = O \left(\sqrt{\frac{\log n}{n h_K^d}} \right).$$

La preuve du Lemme 3.3 est donc terminée. □

Preuve du théorème 3.1 Considérons la décomposition classique suivante

$$\begin{aligned} \sup_{x \in S} \sup_{y \in I} \left| f_n(y|x) - f(y|x) \right| &\leq \frac{1}{\inf_{x \in S} v_n(x)} \left(\sup_{x \in S} \sup_{y \in I} \left| f_n(x, y) - f(x, y) \right| \right. \\ &\quad \left. + \kappa \sup_{x \in S} \left| v_n(x) - v(x) \right| \right), \end{aligned}$$

où κ est la borne supérieure de $f(y|x)$. Alors la preuve du théorème 3.1 découle en appliquant le Lemme 3.1, le Lemme 3.2 et le Lemme 3.3. □

Preuve du corollaire 3.1 Par définition du mode conditionnel, on peut facilement vérifier que

$$\sup_{x \in S} \left| f(\Theta_n(x)|x) - f(\Theta(x)|x) \right| \leq 2 \sup_{x \in S} \sup_{y \in I} \left| f_n(y|x) - f(y|x) \right|. \quad (3.15)$$

Donc, en utilisant le développement de Taylor de $f(\cdot|x)$ au voisinage de $\Theta(x)$ et sous **D1** nous avons

$$f(\Theta_n(x)|x) - f(\Theta(x)|x) = \frac{1}{2} (\Theta_n(x) - \Theta(x))^2 f^{(2)}(\Theta_n^*(x)|x),$$

où $\Theta_n^*(x)$ est compris entre $\Theta_n(x)$ et $\Theta(x)$. Dans ce cas, par (3.15), on trouve

$$(\Theta_n(x) - \Theta(x))^2 \left| f^{(2)}(\Theta_n^*(x)|x) \right| \leq 4 \sup_{x \in S} \sup_{y \in I} \left| f_n(y|x) - f(y|x) \right|.$$

Ainsi, le théorème 3.1, **D1** et **D2** achèvent la preuve. □

Lemme 3.4. Sous les hypothèses **B(i-iv)**, **M1-M2**, **D3-D4**, **K1** et **K3-K4** nous avons

$$\lim n h_K^d h_H^3 \text{Var} (f_n^{0,1}(x, y)) = f(x, y) \int_{\mathbb{R}^d} \int_{\mathbb{R}} (K_d(u) H^{(1)}(v))^2 dudv.$$

Preuve du Lemme 3.4 Notons que

$$\Lambda_i(x, y, X_i, Y_i) = K_d \left(\frac{x - X_i}{h_K} \right) H^{(1)} \left(\frac{y - Y_i}{h_H} \right) - \mathbb{E} \left(K_d \left(\frac{x - X_i}{h_K} \right) H^{(1)} \left(\frac{y - Y_i}{h_H} \right) \right) =: \Lambda_i, \quad (3.16)$$

il est évident que

$$\frac{1}{nh_K^d h_H^2} \sum_{i=1}^n \Lambda_i = f_n^{(0,1)}(x, y) - \mathbb{E}(f_n^{(0,1)}(x, y)).$$

Alors

$$\text{Var} (f_n^{(0,1)}(x, y)) = \frac{1}{(nh_K^d h_H^2)^2} \text{Var} \left(\sum_{i=1}^n \Lambda_i \right)$$

puis

$$\begin{aligned} nh_K^d h_H^3 \text{Var} (f_n^{(0,1)}(x, y)) &= \frac{1}{h_K^d h_H} \text{Var} (\Lambda_1) + \frac{1}{nh_K^d h_H} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \text{Cov} (\Lambda_i, \Lambda_j) \\ &=: \mathcal{I}_{1,n} + \mathcal{I}_{2,n}. \end{aligned}$$

Nous calculons d'abord $\lim \mathcal{I}_{1,n}$, sous les hypothèses **B(i-ii)**, **D3**, **K1** et **K3-K4**. On a

$$\begin{aligned} \lim \mathcal{I}_{1,n} &= \lim \frac{1}{h_K^d h_H} \left[\int_{\mathbb{R}^d} \int_{\mathbb{R}} \left(K_d \left(\frac{x-s}{h_K} \right) H^{(1)} \left(\frac{y-t}{h_H} \right) \right)^2 f(s, t) ds dt \right. \\ &\quad \left. - \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}} K_d \left(\frac{x-s}{h_K} \right) H^{(1)} \left(\frac{y-t}{h_H} \right) f(s, t) ds dt \right)^2 \right] \\ &= \lim \int_{\mathbb{R}^d} \int_{\mathbb{R}} (K_d(u) H^{(1)}(v))^2 f(x - uh_K, y - vh_H) dudv \\ &\quad - h_K^d h_H \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}} K_d(u) H^{(1)}(v) f(x - uh_K, y - vh_H) dudv \right)^2 \\ &= f(x, y) \int_{\mathbb{R}^d} \int_{\mathbb{R}} (K_d(u) H^{(1)}(v))^2 dudv. \end{aligned}$$

Il reste à montrer que $\mathcal{I}_{2,n}$ tend vers 0 lorsque $n \rightarrow \infty$. Nous suivons la même procédure que celle utilisée pour évaluer le terme \mathcal{COV} (voir (3.13)).

$$\begin{aligned} \mathcal{I}_{2,n} &= \frac{1}{nh_K^d h_H} \sum_{r=1}^{n-1} (n-r) \text{Cov}(\Lambda_1, \Lambda_{r+1}) \\ &\leq \frac{n}{nh_K^d h_H} \sum_{r < R} \text{Cov}(\Lambda_1, \Lambda_{r+1}) + \frac{n}{nh_K^d h_H} \sum_{r \geq R} \text{Cov}(\Lambda_1, \Lambda_{r+1}) \\ &=: \mathcal{I}_{21,n} + \mathcal{I}_{22,n}. \end{aligned}$$

D'une part, sous **D4**, **K1** et **K3-K4** pour tous les $i < j$

$$\begin{aligned} \mathcal{I}_{21,n} &\leq \frac{R}{h_K^d h_H} \left[\int_{\mathbb{R}^d} \int_{\mathbb{R}} \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_d \left(\frac{x-u}{h_K} \right) H^{(1)} \left(\frac{y-v}{h_H} \right) K_d \left(\frac{x-s}{h_K} \right) H^{(1)} \left(\frac{y-t}{h_H} \right) \right. \\ &\quad \times f_{1,j-i+1}(u, v, s, t) dudvdsdt - \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_d \left(\frac{x-u}{h_K} \right) H^{(1)} \left(\frac{y-v}{h_H} \right) f(u, v) dudv \\ &\quad \left. \times \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_d \left(\frac{x-s}{h_K} \right) H^{(1)} \left(\frac{y-t}{h_H} \right) f(s, t) dsdt \right] \\ &= O(Rh_K^d h_H). \end{aligned} \tag{3.17}$$

D'autre part, on a

$$\mathcal{I}_{22,n} = n \sum_{r \geq R} Cov \left(\frac{1}{\sqrt{nh_K^d h_H}} \Lambda_1, \frac{1}{\sqrt{nh_K^d h_H}} \Lambda_{r+1} \right).$$

En outre,

$$\begin{aligned} Cov \left(\frac{1}{\sqrt{nh_K^d h_H}} \Lambda_1, \frac{1}{\sqrt{nh_K^d h_H}} \Lambda_{r+1} \right) &= Cov \left(\frac{1}{\sqrt{nh_K^d h_H}} K_d \left(\frac{x-X_1}{h_K} \right) H^{(1)} \left(\frac{y-Y_1}{h_H} \right), \right. \\ &\quad \left. \frac{1}{\sqrt{nh_K^d h_H}} K_d \left(\frac{x-X_{r+1}}{h_K} \right) H^{(1)} \left(\frac{y-Y_{r+1}}{h_H} \right) \right) \\ &=: Cov(g_1(Z_1), g_1(Z_{r+1})), \end{aligned}$$

avec $\|g_1\|_\infty \leq \frac{1}{\sqrt{nh_K^d h_H}} \|K_d\|_\infty \|H^{(1)}\|_\infty < 1$.

Il en résulte que, sous les hypothèses **M1**, **K1** et **K3**, on a,

$$\left| Cov(g_1(Z_1), g_1(Z_{r+1})) \right| \leq \begin{cases} Lip(g_1)^2 \epsilon_r & \text{pour } \psi_1\text{-faiblement dépendent,} \\ 8Lip(g_1) \epsilon_r & \text{pour } \psi_2\text{-faiblement dépendent.} \end{cases}$$

Par ailleurs, nous avons

$$Lip(g_1) \leq \frac{C}{\sqrt{nh_K^d h_H}} (h_H^{-1} + h_K^{-1}) \text{ où } C = \|K_d\|_\infty Lip(H^{(1)}) + \|H^{(1)}\|_\infty Lip(K_d).$$

Par conséquent,

$$\left| Cov(g_1(Z_1), g_1(Z_{r+1})) \right| \leq \begin{cases} \frac{C}{nh_K^d h_H} (h_H^{-1} + h_K^{-1})^2 \epsilon_r & \text{pour } \psi_1\text{-faiblement dépendent,} \\ \frac{C}{\sqrt{nh_K^d h_H}} (h_H^{-1} + h_K^{-1}) \epsilon_r & \text{pour } \psi_2\text{-faiblement dépendent,} \end{cases}$$

donc

$$\mathcal{I}_{22,n} \leq \begin{cases} \frac{C}{h_K^d h_H} (h_H^{-1} + h_K^{-1})^2 U(R) & \text{pour } \psi_1\text{-faiblement dépendent,} \\ \frac{\sqrt{n}C}{\sqrt{h_K^d h_H}} (h_H^{-1} + h_K^{-1}) U(R) & \text{pour } \psi_2\text{-faiblement dépendent,} \end{cases} \tag{3.18}$$

en combinant (3.17) et (3.18), on obtient

$$\mathcal{I}_{2,n} \leq \begin{cases} CRh_K^d h_H + \frac{C}{h_K^d h_H} (h_H^{-1} + h_K^{-1})^2 U(R) & \text{pour } \psi_1\text{-faiblement dépendent,} \\ CRh_K^d h_H + \frac{\sqrt{n}C}{\sqrt{h_K^d h_H}} (h_H^{-1} + h_K^{-1}) U(R) & \text{pour } \psi_2\text{-faiblement dépendent,} \end{cases} \tag{3.19}$$

où R est le même que dans (3.12). Donc, sous **M2**, **B(iv)** : le terme (3.19) devient un $o(1)$. \square

Lemme 3.5. *Sous les hypothèses $\mathbf{B(i-iv)}$, $\mathbf{D3-D4}$, $\mathbf{M1-M2}$, $\mathbf{K1}$ et $\mathbf{K3-K4}$ nous avons*

$$\sqrt{nh_K^d h_H^3} \left(f_n^{(0,1)}(x, y) - \mathbb{E} \left(f_n^{(0,1)}(x, y) \right) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, V(x, y) \right),$$

avec $V(x, y) = f(x, y) \int_{\mathbb{R}^d} \int_{\mathbb{R}} (K_d(u) H^{(1)}(v))^2 dudv < \infty$.

Preuve du Lemme 3.5 La preuve est principalement basée sur le Théorème 1 de [Bardet et al. \(2008\)](#) (voir Théorème 1.3).

Selon (3.16)

$$\begin{aligned} \sqrt{nh_K^d h_H^3} \left(f_n^{(0,1)}(x, y) - \mathbb{E} \left(f_n^{(0,1)}(x, y) \right) \right) &= \sqrt{nh_K^d h_H^3} \frac{1}{nh_K^d h_H^2} \sum_{k=1}^n \Lambda_k \\ &= \frac{1}{\sqrt{nh_K^d h_H}} \sum_{k=1}^n \Lambda_k \\ &= \sum_{k=1}^n U(Z_k) =: \sum_{k=1}^n U_k. \end{aligned}$$

Alors, pour tout $0 < \delta \leq 1$, soit

$$A_n = \sum_{k=1}^n \mathbb{E} \left(|U_k|^{2+\delta} \right).$$

Nous avons donc

$$\begin{aligned} A_n &= n (nh_K^d h_H)^{-\frac{(2+\delta)}{2}} \mathbb{E} \left(|\Lambda_k|^{2+\delta} \right) \\ &= (nh_K^d h_H)^{-\frac{\delta}{2}} (h_K^d h_H)^{-1} \mathbb{E} \left(|\Lambda_k|^{2+\delta} \right) \\ &= (nh_K^d h_H)^{-\frac{\delta}{2}} (h_K^d h_H)^{-1} \left[\mathbb{E} \left(|\Lambda_k|^{2+\delta} \right)^{\frac{1}{2+\delta}} \right]^{2+\delta}. \end{aligned}$$

Par l'inégalité de Minkowski, on obtient

$$\begin{aligned} \mathbb{E} \left(|\Lambda_k|^{2+\delta} \right)^{\frac{1}{2+\delta}} &\leq \left[\mathbb{E} \left(\left| K_d \left(\frac{x - X_k}{h_K} \right) H^{(1)} \left(\frac{y - Y_k}{h_H} \right) \right|^{2+\delta} \right) \right]^{\frac{1}{2+\delta}} \\ &\quad + \left[\mathbb{E} \left(\left| \mathbb{E} \left[K_d \left(\frac{x - X_k}{h_K} \right) H^{(1)} \left(\frac{y - Y_k}{h_H} \right) \right] \right|^{2+\delta} \right) \right]^{\frac{1}{2+\delta}} \\ &\leq 2 \left[\mathbb{E} \left(\left| K_d \left(\frac{x - X_k}{h_K} \right) H^{(1)} \left(\frac{y - Y_k}{h_H} \right) \right|^{2+\delta} \right) \right]^{\frac{1}{2+\delta}} \end{aligned}$$

et alors

$$\begin{aligned} A_n &\leq 2^{2+\delta} (nh_K^d h_H)^{-\frac{\delta}{2}} (h_K^d h_H)^{-1} \mathbb{E} \left(\left| K_d \left(\frac{x - X_k}{h_K} \right) H^{(1)} \left(\frac{y - Y_k}{h_H} \right) \right|^{2+\delta} \right) \\ &\leq 2^{2+\delta} (nh_K^d h_H)^{-\frac{\delta}{2}} \|f\|_\infty \int_{\mathbb{R}^d} \int_{\mathbb{R}} |K_d(s) H^{(1)}(t)|^{2+\delta} ds dt, \end{aligned}$$

ce qui, sous les hypothèses **B(iii)**, **D3**, **K1** et **K3-K4** donne $A_n = o(1)$.

Posons maintenant

$$T(n) = \sum_{k=2}^n \left| \text{Cov} \left(\exp \left(it(U_1 + \dots + U_{k-1}) \right), \exp(itU_k) \right) \right|.$$

Pour déterminer la borne supérieure de $T(n)$, on note

$$\exp \left(it(U_1 + \dots + U_{k-1}) \right) = \sum_{j=1}^{k-1} \left(\exp(itS_j) - \exp(itS_{j-1}) \right) + 1 \text{ avec } S_j = U_1 + \dots + U_j \text{ et } S_0 = 0.$$

Ainsi,

$$\text{Cov} \left(\exp \left(it(U_1 + \dots + U_{k-1}) \right), \exp(itU_k) \right) = \sum_{j=1}^{k-1} \text{Cov} \left(\exp(itS_j) - \exp(itS_{j-1}), \exp(itU_k) \right). \quad (3.20)$$

Soit U_k^* une variable aléatoire indépendante de (U_1, \dots, U_j) et identiquement distribuée comme U_k .

On écrit

$$\left| \text{Cov} \left(\exp(itS_j) - \exp(itS_{j-1}), \exp(itU_k) \right) \right| = \left| \text{Cov} \left(\exp(itS_j) - \exp(itS_{j-1}), \exp(itU_k) - \exp(itU_k^*) \right) \right|$$

et en utilisant $|\exp(ia) - \exp(ib)| \leq |b - a|$, on obtient

$$\begin{aligned} \left| \text{Cov} \left(\exp(itS_j) - \exp(itS_{j-1}), \exp(itU_k) \right) \right| &= \mathbb{E} \left(|tU_j| (|U_k| + |U_k^*|) \right) \\ &= |t| \mathbb{E} \left(|U_j U_k| + |U_j U_k^*| \right). \end{aligned}$$

De plus, pour tout $i \geq 1$ et $j \geq 1$ on a

$$\begin{aligned} \mathbb{E} \left(|U_i U_j| \right) &\leq \frac{1}{nh_k^d h_H} \mathbb{E} \left(\left| K_d \left(\frac{x - X_i}{h_K} \right) H^{(1)} \left(\frac{y - Y_i}{h_H} \right) K_d \left(\frac{x - X_j}{h_K} \right) H^{(1)} \left(\frac{y - Y_j}{h_H} \right) \right| \right) \\ &\quad + \frac{C}{nh_k^d h_H} \left[\mathbb{E} \left(\left| K_d \left(\frac{x - X_i}{h_K} \right) H^{(1)} \left(\frac{y - Y_i}{h_H} \right) \right| \right) \right]^2. \end{aligned}$$

Ensuite, on raisonne comme dans (3.17) et, sous **D3-D4**, **K1** et **K3** on obtient

$$\mathbb{E} \left(|U_i U_j| \right) \leq \frac{1}{nh_k^d h_H} C h_k^{2d} h_H^2,$$

ce qui implique

$$\left| \text{Cov} \left(\exp(itS_j) - \exp(itS_{j-1}), \exp(itU_k) \right) \right| \leq \frac{C h_k^d h_H}{n}. \quad (3.21)$$

D'autre part

$$\text{Cov} \left(\exp(itS_j) - \exp(itS_{j-1}), \exp(itU_k) \right) =: \text{Cov} \left(g_1(Z_1, \dots, Z_j), g_2(Z_k) \right),$$

avec $\|g_1\|_\infty \leq |t| \|U\|_\infty \leq \frac{2|t|}{\sqrt{nh_K^d h_H}} \|K_d\|_\infty \|H^{(1)}\|_\infty < 1$ et $\|g_2\|_\infty = 1$ sous **B(iii)**, **K1** et **K3**.

Alors, comme $Lip(U) \leq C \frac{(h_H^{-1} + h_K^{-1})}{\sqrt{nh_K^d h_H}}$ et $Lip(g_2) = C \frac{(h_H^{-1} + h_K^{-1})}{\sqrt{nh_K^d h_H}}$, sous **M1**

$$\left| Cov\left(\exp(itS_j) - \exp(itS_{j-1}), \exp(itU_k)\right) \right| \leq \begin{cases} C \frac{(h_H^{-1} + h_K^{-1})^2}{nh_K^d h_H} \epsilon_r & \text{pour } \psi_1\text{-faiblement dépendent,} \\ C \frac{(h_H^{-1} + h_K^{-1})}{\sqrt{nh_K^d h_H}} \epsilon_r & \text{pour } \psi_2\text{-faiblement dépendent,} \end{cases} \quad (3.22)$$

où $r = (k - j) \in \{2, \dots, k\}$.

Tout d'abord, concernant le cas ψ_1 -faiblement dépendant, nous avons par (3.20), (3.21) et (3.22)

$$\begin{aligned} T(n) &\leq C \sum_{k=2}^n \sum_{r=2}^k \left(\frac{h_K^d h_H}{n} \wedge \frac{(h_H^{-1} + h_K^{-1})^2}{nh_K^d h_H} \epsilon_r \right) \\ &\leq Cn \sum_{r=2}^n \left(\frac{h_K^d h_H}{n} \wedge \frac{(h_H^{-1} + h_K^{-1})^2}{nh_K^d h_H} \epsilon_r \right) \\ &\leq \sum_{r=2}^{\infty} \left(h_K^d h_H \wedge \frac{(h_H^{-1} + h_K^{-1})^2}{h_K^d h_H} \epsilon_r \right) \\ &\leq \sum_{r < R} \left(h_K^d h_H \wedge \frac{(h_H^{-1} + h_K^{-1})^2}{h_K^d h_H} \epsilon_r \right) + \sum_{r \geq R} \left(h_K^d h_H \wedge \frac{(h_H^{-1} + h_K^{-1})^2}{h_K^d h_H} \epsilon_r \right) \\ &\leq \sum_{r < R} h_K^d h_H + \sum_{r \geq R} \frac{(h_H^{-1} + h_K^{-1})^2}{h_K^d h_H} \epsilon_r \\ &\leq Rh_K^d h_H + \frac{(h_H^{-1} + h_K^{-1})^2}{h_K^d h_H} U(R). \end{aligned}$$

Ensuite, en choisissant R comme dans (3.12) et sous **M2**, nous obtenons $T(n) = o(1)$.

En second lieu, pour le cas ψ_2 -faiblement dépendant, en procédant de la même manière que ci-dessus, nous arrivons à

$$T(n) \leq Rh_K^d h_H + \sqrt{n} \frac{(h_H^{-1} + h_K^{-1})}{\sqrt{h_K^d h_H}} U(R).$$

Donc en choisissant R comme dans (3.12) et sous **M2**, **B(iv)** nous obtenons $T(n) = o(1)$.

Ensuite, en utilisant le Lemme 3.4, toutes les conditions du Théorème de Bardet et al. (2008) (Théorème 1.3) sont satisfaites, ce qui aboutit au résultat. \square

Lemme 3.6. *Sous les hypothèses **B(v)**, **D3** et **K1-K3**, nous avons*

$$\sqrt{nh_K^d h_H^3} \mathbb{E}(f^{(0,1)}(x, \Theta(x))) = o(1).$$

Preuve du Lemme 3.6 On a

$$\mathbb{E}(f^{(0,1)}(x, \Theta(x))) = \frac{1}{h_K^d h_H^2} \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_d \left(\frac{x-u}{h_K} \right) H^{(1)} \left(\frac{\Theta(x)-v}{h_H} \right) f(u, v) dudv.$$

En intégrant par parties par rapport à v , on a

$$\begin{aligned} \mathbb{E}(f^{(0,1)}(x, \Theta(x))) &= \frac{1}{h_K^d h_H} \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_d \left(\frac{x-u}{h_K} \right) H \left(\frac{\Theta(x)-v}{h_H} \right) f^{(0,1)}(u, v) dudv \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}} K_d(s) H(t) f^{(0,1)}(x - sh_K, \Theta(x) - th_H) dsdt. \end{aligned}$$

Le développement de Taylor de $f^{(0,1)}$ au voisinage de $(x, \Theta(x))$ à l'ordre 2 et sous **D3**, **K1-K2** donne

$$\mathbb{E} \left(f^{(0,1)}(x, \Theta(x)) \right) = O(h_K^2 + h_H^2).$$

Par **B(v)**, on arrive au résultat. \square

Lemme 3.7. *Sous les hypothèses **B(i-iii)**, **M1-M2**, **D3** et **K1-K3**, pour n suffisamment grand nous avons*

$$\sup_{x \in S} \left| f_n^{(0,2)}(x, \Theta_n^*(x)) - f^{(0,2)}(x, \Theta(x)) \right| = o(1), \text{ en probabilité.}$$

Preuve du Lemme 3.7 Nous mettons

$$\begin{aligned} \sup_{x \in S} \sup_{y \in I} \left| f_n^{(0,2)}(x, y) - f^{(0,2)}(x, y) \right| &= \sup_{x \in S} \sup_{y \in I} \left| f_n^{(0,2)}(x, y) - \mathbb{E} \left(f_n^{(0,2)}(x, y) \right) \right| \\ &\quad + \sup_{x \in S} \sup_{y \in I} \left| \mathbb{E} \left(f_n^{(0,2)}(x, y) \right) - f^{(0,2)}(x, y) \right| \\ &=: J_1 + J_2. \end{aligned}$$

Donc, en intégrant deux fois par parties et en effectuant le même procédé que pour la preuve du Lemme 3.6, on obtient

$$J_2 \leq \int_{\mathbb{R}^d} \int_{\mathbb{R}} \left| K_d(s) H(t) \left(f^{(0,2)}(x - sh_K, y - th_H) - f^{(0,2)}(x, y) \right) \right| ds dt.$$

Ensuite, le développement de Taylor de $f^{(0,2)}(x - sh_K, y - th_H)$ au voisinage de (x, y) donne

$$\begin{aligned} J_2 &\leq h_K \int_{\mathbb{R}^d} \int_{\mathbb{R}} \left| K_d(s) H(t) \left(s_1 \frac{\partial f^{(0,2)}}{\partial x_1}(x^*, y^*) + \dots + s_d \frac{\partial f^{(0,2)}}{\partial x_d}(x^*, y^*) \right) \right| ds dt \\ &\quad + h_H \int_{\mathbb{R}^d} \int_{\mathbb{R}} \left| K_d(s) H(t) t \frac{\partial f^{(0,2)}}{\partial y}(x^*, y^*) \right| ds dt, \end{aligned}$$

où (x^*, y^*) est compris entre (x, y) et $(x - sh_K, y - th_H)$. Donc, sous **D3**, **K1** et **K2(ii)**, il s'ensuit que

$$J_2 = O(h_K + h_H). \quad (3.23)$$

D'autre part, pour déterminer la borne du terme J_1 , nous procédons comme pour la preuve du Lemme (3.2). En effet, on recouvre le compact S par un nombre fini p_n de boules B_k centrées en x_k dans \mathbb{R}^d telles que

$$\max_{1 \leq k \leq p_n} \|x - x_k\| \leq n^{-\frac{1}{2}} h_k^{\frac{d}{2}+1} h_H^{\frac{5}{2}} =: a_n$$

et I est divisé en q_n intervalles I_l centrés en y_l vérifiant

$$\max_{1 \leq l \leq q_n} \|y - y_l\| \leq (nh_K^d)^{-\frac{1}{2}} h_H^{\frac{7}{2}} =: b_n.$$

Comme S et I sont bornés, il existe deux constantes M et M' telles que, $p_n a_n \leq M$, $q_n b_n \leq M'$ et $MM' \leq C$. Par conséquent, nous avons

$$J_1 \leq \mathcal{G}_{1,n} + \mathcal{G}'_{1,n} + \mathcal{G}_{2,n} + \mathcal{G}'_{2,n} + \mathcal{G}_{3,n}, \quad (3.24)$$

où

$$\begin{aligned}\mathcal{G}_{1,n} &:= \max_{1 \leq k \leq p_n} \sup_{x \in B_k} \sup_{y \in I} |f_n^{(0,2)}(x, y) - f_n^{(0,2)}(x_k, y)| \\ \mathcal{G}'_{1,n} &:= \max_{1 \leq k \leq p_n} \sup_{x \in B_k} \sup_{y \in I} |\mathbb{E}(f_n^{(0,2)}(x, y)) - \mathbb{E}(f_n^{(0,2)}(x_k, y))| \\ \mathcal{G}_{2,n} &:= \max_{1 \leq k \leq p_n} \max_{1 \leq l \leq q_n} \sup_{y \in I_l} |f_n^{(0,2)}(x_k, y_l) - f_n^{(0,2)}(x_k, y)| \\ \mathcal{G}'_{2,n} &:= \max_{1 \leq k \leq p_n} \max_{1 \leq l \leq q_n} \sup_{y \in I_l} |\mathbb{E}(f_n^{(0,2)}(x_k, y_l)) - \mathbb{E}(f_n^{(0,2)}(x_k, y))|\end{aligned}$$

et

$$\mathcal{G}_{3,n} := \max_{1 \leq k \leq p_n} \max_{1 \leq l \leq q_n} |f_n^{(0,2)}(x_k, y_l) - \mathbb{E}(f_n^{(0,2)}(x_k, y_l))|.$$

Donc, sous **K1**, **K3** et par un raisonnement similaire à celui de la preuve du Lemme 3.2 (voir (3.5), (3.6)) nous obtenons

$$\mathcal{G}_{1,n} = O\left(\frac{1}{\sqrt{nh_K^d h_H}}\right) \quad \text{et} \quad \mathcal{G}'_{1,n} = O\left(\frac{1}{\sqrt{nh_K^d h_H}}\right). \quad (3.25)$$

De la même façon

$$\mathcal{G}_{2,n} = O\left(\frac{1}{\sqrt{nh_K^d h_H}}\right) \quad \text{et} \quad \mathcal{G}'_{2,n} = O\left(\frac{1}{\sqrt{nh_K^d h_H}}\right). \quad (3.26)$$

Maintenant, afin de limiter $\mathcal{G}_{3,n}$ nous devons d'abord poser

$$\Psi(x, y) = K_d\left(\frac{x - X_i}{h_K}\right) H^{(2)}\left(\frac{y - Y_i}{h_H}\right) - \mathbb{E}\left(K_d\left(\frac{x - X_i}{h_K}\right) H^{(2)}\left(\frac{y - Y_i}{h_H}\right)\right) =: \Psi_i.$$

Appliquons ensuite le Théorème 1.2 sur les variables Ψ_i . Pour cela, nous calculons d'abord $\sigma_n^2 = \text{Var}\left(\sum_{i=1}^n \Psi_i\right)$. Nous avons

$$\begin{aligned}\sigma_n^2 &= n \text{Var}\left(K_d\left(\frac{x - X_1}{h_K}\right) H\left(\frac{y - Y_1}{h_H}\right)\right) \\ &\quad + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \text{Cov}\left(K_d\left(\frac{x - X_i}{h_K}\right) H\left(\frac{y - Y_i}{h_H}\right), K_d\left(\frac{x - X_j}{h_K}\right) H\left(\frac{y - Y_j}{h_H}\right)\right) \\ &=: \mathcal{S}_1 + \mathcal{S}_2.\end{aligned}$$

En utilisant la même démarche que pour calculer \mathcal{V} (voir (3.7)) et \mathcal{COV} (voir (3.13)), on arrive à $\sigma_n^2 = O(nh_K^d h_H)$.

Ensuite, en appliquant l'inégalité exponentielle de Hwang and Shin (2014) et en procédant de la même manière que pour le résultat de ((3.14), il en déduit que

$$\begin{aligned}\mathbb{P}\left(\mathcal{G}_{3,n} \geq \varepsilon_0 \sqrt{\frac{\log n}{nh_K^d h_H^5}}\right) &= \mathbb{P}\left\{\max_{1 \leq k \leq p_n} \max_{1 \leq l \leq q_n} \frac{1}{nh_K^d h_H^3} \left|\sum_{i=1}^n \Psi_i(x_k, y_l)\right| \geq \varepsilon_0 \sqrt{\frac{\log n}{nh_K^d h_H^5}}\right\} \\ &= (nh_K h_H)^{-1} (nh_H)^{-\frac{9}{2}} O(n^{\frac{13}{2} - C\varepsilon_0^2}) \log n.\end{aligned} \quad (3.27)$$

En conséquence, par les suppositions **B(i-iii)** et pour un choix approprié de ε_0^2 , le terme dans (3.27) est la terme général d'une série de Bertrand convergente. Ainsi, le Lemme de Borel Cantelli nous donne les résultats suivants

$$\mathcal{G}_{3,n} = O\left(\sqrt{\frac{\log n}{nh_K^d h_H^5}}\right),$$

ce qui donne avec (3.24)-(3.26)

$$J_1 = O\left(\sqrt{\frac{\log n}{nh_K^d h_H^5}}\right). \quad (3.28)$$

Ainsi, grâce aux expressions (3.23), (3.28) et sous **B(v)**, il s'ensuit que

$$\sup_{x \in \mathcal{S}} \sup_{y \in I} \left| f_n^{(0,2)}(x, y) - f^{(0,2)}(x, y) \right| = o(1), \text{ p.s} \quad (3.29)$$

Enfin, la continuité de $f^{(0,2)}(\cdot, \cdot)$, le Corollaire 3.1 et (3.29) impliquent la convergence en probabilité de $f_n^{(0,2)}(x, \Theta_n^*(x))$ vers $f^{(0,2)}(x, \Theta(x))$. \square

Preuve du Théorème 3.2 D'après (3.2) nous avons

$$\begin{aligned} \sqrt{nh_K^d h_H^3} (\Theta_n(x) - \Theta(x)) &= - \frac{\sqrt{nh_K^d h_H^3} \left(f_n^{(0,1)}(x, \Theta(x)) - \mathbb{E} \left(f_n^{(0,1)}(x, \Theta(x)) \right) \right)}{f_n^{(0,2)}(x, \Theta_n^*(x))} \\ &\quad - \frac{\sqrt{nh_K^d h_H^3} \mathbb{E} \left(f^{(0,1)}(x, \Theta(x)) \right)}{f_n^{(0,2)}(x, \Theta_n^*(x))}. \end{aligned}$$

Donc, en utilisant les Lemmes 3.5–3.7 et le Théorème de Slutsky, nous achevons la preuve du Théorème 3.2.

Application sur des données réelles

4.1 Introduction

Ce chapitre est une application réelle de notre étude à l'aide des logiciels Eviews10 et Matlab, afin de :

- Illustrer l'efficacité de l'estimateur du mode conditionnel dans le cas d'une analyse prévisionnelle non paramétrique de données réelles ψ -faiblement dépendantes.
- Montrer la supériorité de notre estimateur par rapport à l'estimateur à noyau de la régression

$$\text{classique } r_{NW,n}(\cdot) \text{ en présence d'hétéroscédasticité où : } r_{NW,n}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)}, x \in \mathbb{R}$$

Avant de commencer cette analyse, nous étudions la série de données et nous nous assurons que ces dernières vérifient les hypothèses de notre modèle (la stationnarité et la ψ -dépendance).

4.2 Les données

4.2.1 Description de la base de données

L'ensemble de données est constitué du taux de change quotidien de l'euro par rapport au dollar américain (USD/EUR). Ces données sont obtenues à partir du site web : <https://fr.investing.com/currencies/usd-eur-historical-data>.

L'ensemble des données porte sur la période du 01 janvier 2015 au 01 janvier 2020, soit un total de 1305 observations.

La base contient 6 variables tel que : la date, le taux d'ouverture, taux le plus haut, taux le plus bas, le cours de clôture et le pourcentage de variation.

On s'intéresse particulièrement à la variable de clôture ("Dernie") qui correspond au dernier taux de change calculé chaque jour, dont le graphique est présenté dans la figure 4.1.

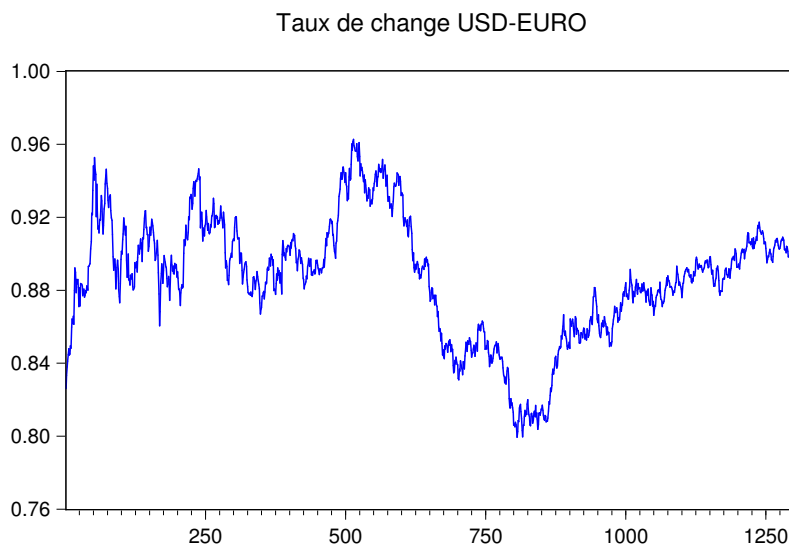


FIGURE 4.1 – L'évolution quotidienne du taux de change USD/EURO (01/01/2015 - 01/01/2020).

4.2.2 Etude de la stationnarité

La forme du graphique 4.1 illustre la non-stationnarité du fait qu'il semble présenter :

- La présence de tendances.
- Une volatilité qui change dans le temps.

La non-stationnarité du processus est confirmée par le test suivant :

Test de Dickey-Fuller augmenté (ADF) :

Le critère de Dickey-Fuller augmenté ou le test ADF est un test statistique permettant de déterminer la stationnarité d'une série chronologique.

Null Hypothesis: SERIES01 has a unit root
 Exogenous: Constant
 Lag Length: 0 (Automatic - based on SIC, maxlag=22)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.855355	0.0510
Test critical values:		
1% level	-3.435153	
5% level	-2.863549	
10% level	-2.567889	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(SERIES01)
 Method: Least Squares
 Date: 10/16/21 Time: 14:57
 Sample (adjusted): 2 1305
 Included observations: 1304 after adjustments

La p-value étant supérieure au seuil de 5% ($0.0510 > 0.05$). Nous acceptons l'hypothèse de base : la série n'est pas stationnaire. Pour la rendre stationnaire, les taux de change des devises sont

transformés en rendements logarithmiques en utilisant la formule de rendement suivante :

$$R_t = \log \left(\frac{V_t}{V_{t-1}} \right) \times 100,$$

où V_t est la valeur de clôture quotidienne du taux de change USD/EURO au temps t .

Remarque 4.1. *L'application de log vise à lisser et à réduire la tendance tandis que la différenciation stabilise le processus.*

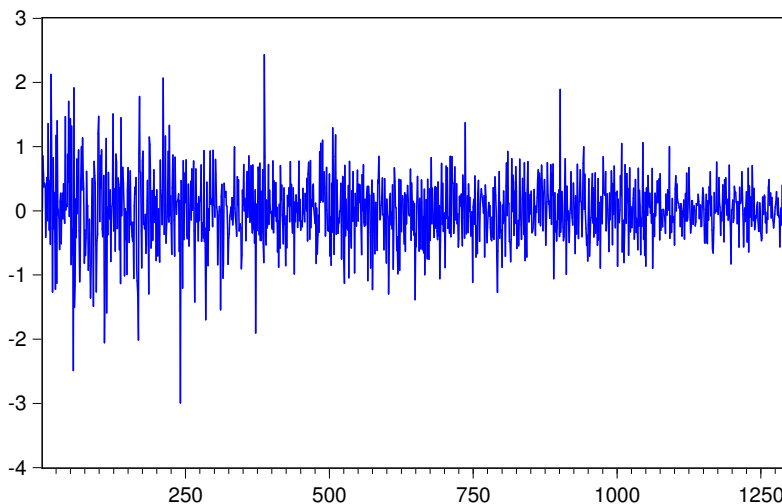


FIGURE 4.2 – Rendements logarithmiques quotidiens des taux de change USD/EURO

La figure de la courbe de différenciation logarithmiques montre une alternance autour du zéro, ce qui montre sa stationnarité. Ce résultat est également confirmé par le test ADF.

Test de Dickey-Fuller augmenté (ADF) :

Null Hypothesis: SERIES02 has a unit root
 Exogenous: Constant
 Lag Length: 4 (Fixed)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-16.26570	0.0000
Test critical values:		
1% level	-3.435173	
5% level	-2.863557	
10% level	-2.567894	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(SERIES02)
 Method: Least Squares
 Date: 10/16/21 Time: 15:42
 Sample (adjusted): 7 1305
 Included observations: 1299 after adjustments

La p-value étant nulle (< 0.05). Nous rejetons l'hypothèse de base : la série de rendements logarithmiques est stationnaire.

4.2.3 Les statistiques descriptives

Un résumé des statistiques pour l'ensemble des données sur les rendements est présenté au graphique suivant. La moyenne est positive, ce qui indique que les rendements des transactions augmentent légèrement au fil du temps. Le coefficient d'asymétrie indique que les rendements ont une distribution asymétrique, et sont décalés vers la gauche. La valeur de l'aplatissement des rendements est de 5.634294, ce qui est supérieur à trois, indiquant que la distribution des rendements suit une distribution à queue lourde, présentant ainsi l'une des caractéristiques importantes des données de séries chronologiques financières, notamment le leptokurtosis. La condition de non-normalité est confirmée par un test de Jarque-Bera qui montre que l'hypothèse nulle de normalité est rejetée au niveau de signification de 5%.

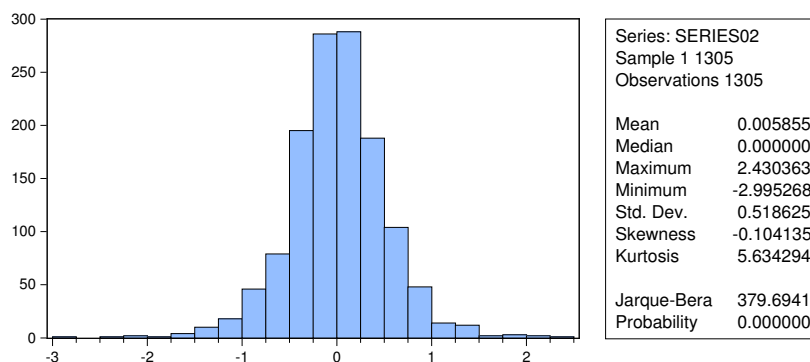


FIGURE 4.3 – Test de normalité sur les rendements

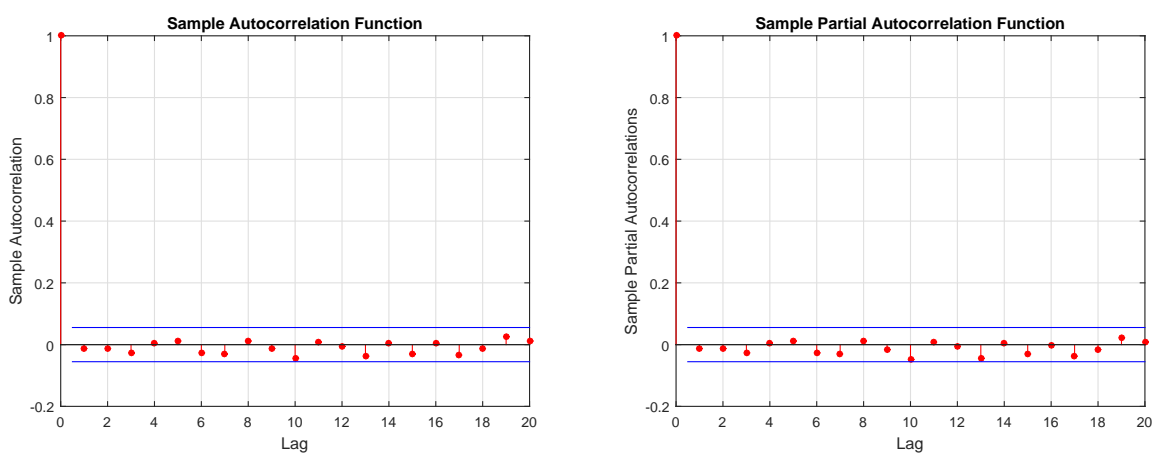


FIGURE 4.4 – Représentation du corrélogramme

Tous ces résultats indiquent que le modèle de prix de rendement que nous venons d'étudier possède des propriétés d'hétéroscédasticité (variabilité de la volatilité ou non-linéarité de la variance) et de non-normalité qui justifient l'utilisation d'une modélisation du type ARCH. Ce type de processus est capable de capter le comportement de la volatilité au cours du temps. Afin de confirmer cette hypothèse, c'est-à-dire la présence d'hétéroscédasticité, nous effectuons le test suivant.

4.2.3.1 Hétéroscédasticité conditionnelle

Heteroskedasticity Test: ARCH			
F-statistic	22.28200	Prob. F(1,1300)	0.0000
Obs*R-squared	21.94022	Prob. Chi-Square(1)	0.0000

Test Equation:			
Dependent Variable: RESID^2			
Method: Least Squares			
Date: 10/16/21 Time: 18:31			
Sample (adjusted): 4 1305			
Included observations: 1302 after adjustments			

- Les deux statistiques Fisher et khi-deux donnent deux p-valeurs inférieures au seuil (5%) : on rejette donc l'hypothèse nulle d'homoscédasticité en faveur de l'alternative d'hétéroscédasticité conditionnelle.

Sur EViews nous pouvons estimer le modèle ARCH approprié (voir la figure 4.5), voici le résultat obtenu qui semble être très significatif.

Dependent Variable: SERIES01				
Method: ML ARCH - Normal distribution (Marquardt / EViews legacy)				
Date: 01/05/22 Time: 23:32				
Sample: 1 1305				
Included observations: 1305				
Convergence achieved after 7 iterations				
Presample variance: unconditional				
GARCH = C(1) + C(2)*RESID(-1)^2 + C(3)*RESID(-2)^2				
Variable	Coefficient	Std. Error	z-Statistic	Prob.
Variance Equation				
C	0.202983	0.009907	20.48816	0.0000
RESID(-1)^2	0.104746	0.027114	3.863216	0.0001
RESID(-2)^2	0.144639	0.032157	4.497931	0.0000
R-squared	-0.000128	Mean dependent var		0.005855
Adjusted R-squared	0.000639	S.D. dependent var		0.518625
S.E. of regression	0.518460	Akaike info criterion		1.487397
Sum squared resid	350.7847	Schwarz criterion		1.499291
Log likelihood	-967.5263	Hannan-Quinn criter.		1.491859
Durbin-Watson stat	2.024566			

FIGURE 4.5 – Estimation du modèle ARCH

Nous constatons que les données peuvent être modélisées par un processus ARCH ou GARCH, qui sont des cas particuliers de processus ψ -faiblement dépendants.

4.3 Prédiction

Nous considérons la relation entre le rendement au temps t (la variable d'intérêt) et le rendement au temps $t - 1$ (la variable explicative ou covariable).

$$\begin{cases} X_t = R_{t-1} \\ Y_t = r(R_t) + \varepsilon_t, R_0 = 0, \varepsilon_t \sim \mathcal{N}(0, 1) \text{ i.i.d.}, t \geq 1. \end{cases}$$

Pour évaluer les performances du prédicteur Θ_n et $r_{NW,n}$ nous avons divisé notre échantillon en deux sous-ensembles. L'échantillon d'apprentissage, de taille $n = 1270$, désigné par $(X_i, Y_i)_{\{i=1, \dots, 1270\}}$ pour lequel les estimateurs sont calculés. Ensuite, pour chaque nouvelle covariable $(X_i)_{\{i=1271, \dots, 1301\}}$, nous calculons les valeurs prédites, par les estimateurs Θ_n et $r_{NW,n}$, correspondant à $(\hat{Y}_i)_{\{i=1271, \dots, 1301\}}$ (les valeurs de rendements pendant le dernier mois) considéré comme l'échantillon test. Dans les calculs des estimateurs, pour les deux noyaux K_d et H , nous avons pris un noyau gaussien, et les fenêtres sont égales à $h_K = h_H = 0.231$.

Nous comparons dans la figure 4.6 les valeurs réelles de $(Y_i)_{\{i=1271, \dots, 1301\}}$ et les prédictions de l'estimateur du mode conditionnel et de l'estimateur de régression.

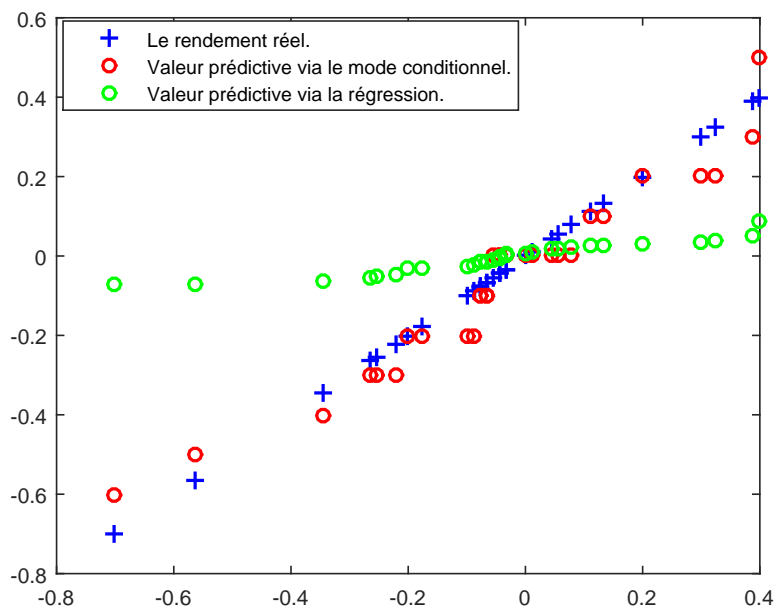


FIGURE 4.6 – Prédiction de rendement par le mode conditionnel et la régression

Pour mesurer aussi la performance de chaque méthode de prédiction, nous examinons la moyenne des erreurs carrées (MSE). Après avoir effectué les calculs, nous trouvons une $MSE = 0.45$ pour notre estimateur, et une $MSE = 0.7$ pour celui de la régression. Nous pouvons donc conclure que notre modèle présente une bonne précision d'estimation et de prédiction par rapport à la régression

et ce résultat est justifié par la présence d'hétéroscédasticité.

Dans la figure suivante, nous traçons l'estimateur de la densité conditionnelle du rendement au temps t sachant le prix du rendement au temps $t - 1$ ($f_n(y | x)$).

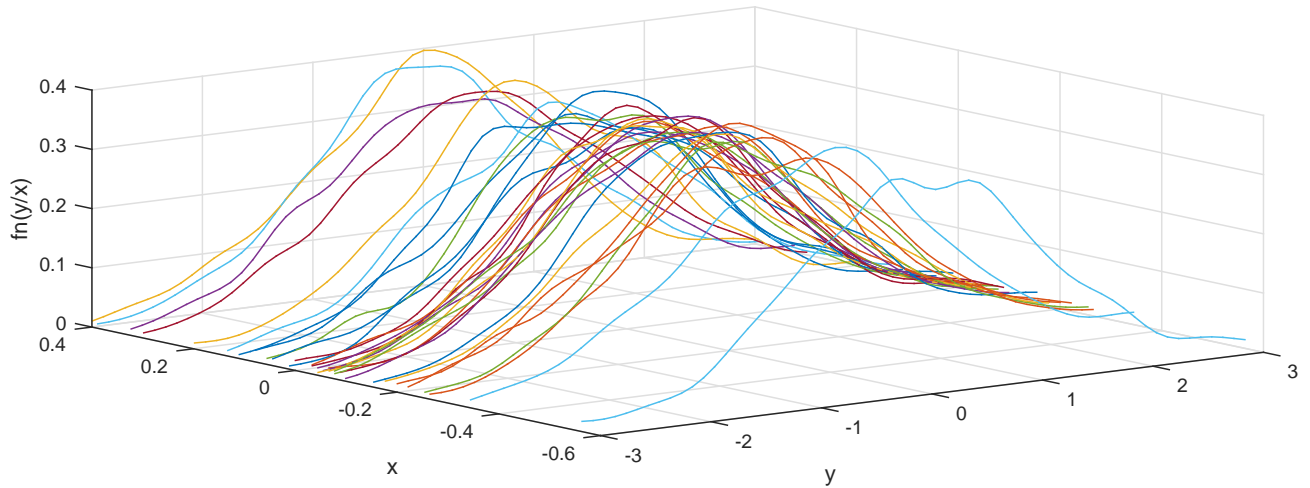


FIGURE 4.7 – L'estimateur de la densité conditionnelle $f_n(y | x)$

Conclusion et perspectives

Conclusion

Bien que la régression classique soit le sujet le plus répandu en statistique non-paramétrique, notamment pour les problèmes de prédiction, elle reste très limitée pour certaines situations. Dans ce travail, nous nous intéressons au problème de l'estimateur du mode conditionnel qui est un outil de prédiction alternative à la régression. Un aperçu des estimateurs non paramétriques a été présenté, en particulier l'estimateur à noyau du mode conditionnel dans un cadre dépendant et indépendant. La principale contribution de notre étude est le type de dépendance le plus général, soit celui de [Doukhan and Louhichi \(1999\)](#).

Dans ce modèle, nous avons déterminé la vitesse de convergence uniforme presque sûre de l'estimateur à noyau du mode conditionnel de l'ordre de $O \left\{ \left(\frac{\log n}{nh_K^d h_H} \right)^{\frac{1}{4}} + h_K + h_H \right\}$. Cette dernière est étroitement liée aux fenêtres et elle est la même que celle obtenue dans les cas i.i.d. Notons que le résultat obtenu nécessite des conditions plus restrictives sur les coefficients de dépendance par rapport aux autres types de dépendance.

Dans ce travail, il est par ailleurs question d'établir la normalité asymptotique de notre estimateur par le noyau de résultats récents en théorie des probabilités pour des données ψ -faiblement dépendantes. Ce résultat nous a permis d'obtenir des intervalles de confiance asymptotiques. Des simulations ont été effectuées pour illustrer la performance de notre estimateur et sa normalité asymptotique pour le cas unidimensionnel et bidimensionnel. Nous avons en outre sélectionné dans ces études les fenêtres optimales pour les deux noyaux. Enfin, l'étude proposée est appliquée aux rendements logarithmiques du taux de change de l'euro par rapport au dollar américain. Une analyse prévisionnelle nous a permis de montrer l'efficacité de notre estimateur et son intérêt.

Perspectives

Le travail réalisé dans cette thèse ouvre le champ à de nombreuses perspectives intéressantes :

- Afin d’obtenir de meilleurs résultats, il serait important d’appliquer la méthode de validation croisée et de plugin pour trouver la fenêtre optimale théorique pour notre estimateur. Une étude comparative pourrait être envisagée.
- Généraliser nos résultats au cas de covariable fonctionnelle. Le cas du mélange est traité dans [Ferraty et al. \(2005\)](#), on peut donc également traiter une adaptation au cas plus général au sens de [Doukhan and Louhichi \(1999\)](#).
- Nous pouvons également envisager d’étendre nos résultats au mode conditionnel dans le cas spatial. Étant donné un processus spatial multidimensionnel stationnaire $Z_i = (X_i, Y_i) \in \mathbb{R}^d \times \mathbb{R}$, $i \in \mathbb{N}^N$ où $N \in \mathbb{N}^*$, nous étudions l’estimateur à noyau de la fonction mode conditionnelle spatiale de la variable réponse Y_i étant donné la variable explicative X_i . Le cas du mélange a été traité dans [Dabo-Niang et al. \(2014\)](#).
- On peut également envisager d’étendre nos résultats au mode conditionnel pour des données incomplètes.
- Dans ce travail, nous nous sommes limités aux estimateurs à noyau. Récemment [Benkhalel et al. \(2020\)](#) ont proposé un estimateur local linéaire de la densité conditionnelle lorsque la variable réponse est censurée tandis que la variable explicative est fonctionnelle. Ils ont établi dans le cas α -mélangeant, la convergence presque complète de l’estimateur construit et ont appliqué leur résultat à l’estimation du mode conditionnel. Il serait alors intéressant d’adapter cette méthode d’estimation et d’étudier ses propriétés pour le cas de données faiblement dépendantes.

Bibliographie

- Abraham, C., Biau, G., and Cadre, B. (2004). On the asymptotic properties of a simple estimate of the mode. *ESAIM : Probability and Statistics*, 8 :1–11.
- Andrews, D. W. (1984). Non-strong mixing autoregressive processes. *Journal of Applied Probability*, 21(4) :930–934.
- Arora, S. and Taylor, J. W. (2016). Forecasting electricity smart meter data using conditional kernel density estimation. *Omega*, 59 :47–59.
- Athreya, K. B. and Pantula, S. G. (1986). Mixing properties of harris chains and autoregressive processes. *Journal of applied probability*, 23(4) :880–892.
- Bardet, J.-M., Doukhan, P., Lang, G., and Ragache, N. (2008). Dependent lindeberg central limit theorem and some applications. *ESAIM : Probability and Statistics*, 12 :154–172.
- Benkhaled, A., Madani, F., and Khardani, S. (2020). Strong consistency of local linear estimation of a conditional density function under random censorship. *Arabian Journal of Mathematics*, 9 :513–529.
- Benrabah, O., Saïd, E. O., and Tatachak, A. (2015). A kernel mode estimate under random left truncation and time series model : asymptotic normality. *Statistical Papers*, 56(3) :887–910.
- Berlinet, A., Gannoun, A., and Matzner-Løber, E. (1998). Normalité asymptotique d’estimateurs convergents du mode conditionnel. *The Canadian Journal of Statistics-La Revue Canadienne de Statistique*, pages 365–380.
- Birkel, T. (1988). Moment bounds for associated sequences. *The annals of Probability*, pages 1184–1193.
- Bosq, D. (1998). Implementation of nonparametric method and numerical applications. In *Nonparametric Statistics for Stochastic Processes*, pages 169–195. Springer.

- Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2) :353–360.
- Bulinski, A. and Shashkin, A. (2007). *Limit theorems for associated random fields and related systems*, volume 10. World Scientific.
- Bulinski, A. and Suquet, C. (2001). Normal approximation for quasi-associated random fields. *Statistics & probability letters*, 54(2) :215–226.
- Bulinski, A. V. and Shabanovich, E. (1998). Asymptotical behaviour for some functionals of positively and negatively dependent random fields. *Fundamentalnaya i Prikladnaya Matematika*, 4(2) :479–492.
- Burton, R. M., Dabrowski, A., and Dehling, H. (1986). An invariance principle for weakly associated random vectors. *Stochastic Processes and their applications*, 23(2) :301–306.
- Cacoullos, T. (1964). Estimation of a multivariate density. Technical report, University of Minnesota.
- Collomb, G., Härdle, W., and Hassani, S. (1986). A note on prediction via estimation of the conditional mode function. *Journal of Statistical Planning and Inference*, 15 :227–236.
- Dabo-Niang, S., Kaid, Z., and Laksaci, A. (2012). On spatial conditional mode estimation for a functional regressor. *Statistics & Probability Letters*, 82(7) :1413–1421.
- Dabo-Niang, S. and Laksaci, A. (2007). Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle. *Comptes Rendus Mathématique*, 344(1) :49–52.
- Dabo-Niang, S., Ould-Abdi, S. A., Ould-Abdi, A., and Diop, A. (2014). Consistency of a nonparametric conditional mode estimator for random fields. *Statistical Methods & Applications*, 23(1) :1–39.
- De Gooijer, J. G. and Gannoun, A. (2000). Nonparametric conditional predictive regions for time series. *Computational Statistics & Data Analysis*, 33(3) :259–275.
- Dedecker, J., Doukhan, P., Lang, G., Rafael, L. R. J., Louhichi, S., and Prieur, C. (2007). Weak dependence. In *Weak dependence : With examples and applications*, pages 9–20. Springer.
- Dedecker, J. and Prieur, C. (2004a). Couplage pour la distance minimale. *Comptes Rendus Mathématique*, 338(10) :805–808.
- Dedecker, J. and Prieur, C. (2004b). Coupling for τ -dependent sequences and applications. *Journal of Theoretical Probability*, 17(4) :861–885.
- Deheuvels, P. (1974). Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité. *CR Acad. Sci. Paris*, 278 :1217–1220.

- Doukhan, P. (1994). *Mixing : properties and examples*, volume 85. Springer Science & Business Media.
- Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic processes and their applications*, 84(2) :313–342.
- Doukhan, P. and Louhichi, S. (2001). Functional estimation of a density under a new weak dependence condition. *Scandinavian Journal of Statistics*, 28(2) :325–341.
- Doukhan, P. and Neumann, M. H. (2007). Probability and moment inequalities for sums of weakly dependent random variables, with applications. *Stochastic Processes and their Applications*, 117(7) :878–903.
- Doukhan, P., Teyssiere, G., and Winant, P. (2006). A larch (∞) vector valued process. In *Dependence in probability and statistics*, pages 245–258. Springer.
- Eddy, W. F. (1980). Optimum kernel estimators of the mode. *The Annals of Statistics*, 8(4) :870–882.
- Eddy, W. F. (1982). The asymptotic distributions of kernel estimators of the mode. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 59(3) :279–290.
- Efromovich, S. (1999). *Nonparametric Curve Estimation : Methods, Theory and Applications*. Springer Science & Business Media.
- Efromovich, S. (2007). Conditional density estimation in a regression setting. *The Annals of Statistics*, 35(6) :2504–2535.
- Efromovich, S. (2010). Oracle inequality for conditional density estimation and an actuarial example. *Annals of the Institute of Statistical Mathematics*, 62(2) :249–275.
- Einbeck, J. and Tutz, G. (2006). Modelling beyond regression functions : an application of multimodal regression to speed–flow data. *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, 55(4) :461–475.
- Esary, J. D., Proschan, F., and Walkup, D. W. (1967). Association of random variables, with applications. *The Annals of Mathematical Statistics*, pages 1466–1474.
- Ezzahrioui, M. (2010). On the asymptotic properties of a nonparametric estimator of the conditional mode for functional dependent data. *Statistica Neerlandica*, 64 :171–201.
- Ezzahrioui, M. and Ould-Saïd, E. (2008). Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *Journal of Nonparametric Statistics*, 20(1) :3–18.

- Faugeras, O. P. (2009). A quantile-copula approach to conditional density estimation. *Journal of Multivariate Analysis*, 100(9) :2083–2099.
- Ferrani, Y., Ould-Saïd, E., and Tatachak, A. (2016). On kernel density and mode estimates for associated and censored data. *Communications in Statistics-Theory and Methods*, 45(7) :1853–1862.
- Ferraty, F., Laksaci, A., and Vieu, P. (2005). Functional time series prediction via conditional mode estimation. *Comptes Rendus Mathématique*, 340(5) :389–392.
- Fortuin, C. M., Kasteleyn, P. W., and Ginibre, J. (1971). Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22(2) :89–103.
- Gorodestkii, V. V. (1977). On the strong mixing property for linear sequences. *Teoriya Veroyatnostei i ee Primeneniya*, 22(2) :421–423.
- Grund, B. and Hall, P. (1995). On the minimisation of \mathbb{L}^p error in mode estimation. *The Annals of Statistics*, 23(6) :2264–2284.
- Gyorfi, L. and Kohler, M. (2007). Nonparametric estimation of conditional distributions. *IEEE Transactions on Information Theory*, 53(5) :1872–1879.
- Hall, P., Racine, J., and Li, Q. (2004). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, 99(468) :1015–1026.
- Holmes, M. P., Gray, A. G., and Isbell, C. L. (2012). Fast nonparametric conditional density estimation. *arXiv preprint arXiv : 1206.5278*.
- Hwang, E. and Shin, D. W. (2014). A note on exponential inequalities of ψ -weakly dependent sequences. *Communications for Statistical Applications and Methods*, 21(3) :245–251.
- Hwang, E. and Shin, D. W. (2016). Kernel estimators of mode under ψ -weak dependence. *Annals of the Institute of Statistical Mathematics*, 68(2) :301–327.
- Hyndman, R. J., Bashtannyk, D. M., and Grunwald, G. K. (1996). Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics*, 5(4) :315–336.
- Ibragimov, I. (1959). Some limit theorems for stochastic processes stationary in the strict sense. In *Dokl. Akad. Nauk SSSR*, volume 125, pages 711–714.
- Ioannides, D. and Matzner-Løber, E. (2002). Nonparametric estimation of the conditional mode with errors-in-variables : Strong consistency for mixing processes. *Journal of Nonparametric Statistics*, 14(3) :341–352.

- Ioannides, D. and Matzner-Løber, E. (2004). A note on asymptotic normality of convergent estimates of the conditional mode with errors-in-variables. *Journal of Nonparametric Statistics*, 16(3-4) :515–524.
- Izbicki, R. and Lee, A. B. (2016). Nonparametric conditional density estimation in a high-dimensional regression setting. *Journal of Computational and Graphical Statistics*, 25(4) :1297–1316.
- Jeon, J. and Taylor, J. W. (2012). Using conditional kernel density estimation for wind power density forecasting. *Journal of the American Statistical Association*, 107(497) :66–79.
- Joag-Dev, K. and Proschan, F. (1983). Negative association of random variables with applications. *The Annals of Statistics*, pages 286–295.
- Khardani, S., Lemdani, M., and Ould-Saïd, E. (2010). Some asymptotic properties for a smooth kernel estimator of the conditional mode under random censorship. *Journal of the Korean Statistical Society*, 39(4) :455–469.
- Khardani, S., Lemdani, M., and Ould-Saïd, E. (2011). Uniform rate of strong consistency for a smooth kernel estimator of the conditional mode for censored time series. *Journal of statistical planning and inference*, 141(11) :3426–3436.
- Khardani, S., Lemdani, M., and Saïd, E. O. (2012). On the strong uniform consistency of the mode estimator for censored time series. *Metrika*, 75(2) :229–241.
- Kolmogorov, A. N. and Rozanov, Y. A. (1960). On strong mixing conditions for stationary gaussian processes. *Theory of Probability & Its Applications*, 5(2) :204–208.
- Konakov, V. (1974). On the asymptotic normality of the mode of multidimensional distributions. *Theory of Probability & Its Applications*, 18(4) :794–799.
- Laksaci, A. and Yousfate, A. (2002). Estimation fonctionnelle de la densité de l’opérateur de transition d’un processus de markov à temps discret. *Comptes Rendus Mathematique*, 334(11) :1035–1038.
- Leclerc, J. and Pierre-Loti-Viaud, D. (2000). Vitesse de convergence presque sûre de l’estimateur à noyau du mode. *Comptes Rendus de l’Academie des Sciences-Series I-Mathematics*, 331(8) :637–640.
- Lehmann, E. L. (1966). Some concepts of dependence. *The Annals of Mathematical Statistics*, 37(5) :1137–1153.
- Louani, D. and Ould-Saïd, E. (1999). Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. *Journal of Nonparametric Statistics*, 11(4) :413–442.
- Martins Rosa, C. (1993). *Prévision robuste sous une hypothèse ergodique*. PhD thesis, Toulouse 1.

- Mehra, K., Ramakrishnaiah, Y., and Sashikala, P. (2000). Laws of iterated logarithm and related asymptotics for estimators of conditional density and mode. *Annals of the Institute of Statistical Mathematics*, 52(4) :630–645.
- Mokkadem, A. and Pelletier, M. (2005). Moderate deviations for the kernel mode estimator and some applications. *Journal of statistical planning and inference*, 135(2) :276–299.
- Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1) :141–142.
- Oliveira, P. E. (2012). *Asymptotics for associated random variables*. Springer Science & Business Media.
- Otneim, H. and Tjøstheim, D. (2018). Conditional density estimation using the local gaussian correlation. *Statistics and Computing*, 28(2) :303–321.
- Ould-Saïd, E. (1993). Estimation non paramétrique du mode conditionnel. application à la prévision. *Comptes rendus de l'Académie des sciences. Série 1, Mathématique*, 316(9) :943–947.
- Ould-Saïd, E. (1997). A note on ergodic processes prediction via estimation of the conditional mode function. *Scandinavian journal of statistics*, 24(2) :231–239.
- Ould-Saïd, E. and Cai, Z. (2005). Strong uniform consistency of nonparametric estimation of the censored conditional mode function. *Nonparametric Statistics*, 17(7) :797–806.
- Ould-Saïd, E. and Tatachak, A. (2007). Asymptotic properties of the kernel estimator of the conditional mode for the left truncated model. *Comptes Rendus Mathématique*, 344(10) :651–656.
- Ould-Saïd, E. and Tatachak, A. (2009a). On the non-parametric estimation of the simple mode under random lefttruncation model. *Rev. Roum. Math. Pure. A*, 54(3) :243–266.
- Ould-Saïd, E. and Tatachak, A. (2009b). Strong consistency rate for the kernel mode estimator under strong mixing hypothesis and left truncation. *Communications in Statistics-Theory and Methods*, 38(8) :1154–1169.
- Ould-Saïd, E. and Tatachak, A. (2011). A nonparametric conditional mode estimate under rlt model and strong mixing condition. *International Journal of Statistics & Economics (formerly known as the "Bulletin of Statistics & Economics")*, 6(S11) :76–92.
- Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3) :1065–1076.

- Pearson, K. (1895). X. contributions to the mathematical theory of evolution-ii. skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London.(A.)*, (186) :343–414.
- Pham, T. D. and Tran, L. T. (1991). Kernel density estimation under a locally mixing condition. In *Nonparametric Functional Estimation and Related Topics*, pages 419–430. Springer.
- Pitt, L. D. (1982). Positively correlated normal variables are associated. *The Annals of Probability*, pages 496–499.
- Prieur, C. (2001). *Applications statistiques de suites faiblement dépendantes et de systemes dynamiques*. PhD thesis, Université de Cergy Pontoise.
- Quintela-Del-Rio, A. and Vieu, P. (1997). A nonparametric conditional mode estimate. *Journal of Nonparametric Statistics*, 8(3) :253–266.
- Rih, S. and Tatachak, A. (2021). Kernel conditional density and mode estimation for psi-weakly dependent observations. *Communications in Statistics-Theory and Methods*, pages 1–27.
- Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants*, volume 31. Springer Science & Business Media.
- Romano, J. P. (1988). On weak convergence and optimality of kernel density estimates of the mode. *The Annals of Statistics*, pages 629–647.
- Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1) :43–47.
- Rosenblatt, M. (1969). Conditional probability density and regression estimators. *Multivariate analysis II*, 25 :31.
- Rossi, V. (2004). *Filtrage non linéaire par noyaux de convolution : application à un procédé de dépollution biologique*. PhD thesis, Montpellier, ENSA.
- Roussas, G. G. (1969). Nonparametric estimation of the transition distribution function of a markov process. *The Annals of Mathematical Statistics*, 40(4) :1386–1400.
- Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, pages 65–78.
- Samanta, M. (1973). Nonparametric estimation of the mode of a multivariate density. *South African Statistical Journal*, 7(2) :109–117.

- Samanta, M. and Thavaneswaran, A. (1990). Non-parametric estimation of the conditional mode. *Communications in Statistics-Theory and Methods*, 19(12) :4515–4524.
- Sart, M. (2017). Estimating the conditional density by histogram type estimators and model selection. *ESAIM : Probability and Statistics*, 21 :34–55.
- Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society : Series B (Methodological)*, 53(3) :683–690.
- Shi, X., Wu, Y., and Miao, B. (2009). A note on the convergence rate of the kernel density estimator of the mode. *Statistics & probability letters*, 79(17) :1866–1871.
- Takeuchi, I., Nomura, K., and Kanamori, T. (2009). Nonparametric conditional density estimation using piecewise-linear solution path of kernel quantile regression. *Neural Computation*, 21(2) :533–559.
- Tran, L. T. (1990). Kernel density estimation on random fields. *Journal of Multivariate Analysis*, 34(1) :37–53.
- Van Ryzin, J. (1969). On strong consistency of density estimates. *The Annals of Mathematical Statistics*, 40(5) :1765–1772.
- Vieu, P. (1996). A note on density mode estimation. *Statistics & probability letters*, 26(4) :297–307.
- Volkonskii, V. and Rozanov, Y. A. (1959). Some limit theorems for random functions. i. *Theory of Probability & Its Applications*, 4(2) :178–197.
- Wen, K. and Wu, X. (2017). Smoothed kernel conditional density estimation. *Economics Letters*, 152 :112–116.
- Withers, C. (1981). Conditions for linear processes to be strong-mixing. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(4) :477–480.
- Youndje, E. (1993). *Estimation non paramétrique de la densité conditionnelle par la méthode du noyau*. PhD thesis, Rouen.