



N° Attribué par la bibliothèque

--	--	--	--	--	--	--	--	--	--



Année univ. : 2017/2018

SUR LA CONSISTANCE DES ESTIMATEURS DE LA DISTRIBUTION CONDITIONNELLE ET SES DÉRIVÉES

Mémoire présenté en vue de l'obtention du diplôme de

Master Académique

Université de Saïda - Dr Moulay Tahar

Discipline : MATHÉMATIQUES

Spécialité : Analyse stochastique, statistique des processus et
applications (ASSPA)

par

Mokhtari Zohra¹

Sous la direction de

Docteur Saâdia Rahmani

Soutenu le 19/06/2018 devant le jury composé de

Mokhtar Kadi	Maître de Conférences B. Univ. Saïda	Président
Saâdia Rahmani	Maître de Conférences A. Univ. Saïda	Encadreur
Rachida Rouane	Maître de Conférences B. Univ. Saïda	Examinatrice
Nawel Hachemi	Maître de Conférences B. Univ. Saïda	Examinatrice

1. e-mail : zohramokhtari4@gmail.com

REMERCIEMENTS

Je souhaite avant tout remercier dieu clément miséricordieux de m'avoir donné le courage et la volonté de réaliser mon objectif.

Mes deuxièmes remerciements, et surtout les plus vifs, vont naturellement à mon encadreur Mademoiselle *Rahmani Saâdia* qui, par ses paroles, ses écrits, sa disponibilité, ses conseils et ses critiques, ont guidé mes réflexions et m'ont accompagnée pas à pas dans mon travail.

Je remercie Docteur *Kadi Mokhtar* pour l'attention qu'il a manifesté à l'égard de ce mémoire, en s'engageant à en être le président de jury.

De même, je suis grandement reconnaissante à Docteur *Rouane Rachida* et Docteur *Hachemi Nawel* de l'intérêt et du temps qu'ils ont bien voulu accorder à l'expertise de ce mémoire, en acceptant d'en être les examinatrices. Soutenir ce travail devant vous, qui m'avez tant appris tout au long de mon parcours universitaire est, pour moi, un privilège certain.

Je suis particulièrement reconnaissante à mes très chers parents pour leur soutien indéfectible au quotidien.

À tous ceux qui ont, de près ou de loin, participé à l'élaboration de ce mémoire, et que je ne pourrai nommer ici, je vous remercie de votre sollicitude aussi minime qu'elle ait pu être.

À tous, merci !

Table des matières

1	Introduction générale	4
1.1	Contexte bibliographique	4
1.2	Préliminaires	7
1.2.1	Définitions	7
1.2.2	Outils	9
1.3	Plan de mémoire	10
2	Le cas indépendant et identiquement distribué	11
2.1	La fonction de répartition conditionnelle	12
2.1.1	Contexte fini-dimensionnel	12
2.1.2	Contexte fonctionnel	14
2.2	La densité conditionnelle et ses dérivées	17
2.2.1	Contexte fini-dimensionnel	17
2.2.2	Contexte fonctionnel	18
2.3	Le mode conditionnel	20
2.3.1	Le modèle et son estimateur	20
2.3.2	Notations et hypothèses	21
2.3.3	Le résultat	21
2.4	Le quantile conditionnel	23

2.4.1	Le modèle et son estimateur	23
2.4.2	Notations et hypothèses	24
2.4.3	Le résultat	24
2.5	Preuve des résultats techniques	26
2.5.1	Démonstrations relatives à la section 2.1	26
2.5.2	Démonstrations relatives à la section 2.2	31
2.5.3	Démonstrations relatives à la section 2.3	36
2.5.4	Démonstrations relatives à la section 2.4	36
2.6	Quelques exemples de probabilités de petites boules	37
2.6.1	Processus de diffusion	37
2.6.2	Mouvement Brownien Fractionnel	39
3	Cas dépendant	41
3.1	Notations et hypothèses	41
3.2	Le résultat	42
3.3	Preuve des résultats techniques	43
4	Application et conclusion	51
4.1	Application	51
4.1.1	Mise en oeuvre sur des données réelles	52
4.2	Conclusion	57
	Bibliographie	58

Chapitre 1

Introduction générale

1.1 Contexte bibliographique

Tous les statisticiens qui manipulent des données s'accorderaient à dire que l'un des problèmes les plus importants qu'ils rencontrent dans leur domaine est l'étude de la relation qui puisse exister entre ces données. Cette étude statistique, a connu un tel développement depuis des dizaines d'années, qu'elle constitue aujourd'hui à elle seule une branche fondamentale de la statistique.

C'est en 1964 que les modèles conditionnels d'estimation non-paramétrique à noyaux auxquels nous nous intéressons au cours de ce mémoire ont été initiés. Ces modèles ont été introduit par les travaux de Nadaraya [16] et ceux de Watson [28].

Un des modèles le plus fréquemment rencontré en statistique non-paramétrique est la fonction de répartition conditionnelle. L'importance de cette fonction provient du fait que la plupart des quantités statistiques utilisées dans la pratique permettant de comprendre la relation liant une variable explicative X et une variable réponse Y sont fonction de celle-ci. En outre, plusieurs outils de prédiction telle que le mode conditionnel, la densité conditionnelle et les quantiles conditionnels sont basées sur l'estimation préliminaire de

ce paramètre fonctionnel.

Historiquement, dans le cas fini-dimensionnel, Roussas [22] semble être le premier à aborder le problème de l'estimation de la fonction de répartition conditionnelle par la méthode à noyau. Il a établi la convergence en probabilité de cet estimateur en considérant des données à caractère markovien.

Le premier article sur l'estimation de la densité conditionnelle a été obtenue par Roussas [22], il a établi la convergence en moyenne quadratique, la convergence en probabilité et la normalité asymptotique. Notons aussi que ce paramètre fonctionnel a été largement étudiée par Youndjé [29] dans le cas de données mélangeantes où indépendantes. Laksaci et Yousfate [14] ont établi la convergence en norme L^p d'un estimateur à noyau de la densité conditionnelle d'un processus markovien stationnaire.

Pour ce qui concerne l'estimation du mode conditionnel, l'étude de ce paramètre à été motivée par sa facilité d'implémentation dans le domaine de la prévision non-paramétrique, (voir l'article de Collomb et al. [6]). Quintela et Vieu [19] ont estimé le mode conditionnel comme étant le point annulant la dérivée d'ordre un de l'estimateur de la densité conditionnelle et ils ont établi la convergence presque-complète de cet estimateur pour le processus de mélange. Ould-said [17] a montré la validité du résultat de Collomb et al. [6] pour des processus ergodiques. Nous renvoyons à Berlinet et al. [1] pour la normalité asymptotique du mode conditionnel d'un processus fortement mélangeant.

L'estimation des quantiles conditionnels fut traitée pour la première fois par Roussas [23] dans le cas de processus de Markov. En 1989, Samanta [25] a établi la normalité asymptotique et la convergence uniforme d'un estimateur à noyau des quantiles conditionnels dans le cas i.i.d. Deux ans plus tard, Roussas en 1991 [24] a considéré le cas où les observations sont issues d'un processus de Markov et il a élaboré un résultat sur la convergence presque sûre d'un estimateur à noyau pour ce paramètre. Nous renvoyons à Berlinet et al. [2] pour une étude générale de la normalité asymptotique de quantiles conditionnels.

La théorie de l'analyse des données fonctionnelles a connu son plus grand essor en 1997 avec la monographie de Ramsay et Silverman [21]. Dans le même esprit, le livre de Bosq [3] a lui aussi fait état de différents aspects théoriques de l'étude des données fonctionnelles, en traitant des processus linéaires en temps continu. Il faut noter que la contribution de Ferraty et Vieu [10] peut être considérée comme déterminante dans le cadre non-paramétrique fonctionnel. Ces auteurs ont étudié les propriétés asymptotiques de plusieurs modèles non-paramétriques tels que l'opérateur de régression, la fonction de répartition conditionnelle, la densité conditionnelle, etc.

Notons que depuis la parution de ce livre, des centaines de recherches ont été dirigées vers cette branche de la statistique qu'est "l'analyse non-paramétrique des données fonctionnelles".

L'estimation non-paramétrique de la fonction de répartition conditionnelle lorsque la variable explicative est à valeurs dans un espace de dimension infini, a été introduite par Ferraty et al. [8]. Ces auteurs ont construit un estimateur à double noyaux pour la fonction de répartition conditionnelle et ils ont précisé la vitesse de convergence presque-complète de cet estimateur. D'autre part, plusieurs auteurs ont étudié l'estimation de la fonction de répartition conditionnelle comme une étude préliminaire de l'estimation des quantiles conditionnels. Laksaci et al. [15] par exemple, ont proposé une méthode d'estimation des quantiles conditionnels, et ils ont établi la convergence presque-complète et la normalité asymptotique de leur estimateur quand les observations sont i.i.d.

Le travail de Ferraty et al. [9] peut être considéré comme le point de départ d'une littérature abondante qui s'est développée sur l'estimation de la densité conditionnelle et de ses dérivées.

Ezzahrioui et Ould-Saïd [7], quant à eux, ont estimé le mode conditionnel par le point qui annule la dérivée de l'estimateur à noyau de la densité conditionnelle, en se concentrant sur la normalité asymptotique de l'estimateur proposé.

1.2 Préliminaires

Dans ce qui suit, nous regroupons quelques définitions ainsi que quelques outils qui seront nécessaires pour l'élaboration des résultats établis dans ce mémoire.

1.2.1 Définitions

La propriété asymptotique que nous allons étudier tout au long de ce mémoire est la convergence presque-complète. Ce critère ainsi que sa vitesse de convergence sont présentés dans les définitions suivantes :

Définition 1.2.1 [10]

On dit que $(X_n)_{n \in \mathbb{N}}$ converge presque-complètement vers une variable aléatoire réelle X , si et seulement si :

$$\forall \varepsilon > 0, \quad \sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| > \varepsilon) < \infty,$$

Et la convergence presque-complète de $(X_n)_{n \in \mathbb{N}}$ vers X est désignée par :

$$\lim_{n \rightarrow \infty} X_n = X, \quad p.co.$$

Définition 1.2.2 [10]

On dit que la vitesse de convergence presque-complète de $(X_n)_{n \in \mathbb{N}}$ vers X est d'ordre u_n telle que $(u_n)_{n \in \mathbb{N}}$ est une suite déterministe de nombres réels positifs qui tend vers zéro si et seulement si :

$$\exists \varepsilon_0, \quad \sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| > \varepsilon_0 u_n) < \infty,$$

Et nous écrivons :

$$X_n - X = O_{p.co.}(u_n).$$

Proposition 1.2.1 [10]

Si $\lim_{n \rightarrow \infty} u_n = 0$, $X_n = O_{p.co.}(u_n)$ et $\lim_{n \rightarrow \infty} Y_n = l_y$, p.co., où l_y est un nombre réel déterministe. Alors :

- i) $X_n Y_n = O_{p.co.}(u_n)$.
- ii) $\frac{X_n}{Y_n} = O_{p.co.}(u_n)$ si $l_y \neq 0$.

Remarque 1.2.1 Ce type de convergence implique à la fois la convergence presque sûre et la convergence en probabilité.

De nombreux résultats de statistique fonctionnelle ont été établis en considérant des échantillons indépendants. Cependant, il est parfois intéressant de considérer et d'étudier des échantillons dépendants afin de pouvoir répondre à des situations où les données ne sont pas indépendantes.

Il y a plusieurs types de modélisation de la dépendance au sein d'un échantillon. Nous nous intéresserons dans ce mémoire à des variables α -mélangeantes. Pour ce but, on rappelle la définition de suite de variables α -mélangeantes

Définition 1.2.3 [10]

Soit $\{\xi_n\}_{n \in \mathbb{Z}}$ une famille de variables aléatoires définie sur l'espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans un espace probabilisable (Ω', \mathcal{A}') . On note pour $-\infty \leq j \leq k \leq +\infty$, par \mathcal{A}_j^k la σ -algèbre engendrée par les variables aléatoires $(\xi_s, j \leq s \leq k)$. on défini le coefficient α de mélange fort par :

$$\alpha(n) = \sup_k \sup_{A \in \mathcal{A}_{-\infty}^k} \sup_{B \in \mathcal{B}_k^{+\infty}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

La suite $(\xi_n)_{n \in \mathbb{Z}}$ est dite α -mélangeante (ou fortement mélangeante), si :

$$\lim_{n \rightarrow \infty} \alpha(n) = 0. \tag{1.1}$$

Parmi les différents types de décroissance des coefficients de α -mélange on distinguera les cas de décroissance arithmétique et géométrique. Nous nous concentrons dans ce mémoire sur le premier type.

Définition 1.2.4 [10]

La suite $(\xi_n)_{n \in \mathbb{N}}$ est dite arithmétiquement (algébriquement) α -mélangeante avec $a > 0$ si :

$$\exists C > 0, \alpha(n) \leq Cn^{-a}. \quad (1.2)$$

1.2.2 Outils

L'outil que nous allons utiliser de manière déterminante dans les problèmes de convergence presque-complète est l'inégalité exponentielle de *Hoeffding* ci-dessous :

Lemme 1.2.1 [13]

Soit $\Delta_1, \dots, \Delta_n$ des variables aléatoires centrés, indépendantes et de même loi, telle qu'il existe deux réels positifs δ_1 et δ_2 vérifiant :

$$|\Delta_1| \leq \delta_1 \text{ et } \mathbb{E}|\Delta_1|^2 \leq \delta_2$$

Alors, pour tout $\varepsilon \in]0, \delta_2/\delta_1[$ on a : $\mathbb{P} \left[n^{-1} \left| \sum_{i=1}^n \Delta_i \right| > \varepsilon \right] \leq 2 \exp \left(\frac{-n\varepsilon^2}{4\delta_2} \right)$.

Le lemme suivant donne l'Inégalité de *Fuk Nageav*, cette inégalité est en fait une extension au cadre de variables fortement mélangeantes de l'Inégalité de *Hoeffding* :

Lemme 1.2.2 [13]

Soit $\{\Delta_i, i \in \mathbb{N}\}$ une famille de variables aléatoires à valeurs dans \mathbb{R} qui vérifient la condition de mélange fort (1.1) avec des coefficients à décroissance algébrique tels que définis en (1.2). On pose

$$s_n^2 = \sum_{i=1}^n \sum_{j=1}^n |\text{Cov}(\Delta_i, \Delta_j)|.$$

Si $\|\Delta\|_\infty < \infty$, $\forall i$, alors on a pour tout $\varepsilon > 0$ et pour tout $r > 1$:

$$\mathbb{P} \left[\left| \sum_{k=1}^n \Delta_k \right| > 4\varepsilon \right] \leq 4 \left(1 + \frac{\varepsilon^2}{rs_n^2} \right)^{-\frac{r}{2}} + 2ncr^{-1} \left(\frac{2r}{\varepsilon} \right)^{a+1}.$$

Le lemme suivant présente une inégalité de covariance qui est très utile pour le calcul de s_n^2 , définie dans le lemme précédent :

Lemme 1.2.3 *Inégalité de Dvdyov-Rio* : Soit $\{\Delta_i, i \in \mathbb{N}\}$ une famille de variables aléatoires à valeurs dans \mathbb{R} qui vérifient la condition de mélange fort (1.1), et telle que $\|\Delta\|_\infty < \infty$, $\forall i$. On a pour tout $i \neq j$:

$$|Cov(\Delta_i, \Delta_j)| \leq 4\|\Delta_i\|_\infty\|\Delta_j\|_\infty\alpha_{|i-j|}.$$

1.3 Plan de mémoire

Dans ce mémoire, nous nous proposons d'étudier les propriétés asymptotiques des paramètres fonctionnels conditionnels dans le cadre non-paramétriques, lorsque la variable explicative est fonctionnelle alors que la réponse est réelle, les paramètres fonctionnels considérés sont la fonction de répartition conditionnelle, la densité conditionnelle ainsi que ses successives dérivées, le mode conditionnel et le quantile conditionnel. Pour ce but, ce mémoire est divisé en quatre chapitres. Après ce chapitre introductif destiné principalement à un historique sur les modèles étudiés dans le cas de dimension finie et infinie, et quelques définitions et outils nécessaires à l'élaboration des résultats étudiés dans ce mémoire, on étudie dans le deuxième chapitre la convergence presque-complète de ses modèles dans le cas où les observation sont indépendants et identiquement distribuées (i.i.d.). Nous traitons dans le troisième chapitre le cas dépendant nous nous intéressons en particulier au cas de mélangeance forte. Le dernier chapitre est consacré à une application et nous le achevons par une conclusion générale.

Chapitre 2

Le cas indépendant et identiquement distribué

Ce chapitre traite de divers modèles non paramétriques comprenant la fonction de répartition conditionnelle ainsi que ses dérivées tels que la densité conditionnelle et ses dérivées, le mode conditionnel et le quantile conditionnel, quand les co-variables sont infinie-dimensionnelles et en considérant des observations indépendantes et identiquement distribuées. Ce chapitre est donc divisé en six sections. Nous établissons la convergence presque-complète d'un estimateur à noyau de la distribution conditionnelle dans la première section. Dans la deuxième section, nous étudions la même propriété asymptotique pour un estimateurs à noyau des dérivées successives de la densité conditionnelle. Le mode conditionnel sera étudié dans la troisième section. La quatrième section est consacrée au quantile conditionnel. la démonstration des résultats techniques sera explicitement donnée dans la cinquième section. Nous achevons ce chapitre par quelques exemples de probabilités de petites boules.

2.1 La fonction de répartition conditionnelle

La régression est certes l'outil le plus utilisé et le plus étudié en estimation, mais elle présente quelques inconvénients. Le premier obstacle qu'elle rencontre est celui des données aberrantes. En effet, la moyenne étant un opérateur sensible aux valeurs atypiques, l'estimation de la régression manque de robustesse dans certaines études, où par exemple elle est face à une distribution multimodale ou encore une distribution asymétrique. Son deuxième inconvénient, et peut-être le plus important, est que la régression se contente de résumer les données en une seule information, et ne peut décrire le comportement du processus au complet. Pour ce genre de besoins, la recherche de la fonction de répartition est plus adaptée que celle de la régression. En effet, la fonction de répartition, tout comme la fonction de densité de probabilité et la fonction caractéristique, a l'avantage de caractériser totalement la loi de la distribution de Y . L'estimation de la fonction de répartition est très utile par exemple pour estimer les chances qu'une observation appartienne à un intervalle donné, ou encore que la valeur à estimer ne dépasse pas un seuil fixé. Elle est aussi très pratique dans l'étude de l'estimation des quantiles.

2.1.1 Contexte fini-dimensionnel

La fonction de répartition de Y étant définie par :

$$F(y) = \mathbb{P}(Y \leq y) ,$$

son estimateur le plus évident n'est autre que "la fonction de répartition empirique" définie par :

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

Mais même si cet estimateur rudimentaire est sans biais et de variance minimale, et qu'il présente donc de bonnes propriétés de convergence uniforme, il a le défaut de ne pas

prendre en compte les structures topologiques de l'espace sur lequel l'étude est effectuée. Le deuxième estimateur logique qu'on pense à introduire a été déduit de l'estimateur à noyau de la densité de probabilité introduit par Rosenblatt et Parzen. Ainsi, l'estimateur à noyau de la fonction de répartition de Y est défini par :

$$\widehat{F}(y) = \frac{1}{n} \sum_{i=1}^n H(h^{-1}(y - Y_i)) ,$$

où $H(x) = \int_{-\infty}^x K(y)dy$ et K est une fonction noyau. La fonction de répartition de Y conditionnelle à X est donnée par :

$$F^x(y) = \mathbb{P}(Y \leq y / X = x) \quad (2.1)$$

Elle peut également être définie comme suit :

$$F^x(y) = \int_{-\infty}^y f^x(t)dt ,$$

où $f^x(y)$ est la densité de probabilité de Y sachant X définie par :

$$f^x(y) = \frac{f(x, y)}{f(x)} , \quad (2.2)$$

de telle sorte que : $f(x, y)$ est la densité de probabilité conjointe du couple (X, Y) et $f(x)$ est la densité de probabilité de la variable X .

Dans le contexte fonctionnel, l'étude de la fonction de répartition conditionnelle n'est pas une extension directe de celle du contexte multivarié. Plus précisément, si le théorème de Jirina nous assure l'existence d'une distribution conditionnelle dans des espaces mesurables, cela n'est pas toujours le cas dans des espaces fonctionnels, où aucune mesure dominante n'est reconnue universellement. Nous nous devons donc de préciser que dans ce qui suit, il est implicitement supposé que la loi conditionnelle de Y sachant X est bien définie.

2.1.2 Contexte fonctionnel

Nous nous intéressons dans ce mémoire dans le cas où \mathcal{F} est fonctionnel. Plus précisément, nous considérons un couple aléatoire (X, Y) où Y est à valeurs dans \mathbb{R} et X est à valeurs dans un espace de dimension infinie $(\mathcal{F}, d(.,.))$ tel que \mathcal{F} est un espace semi-métrique et d est la semi-métrique associée. Nous disons que X est une variable aléatoire fonctionnelle.

À partir d'un échantillon de couples indépendants (X_i, Y_i) , ayant chacun la même distribution que (X, Y) , l'estimateur de la fonction de répartition conditionnelle a été introduit par Ferraty et al. [9]. Pour ce faire, ils se sont inspirés d'une adaptation de l'estimateur introduit par Roussas [23] pour définir un estimateur de (2.1) comme suit :

$$\widehat{F}^x(y) = \frac{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))H(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))} \quad (2.3)$$

Où $h_K = h_{K,n}$ (resp. $h_H = h_{H,n}$) est une suite de nombres réels positifs.

Notations et hypothèses

Avant de lister les conditions nécessaires à établir le premier résultat, nous avons besoin de quelques notations supplémentaires. Désignons par \mathcal{V}_x un voisinage fixé de x , S est un sous-ensemble compact de \mathbb{R} , et notons par $B(x, h_K) = \{x' \in \mathcal{F} / d(x, x') < h_K\}$, nous désignerons par C, C' et C_x des constantes génériques dans \mathbb{R}^{*+} .

$$(H.1) \quad \mathbb{P}(X \in B(x, h_K)) = \phi_x(h_K) > 0.$$

$$(H.2) \quad \forall (y_1, y_2) \in S \times S, \forall (x_1, x_2) \in \mathcal{V}_x \times \mathcal{V}_x,$$

$$|F^{x_1}(y_1) - F^{x_2}(y_2)| \leq C_x(d(x_1, x_2)^{b_1} + |y_1 - y_2|^{b_2}), \quad \text{où } b_1, b_2 \text{ sont des constantes positives.}$$

$$(H.3) \quad \begin{cases} \forall (y_1, y_2) \in \mathbb{R}^2, & |H(y_1) - H(y_2)| \leq C|y_1 - y_2|, \\ \int |t|^{b_2} H^{(1)}(t) < \infty. \end{cases}$$

(H.4) K est une fonction avec support $[0, 1]$ telle que $0 < C_1 < K(t) < C_2 < \infty$.

$$(H.5) \quad \lim_{n \rightarrow \infty} h_K = 0, \quad \lim_{n \rightarrow \infty} \frac{\log n}{n\phi_x(h_K)} = 0.$$

(H.6) $\lim_{n \rightarrow \infty} h_H = 0$ et pour tout $\alpha > 0$ on a $\lim_{n \rightarrow \infty} n^\alpha h_H = 0$.

Remarque 2.1.1 *L'hypothèse (H.1) est une hypothèse de concentration de la variable fonctionnelle et elle joue un rôle important. Ce genre de condition est lié à la semi-métrique $d(\cdot, \cdot)$; elle quantifie et contrôle les probabilités des petites boules. Tandis que (H.2) est une hypothèse de régularité et elle est nécessaire pour évaluer le terme de biais. (H.3) sur le noyau H est une condition technique nécessaire pour simplifier les preuves. Pour ce qui concerne l'hypothèse (H.4), ce noyau contient les noyaux standards utilisés dans la littérature et cette condition est aussi une hypothèse technique. (H.5)-(H.6) Sur les paramètres de lissage sont classiquement utilisées dans le cas de dimension infinie pour combiner entre le biais et la variance.*

Le résultat

Nous sommes maintenant en mesure d'énoncer notre premier résultat qui est la convergence presque-complète de l'estimateur de la fonction de répartition conditionnelle.

Théorème 2.1.1 : *Sous les hypothèses (H.1)-(H.6), nous avons :*

$$\sup_{y \in S} |\widehat{F}^x(y) - F^x(y)| = O(h_K^{b_1}) + O(h_H^{b_2}) + O\left(\sqrt{\frac{\log n}{n\phi_x(h_K)}}\right), \quad p.co. \quad (2.4)$$

Preuve du Théorème 2.1.1 : Pour $i = 1, \dots, n$, nous considérons les quantités

$K_i = K(h_K^{-1}d(x, X_i))$, $H_i = H(h_H^{-1}(y - Y_i))$, et définissons $\widehat{F}_N^x(y)$, et \widehat{F}_D^x par :

$$\widehat{F}_N^x(y) = \frac{1}{n\mathbb{E}K_1} \sum_{i=1}^n K_i H_i(y) \quad \widehat{F}_D^x = \frac{1}{n\mathbb{E}K_1} \sum_{i=1}^n K_i$$

Cette preuve est basée sur la décomposition

$$\begin{aligned} \widehat{F}^x(y) - F^x(y) &= \frac{1}{\widehat{F}_D^x} \{(\widehat{F}_N^x(y) - \mathbb{E}\widehat{F}_N^x(y)) - (F^x(y) - \mathbb{E}\widehat{F}_N^x(y))\} \\ &+ \frac{F^x(y)}{\widehat{F}_D^x} \{\mathbb{E}\widehat{F}_D^x - \widehat{F}_D^x\} \end{aligned} \quad (2.5)$$

et sur les résultats techniques suivants pour lesquels les preuves sont données dans la section 2.5 :

Lemme 2.1.1 : *Sous les hypothèses (H.1) et (H.4)-(H.5), nous avons*

$$\widehat{F}_D^x - \mathbb{E}\widehat{F}_D^x = O\left(\sqrt{\frac{\log n}{n\phi_x(h_K)}}\right), \quad p.co., \quad (2.6)$$

Corollaire 2.1.1 : *Sous les hypothèses du Lemme 2.1.1, nous avons*

$$\sum_{n=1}^{\infty} \mathbb{P}(\widehat{F}_D^x < 1/2) < \infty. \quad (2.7)$$

Lemme 2.1.2 : *Sous les hypothèses (H.1)-(H.5), nous avons*

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |F^x(y) - \mathbb{E}\widehat{F}_N^x(y)| = O(h_K^{b_1}) + O(h_H^{b_2}), \quad p.co. \quad (2.8)$$

Lemme 2.1.3 : *Sous les hypothèses (H.1)-(H.6), nous avons*

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{F}_N^x(y) - \mathbb{E}\widehat{F}_N^x(y)| = O\left(\sqrt{\frac{\log n}{n\phi_x(h_K)}}\right) \quad p.co. \quad (2.9)$$

2.2 La densité conditionnelle et ses dérivées

Comme la densité de probabilité est obtenue en dérivant la fonction de répartition, et qu'inversement l'intégrale de la densité donne la fonction de répartition, il est vrai que la connaissance de l'une implique automatiquement celle de l'autre. Mais la fonction de densité a un net avantage sur le plan visuel, puisqu'elle donne un aperçu concret de quelques caractéristiques que l'autre fonction décrit d'une façon moins explicite, comme la symétrie, les pics ou les creux.

2.2.1 Contexte fini-dimensionnel

Dans le cas où \mathcal{F} est un intervalle de \mathbb{R} , [27] furent les premiers à définir un estimateur non-paramétrique pour la fonction de densité de probabilité, qui n'est autre que "l'Histogramme". Pour construire leur estimateur, ils ont découpé \mathcal{F} en classes C_j de largeurs l_j , $j = 1, \dots, k$. Ils ont alors proposé d'approcher la densité de probabilité f par une fonction en escalier définie sur la classe C_j qui contient le point y par :

$$f_T(y) = \sum_{j=1}^k \frac{p_j}{l_j} \mathbb{1}_{C_j}(y),$$

où p_j est la fréquence des points Y_i dans la classe C_j définie par :

$$p_j = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_j}(Y_i).$$

Cet estimateur présente le même inconvénient que celui du regressogramme. À savoir, le choix arbitraire de la largeur de la classe C_j fait que l'estimateur ne prend pas forcément en compte les points les plus proches de la valeur y où l'on effectue l'estimation. L'histogramme mobile est en ce sens plus approprié que son prédécesseur, puisqu'il est défini sur

une classe C qui est choisie de manière à ce que y soit en son centre, *i.e.* $C = [y - h, y + h]$, où h est une largeur à fixer. L'histogramme mobile est donné par :

$$f_m(y) = \frac{1}{2hn} \sum_{i=1}^n \mathbb{1}_{|y-Y_i| \leq h}.$$

Ce dernier a encore été amélioré par [18] qui a remplacé la fonction discrète $\frac{1}{2}\mathbb{1}_{[-1,1]}(h^{-1}(y - Y_i))$ par une fonction continue décroissante $K(h^{-1}(y - Y_i))$. C'est alors qu'est né l'estimateur non paramétrique de la densité le plus répandu, qui est donné par :

$$f_p(y) = \frac{1}{nh} \sum_{i=1}^n K(h^{-1}(y - Y_i)).$$

Sa version multivariée (cas où \mathcal{F} est un espace de dimension p) est définie comme suit :

$$f_c(y) = \frac{1}{nh^p} \sum_{i=1}^n K(h^{-1}(y - Y_i)).$$

2.2.2 Contexte fonctionnel

En s'inspirant de la définition de la densité conditionnelle de Y sachant X donnée dans (2.2), nous nous intéressons maintenant à l'estimation de la dérivée d'ordre j de la densité conditionnelle $f^{x(j)}$ de Y sachant $X = x$. Nous proposons de définir l'estimateur $\hat{f}^{x(j)}$ de $f^{x(j)}$ comme suit :

$$\hat{f}^{x(j)}(y) = \frac{h_H^{-j-1} \sum_{i=1}^n K(h_K^{-1}d(x, X_i)) H^{(j+1)}(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))}. \quad (2.10)$$

Une estimation similaire a été déjà introduite dans le cas particulier où X est une variable aléatoire réelle par Rosenblatt en 1969. Il a été aussi largement étudié par Youndjé [29].

Notations et hypothèses

Afin d'établir la convergence presque-complète de la $j^{\text{ème}}$ dérivée de la densité conditionnelle, nous avons besoin des hypothèses supplémentaires suivantes :

$$(H.7) \quad \forall (y_1, y_2) \in S \times S, \forall (x_1, x_2) \in \mathcal{V}_x \times \mathcal{V}_x,$$

$$|f^{x_1(j)}(y_1) - f^{x_2(j)}(y_2)| \leq C_x(d(x, X_i)^{b_1} + |y_1 - y_2|^{b_2}).$$

$$(H.8) \quad \begin{cases} \forall (y_1, y_2) \in \mathbb{R}^2, & |H^{(j+1)}(y_1) - H^{(j+1)}(y_2)| \leq C|y_1 - y_2|, \\ \exists v > 0, \forall j' \leq j, & \lim_{y \rightarrow \infty} |y|^{1+v} |H^{(j'+1)}(y)| = 0. \end{cases}$$

$$(H.9) \quad \lim_{n \rightarrow \infty} h_K = 0, \quad \lim_{n \rightarrow \infty} \frac{\log n}{nh_H^{2j+1} \phi_x(h_K)} = 0.$$

Le résultat

Le résultat suivant concerne le comportement asymptotique de l'estimateur $\widehat{f}^{x(j)}$.

Théorème 2.2.1 : *Sous les hypothèses (H.1) et (H.4)-(H.9), nous avons :*

$$\sup_{y \in S} |\widehat{f}^{x(j)}(y) - f^{x(j)}(y)| = O(h_K^{b_1}) + O(h_H^{b_2}) + O\left(\sqrt{\frac{\log n}{nh_H^{2j+1} \phi_x(h_K)}}\right), \quad p.co. \quad (2.11)$$

Preuve du Théorème 2.2.1 : Cette preuve est basée sur la décomposition

$$\begin{aligned} \widehat{f}^{x(j)}(y) - f^{x(j)}(y) &= \frac{1}{\widehat{F}_D^x} \{(\widehat{f}_N^{x(j)}(y) - \mathbb{E}\widehat{f}_N^{x(j)}(y)) - (f^{x(j)}(y) - \mathbb{E}\widehat{f}_N^{x(j)}(y))\} \\ &+ \frac{f^{x(j)}(y)}{\widehat{F}_D^x} \{\mathbb{E}\widehat{F}_D^x - \widehat{F}_D^x\} \end{aligned} \quad (2.12)$$

où

$$\widehat{f}_N^{x(j)}(y) = \frac{1}{nh_H^{j+1} \mathbb{E}K_1} \sum_{i=1}^n K(h_K^{-1}d(x, X_i)) H^{(j+1)}(h_H^{-1}(y - Y_i))$$

et sur les résultats techniques suivants :

Lemme 2.2.1 : Sous les hypothèses (H.1),(H.4) et (H.7)-(H.9), nous avons :

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |f^{x(j)}(y) - \mathbb{E} \widehat{f}_N^{x(j)}(y)| = O(h_K^{b_1}) + O(h_H^{b_2}), \quad p.co. \quad (2.13)$$

Lemme 2.2.2 : Sous les hypothèses (H.1),(H.4) et (H.6)-(H.9), nous avons :

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{f}_N^{x(j)}(y) - \mathbb{E} \widehat{f}_N^{x(j)}(y)| = O\left(\sqrt{\frac{\log n}{nh_H^{2j+1} \phi_x(h_K)}}\right) \quad p.co. \quad (2.14)$$

Maintenant, soit $\widehat{f}^x = \widehat{f}^{x(0)}$, il est évident que le théorème précédent nous permet d'obtenir le corollaire suivant.

Corollaire 2.2.1 : Sous les hypothèses du Théorème 2.2.1 pour $j = 0$, et supposons que (H.1),(H.3),(H.4) et (H.6) sont satisfaites, nous avons

$$\sup_{y \in S} |\widehat{f}^x(y) - f^x(y)| = O(h_K^{b_1}) + O(h_K^{b_2}) + \left(\sqrt{\frac{\log n}{nh_H \phi_x(h_K)}}\right), \quad p.co. \quad (2.15)$$

2.3 Le mode conditionnel

2.3.1 Le modèle et son estimateur

Soit $\{(X_i, Y_i), i \geq 1\}$ des couples de variables aléatoires i.i.d., tel que X_i prend les valeurs dans l'espace semi-métrique $(\mathcal{F}, d(\cdot, \cdot))$ et $Y_i \in \mathbb{R}$.

On définit le mode conditionnel noté par θ comme suit :

$$f(\theta) = \sup_{y \in \mathbb{R}} f^x(y) \quad (2.16)$$

Soit S un ensemble compact tel que $S = [\theta - \xi, \theta + \xi]$ est choisi de telle sorte qu'il n'y ait qu'un unique mode θ . Cet estimateur est basé sur l'estimation fonctionnelle de la densité

conditionnelle. L'estimateur $\widehat{\theta}$ du mode conditionnel θ est défini comme :

$$\widehat{f}^x(\widehat{\theta}) = \sup_{y \in S} \widehat{f}^x(y) \quad (2.17)$$

L'estimateur $\widehat{\theta}$ n'est pas nécessairement unique.

2.3.2 Notations et hypothèses

Afin d'établir la convergence presque-complète de l'estimateur $\widehat{\theta}$, on introduit les hypothèses suivantes :

$$(H.10) \quad \exists \xi > 0, f^x \nearrow \text{ dans } [\theta - \xi, \theta] \text{ et } f^x \searrow \text{ dans } [\theta, \theta + \xi].$$

$$(H.11) \quad f^x \text{ est } j\text{-fois continûment différentiable par rapport à } y \text{ sur } [\theta - \xi, \theta + \xi].$$

$$(H.12) \quad \begin{cases} f^{x(l)}(\theta) = 0 \text{ si } 1 \leq l < j, \\ \text{et } |f^{x(j)}(\theta)| > 0. \end{cases}$$

Signalons que ces conditions ont une grande influence sur la vitesse de convergence de l'estimateur $\widehat{\theta}$ (dans le théorème ci-dessous).

Remarque 2.3.1 *L'hypothèse (H.10) a été supposée pour assuré l'unicité du mode conditionnel, tandis que l'hypothèses (H.11) est une condition de régularité qui caractérise l'espace fonctionnel de notre modèle, de plus elle est nécessaire pour évaluer le terme de biais qui apparait dans la vitesse de convergence. Pour ce qui concerne l'hypothèse (H.12), c'est une condition technique pour simplifier les preuves.*

2.3.3 Le résultat

Le théorème suivant donne un résultat de convergence presque-complète de l'estimateur du mode conditionnel :

Théorème 2.3.1 : Si les conditions du Théorème 2.2.1 sont vérifiées avec (H.10)-(H.12), nous avons :

$$\widehat{\theta} - \theta = O\left(h_{\frac{b_1}{K}}^{\frac{b_1}{j}}\right) + O\left(h_{\frac{b_2}{H}}^{\frac{b_2}{j}}\right) + O\left(\left(\sqrt{\frac{\log n}{nh_H\phi_x(h_K)}}\right)^{\frac{1}{j}}\right), \quad p.co.$$

Preuve du Théorème 2.3.1 : D'une part, nous avons

$$\begin{aligned} |f^x(\widehat{\theta}) - f^x(\theta)| &= |f^x(\widehat{\theta}) - \widehat{f}^x(\widehat{\theta}) + \widehat{f}^x(\widehat{\theta}) - f^x(\theta)| \\ &\leq |f^x(\widehat{\theta}) - \widehat{f}^x(\widehat{\theta})| + |\widehat{f}^x(\widehat{\theta}) - f^x(\theta)| \\ &\leq \sup_{y \in S} |f^x(y) - \widehat{f}^x(y)| + \sup_{y \in S} |\widehat{f}^x(y) - f^x(y)| \\ &\leq 2 \sup_{y \in S} |\widehat{f}^x(y) - f^x(y)| \end{aligned} \quad (2.18)$$

D'autre part, en utilisant le développement de *Taylor* de la fonction f^x :

$$f^x(\widehat{\theta}) = f^x(\theta) + \frac{1}{j!} f^{x(j)}(\theta^*)(\theta - \widehat{\theta})^j$$

Où θ^* se situe entre $\widehat{\theta}$ et θ . On a :

$$(\widehat{\theta} - \theta)^j = \frac{j!}{f^{x(j)}(\theta^*)} (f^x(\widehat{\theta}) - f^x(\theta))$$

Ainsi, d'après (2.18), il vient :

$$|\widehat{\theta} - \theta|^j \leq \frac{j!}{f^{x(j)}(\theta^*)} \sup_{y \in S} |\widehat{f}^x(y) - f^x(y)|$$

Nous aurions :

$$(\widehat{\theta} - \theta)^j = O\left(\sup_{y \in S} |\widehat{f}^x(y) - f^x(y)|\right), \quad p.co. \quad (2.19)$$

Nous concluons en combinant (2.19), le Théorème 2.2.1 et les deux résultats techniques suivants :

Lemme 2.3.1 : Si les conditions de Théorème 2.2.1 sont vérifiées avec (H.10), nous avons :

$$\widehat{\theta} - \theta \rightarrow 0, \quad p.co.$$

Corollaire 2.3.1 : En utilisant le résultat (2.18), on a :

$$\exists \tau > 0, \sum_{n=1}^{n=\infty} \mathbb{P}(f^{x(j)}(\theta^*) < \tau) < \infty.$$

2.4 Le quantile conditionnel

Définition 2.4.1 Soit (X, Y) un couple de variables aléatoires i.i.d, tels que X prend ses valeurs dans un espace semi-métrique $(\mathcal{F}, d(.,.))$ et $Y \in \mathbb{R}$. Pour $\alpha \in]0, 1[$, le quantile conditionnel d'ordre α noté $t_\alpha(x)$ est une solution de l'équation

$$t_\alpha(x) = \inf\{y \in \mathbb{R} / F^x(y) \geq \alpha\} \quad (2.20)$$

Le quantile conditionnel $t_\alpha(x)$ peut être défini aussi comme solution du problème de minimisation suivant :

$$t_\alpha(x) = \arg \min_{\theta \in \mathbb{R}} \mathbb{E}[\rho_\alpha(Y - \theta) | X = x].$$

où ρ_α est la fonction définie, pour tout $z \in \mathbb{R}$, par : $\rho_\alpha(z) = |z| + (2\alpha - 1)z$.

Dans ce mémoire, nous nous sommes intéressés par (2.20).

2.4.1 Le modèle et son estimateur

On définit le quantile conditionnel d'ordre α noté t_α , comme étant une solution de l'équation :

$$F^x(t_\alpha) = \alpha.$$

On estime le quantile conditionnel t_α par \widehat{t}_α tel que :

$$\widehat{F}^x(\widehat{t}_\alpha) = \alpha.$$

2.4.2 Notations et hypothèses

Pour obtenir la convergence presque-complète de l'estimateur de quantile conditionnel, on aura besoin des hypothèses suivantes :

(H.13) La fonction de répartition $F^x(\cdot)$ est strictement croissante.

(H.14) Le noyau H est strictement croissant.

(H.15) La fonction de répartition $F^x(\cdot)$ est j -fois continûment différentiable dans un voisinage de t_α .

$$(H.16) \left\{ \begin{array}{l} F^{x(l)}(t_\alpha) = 0 \text{ si } 1 \leq l < j, \\ \text{et } |F^{x(j)}(t_\alpha)| > 0. \end{array} \right.$$

Remarque 2.4.1 *Avant de passer à notre principal résultat de cette section, mentionnons que les hypothèses (H.13) et (H.14) ont été supposées pour assurer l'unicité du quantile conditionnel et son estimateur respectivement. Pour surmonter la difficulté qui apparaît lorsqu'on estime le quantile conditionnel t_α liée au lissage de la courbe de la fonction de répartition conditionnel F^x autour de t_α les conditions (H.15) et (H.16) ont été supposées.*

2.4.3 Le résultat

Le théorème suivant concerne le comportement asymptotique de l'estimateur du quantile conditionnel.

Théorème 2.4.1 : *Si les hypothèses du Théorème 2.2.1 sont satisfaites avec (H.13)-(H.16), nous avons :*

$$\hat{t}_\alpha - t_\alpha = O\left(h_K^{\frac{b_1}{j}}\right) + O\left(h_H^{\frac{b_2}{j}}\right) + O\left(\left(\sqrt{\frac{\log n}{n\phi_x(h_K)}}\right)^{\frac{1}{j}}\right) \quad (2.21)$$

Preuve du Théorème 2.4.1 En utilisant le développement de *Taylor* de la fonction \widehat{F}^x :

$$\begin{aligned}\widehat{F}^x(t_\alpha) - \widehat{F}^x(\widehat{t}_\alpha) &= \sum_{l=1}^{j-1} \frac{(t_\alpha - \widehat{t}_\alpha)^l}{l!} \widehat{F}^{x(l)}(t_\alpha) \\ &\quad + \frac{(t_\alpha - \widehat{t}_\alpha)^j}{j!} \widehat{F}^{x(j)}(t^*)\end{aligned}$$

Où t^* est entre t_α et \widehat{t}_α , et d'après la première partie de (H.16), on a :

$$\begin{aligned}\widehat{F}^x(t_\alpha) - \widehat{F}^x(\widehat{t}_\alpha) &= \sum_{l=1}^{j-1} \frac{(t_\alpha - \widehat{t}_\alpha)^l}{l!} \left(\widehat{f}^{(l-1)}(t_\alpha) - f^{x(l-1)}(t_\alpha) \right) \\ &\quad + \frac{(t_\alpha - \widehat{t}_\alpha)^j}{j!} \widehat{f}^{x(j-1)}(t^*).\end{aligned}\tag{2.22}$$

Remarquons que l'expression (2.22) implique que :

$$(t_\alpha - \widehat{t}_\alpha)^j = \frac{j!}{\widehat{f}^{x(j-1)}(t^*)} \left((\widehat{F}^x(t_\alpha) - F^x(t_\alpha)) - \left(\sum_{l=1}^{j-1} (t_\alpha - \widehat{t}_\alpha)^l (\widehat{f}^{x(l-1)}(t_\alpha) - f^{x(l-1)}(t_\alpha)) \right) \right).$$

Donc

$$\begin{aligned}(t_\alpha - \widehat{t}_\alpha)^j &= O(\widehat{F}^x(t_\alpha) - F^x(t_\alpha)) \\ &\quad + O\left(\sum_{l=1}^{j-1} (t_\alpha - \widehat{t}_\alpha)^l (\widehat{f}^{x(l-1)}(t_\alpha) - f^{x(l-1)}(t_\alpha)) \right), \quad p.co.\end{aligned}\tag{2.23}$$

En comparant le taux de convergence donné dans les Théorème 2.1.1 et 2.2.1, nous constatons que le terme dominant dans l'équation (2.23) est le premier terme. Donc nous avons :

$$(t_\alpha - \widehat{t}_\alpha)^j = O(\widehat{F}^x(t_\alpha) - F^x(t_\alpha)), \quad p.co.\tag{2.24}$$

Par conséquent du Théorème 2.2.1, il suffit de montrer les deux résultats techniques suivants :

Lemme 2.4.1 *Si les conditions du Théorème 2.1.1 sont satisfaites avec (H.13) et (H.14), on a :*

$$\widehat{t}_\alpha - t_\alpha \rightarrow 0, \quad p.co.$$

Corollaire 2.4.1 *Sous les hypothèses (H.15) et (H.12), on a :*

$$\exists \tau > 0, \sum_{n=1}^{n=\infty} \mathbb{P}(\widehat{f}^{x(j-1)}(t^*) < \tau) < \infty$$

2.5 Preuve des résultats techniques

2.5.1 Démonstrations relatives à la section 2.1

preuve du Lemme 2.1.1 On a :

$$\begin{aligned} \widehat{F}_D^x - \mathbb{E}\widehat{F}_D^x &= \frac{1}{n} \sum_{i=1}^n \left(\frac{K_i}{\mathbb{E}K_i} - \mathbb{E} \left(\frac{K_i}{\mathbb{E}K_i} \right) \right) \\ &= \frac{1}{n} \sum_{i=1}^n \left(\frac{K_i}{\mathbb{E}K_1} - 1 \right) \\ &= \frac{1}{n} \sum_{i=1}^n (\widetilde{\Delta}_i - 1) \end{aligned}$$

Telle que $\forall i = 1 \dots n$, $\widetilde{\Delta}_i = \frac{K_i}{\mathbb{E}K_1}$, les $\widetilde{\Delta}_i$ sont des variables indépendantes et identiquement distribuées centrées.

D'après (H.1) et (H.5), on a : $C\phi_x(h_K) < \mathbb{E}K_1 < C'\phi_x(h_K)$, donc nous pouvons obtenir directement :

$$|\widetilde{\Delta}_i| < C/\phi_x(h_K) = \delta_1 \quad \text{et} \quad \mathbb{E}|\widetilde{\Delta}_i|^2 < C'/\phi_x(h_K) = \delta_2.$$

Ainsi, on applique le Lemme de Hoeffding [13] pour tout $\eta \in [0, \delta_2/\delta_1]$, on aura :

$$\mathbb{P}\left(\widehat{F}_D^x - \mathbb{E}\widehat{F}_D^x > \eta\sqrt{\frac{\log n}{n\phi_x(h_K)}}\right) = \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n(\widetilde{\Delta}_i - 1) > \eta\left(\sqrt{\frac{\log n}{n\phi_x(h_K)}}\right)\right)$$

On pose $\varepsilon = \eta\sqrt{\frac{\log n}{n\phi_x(h_K)}}$, il vient :

$$\begin{aligned} \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n(\widetilde{\Delta}_i - 1) > \varepsilon\right) &\leq 2\exp\left(\frac{-n\varepsilon^2}{4\delta_2}\right) \\ &\leq 2\exp\left(\frac{-n\left(\frac{\eta^2 \log n}{n\phi_x(h_K)}\right)}{4C'}\right) \\ &\leq 2\exp\left(\log n^{-\eta^2/4C'}\right) \\ &\leq 2n^{-\eta^2/4C'} \\ &\leq C'n^{-C\eta^2} \end{aligned}$$

■

Preuve du Corollaire 2.1.1 : Nous avons :

$$\mathbb{P}(|\widehat{F}_D^x| \leq 1/2) \leq \mathbb{P}(|\widehat{F}_D^x - 1| > 1/2).$$

On a $\widehat{F}_D^x - 1 = \widehat{F}_D^x - \mathbb{E}\widehat{F}_D^x$ et d'après le preuve du lemme 2.1.1 on a $\mathbb{P}(|\widehat{F}_D^x - 1|) < \infty$.

Donc $\mathbb{P}(|\widehat{F}_D^x| < 1/2) < \infty$. ■

Preuve du Lemme 2.1.2 : On a :

$$\mathbb{E}\widehat{F}_N^x(y) - F^x(y) = \frac{1}{\mathbb{E}K_1}\mathbb{E}\sum_{i=1}^n(K_i H_i(y)) - F^x(y) \quad (2.25)$$

En conditionnant par X , on obtient :

$$\mathbb{E}\widehat{F}_N^x - F^x(y) = \frac{1}{\mathbb{E}K_1}\mathbb{E}(K_1[\mathbb{E}(H_1(y)|X) - F^x(y)]).$$

D'une part, nous avons :

$$\mathbb{E}(H_1(y)|X) = \int_{-\infty}^{+\infty} H(h_H^{-1}(y-z))f^X(z)dz,$$

Une intégration par partie conduit à :

$$\begin{aligned} \mathbb{E}(H_1(y)|X) &= [H(h_H^{-1}(y-z))F^X(z)]_{-\infty}^{+\infty} + \frac{1}{h_H} \int_{-\infty}^{+\infty} H^{(1)}(h_H^{-1}(y-z))F^X(z)dz \\ &= \frac{1}{h_H} \int_{-\infty}^{+\infty} H^{(1)}(h_H^{-1}(y-z))F^X(z)dz \end{aligned}$$

Maintenant, en utilisant le changement de variable $t = \frac{y-z}{h_H}$, on trouve :

$$|\mathbb{E}(H_1(y)|X) - F^x(y)| \leq \int_{-\infty}^{+\infty} H^{(1)}(t)|F^X(y-h_H t) - F^x(y)|dt.$$

D'autre part, d'après (H.2), on a :

$$|\mathbb{E}(H_1(y)|X) - F^x(y)| \leq C_x \int_{-\infty}^{+\infty} H^{(1)}(t)(h_K^{b_1} + |t|^{b_2} h_H^{b_2})dt. \quad (2.26)$$

En utilisant la condition (H.3), (2.13) est une conséquence directe de (2.25) et (2.26) et du Corollaire 2.1.1.

Donc

$$\frac{1}{\widehat{F}_D^x} \left| \sup_{y \in S} |\widehat{F}^x(y) - \mathbb{E}\widehat{F}_N^x(y)| \right| = O(h_K^{b_1}) + O(h_H^{b_2})$$

■

Preuve du Lemme 2.1.3 : L'idée est de recouvrir le compact S par un nombre fini d'intervalles S_k de longueurs égales. On note τ_n le nombre de ces intervalles, l_n la longueur

de chacun d'entre eux et t_k leurs centres. On a : $S \subset \bigcup_{k=1}^{k=\tau_n} S_k$ où $S_k =]t_k - l_n, t_k + l_n]$. On

pose $t_y = \arg \min_{t \in t_1, \dots, t_{\tau_n}} |y - t|$, nous avons :

$$\begin{aligned} \frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{F}_N^x(y) - \mathbb{E}\widehat{F}_N^x(t_y)| &\leq \underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{F}_N^x(y) - \widehat{F}_N^x(t_y)|}_{T_1} + \underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{F}_N^x(t_y) - \mathbb{E}\widehat{F}_N^x(t_y)|}_{T_2} + \\ &\quad \underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\mathbb{E}\widehat{F}_N^x(t_y) - \mathbb{E}\widehat{F}_N^x(y)|}_{T_3}. \end{aligned} \tag{2.27}$$

Concernant (T_1) : d'après la condition (H.8), on a :

$$\begin{aligned} \frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{F}_N^x(y) - \widehat{F}_N^x(t_y)| &\leq \frac{1}{\widehat{F}_D^x} \sup_{y \in S} \frac{1}{n\mathbb{E}k_1} \sum_{i=1}^n |H_i(y) - H_i(t_y)| K_i \\ &\leq \frac{1}{\widehat{F}_D^x} \sup_{y \in S} \frac{C|y - t_y|}{h_H} \left(\frac{1}{n\mathbb{E}K_1} \sum_{i=1}^n K_i \right) \\ &\leq C \frac{l_n}{h_H} \end{aligned} \tag{2.28}$$

On pose maintenant $l_n = n^{-\alpha-1/2}$

D'après (H.6), nous avons :

$$l_n/h_H = o\left(\sqrt{\frac{\log n}{n\phi_x(h_K)}}\right)$$

Ainsi, pour n assez grand, nous écrivons :

$$\mathbb{P}\left(\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{F}_N^x - \widehat{F}_N^x| > \frac{\eta}{3} \sqrt{\frac{\log n}{n\phi_x(h_K)}}\right) = 0 \tag{2.29}$$

Concernant (T_2) :

$$\begin{aligned} & \mathbb{P} \left(\sup_{y \in S} |\widehat{F}_N^x(t_y) - \mathbb{E}\widehat{F}_N^x(y)| > \frac{\eta}{3} \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) \\ &= \mathbb{P} \left(\max_{t_y \in \{t_1, \dots, t_{\tau_n}\}} |\widehat{F}_N^x(t_y) - \mathbb{E}\widehat{F}_N^x(t_y)| > \frac{\eta}{3} \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) \\ &\leq \tau_n \max_{t_y \in \{t_1, \dots, t_{\tau_n}\}} \mathbb{P} \left(|\widehat{F}_N^x(t_y) - \mathbb{E}\widehat{F}_N^x(t_y)| > \frac{\eta}{3} \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right). \end{aligned}$$

Soit $\Delta_i = \frac{1}{\mathbb{E}K_1} (K_i H_i(t_y) - \mathbb{E}(K_i H_i(t_y)))$. En utilisant des arguments similaires à ceux de la preuve du Lemme 2.1.1 et puisque $H \leq 1$, nous déduisons que $|\Delta_1| \leq C\phi_x(h_K)$. Nous appliquons maintenant l'inégalité exponentielle de Hoeffding [13] pour obtenir :

$$\mathbb{P} \left(|\widehat{F}_N^x(t_y) - \mathbb{E}\widehat{F}_N^x(t_y)| > \frac{\eta}{3} \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) \leq 2 \exp(-C\eta^2 \log n).$$

En choisissant η de telle sorte que $C\eta^2 = 3/2 + 2\alpha$, on obtient :

$$\begin{aligned} \tau_n \max_{t_y \in \{t_1, \dots, t_{\tau_n}\}} \mathbb{P} \left(|\widehat{F}_N^x(t_y) - \mathbb{E}\widehat{F}_N^x(t_y)| > \frac{\eta}{3} \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) &\leq C\tau_n n^{-3/2-2\alpha} \\ &\leq \frac{C}{l_n} n^{-3/2-2\alpha}. \end{aligned}$$

En gardant le même choix de $l_n = n^{-\alpha-1/2}$, nous déduisons de l'inégalité précédente que :

$$\mathbb{P} \left(\sup_{y \in S} |\widehat{F}_N^x(t_y) - \mathbb{E}\widehat{F}_N^x(t_y)| > \frac{\eta}{3} \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) \leq Cn^{-1-\alpha}.$$

Finalement, d'après le Corollaire 2.1.1, on obtient :

$$\mathbb{P} \left(\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{F}_N^x(t_y) - \mathbb{E}\widehat{F}_N^x(t_y)| > \frac{\eta}{3} \sqrt{\frac{\log n}{n\phi_x(h_K)}} \right) \leq Cn^{-1-\alpha}. \quad (2.30)$$

Concernant (T_3) : d'après (2.28), nous avons :

$$\sup_{y \in S} |\mathbb{E}\widehat{F}_N^x(t_y) - \mathbb{E}\widehat{F}_N^x(y)| \leq C \frac{l_n}{h_H}.$$

Et d'après T_1 , nous écrivons pour n assez grand :

$$\mathbb{P}\left(\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\mathbb{E}\widehat{F}_N^x(t_y) - \mathbb{E}\widehat{F}_N^x(y)| > \frac{\eta}{3} \sqrt{\frac{\log n}{n\phi_x(h_K)}}\right) = 0 \quad (2.31)$$

Donc la preuve du Lemme 2.1.3 est une conséquence de (2.27), (2.29), (2.30) et (2.31). ■

2.5.2 Démonstrations relatives à la section 2.2

Preuve du Lemme 2.2.1 Soit $H_i^{(j+1)}(y) = H^{(j+1)}(h_H^{-1}(y - Y_i))$, nous avons

$$\mathbb{E}\widehat{f}_N^{x(j)}(y) - f^{x(j)}(y) = \frac{1}{h_H^{j+1}\mathbb{E}K_1} \mathbb{E}(K_1 H_1^{(j+1)}(y)) - f^{x(j)}(y)$$

En conditionnant par X , on obtient :

$$\mathbb{E}\widehat{f}_N^{x(j)}(y) - f^{x(j)}(y) = \frac{1}{h_H^{j+1}\mathbb{E}K_1} \mathbb{E}(K_1 [\mathbb{E}(H_1^{(j+1)}(y)|X) - h_H^{j+1} f^{x(j)}(y)]). \quad (2.32)$$

Dans un premier temps, nous avons :

$$\begin{aligned} \mathbb{E}(H_1^{(j+1)}(y)|X) &= \int_{\mathbb{R}} H^{(j+1)}(h_H^{-1}(y - z)) f^X(z) dz \\ &= - \sum_{l=1}^j h_H^l [H^{(j-l+1)}(h_H^{-1}(y - z)) f^{X^{(l-1)}}(z)]_{-\infty}^{+\infty} \\ &\quad + h_H^j \int_{\mathbb{R}} H^{(1)}(h_H^{-1}(y - z)) f^{X^{(j)}}(z) dz \end{aligned}$$

Donc

$$\mathbb{E}(H_1^{(j+1)}(y)|X) = h_H^j \int_{\mathbb{R}} H^{(1)}(h^{-1}(y - z)) f^{X^{(j)}}(z) dz \quad (2.33)$$

D'autre part le changement de variable $t = (y - z)/h_H$ permet d'écrire :

$$|\mathbb{E}(H_1^{(j+1)}(y)|X) - h_H^{j+1} f^{x(j)}(y)| \leq h_H^{j+1} \int_{\mathbb{R}} H^{(1)}(t) |f^{X(j)}(y - h_H t) - f^{x(j)}(y)| dt.$$

Finalement, d'après (H.7), on a :

$$|\mathbb{E}(H_1^{(j+1)}(y)|X) - h_H^{j+1} f^{x(j)}(y)| \leq C_x h_H^{j+1} \int_{\mathbb{R}} H^{(1)}(t) (h_K^{b_1} + |t|^{b_2} (h_H^{b_2})) dt.$$

La condition (H.3), L'équation (2.32) et le Corollaire 2.1.1 achève le résultat. \blacksquare

Preuve du Lemme 2.2.2 : En utilisant la même idée de recouvrir le compact S comme la preuve du Lemme 2.1.3, on a :

$$\begin{aligned} \frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{f}_N^x(y) - \mathbb{E}\widehat{f}_N^x(y)| &\leq \underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{f}_N^x(y) - \widehat{f}_N^x(t_y)|}_{T_1} + \underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{f}_N^x(t_y) - \mathbb{E}\widehat{f}_N^x(t_y)|}_{T_2} \\ &\quad + \underbrace{\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\mathbb{E}\widehat{f}_N^x(t_y) - \mathbb{E}\widehat{f}_N^x(y)|}_{T_3}. \end{aligned} \tag{2.34}$$

Concernant (T_1) : d'après (H.8), on a :

$$\begin{aligned} \frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{F}_N^x(y) - \widehat{F}_N^x(t_y)| &\leq \frac{1}{\widehat{F}_D^x} \sup_{y \in S} \frac{1}{nh_H^{j+1} \mathbb{E}K_1} \sum_{i=1}^n |H_i^{(j+1)}(y) - H_i^{(j+1)}(t_y)| K_i \\ &\leq \frac{1}{\widehat{F}_D^x} \sup_{y \in S} \frac{C|y - t_y|}{h_H} \left(\frac{1}{nh_H^{j+1} \mathbb{E}K_1} \sum_{i=1}^n K_i \right) \\ &\leq C \frac{l_n}{h_H^{j+2}} \end{aligned} \tag{2.35}$$

Cette fois-ci, on pose $l_n = n^{-\frac{3\alpha}{2} - \frac{1}{2}}$, l'hypothèse (H.6) et la deuxième partie de (H.9) permet d'écrire :

$$l_n/h_H^{j+2} = o\left(\sqrt{\frac{\log n}{nh_H^{2j+1} \phi_x(h_K)}}\right) \tag{2.36}$$

Ainsi, pour n assez grand :

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{f}_N^{x(j)}(y) - \widehat{f}_N^{x(j)}(t_y)| = O\left(\sqrt{\frac{\log n}{nh_H^{2j+1}\phi_x(h_H)}}\right).$$

Donc

$$\mathbb{P}\left(\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{f}_N^{x(j)}(y) - \widehat{f}_N^{x(j)}(t_y)| > \frac{\eta}{3} \sqrt{\frac{\log n}{nh_H^{2j+1}\phi_x(h_K)}}\right) = 0. \quad (2.37)$$

Concernant (T_2) : On commence par

$$\widehat{f}_N^{x(j)}(t_y) - \mathbb{E}\widehat{f}_N^{x(j)}(t_y) = \frac{1}{n} \sum_{i=1}^n \left[\frac{H_i^{(j+1)}(t_y)K_i}{h_H^{j+1}\mathbb{E}K_1} - \frac{\mathbb{E}(H_i^{(j+1)}(t_y)K_i)}{h_H^{j+1}\mathbb{E}K_1} \right] \quad (2.38)$$

$$\text{On pose} \quad \Delta_i^* = \frac{H_i^{(j+1)}(t_y)K_i}{h_H^{j+1}\mathbb{E}K_1} - \frac{\mathbb{E}(H_i^{(j+1)}(t_y)K_i)}{h_H^{j+1}\mathbb{E}K_1},$$

On a $|\Delta_i^*| \leq Ch_H^{-j-1}\phi_x(h_K)^{-1}$. Maintenant, nous montrons que :

$$\mathbb{E}\Delta_i^{*2} = O(h^{-2j-1}\phi_x(h_K)^{-1}). \quad (2.39)$$

Premièrement, nous pouvons écrire :

$$\mathbb{E}\Delta_i^{*2} \leq \mathbb{E}\left(\frac{\mathbb{E}(H_i^{(j+1)2}(t_y)K_i^2)}{h_H^{j+1}\mathbb{E}K_1}\right) \quad \text{et} \quad \mathbb{E}\left(H_1^{(j+1)2}(t_y)K_1^2\right) = \mathbb{E}\left(K_1^2\mathbb{E}(H_1^{(j+1)2}(t_y)|X)\right).$$

L'hypothèse (H.8) implique que :

$$\begin{aligned}
\left| \frac{1}{h_H} \mathbb{E}(H_1^{(j+1)^2}(t_y)|X) - f^X(t_y) \int_{\mathbb{R}} H^{(j+1)^2}(y) dy \right| &= \left| \int_{\mathbb{R}} \frac{1}{h_H} H^{(j+1)^2} \left(\frac{u}{h_H} \right) (f^X(t_y - u) - f^X(t_y)) du \right| \\
&\leq \left| \int_{|u| \leq A} \frac{1}{h_H} H^{(j+1)^2} \left(\frac{u}{h_H} \right) (f^X(t_y - u) - f^X(t_y)) du \right| \\
&\quad + \left| \int_{|u| > A} \frac{1}{h_H} H^{(j+1)^2} \left(\frac{u}{h_H} \right) (f^X(t_y - u) - f^X(t_y)) du \right| \\
&\leq \underbrace{C \sup_{|u| \leq A} |f^X(t_y - u) - f^X(t_y)|}_{B_1} + \underbrace{\sup_{|y| \leq A/h_H} H^{(j+1)^2}(y)}_{B_2} \\
&\quad + \underbrace{f^X(t_y) \int_{|y| \leq A/h_H} H^{(j+1)^2}(y) dy}_{B_3}.
\end{aligned}$$

Nous déduisons d'après (H.8) que :

$$\forall \varepsilon > 0, \forall A > 0, \exists n_{A,\varepsilon}, \forall n \geq n_{A,\varepsilon}, B_2 + B_3 < \varepsilon.$$

De plus, comme f^x est continue, on a :

$$\forall \varepsilon > 0, \exists A > 0, \forall n \leq n_{A,\varepsilon}, B_1 < \varepsilon.$$

Ainsi, nous obtenons :

$$\lim_{n \rightarrow \infty} \frac{1}{h_H} \mathbb{E} \left(H_1^{(j+1)^2}(t_y)^2 | X \right) = f^X(t_y) \int_{\mathbb{R}} H^{(j+1)^2}(y) dy.$$

Finalement, puisque $0 < C\phi_x(h_K) < \mathbb{E}K_1$ et $\mathbb{E}K_1^2 < C'\phi_x(h_K)$, nous avons (2.39). En appliquant l'Inégalité de Hoeffding [13], on trouve :

$$\mathbb{P} \left(|\widehat{f}_N^{x(j)}(t_y) - \mathbb{E}\widehat{f}_N^{x(j)}(t_y)| > \eta \sqrt{\frac{\log n}{nh^{2j+1}\phi_x(h_K)}} \right) \leq 2 \exp \left\{ -\eta C \frac{\eta^2 \log n}{nh_H^{2j+1}\phi_x(h_K)} h_H^{2j+1} \phi_x(h_K) \right\}$$

Ce qui entraîne à :

$$\mathbb{P} \left(\left| \widehat{f}_N^{x(j)}(t_y) - \mathbb{E} \widehat{f}_N^{x(j)}(t_y) \right| > \eta \sqrt{\frac{\log n}{nh^{2j+1} \phi_x(h_K)}} \right) \leq C' n^{-C\eta^2}$$

Ainsi, d'une manière analogue à ce qui a été prouvé dans le Lemme 2.1.3, on en déduit que :

$$\mathbb{P} \left(\sup_{y \in S} \left| \widehat{f}_N^{x(j)}(t_y) - \mathbb{E} \widehat{f}_N^{x(j)}(t_y) \right| > \eta \sqrt{\frac{\log n}{nh_H^{2j+1} \phi_x(h_K)}} \right) \leq \frac{C'}{l_n} n^{-C\eta^2}.$$

En choisissant η de telle sorte que $C\eta^2 = 5\alpha/3 + 3/2$, on obtient :

$$\sup_{y \in S} \left| \widehat{f}_N^{x(j)}(t_y) - \mathbb{E} \widehat{f}_N^{x(j)}(t_y) \right| = O \left(\sqrt{\frac{\log n}{nh_H^{2j+1} \phi_x(h_K)}} \right).$$

Donc

$$\mathbb{P} \left(\frac{1}{\widehat{F}_D^x} \sup_{y \in S} \left| \widehat{f}_N^{x(j)}(t_y) - \mathbb{E} \widehat{f}_N^{x(j)}(t_y) \right| > \frac{\eta}{3} \sqrt{\frac{\log n}{nh_H^{2j+1} \phi_x(h_K)}} \right) = 0. \quad (2.40)$$

Concernant T_3 : d'après (2.35), on a :

$$\sup_{y \in S} \left| \mathbb{E} \widehat{f}_N^x(t_y) - \mathbb{E} \widehat{f}_N^x(y) \right| \leq C \frac{l_n}{h_H^{j+2}}$$

Et d'après T_1 , nous écrivons pour n assez grand :

$$\sup_{y \in S} \left| \mathbb{E} \widehat{f}_N^x(t_y) - \mathbb{E} \widehat{f}_N^x(y) \right| = O \left(\sqrt{\frac{\log n}{nh_H^{2j+1} \phi_x(h_H)}} \right).$$

Donc on a :

$$\mathbb{P} \left(\frac{1}{\widehat{F}_D^x} \sup_{y \in S} \left| \mathbb{E} \widehat{f}_N^x(t_y) - \mathbb{E} \widehat{f}_N^x(y) \right| > \frac{\eta}{3} \sqrt{\frac{\log n}{nh_H^{2j+1} \phi_x(h_K)}} \right) = 0. \quad (2.41)$$

Finalement, la preuve du Lemme 2.2.2 est une conséquence du Corollaire 2.1.1 et les équations (2.37), (2.40) et (2.41). ■

2.5.3 Démonstrations relatives à la section 2.3

Preuve du Lemme 2.3.1 : Puisque f^x est continue, d'après (H.7) et en utilisant (H.10), nous avons :

$$\forall \varepsilon > 0, \exists \delta(\varepsilon) > 0, \forall y \in (\theta - \xi, \theta + \xi), |f^x(y) - f^x(\theta)| \leq \delta(\varepsilon) \Rightarrow |y - \theta| \leq \varepsilon.$$

Comme $\widehat{\theta} \in (\theta - \xi, \theta + \xi)$, nous avons :

$$\exists \delta(\varepsilon) > 0, |f^x(\widehat{\theta}) - f^x(\theta)| \leq \delta(\varepsilon) \Rightarrow |\widehat{\theta} - \theta| \leq \varepsilon.$$

Donc, on obtient directement :

$$\forall \varepsilon > 0, \exists \delta(\varepsilon) > 0, \mathbb{P}(|\widehat{\theta} - \theta| > \varepsilon) \leq \mathbb{P}(|f^x(\widehat{\theta}) - f^x(\theta)| \geq \delta(\varepsilon)).$$

Ainsi, d'après (2.18), on a :

$$\mathbb{P}(|\widehat{\theta} - \theta| > \varepsilon) \leq 2 \sup_{y \in S} \mathbb{P}(|\widehat{f}^x(y) - f^x(y)| \geq \delta(\varepsilon))$$

Donc la preuve de ce lemme est une conséquence directe du Théorème 2.2.1. ■

Preuve du Corollaire 2.3.1 : La preuve de ce corollaire est une conséquence directe de la seconde partie de l'hypothèse (H.12). Donc on obtient directement :

$$\exists \tau > 0, \mathbb{P}(\widehat{f}^{x(j)}(\theta^*) < \tau) < \infty.$$

■

2.5.4 Démonstrations relatives à la section 2.4

Preuve du Lemme 2.4.1 : Puisque F^x est continue d'après (H.3) et selon (H.14), nous avons :

$$\forall \varepsilon > 0, \exists \delta(\varepsilon) > 0, \forall y, |F^x(y) - F^x(t_\alpha)| \leq \delta(\varepsilon) \Rightarrow |y - t_\alpha| \leq \varepsilon.$$

Cela conduit directement à :

$$\begin{aligned} \forall \varepsilon > 0, \exists \delta(\varepsilon) > 0, \mathbb{P}(|\hat{t}_\alpha - t_\alpha| > \varepsilon) &\leq \mathbb{P}(|\hat{F}^x(\hat{t}_\alpha) - \hat{F}^x(t_\alpha)| \geq \delta(\varepsilon)) \\ &= \mathbb{P}(|F^x(t_\alpha) - \hat{F}^x(t_\alpha)| \geq \delta(\varepsilon)) \end{aligned}$$

Donc la preuve de ce lemme est une conséquence directe du Théorème 2.1.1. ■

Preuve du Corollaire 2.4.1 : La fonction $f^{x(j-1)}$ est continue d'après la condition (H.15). Donc, la preuve de ce corollaire est une conséquence direct de la second partie de l'hypothèse (H.12), on obtient facilement :

$$\exists \tau > 0, \mathbb{P}(\hat{f}^{x(j-1)}(t^*) < \tau) < \infty. \quad \blacksquare$$

2.6 Quelques exemples de probabilités de petites boules

Nous avons déjà souligné que les vitesses de convergence dépendent de la quantité $P(d(x, X) < h_K)$. Ces vitesses peuvent être alors précisées en utilisant les résultats sur les probabilités de petites boules pour des processus standards. Dans ce paragraphe on donne l'expression de la mesure de probabilité de petite boule pour deux modèles différents de processus à temps continue. On va également donné pour chaque cas la vitesse de convergence de l'estimateur basé sur le processus en question.

2.6.1 Processus de diffusion

Considérons l'espace fonctionnel \mathcal{F} comme étant l'espace de Cameron-Martin¹ de l'espace $C([0, 1]; \mathbb{R}^p)$ ². Prenons comme métrique d celle associée à la norme infinie :

$$\|x\|_{sup} = \sup_{t \in [0, 1]} |x(t)| \quad , \quad \forall x \in C([0, 1]; \mathbb{R}^p) \quad ,$$

1. Soit E un espace localement convexe avec une mesure μ . [4] a défini l'espace de Cameron-Martin de E par $E^{CM} = \{x \in E / \mu(\cdot - x) \sim \mu(\cdot)\}$.

2. $C([0, 1]; \mathbb{R}^p)$ est l'ensemble des fonctions continues à valeurs dans \mathbb{R}^p définies sur $[0, 1]$.

et notons par P_w la mesure de probabilité de Wiener³ définie sur $C([0, 1]; \mathbb{R}^p)$.

[4] a montré que la mesure de Wiener d'une petite boule de probabilité centrée était de la forme

$$P_w(\|x\|_{sup} \leq h_K) \sim \frac{4}{\pi} e^{-\pi^2/8h_K^2}.$$

Il s'en suit, d'après la définition de l'espace de Cameron-Martin \mathcal{F} , que :

$$P_w(\|x - x_0\|_{sup} \leq h_K) \sim C_{x_0} \frac{4}{\pi} e^{-\pi^2/8h_K^2}, \quad \forall x_0 \in \mathcal{F}. \quad (2.42)$$

À présent, notons par w_t le processus de Wiener, et considérons le processus de diffusion X_t^{diff} suivant :

$$X_t^{diff} = \int_0^t \Theta(s, X^{diff}) ds + w_t,$$

où la fonction Θ est telle que l'équation précédente a une unique solution X^{diff} et elle vérifie l'hypothèse suivante :

$$\int_0^1 \Theta^2(s, X^{diff}) ds < \infty \quad p.s.$$

alors le processus X_t^{diff} a une mesure de probabilité absolument continue par rapport à la mesure P_w . Ainsi, sous l'hypothèse précédente, (2.42) implique que

$$\mathbb{P}(\|X^{diff} - x_0\|_{sup} \leq h_K) \sim C_{x_0} e^{-\pi^2/8h_K^2}.$$

Finalement, il est facile de voir que notre condition cruciale (H.1) est trivialement satisfaite pour le processus de diffusion en choisissant notre fonction sous la forme

$$\phi_x^{diff}(h_K) \sim C_x e^{-\pi^2/8h_K^2}$$

3. Soient $(t_i)_{i=1,k} \in [0, 1]$ tels que $0 < t_1 < t_2 < \dots < t_k \leq 1$, et $(A_i)_{i=1,k}$ une famille de Borels de $[0, 1]$. En considérant la densité $f_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}$, la mesure de Wiener. P_w est définie pour tout $X \in C([0, 1]; \mathbb{R}^p)$ par :

$$P_w(X(t_i) \in A_i, i = 1, k) = \int_{A_1} \dots \int_{A_k} f_{t_1}(x_1) f_{t_2-t_1}(x_2 - x_1) \dots f_{t_k-t_{k-1}}(x_k - x_{k-1}) dx_k \dots dx_1.$$

Il suffit alors de choisir les paramètres tels que

$$h_K \sim \eta_H (\log n)^{-1/2} \quad \text{pour } \eta_H \text{ assez grand}$$

$$\text{et } h_H \sim n^{-\varepsilon} \text{ pour } \varepsilon > 0$$

pour voir que les conditions (H.5) et (H.6) sont satisfaites. Finalement, nous arrivons à une vitesse de convergence de l'estimateur \widehat{F} dans notre Théorème 2.1.1 d'ordre

$$O_{p.co.} (\log n)^{-b_1/2}, \text{ p.co.}$$

2.6.2 Mouvement Brownien Fractionnel

Prenons maintenant l'espace $C([0, 1]; \mathbb{R})$, et \mathcal{F} l'espace de Cameron-Martin qui lui est associé. La métrique d utilisée ici est toujours celle associée à la norme infinie.

Considérons X^{MBF} le mouvement Brownien Fractionnel⁴ de paramètre α , $0 < \alpha < 2$.

Cette fois-ci, nous avons :

$$C_{x_0} e^{h_K^{-2/\alpha}} \leq \mathbb{P}(\|X^{MBF} - x_0\|_{sup} \leq h_K) \leq C'_{x_0} e^{h_K^{-2/\alpha}}, \forall x_0 \in \mathcal{F}, \quad (\text{voir Ferraty et al. [9]}).$$

Finalement, il est facile de voir que notre condition cruciale (H.1) est trivialement satisfaite pour le processus de mouvement Brownien Fractionnel en choisissant notre fonction sous la forme

$$\phi_x^{MBF}(h_K) \sim C_x e^{h_K^{-2/\alpha}}.$$

En choisissant ensuite les paramètres de lissage tels que

$$h_K \sim C(\log n)^{-\alpha/2} \quad h_H \sim n^{-\varepsilon},$$

4. Un mouvement Brownien Fractionnel β d'ordre α est un processus Gaussien centré tel que $\beta_0 = 0$ et $\forall t \neq s \in [0, 1], \mathbb{E}|\beta_t - \beta_s| = |t - s|^\alpha$.

pour voir que les conditions (H.5) et (H.6) sont vérifiées. Finalement, nous arrivons à une vitesse de convergence dans notre Théorème 2.1.1 d'ordre

$$O_{p.co.} (\log n)^{-\alpha b_1/2}, p.co.$$

Nous remarquons que pour les deux exemples citées, la mesure de concentration est de la forme :

$$\phi_x(h) \sim C_x e^{Ch^{-2\alpha}}.$$

Chapitre 3

Cas dépendant

Dans ce chapitre nous établissons la convergence presque-complète de l'estimateur à noyau du mode conditionnel défini dans (2.17) en considérant le cas où les observations sont dépendantes. Ce chapitre est organisé comme suit : on présente dans la première section les notations et les hypothèses nécessaires. Dans la deuxième section nous énonçons le résultat principal de ce chapitre. La troisième section est consacrée aux démonstrations des résultats élaborés dans la section précédente.

3.1 Notations et hypothèses

L'estimateur à noyau du mode conditionnel dans ce chapitre, est construit à l'aide d'un échantillon de variables fortement mélangées. Pour cela, on considère le modèle (2.16) et on garde les mêmes notations, ainsi, que les mêmes hypothèses de la section 2.3 du chapitre précédent (Cas i.i.d.) et on ajoute les conditions suivantes :

$$(M.1) \sup_{i \neq j} \mathbb{P}((X_i, X_j) \in B(x, r) \times B(x, r)) = \phi_x(r)\psi_x(r).$$

(M.2) Les coefficients de mélange de la suite (X_i, Y_i) vérifient la condition :

$$\exists a > (5 + \sqrt{17})/2, \exists c > 0, \forall n, \alpha_n \leq cn^{-a}.$$

$$(M.3) \quad \lim_{n \rightarrow \infty} h_H = 0 \text{ et } \exists \beta_1 \leq \frac{4}{(a+1)(a-2)} \text{ telle que } \lim_{n \rightarrow \infty} n^{\beta_1} h_H = \infty.$$

$$(M.4) \quad \begin{cases} \lim_{n \rightarrow \infty} h_K = 0, & \lim_{n \rightarrow \infty} \frac{\log n}{nh_H \chi_x(h_K)}, \\ \text{et } \exists \beta_2 > 0, c_1 > 0, c_2 > 0, c_2 n^{\left(\frac{3-a}{a+1}\right) + \beta_2} \leq \chi_x(h_K) \leq c_1 n^{\frac{1}{1-a}}. \end{cases}$$

Telle que la fonction χ_x est définie par $\chi_x = \max\{\phi_x(h), \psi_x(h)\}$.

Remarque 3.1.1 *la première condition concerne la distribution des deux couples distincts (X_i, Y_i) et (X_j, Y_j) c'est en fait une condition de concentration dans le cas dépendant. La deuxième condition précise le type de dépendance qui est la mélangeance algébrique. Les deux dernières hypothèses concernent les paramètres de lissage h_K et h_H .*

3.2 Le résultat

Nous sommes maintenant en mesure d'énoncer notre résultat principal de ce chapitre.

Théorème 3.2.1 *Sous les hypothèses (H.1), (H.3), (H.4), (H.7), (H.10)-(H.12) et (M.1)-(M.4), nous avons :*

$$\hat{\theta} - \theta = O\left(h_K^{\frac{b_1}{j}}\right) + O\left(h_H^{\frac{b_2}{j}}\right) + O\left(\left(\sqrt{\frac{\log n}{nh_H \chi_x(h_K)}}\right)^{\frac{1}{j}}\right), \quad p.co.$$

Preuve du Théorème 3.2.1 : Cette preuve est basée sur la décomposition (2.12), le Lemme 2.2.1, le Corollaire 2.1.1, le Corollaire 2.2.1 et les résultats techniques suivants :

Lemme 3.2.1 : *Sous les hypothèses (H.1), (H.3), (H.4) et (M.1)-(M.4), nous avons :*

$$\hat{F}_D^x - \mathbb{E}\hat{F}_D^x = O\left(\sqrt{\frac{\log n}{n\chi_x(h_K)}}\right), \quad p.co.$$

Lemme 3.2.2 : Sous les hypothèses (H.1), (H.3), (H.4), (H.7) et (M.1)-(M.4), nous avons :

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{f}_N^x(y) - \mathbb{E}\widehat{f}_N^x(y)| = O\left(\sqrt{\frac{\log n}{nh_H \chi_x(h_K)}}\right), \quad p.co.$$

Lemme 3.2.3 : Sous les hypothèses du Lemme 3.2.1, (H.10) et (H.12), nous avons :

$$\exists c > 0, \sum_{n=1}^{\infty} \mathbb{P}(f^{x(j)}(\theta^*) < c) < \infty.$$

3.3 Preuve des résultats techniques

Preuve du Lemme 3.2.1 : Nous avons :

$$\widehat{F}_D^x - \mathbb{E}\widehat{F}_D^x = \frac{1}{n} \sum_{i=1}^n \left(\frac{K_i}{\mathbb{E}K_1} - 1 \right) = \frac{1}{n\mathbb{E}K_1} \sum_{i=1}^n \widetilde{\Delta}_i.$$

où

$$\widetilde{\Delta}_i = K_i - \mathbb{E}(K_i)$$

L'outil principal de cette preuve est l'application de l'inégalité de Fuk Nageav [13]. Pour cela, nous devons calculer le comportement asymptotique de la quantité s_n^2 comme suit :

$$s_n^2 = \sum_{i=1}^n \sum_{j=1}^n |Cov(\widetilde{\Delta}_i, \widetilde{\Delta}_j)| = s_n^{2*} + \sum_{i=1}^n Var(\widetilde{\Delta}_i). \quad (3.1)$$

où

$$s_n^{2*} = \sum_{i \neq j}^n |Cov(\widetilde{\Delta}_i, \widetilde{\Delta}_j)|.$$

Premièrement, nous définissons les ensembles S_1 et S_2 comme suit :

$$S_1 = \{(i, j) \text{ telle que } 1 \leq j - i \leq m_n\}$$

$$S_2 = \{(i, j) \text{ telle que } m_n + 1 \leq j - i \leq n - 1\}$$

Où la suite m_n est choisie telle que $m_n \rightarrow \infty$. Et en denotant $J_{1,n}$ et $J_{2,n}$ comme les sommes des covariances sur les ensembles S_1 et S_2 , respectivement. Donc :

$$\begin{aligned} J_{1,n} &= \sum_{S_1} |Cov(\tilde{\Delta}_i, \tilde{\Delta}_j)| \\ &= \sum_{S_1} |Cov(K_i - \mathbb{E}(K_i), K_j - \mathbb{E}(K_j))| \\ &= \sum_{S_1} |\mathbb{E}(K_i K_j) - \mathbb{E}(K_i)\mathbb{E}(K_j)| \end{aligned}$$

D'après (H.1), (H.3) et (M.1), nous pouvons écrire :

$$J_{1,n} \leq Cn m_n (\phi_x(h_K) \psi_x(h_K) + (\phi_x(h_K))^2). \quad (3.2)$$

D'autre part, pour $J_{2,n}$ on a :

$$J_{2,n} = \sum_{S_2} |Cov(\tilde{\Delta}_i, \tilde{\Delta}_j)|$$

Pour étudier la somme sur S_2 , nous utilisons l'inégalité de Dvydov-Rio dans le cas L^∞ (ce qui est possible puisque $\tilde{\Delta}_i$ est bornée). Cela conduit, pour tous $i \neq j$ à :

$$|Cov(\tilde{\Delta}_i, \tilde{\Delta}_j)| \leq C\alpha(|i - j|)$$

Et donc nous obtenons :

$$J_{2,n} \leq n^2(\alpha(m_n)) \leq Cn^2 m_n^{-\alpha}. \quad (3.3)$$

En utilisant les deux résultats (3.2) et (3.3), puis en utilisant la définition de χ_x et finalement en choisissant :

$$m_n = \frac{1}{\chi_x(h_K)},$$

Nous pouvons écrire, pour n assez grand :

$$\begin{aligned}
s_n^{2*} &= J_{1,n} + J_{2,n} \\
&\leq C[nm_n\chi_x^2(h_K) + n^2m_n^{-\alpha}] \\
&\leq O(n\chi_x(h_K)).
\end{aligned} \tag{3.4}$$

Maintenant, pour tout $i = 1, \dots, n$ nous pouvons écrire :

$$\text{Var}(\tilde{\Delta}_i) = \mathbb{E}(K_i^2) - (\mathbb{E}(K_i))^2.$$

Et d'après (H.1), on a :

$$\text{Var}(\tilde{\Delta}_i) \leq C(\phi_x(h_K) + (\phi_x(h_K))^2).$$

Donc

$$\text{Var}(\tilde{\Delta}_i) = O(\phi_x(h_K)).$$

Enfin, ce dernier résultat combiné avec (3.1) et (3.4) conduit directement à :

$$s_n^2 = o(n\phi_x(h_K)) + O(n\phi_x(h_K)).$$

Ainsi, nous sommes maintenant en mesure d'appliquer l'inégalité classique de Fuk Nagaev pour obtenir : $\forall \lambda, r > 0$:

$$\begin{aligned}
\mathbb{P} \left\{ |F_D^x - \mathbb{E}F_D^x| > 4 \frac{\lambda}{n\mathbb{E}K_1} \right\} &\leq \mathbb{P} \left\{ |F_D^x - \mathbb{E}F_D^x| > 4 \frac{\lambda}{n\chi_x(h_K)} \right\} \\
&\leq A_1 + A_2,
\end{aligned}$$

Où $A_1 = 4 \left(1 + \frac{\lambda^2}{r s_n^2} \right)$ et $A_2 = 4 \frac{cn}{r} \left(\frac{r}{\lambda} \right)^{(a+1)}$.

En posant $\lambda = (\eta/4) \sqrt{n\chi_x(h_K) \log n}$, on aura :

$$A_2 = cnr^a \lambda^{-(a+1)} = cnr^a (n\phi_x(h_K)) (\log n)^{-(a+1)/2}$$

En prenant $r = C(\log n)^2$ il s'ensuit que :

$$A_2 = C(\log n)^{(2a)-(a+1)/2} n^{(1-(a+1)/2)} \chi_x(h_K)^{-(a+1)/2}.$$

Ensuite, d'après la deuxième partie de (M.4), on a :

$$\chi_x(h_K) \leq C_2 n^{-(\frac{3-a}{2}) - \beta_2(\frac{a+1}{2})}.$$

D'où

$$\begin{aligned} A_2 &\leq C(\log n)^{(2a-(a+1)/2)} n^{(1-(a+1)/2)} n^{-(\frac{3-a}{2} + \beta_2(\frac{a+1}{2}))} \\ &\leq C(\log n)^{(2a-(a+1)/2)} n^{-1 - \beta_2(\frac{a+1}{2})}. \end{aligned}$$

Donc finalement, $\exists \nu > 0$ telle que :

$$A_2 \leq C n^{-1-\nu} \tag{3.5}$$

Concernant A_1 on a :

$$\begin{aligned} A_1 &= 4 \left(1 + \frac{\eta^2 n \chi_x(h_K) \log n}{16 r s_n^2} \right)^{-r/2} \\ &\leq C \exp \left(-r/2 \log \left(1 + \frac{\eta^2 n \chi_x(h_K) \log n}{16 r s_n^2} \right) \right) \end{aligned}$$

En gardant le même choix de $r = C(\log n)^2$, nous pouvons écrire :

$$A_1 \leq C \exp \left(-\eta^2 \frac{\log n}{32} \right) = C n^{-\eta^2/32} \tag{3.6}$$

Finalement, avec une combinaison des résultats (3.5) et (3.6), $\exists \nu > 0$ telle que :

$$\mathbb{P} \left(\left| \widehat{F}_D^x - \mathbb{E} \widehat{F}_D^x \right| > \eta \sqrt{\frac{\log n}{n \phi_x(h_K)}} \right) \leq C' n^{-1-\nu}.$$

■

Preuve du Lemme 3.2.2 : En utilisant la même idée de recouvrir le compact S comme la preuve du Lemme 2.1.3 et en se basant sur la décomposition (2.34), on a :

Concernant T_1 : En suivant la même démarche que la preuve du Lemme 2.2.2 en prenant $j = 0$, on a d'après la définition de $\widehat{f}_N^x(\cdot)$:

$$\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{f}_N^x(y) - \widehat{f}_N^x(t_y)| \leq C \frac{l_n}{h_H^2}.$$

Maintenant on pose :

$$l_n = n^{-\frac{3}{2}\beta_1 - \frac{1}{2}}$$

Et d'après (M.3), nous avons :

$$l_n/h_H^2 = o(\sqrt{\log n(nh_H\chi_x(h_K))^{-1}}).$$

Ainsi, pour n assez grand, nous pouvons écrire :

$$\mathbb{P} \left(\frac{1}{\widehat{F}_D^x} \sup_{y \in S} |\widehat{f}_N^x(y) - \widehat{f}_N^x(t_y)| > \frac{\eta}{3} \sqrt{\frac{\log n}{nh_H\chi_x(h_K)}} \right) = 0 \quad (3.7)$$

Concernant T_2 :

$$\widehat{f}_N^x(t_y) - \mathbb{E}\widehat{f}_N^x(t_y) = \frac{1}{nh_H\mathbb{E}K_1} \sum_{i=1}^n \{H_i(t_y)K_i - \mathbb{E}(H_i(t_y)K_i)\}. \quad (3.8)$$

On pose $\Delta_i^* = \{H_i(t_y)K_i - \mathbb{E}(H_i(t_y)K_i)\}$. En suivant les même étapes que la preuve du Lemme 3.2.1, nous notons maintenant :

$$s_n'^2 = \sum_{i,j=1}^n |Cov(\Delta_i^*, \Delta_j^*)| = \sum_{i \neq j=1}^n |Cov(\Delta_i^*, \Delta_j^*)| + \sum_{i=1}^n Var(\Delta_i^*).$$

Tout d'abord, nous recherchons le comportement asymptotique de la quantité $s_n'^{2*}$ définie comme suit :

$$s_n'^{2*} = \sum_{i \neq j=1}^n |Cov(\Delta_i^*, \Delta_j^*)|.$$

Nous divisons la somme en deux ensembles définis par :

$$S_1^{*'} = \{(i, j) \text{ telle que } 0 \leq |i - j| \leq m_n'\}$$

et

$$S_2^{*'} = \{(i, j) \text{ telle que } m_n' + 1 \leq |i - j| \leq n - 1\}$$

Où $m_n' \rightarrow \infty$. Et en denotant $J_{1,n}'$ et $J_{2,n}'$ comme les sommes des covariances sur les ensembles $S_1^{*'}$ et $S_2^{*'}$, respectivement. Puisque H et K sont bornés, on a :

$$\begin{aligned} J_{1,n}' &= \sum_{S_1^{*'}} |Cov(\Delta_i^*, \Delta_j^*)| \\ &\leq \sum_{S_1^{*'}} (C(\mathbb{P}(X_i, X_j) \in B(x, h_{Kn}) \times B(x, h_{kn}))) + C\mathbb{P}(X_i \in B(x, h_{Kn}))\mathbb{P}(X_j \in B(x, h_K)) \\ &\leq Cnm_n'[\psi_x(h_K)\phi_x(h_K) + \phi_x^2(h_K)] \end{aligned} \quad (3.9)$$

De plus $\forall i \neq j$ on a $Cov(\Delta_i^*, \Delta_j^*) \leq C\alpha(|i - j|)$. Alors :

$$\begin{aligned} J_{2,n}' &= \sum_{S_2^{*'}} |Cov(\Delta_i^*, \Delta_j^*)| \\ &\leq \sum_{S_2^{*'}} (C\alpha(|i - j|)) \\ &\leq Cn^2\alpha(|i - j|) \\ &\leq Cn^2m_n'^{-\alpha}. \end{aligned} \quad (3.10)$$

On pose $m_n' = \frac{1}{h_H\chi_x(h_K)}$, et nous concluons comme pour la preuve du Lemme 3.2.1 que :

$$s_n'^2 = O(nh_H\chi_x(h_K)).$$

$$\begin{aligned}
& \mathbb{P} \left(\sup_{y \in S} |\widehat{f}_N^x(t_y) - \mathbb{E}\widehat{f}_N^x(t_y)| > \frac{\varepsilon}{3} \left(\frac{1}{C_1 n h_H \chi_x(h_K)} \right) \right) \\
&= \mathbb{P} \left(\max_{t_y \in \{t_1, \dots, t_{\tau_n}\}} |\widehat{f}_N^x(t_y) - \mathbb{E}\widehat{f}_N^x(t_y)| > \frac{\varepsilon}{3} \left(\frac{1}{C_1 n h_H \chi_x(h_K)} \right) \right) \\
&\leq \tau_n \max_{t_y \in \{t_1, \dots, t_{\tau_n}\}} \mathbb{P} \left(|\widehat{f}_N^x(t_y) - \mathbb{E}\widehat{f}_N^x(t_y)| > \frac{\varepsilon}{3} \left(\frac{1}{C_1 n h_H \chi_x(h_K)} \right) \right) \\
&\leq \tau_n \max_{t_y \in \{t_1, \dots, t_{\tau_n}\}} \mathbb{P} \left(|\widehat{f}_N^x(t_y) - \mathbb{E}\widehat{f}_N^x(t_y)| > \frac{\varepsilon}{3} \left(\frac{1}{C_1 n h_H \mathbb{E}K_1} \right) \right) \\
&\leq Cl_n(A'_1 + A'_2)
\end{aligned}$$

où

$$A'_1 = 4 \left(1 + \frac{\varepsilon^2}{r s_n'^2} \right)^{-r/2} \quad \text{et} \quad A'_2 = 4ncr^{-1} \left(\frac{r}{\varepsilon} \right)^{a+1}.$$

Pour obtenir le résultat on pose $\varepsilon = C_1 \eta \sqrt{\frac{\log n}{nh_H \chi_x(h_K)}}$ et $r = C(\log n)2$, et en utilisant le fait que $s_n'^2 = O(nh_H \chi_x(h_K))$. Le calcul de A'_1 et A'_2 est exactement le même que A_1 et A_2 dans la preuve du Lemme 3.2.1. Donc :

$$A'_1 = Cn^{-\eta/32} \quad \text{et} \quad A'_2 = Cn^{-1-\nu}$$

De plus en gardant le même choix de $l_n = n^{-\frac{3}{2}\beta_1 - \frac{1}{2}}$, on obtient pour η assez grand :

$$l_n^{-1}(A'_1 + A'_2) = Cn^{-1-\nu}$$

Finalement, on a :

$$\sup_{y \in S} |\widehat{f}_N^x(t_y) - \mathbb{E}\widehat{f}_N^x(t_y)| = O \left(\sqrt{\frac{\log n}{nh_H \chi_x(h_K)}} \right),$$

et donc

$$\mathbb{P} \left(\frac{1}{\widehat{F}_D^x} \sup_{y \in S} \left| \widehat{f}_N^x(t_y) - \mathbb{E} \widehat{f}_N^x(t_y) \right| > \frac{\eta}{3} \sqrt{\frac{\log n}{nh_H \chi_x(h_K)}} \right). \quad (3.11)$$

Concernant T_3 : D'après (H.3) et (H.4), nous avons :

$$\sup_{y \in S} |\mathbb{E} \widehat{f}_N^x(t_y) - \mathbb{E} \widehat{f}_N^x(y)| \leq C \frac{l_n}{h_H^2}.$$

Et d'après T_1 , nous écrivons pour n assez grand :

$$\mathbb{P} \left(\frac{1}{\widehat{F}_D^x} \sup_{y \in S} \left| \mathbb{E} \widehat{f}_N^x(t_y) - \mathbb{E} \widehat{f}_N^x(y) \right| > \frac{\eta}{3} \sqrt{\frac{\log n}{nh_H \chi_x(h_K)}} \right). \quad (3.12)$$

Finalement, la preuve du Lemme 3.2.2 est une conséquence du Corollaire 2.1.1 et la décomposition (2.34) et les équations (3.7), (3.11) et (3.12). ■

Preuve du Lemme 3.2.3 : Pour établir ce lemme, il suffit de montrer que :

$$\widehat{\theta} - \theta \rightarrow 0 \quad p.co. \quad (3.13)$$

En utilisant les mêmes arguments que dans la preuve du Lemme 2.3.1, on montre :

$$\forall \varepsilon > 0, \exists \delta(\varepsilon) > 0, \mathbb{P}(|\widehat{\theta} - \theta| > \varepsilon) \leq \mathbb{P}(|f^x(\widehat{\theta}) - f^x(\theta)| > \delta(\varepsilon)).$$

Alors, $\widehat{\theta} - \theta$ tend vers 0

Finalement, la preuve du Lemme 3.2.3 se déduit directement de la continuité de f^x , la condition (H.10) et l'équation (3.13). ■

Chapitre 4

Application et conclusion

Dans ce chapitre nous étudions un exemple d'application sur la prévision de la concentration de l'ozone, ainsi qu'une comparaison avec l'approche standard de la régression. Nous achevons ce chapitre par une conclusion générale.

4.1 Application

On peut obtenir aisément un échantillon de variables fonctionnelles à partir d'un processus à temps continu. En effet, soit $(Z_t)_{t \in [0, b[}$ un processus réel stationnaire à temps continu. A partir de Z_t on construit N variables fonctionnelles $(X_i)_{i=1, \dots, N}$ définies de la façon suivante :

$$\forall t \in [0, b[, X_i(t) = Z_{N^{-1}((i-1)b+t)}.$$

Si on souhaite prédire une caractéristique réelle (notée Y) de X_N sachant la courbe X_{N-1} , on considère les observations (X_i, Y_i) où Y_i est la caractéristique qu'on veut prévoir à l'instant i . En utilisant le mode conditionnel comme outil de prévision, on peut prévoir Y par $\hat{\theta}(X_{N-1})$.

4.1.1 Mise en oeuvre sur des données réelles

Dans ce paragraphe on examine un exemple pour lequel la prévision via le mode conditionnel est meilleure que celle obtenue par la régression. Cet exemple concerne un problème de pollution. Il s'agit plus particulièrement de l'étude de la courbe de la concentration de l'ozone sur le pôle nord et l'objectif est de prévoir le maximum de la concentration maximale de l'ozone dans une journée sachant la courbe de cette concentration pendant la journée précédente. Plus précisément, les observations ont été obtenues selon les données suivantes :

Country	Norway
Code	NO42
Database code	NO0042R
Geographical coordinates	78° 54' N, 11° 53' E
EMEP coordinates (50 km)	69.02, 104.35
Altitude above sea level	474
In operation since	September 1989
Closest climatological station	99910 Ny-Alesund
Main wind direction	E-S-E

Nous avons récupéré ces données de la concentration de l'ozone de l'année 2003 (livrée par heure) sur le site <http://www.nilu.no/projects/cc/sitedescriptions/>. Elles sont représentées par la figure 4.1 :

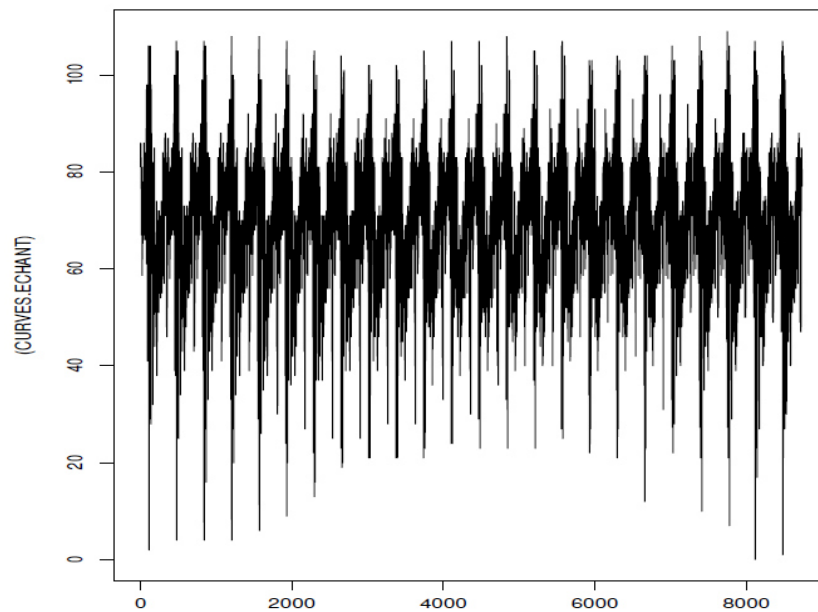


FIGURE 4.1 – Les données de la concentration de l’ozone (2003).

En reprenant les notations du paragraphe précédant, on peut représenter la concentration de l’ozone par un processus stochastique (Z_t) . En faisant un découpage par morceau (1 morceau=1 jour), on construit 365 variables aléatoires fonctionnelles (courbes) notées $X_{i=1,\dots,365}$. Elles sont représentées ci-après :

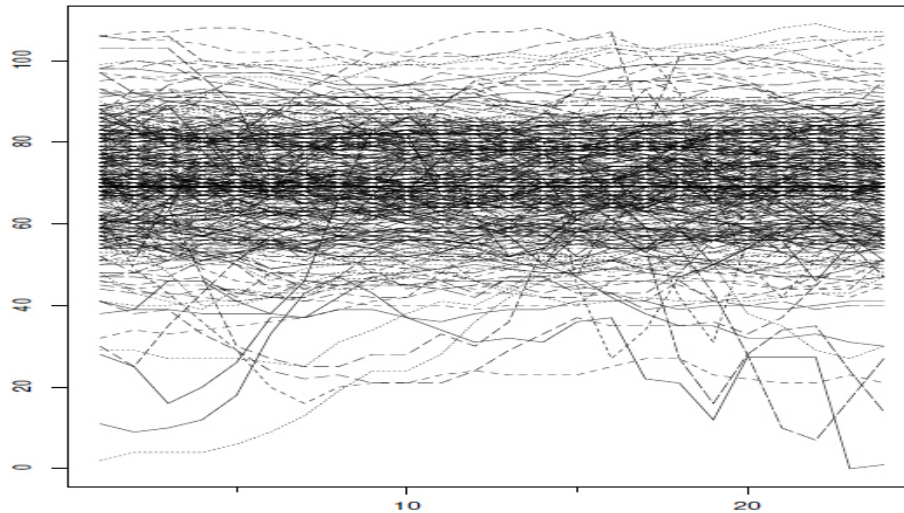


FIGURE 4.2 – Les 365 variables aléatoires fonctionnelles.

Notre objectif est de prévoir la valeur maximale de la concentration de l’ozone sachant la courbe de la concentration de l’ozone de la journée précédente. Ainsi la valeur réponse, notée Y , est $Y_i = \sup_{t \in [1, 24]} X_{i+1}(t)$.

Algorithme

Les estimations sont obtenus en choisissant le paramètre de lissage localement par la méthode de validation croisée sur le nombre des voisins les plus proches et on considère un noyau quadratique défini par :

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2) & \text{si } x \in [-1, 1] \\ 0 & \text{ailleurs.} \end{cases}$$

Le problème du choix de la métrique est aussi très important dans ce jeux des données. Dans notre cas nous avons choisi la norme $L^2_{[1, 24]}$ calculée en utilisant l’analyse en composantes principales fonctionnelles. On procède par l’algorithme suivant :

- Étape 1 : On divise nos observations en deux paquets :
 - $(X_j, Y_j)_{j=1, \dots, n}$ échantillon d'apprentissage,
 - $(X_i, Y_i)_{i=1, \dots, I}$ échantillon de test.
- Étape 2 : Pour chaque X_{new} dans l'échantillon de test, on prévoio sa valeur réponse Y_{new} par :
 - Méthode 1 : La régression

$$\hat{Y}_{new}^{reg} = r_h(X_{new}),$$

avec

$$r_h(x) = \frac{\sum_{j=1}^n K(h_K^{-1}d(x, X_j))Y_j}{\sum_{j=1}^n K(h_K^{-1}d(x, X_j))}$$

- Méthode 2 : Le mode conditionnel

$$\hat{Y}_{new}^{mode} = \hat{\theta}(X_{new}),$$

avec

$$\hat{\theta}(x) = \arg \max_y \hat{f}_y^x$$

- Étape 3 : Pour chaque courbe X_i de l'échantillon de test, on pose :

$$i_{reg} = \arg \min_{j=1, \dots, n} d(X_i, X_j)$$

et

$$i_{mode} = \arg \min_{X_j \in \mathcal{J}} d(X_i, X_j),$$

où \mathcal{J} est l'ensemble des courbes de l'échantillon d'apprentissage appartenant à la même saison que celle de X_i .

- Étape 4 : On prend

$$\widehat{Y}_{new}^{reg} = r_{h_{i_{reg}}^{opt}}(X_{new})$$

et

$$\widehat{Y}_{new}^{mode} = \widehat{\theta}(X_{i_{mode}}).$$

- Étape 5 : L'erreur utilisée pour comparer les deux méthodes est

$$\frac{1}{n} \sum_{i=1}^n (Y_i - \widehat{T}(X_i))^2,$$

où \widehat{T} désigne l'estimateur utilisé : la régression ou le mode conditionnel.

- Étape 6 : On repartage nos observations en deux paquets plusieurs fois et on répète les mêmes étapes 1 – 6.
- Étape 7 : On compare les deux méthodes en présentant notre résultat par un graphe de type boîte à moustaches.

Résultats

On répète 36 fois cette procédure et à chaque fois l'erreur de prédiction est calculée. Nous avons constaté que les erreurs obtenues en faisant la prévision via le mode conditionnel sont dans l'ensemble plus faibles que celles de la régression (voir le graphe ci-dessous).

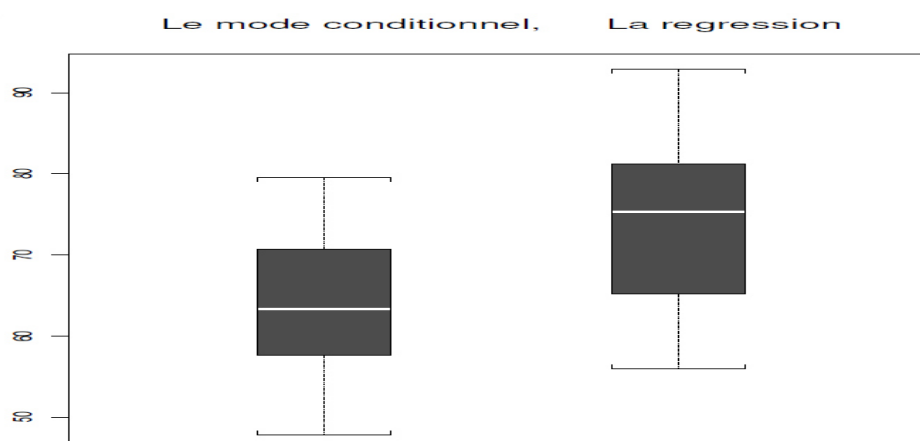


FIGURE 4.3 – Les erreurs de prévision obtenues par les deux méthodes.

4.2 Conclusion

Dans ce mémoire, nous nous sommes intéressés à l'estimation non-paramétrique des modèles conditionnels qui sont liés à la fonction de répartition conditionnelle comme la densité conditionnelle et ses dérivées, le mode conditionnel et le quantile conditionnel; quand la variable explicative est fonctionnelle en utilisant la méthode du noyau. Nous avons fixé comme objectif la convergence presque-complète, dans le cas où les observations sont indépendantes et identiquement distribuées. Dans chacun des cas les vitesses de convergence ont été mentionnées.

Ensuite, nous avons établi la même propriété asymptotique pour un estimateur à noyau du mode conditionnel en considérant des observations fortement mélangeantes.

Afin de confirmer l'efficacité des estimateurs étudiés nous avons terminé notre travail par une application de la prévision en série chronologique. En effet, La régression non-paramétrique a toujours été un outil important dans la prévision. Cependant, il est connu que le mode conditionnel offre une bonne alternative au problème de prévision. Nous avons

donc présenté un exemple sur des données réelles dont l'objectif est l'étude comparative entre prévision par la régression et le mode conditionnel. Cette comparaison a démontré l'efficacité et la supériorité de la prévision en utilisant le mode conditionnel.

On peut aussi remarquer que grâce aux hypothèses supplémentaires du cas dépendant nous avons réussi à établir des vitesses de convergence semblables au cas i.i.d.

Finalement, il faut mentionner que la dimensionnalité des observations (resp. du modèle) est bien exploitée dans l'expression de la vitesse de convergence. On trouvera la dimensionnalité du modèle dans la partie biais, tandis que, la dimensionnalité de la variable fonctionnelle est dans la partie dispersion à travers la propriété de concentration de la mesure de probabilité de la variable fonctionnelle qui est étroitement liée à l'espace fonctionnel de la variable explicative.

Bibliographie

- [1] Berlinet, A., Gannoun, A. and Matzner-Lober, E. (1998). Propriétés asymptotiques d'estimateurs convergents des quantiles conditionnels. *C.R. Acad. Sci., Paris, Sér. I, Math.* **326**, 611-614.
- [2] Berlinet, A., Gannoun, A. and Matzner-Lober, E. (2001). Asymptotic normality of convergent estimates of conditional quantiles. *Statistics.* **35**, 139-169.
- [3] Bosq, D. (2000). *Linear processes in function spaces. Theory and applications.* Lecture Notes in Statistics. **149**. Springer-Verlag.
- [4] Bogachev V.I. (1998). *Gaussian measures*, volume 62. American Mathematical Society Providence.
- [5] Cardot, H., Crambes, C. and Sarda, P. (2004). Spline estimation of conditional quantities for functional covariates. *C. R., Math., Acad.Sci. Paris.* **339**, 141-144.
- [6] Collomb, G., Hardle, W. and Hassani, S. (1987). A note on prediction via estimation of the conditional mode function. *J. Stat. Plann. Inference.* **15**, 227-236.
- [7] Ezzahrioui, M. and Ould Saïd, E. (2008), Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data, *J. of Nonparametric Stat.* **20**, 3-18.
- [8] Ferraty, F., Laksaci, A. and Vieu, P. (2005). Functional times series prediction via conditional mode. *C. R., Math., Acad. Sci. Paris.* **340**, 389-392.

-
- [9] Ferraty F., Laksaci A., et Vieu P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statistical Inference for Stochastic Processes*. **9**, 47-76.
- [10] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis theory and practice*. Springer-Verlag.
- [11] Ferraty, F., Laksaci, A., Tadj, A. and Vieu, P. (2011). Kernel regression with functional response. *Electronic Journal of Statistics*. **5**, 159-171.
- [12] Gannoun, A., Saracco, J. and Yu, K. (2003). Nonparametric time series prediction by conditional median and quantiles. *Journal of statistical Planning and inference*. **117**, 207-223.
- [13] Hoeffding, W.(1963). probability inequalities for sums of bounded random variables. *Journal of American Statistical Association*. **58**, 13-30.
- [14] Laksaci, A. et Yousfate, A. (2002). Estimation fonctionnelle de la densité de l'opérateur de transition d'un processus de Markov à temps discret. *C. R. Math. Acad. Sci. Paris*. **334**, 1035-1038.
- [15] Laksaci, A., Lemdani, M. and Ould-Saïd, E. (2008). A generalized L1- approach for a kernel estimator of conditional quantile with functional regressors : consistency and asymptotic normality. *Stat. Probab. Lett.*. **79**, 1065-1073.
- [16] Nadaraya E.A. (1964). On estimating regression. *Theory Proba. Appl.* **10**, 186-196.
- [17] Ould-Saïd, E. (1997). A note on ergodic processes prediction via estimation of the conditional mode function. *Scand. J. Stat.* **24**, 231-239.
- [18] Parzen E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*. **33**, 1065-1076.
- [19] Quintela-Del-Rio, A. and Vieu, P. (1997). A nonparametric conditional mode estimate. *J. Nonparametric Statis.* **8**, 253-266.

-
- [20] Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*, Springer, Series in Statistics. New-York.
- [21] Ramsay, J., Silverman, B.W. (1997). *Functional Data Analysis*. Springer-Verlag, New York.
- [22] Roussas, George G. (1968). On some properties of nonparametric estimates of probability density function. *Bull. Soc. Math. Grèce (NS)*. **9**, 29-43.
- [23] Roussas, George G. (1969). Nonparametric estimation of the transition distribution function of a Markov process, *Ann. Math. Statist.* **40**, 1386-1400.
- [24] Roussas, George G. (1991). Recursive estimation of the transition distribution function of a Markov process asymptotic normality. *Statistic and probability Letters*. **11**, 435-447.
- [25] Samanta, M. (1989). Nonparametric estimation of conditional quantiles. *Statistic and probability Letters*. **7**, 407-412.
- [26] Stone, C. J. (1977). Consistent nonparametric regression. *Discussion. Ann. Stat.* **5**, 595-645.
- [27] Tukey J.W. et collab. (1961). Curves as parameters, and touch estimation. Dans *Proceedings of the 4th Berkeley Symposium*. **31**, 681-694.
- [28] Watson, G. (1964). Smooth regression analysis. *Sankhya Ser. A*. **26**, 359-372.
- [29] Youndjé, E. (1993). *Estimation non-paramétrique de la densité conditionnelle par la méthode du noyau*. PhD thesis (in French), Rouen University
- [30] Young, Z. and Hua, L. (2000). Asymptotic normality for L^1 norm kernel estimator of conditional median under mixing dependence. *J. of Multivariate Anal.* **73**, 136-154.
- [31] Zhang, J. (2001). Estimation of conditional quantiles using artificial neural network-
Estimation de quantiles conditionnels par réseaux de neurones. **332**, 569-574.