



HAL
open science

Choix optimal du paramètre de lissage dans l'estimation non paramétrique de la fonction de densité pour des processus stationnaires à temps continu

Khadijetou El Heda

► **To cite this version:**

Khadijetou El Heda. Choix optimal du paramètre de lissage dans l'estimation non paramétrique de la fonction de densité pour des processus stationnaires à temps continu. Statistiques [math.ST]. Université du Littoral Côte d'Opale, 2018. Français. NNT : 2018DUNK0484 . tel-02001263

HAL Id: tel-02001263

<https://theses.hal.science/tel-02001263>

Submitted on 31 Jan 2019

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

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

UNIVERSITÉ DU LITTORAL CÔTE D'OPALE

Ecole Doctorale EDSPI Lille

THÈSE DE DOCTORAT

Présentée pour obtenir

Le titre de docteur de l'Université
du Littoral Côte d'Opale

Discipline : Mathématiques appliquées

Option : Statistique

présentée par

Khadijetou EL HEDA

**Choix optimal du paramètre de lissage
dans l'estimation non paramétrique de la
fonction de densité pour des processus
stationnaires à temps continu.**

Soutenue le 25 octobre 2018 devant le jury composé de :

M. LOUANI, D.	Professeur	Université de Reims, France	Co-directeur
M. OULD SAÏD E.	Professeur	ULCO, France	Co-directeur
M. DEMONGEOT, J.	Professeur	Université de Grenoble, France	Rapporteur
M. VIEU, P.	Professeur	Université de Toulouse, France	Rapporteur
M. SCHNEIDER, D.	Professeur	ULCO, France	Examineur
Mme. YAO, A. F.	Professeur	Université de Clermont-ferrant, France	Examinatrice

Remerciements

Mes premiers remerciements, vont à mes deux directeurs de thèse, les Professeurs Djamal LOUANI et Elias OULD SAÏD.

Je remercie infiniment le Professeur Elias OULD SAÏD, qui, par la confiance qu'il m'a accordée, m'a permis d'entreprendre ce travail de thèse. Cette thèse qui a été en amont et en aval portée par le soutien sans faille du Professeur Djamal LOUANI, que je tiens particulièrement à remercier. L'aboutissement de cette thèse n'aurait pu être, sans son indéfectible appui, son temps et ses conseils inestimables durant toutes ces années de travail.

J'adresse également mes sincères remerciements aux Professeurs Philippe VIEU et Jacques DEMONGEOT, qui ont accepté d'être les rapporteurs de cette thèse. Par leurs commentaires éclairés, ils ont aiguillé mes travaux dans une perspective plus vaste. Je remercie aussi les Professeurs, Dominique SCHNEIDER et Anne Fraçoise YAO qui me font le grand honneur de participer à mon jury de soutenance.

A mon père, le Docteur Ahmed ould EL HEDA, qui m'a soutenu par sa bienveillance et son amour indéfectible depuis le début de ma scolarité. Je lui adresse mes remerciements infinis.

A ma mère, Madame aziza ABDELWEDOUD, qui m'a permise, par son soutien constant, d'appréhender la thèse avec beaucoup de sérénité. Je lui adresse mes remerciements les plus affectueux.

A mon frère, Monsieur Dy ould ZEIN, que je remercie amplement pour son aide et ses encouragements ainsi qu'à ma soeur, Madame Mina sidiya, pour son amour et son soutien fraternel.

En fin, je tiens particulièrement à remercier grandement mes amies, Jyda, Selma, Sarah, Aicha, Najia, Meimouna, Tekber, Sahar, Sana et Emal. Votre amitié sincère et votre présence ininterrompue m'ont assurément permises d'avancer avec confiance.

Résumé

Les travaux de cette thèse portent sur le choix du paramètre de lissage dans le problème de l'estimation non paramétrique de la fonction de densité associée à des processus stationnaires ergodiques à temps continu. La précision de cette estimation dépend du choix de ce paramètre.

La motivation essentielle est de construire une procédure de sélection automatique de la fenêtre et d'établir des propriétés asymptotiques de cette dernière en considérant un cadre de dépendance des données assez général qui puisse être facilement utilisé en pratique.

Cette contribution se compose de trois parties. La première partie est consacrée à l'état de l'art relatif à la problématique qui situe bien notre contribution dans la littérature.

Dans la deuxième partie, nous construisons une méthode de sélection automatique du paramètre de lissage liée à l'estimation de la densité par la méthode du noyau. Ce choix issu de la méthode de la validation croisée est asymptotiquement optimal.

Dans la troisième partie, nous établissons des propriétés asymptotiques, de la fenêtre issue de la méthode de la validation croisée, données par des résultats de convergence presque sûre.

Mots-clefs : Paramètre de Lissage, estimation non-paramétrique, estimateur à noyau, consistance, convergence presque sûre, ergodicité, stationarité, temps continu, densité, vitesse de convergence.

Abstract

The work this thesis focuses on the choice of the smoothing parameter in the context of non-parametric estimation of the density function for stationary ergodic continuous time processes. The accuracy of the estimation depends greatly on the choice of this parameter.

The main goal of this work is to build an automatic window selection procedure and establish asymptotic properties while considering a general dependency framework that can be easily used in practice.

The manuscript is divided into three parts. The first part reviews the literature on the subject, set the state of the art and discusses our contribution in within.

In the second part, we design an automatical method for selecting the smoothing parameter when the density is estimated by the kernel method. This choice stemming from the cross-validation method is asymptotically optimal.

In the third part, we establish an asymptotic properties pertaining to consistency with rate for the resulting estimate of the window-width.

Keywords : smoothing parameter, non parametric estimation, kernel estimator, consistence, almost surely consistence, ergodicity, stationarity, continuous time, density, asymptotic normality.

Publications

- [1] Khadijetou El heda, Djamel Louani. Optimal bandwidth selection in kernel density estimation for continuous time dependent process. (2018). *Statistics & Probability Letters* , **138**, p. 9-19.
- [2] Khadijetou El heda, Djamel Louani. Convergence properties of the optimal bandwidth estimate in kernel density estimation for continuous time stationary processes (2018). En préparation.

Conférences

- [1] Optimal bandwidth selection in kernel density estimation for continuous time dependent process. *International Workshop on Perspectives On High-dimensional Data Analysis (HDDA-VIII), Marrakech. 9-13 Avril 2018.*
- [2] Choix de la fenêtre optimale dans l'estimation non paramétrique de la densité pour des processus stationnaires ergodiques à temps continu. *SFDS 50^{èmes} journées de Statistiques, EDF Lab Paris Saclay. 28-1 Juin 2018.*

Exposé

- [1] Choix de la fenêtre optimale dans l'estimation non paramétrique de la densité pour des processus stationnaires ergodiques à temps continu. *Séminaire au Laboratoire Paul Painlevé, Lille. 27 Juin, 2018.*

Table des matières

1	Introduction	1
1.1	Estimation non paramétrique en temps discret	5
1.1.1	Historique des méthodes de sélection du paramètre de lissage .	6
1.1.2	Méthode de la validation croisée	9
1.2	Théorie ergodique pour des processus stationnaires	14
1.3	Cadre non paramétrique en temps continu	15
1.4	Plan de la thèse	18
1.4.1	Chapitre 2. Construction d'une procédure de sélection du paramètre de lissage optimal	19
1.4.2	Chapitre 3. Vitesses de convergence du paramètre de lissage optimal	21
2	Procédure de sélection automatique du paramètre de lissage optimal en estimation de la densité d'un processus à temps continu	23
2.1	Introduction	24
2.2	Résultats	25
2.3	Preuves	29
2.3.1	Preuve de la proposition 2.1	31
2.3.2	Preuve du théorème 2.1	33
3	Vitesses de convergence du paramètre de lissage optimal en estimation de la densité d'un processus à temps continu	41
3.1	Introduction	41
3.2	Résultats	43
3.3	Preuves	47
3.3.1	Preuve de la Proposition 3.1	47

3.3.2	Preuve du Théorème 3.1	54
3.3.3	Preuve du Théorème 3.2	57
3.3.4	Preuve du Corollaire 3.1	58
3.3.5	Preuve du Corollaire 3.2	59
3.3.6	Preuve du Théorème 3.3	60
3.3.7	Preuve du Théorème 3.4	64
	Annexe	67
	Perspectives de recherche	71
	Bibliographie	73

Chapitre 1

Introduction

L'inférence statistique se décompose en deux branches principales, l'estimation et les tests d'hypothèses. Lorsque l'on est en présence d'un nombre fini de paramètres à estimer, on parle d'estimation paramétrique. Alors que si l'objectif est d'estimer une fonctionnelle ou un opérateur qui ne peuvent être représentés par un nombre fini de paramètres, on a recours à des méthodes d'estimation dites non paramétriques. Outre les différences de formes entre ces deux types d'estimation, il y a aussi une incidence sur les vitesses de convergence. De façon générale, les vitesses de convergence en estimation non paramétrique sont moindres comparées aux vitesses paramétriques. L'estimation non paramétrique est tributaire du paramètre de lissage qui joue un rôle fondamental. Le sur-lissage ou le sous-lissage aboutissent tous les deux à des estimateurs de mauvaise qualité en égard au critère utilisé et à la fonction cible à estimer. Il est alors d'une importance capitale que de construire une procédure permettant un choix optimal pour un critère donné de ce paramètre. L'erreur quadratique moyenne (EQM), L'erreur quadratique moyenne intégrée (EQMI), l'erreur quadratique intégrée (EQI) et la validation croisée sont parmi les critères les plus utilisées pour construire des procédures de choix optimaux de ce paramètre.

Les premiers travaux relatifs à l'estimation non paramétrique remontent à Rosenblatt (1956) et Parzen(1962) où la méthode du noyau de convolution est utilisée pour estimer la fonction de densité de probabilité. De nombreuses contributions ont alors été consacrées à l'estimation dans le cadre de données à temps discret des fonctionnelles telles que la régression, la fonction taux de hasard. Nous citons

pour cela des travaux de Nadaraya (1964) et Watson (1964), Rosenblatt (1971), Prasaka Rao (1983), Silverman (1986). Le cadre de données indépendantes est loin de couvrir toute la réalité du terrain. Une littérature abondante est consacrée au cas dépendant où les divers types de mélange ainsi que l'ergodicité sont considérés. Nous renvoyons entre autres aux travaux de Masry (1983), Roussas(1990), Härdle (1990), Tran (1990).

Le choix du paramètre de lissage a fait l'objet de nombreuses contributions dans la littérature. Dans le cadre de données discrètes indépendantes et identiquement distribuées, la fenêtre optimale théorique relative à l'estimateur à noyau de la densité minimisant l'erreur quadratique moyenne intégrée a été obtenue par Parzen (1962). Cependant, cette fenêtre est inconnue en pratique. Plusieurs travaux se sont intéressés à la procédure de choix d'une fenêtre pouvant approximer celle-ci. Nous citons tout d'abord le travail de Stone (1984). Il décrit une règle de sélection de la fenêtre en utilisant la méthode de la validation croisée. Sous l'hypothèse que la densité marginale est bornée, ce choix est ainsi asymptotiquement optimal. Notons que ce critère a été introduit par Rudemo (1982) et Bowman (1984). Härdle & Marron (1985) ont considéré le choix de la fenêtre dans le cas de la fonction de régression et de la fonction du taux de hasard et ont établi l'équivalence asymptotique entre l'EQI, l'EQMI et l'EQM pour les histogrammes, les estimateurs par la méthode de noyau et la méthode des séries orthogonales. Il existe plusieurs règles de sélections automatiques de données pour gérer le problème du choix optimal du paramètre de lissage. Les méthodes dites plug-in minimisant à la fois l'EQMI et l'EQM, voir Silverman (1986), Heidenreich (2013) et Hall & Marron (1987) pour plus de détails. Chac'on et al. (2007) ont montré l'existence d'un minimiseur exact de l'EQMI par rapport à l'estimateur à noyau de la densité en fonction du paramètre de lissage et ont donné des propriétés limites de ce paramètre optimal. Un autre travail de Chac'on et al (2012) donne des propriétés du paramètre optimal qui minimise l'EQM de l'estimateur à noyau de la densité. Tenreiro (2017) présente dans son travail une version modifiée du critère de validation croisée en introduisant une pondération dans la validation croisée (PVC). Des résultats de simulations numériques montrent que la méthode PVC est plus performante que la méthode standard pour les cas de densité "facile à estimer" et "difficile à estimer". En prenant appui sur le travail de

Hall, P. (1982) montrant l'équivalence asymptotique entre l'EQMI et l'EQI, Hall, P. (1983) établit l'optimalité asymptotique du paramètre de lissage issu du critère de la validation croisée. Hall, P. & Marron, J. (1987) ont eux évalué les distances entre les différents paramètres résultant de la minimisation des critères EQI, EQMI et la validation croisée pour ensuite établir des vitesses de convergence. Des résultats similaires ont été établis par Rice, J. (1984) dans le cas de l'estimation de la régression. Dans le cas des données α -mélangeantes, Hart & Vieu (1990) ont introduit le critère de la validation croisée en considérant "a sequence-leave-out" dans le cas de l'estimation de la densité. Ils ont montré sous des conditions peu restrictives que le paramètre minimisant le critère introduit est asymptotiquement optimal. Le cas de la fonction taux de hasard pour des données α -mélangeantes est traité dans le travail de Esévez-pérez, Quintela-del-río et Vieu (2002). Des résultats de convergence du paramètre qui en ressort ont été établis. Sans vouloir être exhaustif sur la question, nous citons aussi les travaux de Sarda & Vieu (1991), Youndjé et al (1996), Tea Yoon Kim & Denis (1997) pour traiter le problème du choix optimal du paramètre de lissage. Dans le cadre des données fonctionnelles, Rachdi & Vieu (2007) ont considéré l'estimation de la fonction de régression pour construire un paramètre de lissage asymptotiquement optimal pour le critère de la validation croisée.

L'estimation non paramétrique relative à des processus à temps continu a reçu un intérêt particulier ces dernières décennies. De nombreux résultats traitant de la convergence, des vitesses convergence et de la normalité asymptotique ont été établis. Pour un large panorama sur la question, nous renvoyons à Bosq (1998) et les références qui y sont citées. Comme dans le cas discret, la qualité de l'estimation est fortement impactée par le choix du paramètre de lissage. L'objet de cette thèse est de construire une procédure de sélection automatique du paramètre de lissage et établir les propriétés d'optimalité pour le critère de validation croisée et de convergence. Dans un souci de plus de précision, nous détaillons ci-après les résultats les plus significatifs sur la question.

Le présent manuscrit présente des résultats de recherches originaux portant sur la présentation dans un premier lieu de la fenêtre optimale théorique h_{opt} minimisant l'EQMI relative à l'estimation non paramétrique de la fonction de densité associée

à des processus à temps continu. L'expression de cette fenêtre est déterministe et fait intervenir des quantités inconnues. Notons qu'en pratique, la fenêtre dépend des données aléatoires, nous considérons alors l'EQI et construisons à partir de ce critère, une méthode de sélection automatique du paramètre de lissage. La fenêtre \hat{h} issue de cette méthode de sélection est asymptotiquement optimale. Dans un second lieu, nous évaluons l'ordre de grandeur de h_{opt} minimisant l'EQMI pour ensuite établir des vitesses de convergence des paramètres h_0 et \hat{h} , minimisant respectivement l'EQI et le critère de validation croisée, vers h_{opt} . L'ensemble de cette thèse est composé de trois chapitres.

Dans l'introduction, nous présentons l'historique des travaux réalisés auparavant sur le choix du paramètre de lissage dans le cadre de l'estimation non paramétrique en temps discret ainsi que les différentes méthodes de sélection de ce paramètre. Nous présentons en suite le cadre stationnaire ergodique tant en temps discret qu'en temps continu. Nous revenons sur la méthode de sélection automatique pour le critère de validation croisée présentée par Stone (1984) pour des données indépendantes et celles exposées par Hart & Vieu (1990) pour des données α -mélangeantes. Afin de pouvoir établir nos résultats, nous présentons tout d'abord le cadre non paramétrique en temps continu et développons les propriétés qui nous ont servies pour aboutir à ces résultats. Nous finissons par un résumé des principaux résultats.

Le chapitre 2 est consacré au choix optimal du paramètre de lissage " h " dans l'estimation non paramétrique de la fonction de densité associée à un processus stationnaire ergodique à temps continu. Dans un premier temps, nous présenterons l'expression exacte du paramètre de lissage h_{opt} . Du fait que le critère d'évaluation EQMI aboutit à un choix non aléatoire dépendant de paramètres inconnues, la solution est de considérer l'EQI. Notre objectif est ainsi de chercher à minimiser l'EQI suivant le paramètre de lissage h . Sachant que l'EQI dépend de la fonction de densité qui est "inconnue", nous construisons alors une procédure évitant de traiter cette quantité. A cet égard, nous estimons la fonction de densité et introduisons le critère de validation croisée adapté à ce processus. En utilisant un premier résultat qui permet d'obtenir l'équivalence asymptotique entre l'EQI et le critère de validation croisée introduit, la règle de sélection de la fenêtre optimale est construite ainsi en minimisant ce critère par rapport à " h ". Finalement notre principal résultat montre

que \hat{h} , issue de cette méthode de sélection, est asymptotiquement optimal.

Dans le chapitre 3, nous présentons dans un premier temps l'ordre de grandeur de la fenêtre optimale h_{opt} minimisant l'EQMI relative à la densité f d'un processus stationnaire ergodique à temps continu. En utilisant l'équivalence asymptotique entre l'EQI et l'EQMI associées à f et le critère de la validation croisée, nous établissons dans un deuxième temps que les fenêtres minimisant ces différents critères ont le même ordre de grandeur. Ces résultats font appel à des vitesses de convergence de l'estimateur à noyau de la densité plus rapides que celles présentées par Didi & Louani (2013). Ce qui conduit aux critères d'optimalité $\frac{h_0}{h_{opt}} \rightarrow 1$ et $\frac{\hat{h}}{h_{opt}} \rightarrow 1$ presque sûrement. En fin, nous présentons des vitesses de convergence des fenêtres h_0 et \hat{h} ainsi que celles de $EQI(\hat{h})$ et $EQI(h_{opt})$.

1.1 Estimation non paramétrique en temps discret

Soit X_1, \dots, X_n , n variables aléatoires réelles, indépendantes, identiquement distribuées et de densité marginale f . La méthode d'estimation par histogramme est historiquement la première méthode pour estimer une fonction de densité. L'estimateur se présente dans les termes suivants. En considérant, une partition $(b_i)_i$ d'amplitude $h = h_n$ d'un intervalle T de \mathbb{R} , l'estimateur de la densité f est défini, pour tout $x \in I$, par

$$f_{n,h}(x) = \frac{1}{nh} \sum_{j=0}^n \mathbb{1}_{\{X_j \in b_i, x \in b_i\}},$$

où $\mathbb{1}$ désigne la fonction indicatrice. L'estimateur à noyau de Parzen-Rosenblatt de f est lui défini, pour tout $x \in \mathbb{R}$, par

$$f_{n,h}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right),$$

où K est une fonction positive d'intégrale 1 et h le paramètre de lissage. Le paramètre de lissage optimal au sens du critère EQMI donné par Parzen (1962) et établi sous les hypothèses

- K est une fonction positive, $\int K = 1$ et $\int x^2 K(x) \neq 0$,

— f est bornée, deux fois dérivable, de dérivées bornées,

est de la forme

$$h_{opt} = \left\{ \frac{\int K^2}{(\int x^2 K(x))^2 \int f''^2} \right\}^{\frac{1}{5}} n^{-\frac{1}{5}}. \quad (1.1)$$

Cette expression ne peut être utilisée en pratique puisqu'elle dépend de quantités inconnues. Nous présentons dans la section suivante les différentes procédures de sélection du paramètre de lissage approprié.

1.1.1 Historique des méthodes de sélection du paramètre de lissage

•**Choix subjectif** Cette approche consiste à représenter sur un graphe plusieurs estimateurs correspondant à plusieurs paramètres de lissage et ensuite choisir parmi ces estimateurs, le paramètre de lissage correspondant à l'estimateur le plus proche graphiquement de la fonction de densité. Pour plus de détail voir Silverman (1986).

•**Forme de la fenêtre dans le cas de la loi normale** 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,

$$\hat{h} \approx 1.06\sigma n^{-\frac{1}{5}}.$$

Il suffit d'estimer σ pour avoir une forme utilisable en pratique.

•**Méthode du graphe test** Cette méthode subjective a été introduite par Silverman (1978a). Son idée vient du comportement de l'estimateur de la dérivée seconde si celle-ci était proche de la vraie dérivée seconde. Le résultat est présenté, en supposant que

— Le noyau K est symétrique, deux fois différentiable et $\int |z|^2 K(z) dz \neq 0$.

- La densité f est de dérivée seconde bornée et uniformément continue.
- $\hat{h} = \hat{h}(n)$ est choisie de façon à minimiser l'erreur maximale dans l'estimation de la densité, ceci se traduit, pour $n \rightarrow \infty$, par

$$\frac{\sup |f''_{n,h} - \mathbb{E}f''_{n,h}|}{|\mathbb{E}f''_{n,h}|} \rightarrow k,$$

où k est donné par

$$k = \frac{1}{2} \left\{ \int |z^2 K(z) dz| \right\} \left\{ \frac{\int (K'')^2}{\int K^2} \right\}^{\frac{1}{2}}.$$

La méthode proposée par Silverman est de représenter graphiquement $f''_{n,h}$ pour différentes valeurs de h . Le meilleur "graphe test" doit avoir de rapides fluctuations sans pour autant dissimuler les variations systématiques, voir Silverman (1978a) pour plus de détails.

• **Méthode plug-in** Cette méthode consiste à estimer la fenêtre

$$h_{opt} = A(K)B(f)n^{-\frac{1}{5}}$$

donnée dans la formule (1.1) de Parzen par h_1 de la forme

$$h_1 = A(K)B(f_{n,h_0})n^{-\frac{1}{5}},$$

construit à partir d'un estimateur à noyau f_{n,h_0} de f correspondant à un choix initial h_0 de la fenêtre. Cette approche a été proposée par Woodroof (1970). La fenêtre h_1 est toute fois sensible au choix initial de h_0 , voir Scott et Factor (1981) pour plus de détails. Des simulations numériques montrent que pour des densités bimodales, à queues de distributions chargées, cette méthode présente des résultats décevants, voir Bowman (1985).

Les méthodes citées jusqu'à présent ne sont pas automatiques, elle restent imprécises voir grossières dans certains cas.

Nous présentons maintenant des méthodes de sélections automatiques.

- Une règle de sélection de la fenêtre est une fonction $\hat{h} = \hat{h}(n)$ du processus (X_1, \dots, X_n) à valeurs dans \mathbb{R}^+ .

La méthode de sélection suivante rentre dans le cadre paramétrique, nous la citons tout de même

- **Méthode de vraisemblance de la validation croisée** Elle a été introduite par Duin (1976) et Habbema, Herman et Van Der Brock (1974). Elle consiste à prolonger l'idée de la vraisemblance pour juger l'estimation. En supposant Y indépendant du processus X_i de densité f . \hat{f} est la famille paramétrique d'estimateurs dépendant de la fenêtre h . L'estimateur $\log \hat{f}_{-i}$ de $\log f$ est construit à partir des données $\{Y, X_j, j = 1, \dots, N, j \neq i\}$. Le critère de vraisemblance de validation croisée est alors donné par

$$CV(h) = \frac{1}{n} \sum_{i=1}^n \log \hat{f}_{-i}(X_i).$$

Choisir "h" maximisant ce critère conduit à un estimateur proche de la densité au sens de la distance de Kullback- Leibler définie par

$$I(\hat{f}, f) = \int f \log \left(\frac{f}{\hat{f}} \right) dx.$$

Si la fonction de densité n'est pas à support compact et le noyau K à support compact,

$$I(\hat{f}, f) \rightarrow \infty, \quad \text{lorsque } n \rightarrow \infty,$$

nous renvoyons à Silverman (1986). Cette méthode exclue le cas de la loi normale, ce qui présente un gros handicap.

Nous introduisons maintenant la méthode de validation croisée.

- **Méthode de validation croisée**

Cette méthode a été introduite par Rudemo (1982) et Bowman (1984). Pour un estimateur à noyau $f_{n,h}$ de la fonction de densité f , l'EQI est définie par

$$\begin{aligned} L_{n,h} &= \int [f_{n,h}(x) - f(x)]^2 dx \\ &= \int f_{n,h}^2(x) dx - 2 \int f_{n,h}(x) f(x) dx + \int f^2(x) dx. \end{aligned}$$

L'idée est de construire un estimateur de $\int f_{n,h}^2 - 2 \int f_{n,h} f$. Cet estimateur, construit à partir des données $(X_i)_{1 \leq i \leq n}$, est donné par le critère de validation croisée suivant,

$$M_{n,h} = \int f_{n,h}^2 - \frac{2}{n(n-1)h} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{X_i - X_j}{h}\right).$$

La règle de sélection de la fenêtre se traduit par la minimisation de ce critère,

$$\hat{h} = \operatorname{argmin}_h M_{n,h}.$$

En utilisant le résultat de Stone (1984), \hat{h} est asymptotiquement optimal. Cette méthode sera détaillée dans la section suivante.

1.1.2 Méthode de la validation croisée

Choix du paramètre de lissage pour des données indépendantes

L'estimateur à noyau f_{nh} de la densité f est donné par

$$f_{n,h} = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right),$$

où $h \in H_n \subset \mathbb{R}^+$. l'EQI relative à $f_{n,h}$ est donnée par

$$\begin{aligned} L_{n,h} &= \|f_{n,h} - f\|^2 & (1.2) \\ &:= \int [f_{n,h}(x) - f(x)]^2 dx \\ &= \int f_{n,h}^2(x) dx - 2 \int f_{n,h}(x) f(x) dx + \int f^2(x) dx. \end{aligned}$$

Le paramètre de lissage théorique optimal $h_0 \in \mathbb{R}^+$ au regard de l'EQI est donné par

$$h_0 = \operatorname{argmin}_h L_{n,h}.$$

Minimiser $L_{n,h}$ par rapport à h revient à minimiser la quantité

$$\int f_{n,h}^2(x) dx - 2 \int f_{n,h}(x) f(x) dx.$$

Puisque cette quantité est inconnue, ceci conduit à l'introduction de l'estimateur "one-leave-out", $f_{n,-i}$ de la fonction de densité f construit à partir des données $(X_j)_{1 \leq j \leq n, j \neq i}$, défini par

$$f_{n,-i}(x) = \frac{1}{(n-1)h} \sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{x - X_j}{h}\right).$$

Le critère de validation croisée est alors donné par

$$M_{n,h} = \int f_{n,h}^2 - \frac{2}{n} \sum_{i=1}^n f_{n,-i}(X_i).$$

- Equivalence asymptotique entre l'EQI et le critère de validation croisée

$$\begin{aligned} \mathbb{E} \left[\int f_{n,h}^2 - 2 \int f_{n,h} f \right] &= \int \mathbb{E} [f_{n,h}^2(x)] dx - 2 \int [\mathbb{E} f_{n,h}(x)] f(x) dx \\ &= \int \mathbb{E} [f_{n,h}^2(x)] dx - 2 \int \mathbb{E} [f_{n,h}(x)] f(x) dx \\ &= \int \mathbb{E} [f_{n,h}^2(x)] dx - \frac{2}{h} \mathbb{E} \left[K \left(\frac{X_i - X_j}{h} \right) \right] \\ &= \int \mathbb{E} [f_{n,h}^2(x)] dx - \frac{2}{n(n-1)h} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathbb{E} \left[K \left(\frac{X_i - X_j}{h} \right) \right] \\ &= \mathbb{E} [M_{n,h}]. \end{aligned}$$

Minimiser $L_{n,h}$ par rapport à h est équivalent à minimiser le critère de validation croisée $M_{n,h}$ par rapport à h . La règle de sélection est donnée par

$$\hat{h} = \operatorname{argmin}_h M_{n,h}.$$

- Critère d'optimalité

— La fenêtre \hat{h} est dite asymptotiquement optimale si

$$\lim_{n \rightarrow \infty} \frac{L_{n,\hat{h}}}{\inf_h L_{n,h}} = 1, \quad p.s.$$

Théorème (Stone (1984)). *Sous l'hypothèse que f soit bornée, \hat{h} est asymptotiquement optimal.*

Ce résultat est établi sous la condition suivante

— K est un noyau positif, à support compact, symétrique et Hölder-continue, ie, pour tout $(x, y) \in \mathbb{R}^2$, il existe deux constantes $\lambda > 0$ et $D > 0$ tel que $|K(x) - K(y)| \leq D|x - y|^\lambda$.

Stone démontre le résultat premièrement pour $h \in H_N := \{h_1, \dots, h_N\}$. $H_N \subset H_n$ et satisfaisant la condition suivante

$$\#H_N \leq An^a, \quad A, a > 0.$$

Pour vérifier que \hat{h} est asymptotiquement optimal, il suffit de démontrer que,

$$\lim_{n \rightarrow \infty} \max_{h, h' \in H_N} \frac{|L_{n, h'} - L_{n, h} - (M_{n, h'} - M_{n, h})|}{L_{n, h} + L_{n, h'}} = 0, \quad p.s. \quad (1.3)$$

Démontrer l'assertion (1.3) revient à démontrer

$$(1) \liminf_n \min_{h \in H_N} \left(\frac{L_{n, h}}{\|\mathbb{E}[f_{n, h}] - f\|^2 + \frac{1}{nh}} \right) > 0 \quad p.s.$$

$$(2) \lim_{n \rightarrow \infty} \max_{h, h' \in H_N} \frac{|L_{n, h'} - L_{n, h} - (M_{n, h'} - M_{n, h})|}{\|\mathbb{E}[f_{n, h}] - f\|^2 + \|\mathbb{E}[f_{n, h'}] - f\|^2 + \frac{1}{nh} + \frac{1}{nh'}} = 0, \quad p.s.$$

Grâce à un résultat, donné dans le lemme 1 de Stone (1984), les dénominateurs de (1) et (2) sont contrôlés, pour h relativement petit, par

$$\|\mathbb{E}[f_{n, h}] - f\|^2 > Ch^r,$$

où C et r sont deux constantes positives. Par un développement approprié de $L_{n, h}$, l'utilisation de l'inégalité de Bernstein, la preuve de (1) et (2) fait appel à la méthode de "Poissonization", voir aussi Rosenblatt (1975). Par l'argument de Hölder-continuité de K , Stone montre le résultat pour H_n .

Choix du paramètre de lissage pour des données dépendantes

Nous introduisons tout d'abord la notion de processus α -mélangeants. Dans ce qui suit, $\sigma(X_1, \dots, X_n)$ désigne la tribu engendrée par les variables aléatoires X_1, \dots, X_n .

Définition 1.1. Un processus $(X_i)_i$ est dit α -mélangeant s'il existe, un coefficient $\alpha(m)$ vérifiant, pour tout entiers positifs k et m , tout ensemble $A \in \sigma(X_1, \dots, X_k)$ et tout $B \in \sigma(X_{m+k}, \dots)$, les propriétés suivantes

$$— |P(A \cap B) - P(A)P(B)| \leq \alpha(m).$$

$$— \lim_{m \rightarrow \infty} \alpha(m) = 0.$$

Nous considérons ici l'EQI pondérée relative à l'estimateur à noyau de f et définie par

$$EQI(f) = \int [f_{n,h}(x) - f(x)]^2 w(x) dx,$$

où w est une fonction positive à support compact. Le paramètre h es pris ici dans un ensemble H_n .

Pour minimiser l'EQI, Hart & Vieu (1990) ont défini le critère de la validation croisée en introduisant, un estimateur $f_n^{(i)}$ de f , appelé l'estimateur de "sequence-leave-out", basé sur les données $\{X_j\}$, tel que $|j - i| > l_n$ défini par

$$f_n^{(i)}(x) = \frac{1}{n l_n h} \sum_{|j-i|>l_n} K\left(\frac{x - X_j}{h}\right),$$

où l_n est une suite d'entiers positifs appelée "sequence-leave-out" et $n l_n$ tel que, $n l_n$ est le nombre d'index de données (X_i, X_j) vérifiant $|j - i| > l_n$. Le critère de la validation croisée est défini par

$$CV(h) = \int f_{n,h}^2(x) w(x) dx - \frac{2}{n} \sum_{i=1}^n f_n^{(i)}(X_i) w(X_i),$$

où

$$\hat{h}_{l_n} = \operatorname{argmin}_{h \in H_n} CV(h).$$

Pour établir leur résultat, Hart & Vieu (1990) ont considéré les hypothèses suivantes

(H_a) Le noyau K est supposé borné, symétrique, à support compact, $\int x^\nu K(x) < \infty$, $\int x^k K(x) dx = 0$, $k = 1, \dots, \nu - 1$ et Lipschitz continue, ie, Hölder continue avec $\lambda = 1$.

(H_b) $H_n := [An^{-a}, Bn^{-b}]$, $0 < b \leq \frac{1}{2\nu + 1} \leq a < \frac{2}{1 + 4\nu}$, où A et B sont deux constantes positives.

(H_c) La suite de "leave-out" $\{l_n\}_{\mathbb{N}^*}$ vérifie, $l_n \leq n^{\theta_1}$ où $\theta_1 < 1 - \frac{a(1 + 4\nu)}{2}$.

(H_d) Le coefficient de mélange vérifie, $\sup_{j > n^{\theta_1}} \alpha(j) = o(n^{-\theta_2})$, où $\theta_2 := \theta_2(a, b, \nu, \theta_1)$ (voir Hart & Vieu(1990)).

(H_e) La densité f est bornée, admet ν dérivées continues pour $\nu \in \mathbb{N}^*$ et $\max(f(x), f(-x)) \rightarrow 0$, quand $x \rightarrow \infty$.

Théorème (Hart & Vieu(1990)). *Sous les Hypothèses (H_a) – (H_e), \hat{h}_{l_n} est asymptotiquement optimal.*

Vérifier la propriété de l'optimalité asymptotique revient à vérifier (voir Marron (1987)) que

$$\sup_{h \in H_n} \frac{|D_{T,l_n}(h)|}{EQI(h)} \rightarrow 0, \quad p.s, \quad (1.4)$$

où

$$D_{T,l_n}(h) = \frac{1}{n} \sum_{i=1}^n f_n^{(i)}(X_i)w(X_i) - \int f_{n,h}(x)f(x)w(x)dx - \sum_{i=1}^n f(X_i)w(X_i) + \int f^2(x)w(x)dx.$$

La preuve du théorème se décompose en trois étapes.

1. La première étape consiste à établir que

$$\sup_{h \in H_n} \frac{|D_{T,l_n}(h) - D_{T,l_n^*}(h)|}{EQI(h)} \rightarrow 0 \quad p.s,$$

où l_n est contrôlée par l_n^* qui croît algébriquement.

2. La deuxième étape consiste à démontrer l'assertion (1.4) pour l_n^* et h' appartenant à un ensemble fini H'_n de H_n , ie,

$$\sup_{h \in H'_n} \frac{|D_{T,l_n^*}(h')|}{EQI(h')} \rightarrow 0 \quad p.s.$$

3. La dernière étape consiste à établir, pour $h \in H_n$ proche de $h' \in H'_n$, que

$$\sup_{h \in H_n} \frac{|D_{T,l_n^*}(h) - D_{T,l_n^*}(h')|}{EQI(h)} \rightarrow 0 \quad p.s.$$

L' EQI se décompose en deux termes, le terme en biais et le terme en variance (voir Rosenblatt,(1971)), le terme en biais est de l'ordre de $h^{2\nu}$, ainsi par le théorème 2 de Vieu (1989), pour n suffisamment grand et tout $h \in H_n$, il existe une constante finie C_b tel que,

$$EQI(h) \geq C_b h^{2\nu}.$$

L'étape 1 découle directement en considérant l'estimateur relatif au noyau $\frac{K}{\int |K|}$ et par la proposition 4.1 de Roussas (1988). L'étape 2 se traite par des inégalités exponentielles pour des données α -mélangeantes (voir Théorème A.2 de Roussas (1988) ainsi que l'inégalité de Bienaymé). L'étape 3 se déduit directement du fait que le noyau K est Lipschitzien et de la proximité entre $h \in H_n$ et $h' \in H'_n$.

1.2 Théorie ergodique pour des processus stationnaires

L'étude des propriétés des processus à temps continu ne peut être envisagée en dehors d'un cadre de dépendance des données. Les dépendances faibles ont été considérées par de nombreux auteurs dans les cadres à la fois de processus à temps discret que de processus à temps continu. Nous considérons dans cette thèse le cadre ergodique qui est plus général que le cadre de dépendances faibles. La théorie ergodique doit ses origines à la mécanique statistique. Birkoff (1931) et Von Neumann (1932) sont les pionniers dans ce domaine. Son utilisation dans l'analyse des systèmes dynamiques a permis d'établir de nombreux résultats. Il est fait état de l'équivalence entre le comportement moyen des systèmes dynamiques et la moyenne temporelle des comportements d'un système dynamique. Nous renvoyons à Peksir (2000), pour le théorème ponctuel de Birkoff et le théorème ergodique de Von Neumann.

La stationnarité classique est décrite dans le concept probabiliste par une mesure préservant la transformation τ . On considère ainsi, pour une fonction f intégrable, les moyennes prises sur les séquences $f, f \circ \tau, f \circ \tau^2, \dots$. Elle peut être définie d'une façon équivalente comme suit.

Définition 1.2. Soit $Y = (Y_0, Y_1, \dots)$ un processus défini sur un espace probabilisé (E, \mathcal{F}, P) . La distribution du processus est déterminée par la probabilité des événements de type $\{Y_{t_1} \in F_1, \dots, Y_{t_n} \in F_n\}$. Y est stationnaire si, pour tout $n \in \mathbb{N}$, $F_1, \dots, F_n \in \mathcal{F}$, $t_1, \dots, t_n, s \in \mathbb{Z}$

$$P(Y_{t_1} \in F_1, \dots, Y_{t_n} \in F_n) = P(Y_{t_1+s} \in F_1, \dots, Y_{t_n+s} \in F_n).$$

Les définitions qui suivent permettent de définir l'ergodicité des processus à temps continu.

Définition 1.3 (Ensembles invariants). Soit $\{Y_t\}_{t \in \mathbb{R}^+}$ un processus à temps continu défini sur un espace mesuré (E, \mathcal{F}, P) . Pour $\delta > 0$, soit T^δ une transformation δ -shift, ie, $(T^\delta \circ T^s)(Y) = Y_{\delta+s}$. Un ensemble mesurable A est dit δ -invariant, s'il ne change pas sous une transformation δ -shift ($T^\delta(A) = A$).

Définition 1.4 (δ -ergodicité). $Y = \{Y_t\}_{t \in \mathbb{R}^+}$ est dit δ -ergodique, si tout ensemble mesurable δ -invariant lié au processus Y , a une probabilité de 0 ou 1.

Définition 1.5 (Ergodicité). Un processus $Y = \{Y_t\}_{t \in \mathbb{R}^+}$ est dit ergodique s'il est δ -ergodique pour tout $\delta > 0$.

Nous énonçons le théorème ponctuel ergodique de Birkoff lié aux processus stationnaires à temps discret, (voir Krengel (1983), théorème 4.4 p.26).

Théorème (Birkoff). *Si Y_0, Y_1, \dots est un processus réel stationnaire, Y_0 intégrable et \mathcal{F} une σ -algèbre des ensembles invariants alors*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^n Y_i = \mathbb{E}[Y_0 / \mathcal{F}], \quad p.s.$$

Si en plus le processus est ergodique,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^n Y_i = \mathbb{E}[Y_0], \quad p.s.$$

Théorème (ergodique de Birkoff en temps continu). *Si $(Y_t)_{t \in [0, T]}$ est un processus réel stationnaire, alors*

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T Y_t dt = \mathbb{E}[Y_0], \quad p.s.$$

1.3 Cadre non paramétrique en temps continu

Pour $T > 0$, soit $(X_t, 0 \leq t \leq T)$ un processus à temps continu de densité marginale f . L'estimateur à noyau de la densité est défini, pour tout $x \in \mathbb{R}$, par

$$f_{T,h}(x) = \frac{1}{Th} \int_0^T K\left(\frac{x - X_t}{h}\right) dt,$$

où $h = h(T)$ est le paramètre de lissage, supposé évoluant dans l'ensemble $H_T := [a_T, b_T] \subset \mathbb{R}_+$, K est une fonction mesurable, positive d'intégrale 1.

Dans ce travail, nous faisons appel à un résultat de convergence ponctuelle de l'estimateur à noyau de la fonction de densité dû à Didi & Louani (2013), voir le théorème 2.1, dans le quel ils utilisent une version fonctionnelle du théorème ergodique de Birkoff, introduite dans le cas discret par Delecroix (1987) et une hypothèse similaire à celle de Wu (2003). Considérons, pour un réel positif δ tel que $n = \frac{T}{\delta} \in \mathbb{N}$,

la partition $(T_j)_{1 \leq j \leq n}$ de pas δ de l'intervalle $[0, T]$. Pour $t \in [0, T]$, $f^{\mathcal{F}_{t-\delta}}$ est la densité conditionnelle sachant la tribu $\mathcal{F}_{t-\delta} := \sigma(X_{t-\delta})$. La projection \mathcal{P}_k est définie pour une v.a réelle ξ et $k \in \mathbb{N}$, par $\mathcal{P}_k \xi = \mathbb{E}[\xi | \mathcal{F}_k] - \mathbb{E}[\xi | \mathcal{F}_{k-1}]$, où $\mathbb{E}[\xi | \mathcal{F}_k]$ est l'espérance conditionnelle de ξ sachant la σ -algèbre \mathcal{F}_k , avec $\mathcal{F}_k := \mathcal{F}_{T_k}$. Pour établir ce résultat, Didi & Louani (2013) utilisent les hypothèses suivantes :

(A) (i) K est un noyau positif d'intégrale 1.

(ii) $\int |z|K(z)dz < \infty$.

(B) (i) La densité f est différentiable avec une dérivée bornée.

(ii) Pour tout $\delta > 0$, la densité conditionnelle $f^{\mathcal{F}_{t-\delta}}$ est différentiable avec une dérivée presque sûrement bornée.

(iii) Pour tout $t \in [0, T]$, $\delta > 0$ et $x \in \mathbb{R}$, la fonction $f^{\mathcal{F}_{t-\delta}}(x)$ est presque sûrement bornée par une fonction déterministe $b_{t,\delta}(x)$, qui vérifie la propriété ergodique, pour tout $\delta > 0$, $\frac{1}{T} \int_0^T b_{t,\delta}(x) dt \rightarrow D(x) \neq 0$, quand $T \rightarrow \infty$.

(C) Pour tout $\delta > 0$, $\sup_x \int_{\mathbb{R}^+} \|\mathcal{P}_1 f_t^{\mathcal{F}_{t-\delta}}(x)\|^2 dt < \infty$.

Théorème (Didi & Louani (2013)). *Sous les hypothèses (A), (B) et (C), si $\frac{Th_T}{\log T} \rightarrow \infty$ lorsque $T \rightarrow \infty$, alors, pour tout $x \in \mathbb{R}$,*

$$f_{T,h}(x) - f(x) = O(h_T) + O_{p.s} \left[\frac{\log T}{Th} \right]^{\frac{1}{2}}, \quad \text{lorsque } T \rightarrow \infty.$$

Pour établir nos résultats, nous énonçons un résultat de Stone (1984), qui nous permettra de contrôler la norme L^2 du biais, lorsque le paramètre h est relativement petit.

Lemma 1 (Stone-1984). *Pour $h > 0$ relativement petit, il existe deux constantes positives C et r tel que*

$$\begin{aligned} \|\mathbb{E}[f_{n,h}] - f\|^2 &:= \int [\mathbb{E}[f_{n,h}](x) - f(x)]^2 dx \\ &\geq Ch^r + \frac{1}{Th}. \end{aligned}$$

Ce résultat a été établi dans le cas de données indépendantes et identiquement distribuées. Cependant,

$$\begin{aligned}\mathbb{E}[f_{T,h}] &:= \frac{1}{h} \int K\left(\frac{x-y}{h}\right) f(y) dy \\ &= \mathbb{E}[f_{n,h}],\end{aligned}$$

et le lemme 1 de Stone (1984) reste valide dans le cas continu.

Afin de présenter nos résultats, nous introduisons notre cadre de travail et quelques propriétés qui nous seront utiles. Soit $(X_t, 0 \leq t \leq T)$, $T \in \mathbb{R}^+$ un processus stationnaire ergodique de densité marginale f . Considérons l'estimateur à noyau de f défini, pour tout $x \in \mathbb{R}$, par

$$f_{T,h}(x) = \frac{1}{Th} \int_0^T K\left(\frac{x - X_t}{h}\right) dt,$$

où h est le paramètre de lissage appartenant à $H_T := [a_T, b_T] \subset \mathbb{R}_+$. Considérons la densité jointe f_{X_s, X_t} du vecteur (X_s, X_t) et la densité marginale f_{X_s} de X_s . Introduisons à présent la fonction $g_{s,t}$ définie par

$$g_{s,t} := f_{X_s, X_t} - f_{X_s} f_{X_t}.$$

Comme le processus est stationnaire, nous en déduisons que

$$g_{s,t} = g_{0,|s-t|}, \quad g_{0,|s-t|} := g_{|s-t|} \quad .$$

Diverses mesures d'écart entre l'estimateur de la fonction de densité $f_{T,h}$ et la densité réelle f ont été évaluées. Le premier critère d'évaluation étudié) est l'EQMI (Rosenblatt (1956)). Ce critère a été largement étudié dans la littérature, il est donné par

$$M_h = \mathbb{E} \int [f_{T,h}(x) - f(x)]^2 dx.$$

Soit $K^{(2)}$ le produit de convolution de K avec lui même. Notons que $K^{(2)}$ a les mêmes propriétés que K . Par ailleurs, pour tout $t > 0$, soit $\mathcal{F}_t = \sigma((X_s) : 0 \leq s \leq t)$ la σ -algèbre générée par $(X_s)_{0 \leq s \leq t}$. Pour une variable aléatoire ξ et tout $k \in \mathbb{N}$, la projection \mathcal{P}_k est définie par $\mathcal{P}_k \xi = \mathbb{E}[\xi | \mathcal{F}_k] - \mathbb{E}[\xi | \mathcal{F}_{k-1}]$, où $\mathbb{E}[\xi | \mathcal{F}_k]$ est l'espérance

conditionnelle de ξ sachant la σ -algèbre $\mathcal{F}_k := \mathcal{F}_{T_k}$.

Forme du paramètre de lissage optimal et son ordre de grandeur

Sous des hypothèses adaptées sur le noyau K , la densité f et la fonction $g_{s,t}$, la forme exacte du paramètre de lissage optimal minimisant l'EQMI est donnée par

$$h_{opt} = \left\{ \frac{2 \int K^2}{\int z^2 (\frac{1}{2} K^{(2)} - K)(z) dz \int f'' f} \right\}^{\frac{1}{3}} \left\{ \frac{\iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T [T-u] g_u(y, z) du \right\} dy dz}{T^2} \right\}^{\frac{1}{3}}.$$

Sous des hypothèses appropriées sur le noyau K , la densité f et la fonction $g_{s,t}$, l'ordre de grandeur de h_{opt} est donné par

$$h_{opt} = \left\{ \frac{4 \int K^2 \int_0^\infty \int g_u(z, z) dz du}{\int z^2 (\frac{1}{2} K^{(2)} - K)(z) dz \int f'' f} \right\}^{\frac{1}{3}} \left\{ \frac{a_T^\alpha}{T} \right\}^{\frac{1}{3}} + o\left(\frac{a_T^\alpha}{T}\right)^{\frac{1}{3}}.$$

Le paramètre de lissage h_{opt} étant déterministe, il est plus judicieux de considérer l'EQI au lieu de l'EQMI, puisque le paramètre de lissage doit être aléatoire en pratique. Dans ce travail nous considérons donc l'EQI définie par

$$\begin{aligned} L_{T,h} &= \int [f_{T,h}(x) - f(x)]^2 dx \\ &= \int f_{T,h}^2(x) dx - 2 \int f_{T,h}(x) f(x) dx + \int f^2(x) dx. \end{aligned}$$

Minimiser cette erreur par rapport à h revient à minimiser

$$\int f_{T,h}^2(x) dx - 2 \int f_{T,h}(x) f(x) dx.$$

Puis que cette dernière expression dépend de la fonction de densité inconnue f , nous devons donc construire une procédure de sélection évitant de traiter des quantités inconnues en pratique. Nous présentons dans la section suivante le plan détaillé de ce manuscrit.

1.4 Plan de la thèse

Le manuscrit de cette thèse est composé de trois chapitres. L'état de l'art ainsi qu'une revue bibliographique sont données en introduction. Le chapitre 2 est consacré à la construction d'une procédure de sélection automatique du paramètre de

lissage qui réalise le minimum pour le critère de validation croisée. Le chapitre 3 établit des vitesses de convergence vers le paramètre optimal pour l'EQMI des paramètres issus du critère de validation croisée et l'EQI.

1.4.1 Chapitre 2. Construction d'une procédure de sélection du paramètre de lissage optimal

Le chapitre 2 introduit la méthode de sélection par validation croisée du paramètre de lissage optimal dans l'estimation non paramétrique de la fonction de densité. L'EQMI définie par

$$M_h = \mathbb{E} \int [f_{T,h}(x) - f(x)]^2 dx,$$

est largement utilisée pour l'évaluation de l'écart entre la densité et son estimateur. Sous des conditions de régularité du noyau, de la fonction de densité et de ses dérivées premières et secondes ainsi que des hypothèses adéquates sur la fonction $g_{s,t}$, nous avons obtenu l'expression exacte du paramètre de lissage h_{opt} minimisant l'EQMI. Ce résultat est obtenu grâce à une décomposition adaptée de l'espace \mathbb{R}^2 et l'usage d'un développement de Taylor. h_{opt} dépend des quantités inconnues f et $g_{s,t}$. Il est donc impossible de le calculer en pratique. De ce fait, nous devons construire une procédure qui nous permettra d'éviter de traiter des quantités inconnues. Nous considérons alors l'EQI

$$L_{T,h} = \int [f_{T,h}(x) - f(x)]^2 dx$$

comme critère d'évaluation et introduisons le critère de la validation croisée. Nous avons établi, sous des hypothèses appropriées, l'équivalence asymptotique entre ces deux critères. Minimiser l'EQI revient donc à minimiser le critère de la validation croisée. Sous des hypothèses appropriées, le résultat principal établit que le paramètre \hat{h} réalisant le minimum pour le critère de validation croisée est asymptotiquement optimal.

La proposition suivante donne la forme du paramètre de lissage optimal pour l'EQMI.

Proposition 1. *Sous des hypothèses adaptées de K , f , f' , f'' et g_u , la fenêtre optimale théorique par rapport à l'erreur moyenne quadratique intégrée est donnée*

par

$$h_{opt} = \left\{ \frac{2 \int K^2}{\int z^2 (\frac{1}{2} K^{(2)} - K)(z) dz \int f'' f} \right\}^{\frac{1}{3}} \left\{ \frac{\iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T [T-u] g_u(y, z) du \right\} dy dz}{T^2} \right\}^{\frac{1}{3}}.$$

Le critère de validation croisée approprié au cas de l'estimation de la densité pour des processus à temps continu est défini comme suit.

$$M_{T,h} = \int f_{T,h}^2(x) dx - \frac{2}{T^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int_{T_{i-1}}^{T_i} \int_{T_{j-1}}^{T_j} K_h(X_s - X_t) ds dt,$$

où, pour $n \in \mathbb{N}$, $\delta = \frac{T}{n}$ et $T_j = j\delta$, $1 \leq j \leq n$. Ce critère fait intervenir un estimateur s'identifiant à l'estimateur de "one-leave-out" de f en temps discret, construit à partir des données $(X)_{t \in \cup_{\substack{j=1 \\ j \neq i}}^n [T_{j-1}, T_j]}$, donné par

$$f_{T,-i\delta}(x) := \frac{1}{(n-1)\delta} \sum_{\substack{j=1 \\ j \neq i}}^n \int_{T_{j-1}}^{T_j} K_h(x - X_t) dt.$$

Un critère asymptotiquement équivalent à $M_{T,h}$ est donné par

$$\frac{1}{n} \sum_{i=1}^n \int f_{T,-i}^2(x) dx - \frac{2}{T} \sum_{i=1}^n \int_{T_{i-1}}^{T_i} f_{T,-i}(X_s) ds.$$

La proposition suivante établit l'équivalence asymptotique entre l'EQI et le critère de la validation croisée $M_{T,h}$ et revêt une grande importance dans la preuve de notre résultat principal.

Proposition 2. *Sous des hypothèses appropriées,*

$$\lim_{T \rightarrow \infty} \mathbb{E} \left[L_{T,h} - \int f^2(x) dx \right] = \lim_{T \rightarrow \infty} \mathbb{E} [M_{T,h}].$$

Avant d'énoncer notre résultat principal, posons tout d'abord les hypothèses relatives à h . Il existe un réel r , $0 < r < 1$, tel que

- $T a_T^{2r+1} / \log T \rightarrow \infty$ quand $T \rightarrow \infty$.
- $\frac{b_T}{a_T^r} \rightarrow 0$ quand $T \rightarrow \infty$.

Théorème 1.6. *Sous des hypothèses adaptées, la règle de sélection de la fenêtre, \hat{h} , est asymptotiquement optimale.*

1.4.2 Chapitre 3. Vitesses de convergence du paramètre de lissage optimal

L'objet du chapitre 3 est d'étudier le comportement asymptotique des critères l'EQI, l'EQMI, $M_{T,h}$ et d'évaluer les distances $L_{T,h_{opt}} - L_{T,h_0}$, $L_{T,\hat{h}} - L_{T,h_0}$ et $L_{T,h_{opt}} - L_{T,\hat{h}}$ ainsi que les vitesses de convergence presque sûres de h_0 vers h_{opt} et \hat{h} vers h_{opt} , où h_0 , h_{opt} et \hat{h} les fenêtres minimisant respectivement l'EQI, l'EQMI et $M_{T,h}$. Le lemme suivant présente l'expression donnant l'ordre de grandeur la fenêtre optimale minimisant l'EQMI.

Proposition 1.1. Sous des hypothèses adaptées,

$$h_{opt} = \left\{ \frac{4 \int K^2 \int_0^\infty \int g_u(z, z) dz du}{\int z^2 (\frac{1}{2} K^{(2)} - K)(z) dz \int f'' f} \right\}^{\frac{1}{3}} \left\{ \frac{a_T^\alpha}{T} \right\}^{\frac{1}{3}} + o \left(\frac{a_T^\alpha}{T} \right)^{\frac{1}{3}}.$$

Afin d'établir une équivalence asymptotique entre les différents critères l'EQI, l'EQMI et $M_{T,h}$, nous énonçons les théorèmes suivants.

Théorème 1.7. *Sous des hypothèses convenables, il existe $m < 1$ et $q \leq 2$, tel que*

$$L_{T,h} = M_h + O_{p.s.} (h)^{\min(\frac{2}{q}-1, 2)} + O_{p.s.} \left[\frac{\log T}{T h^{2-\frac{1}{m}}} \right]^{\frac{1}{2}}. \quad (1.5)$$

Théorème 1.8. *Sous des hypothèses adaptées, pour tout h de l'ordre de grandeur de $\left(\frac{a_T^\alpha}{T}\right)^{\frac{1}{3}}$, il existe $m < 1$, tel que*

$$M_{T,h} = L_{T,h} - \frac{2}{T} \int_0^T f(X_t) dt + \int f^2(x) dx + O \left(\frac{a_T^\alpha}{T} \right)^{\frac{2}{3}} + O_{p.s.} \left[\frac{\log T}{\left(T^{1+\frac{1}{m}} a_T^{\alpha(2-\frac{1}{m})} \right)^{\frac{1}{3}}} \right]^{\frac{1}{2}}.$$

Critère d'optimalité

Un estimateur \tilde{h} de la fenêtre h_{opt} est dit asymptotiquement aussi "bon" que h_{opt} si

$$\frac{L_{T,\tilde{h}}}{M_{h_{opt}}} \longrightarrow 1, \quad p.s. \quad (1.6)$$

Sous les conditions des Théorèmes 1.7 et 1.8, nous avons que h_0 et \hat{h} sont asymptotiquement aussi bons que h_{opt} . Par conséquent, pour $\tilde{h} \in \{h_0, \hat{h}\}$, nous obtenons

$$\frac{\tilde{h}}{h_{opt}} \longrightarrow 1, \quad p.s. \quad (1.7)$$

Nous présentons les théorèmes 3.3 et 3.4 suivants permettant de donner les vitesses de convergence respectives de h_0 et \hat{h} .

Théorème 1.9. *Sous des hypothèses adaptées, il existe $m < 1$, tel que*

$$h_0 - h_{opt} = O\left(\frac{a_T^\alpha}{T}\right)^{\min\{\frac{2}{3q}-\frac{2}{3}, \frac{1}{3}\}} + O_{p.s}\left[\frac{\log T}{\left(T^{\frac{1}{m}-1}a_T^{\alpha(4-\frac{1}{m})}\right)^{\frac{1}{3}}}\right]^{\frac{1}{2}}. \quad (1.8)$$

- Pour $q < \frac{2}{3}$ et $m < \frac{1}{3}$, une conséquence directe du théorème 3.3 est

$$L_{T,h_{opt}} - L_{T,h_0} = O\left(\frac{a_T^\alpha}{T}\right)^{\frac{2}{3}} + O_{p.s}\left[\frac{\log T}{\left(T^{\frac{1}{m}+1}a_T^{\alpha(2-\frac{1}{m})}\right)^{\frac{1}{3}}}\right]^{\frac{1}{2}} + O_{p.s}\left[\frac{\log T}{\left(T^{\frac{1}{m}-1}a_T^{\alpha(4-\frac{1}{m})}\right)^{\frac{1}{3}}}\right]. \quad (1.9)$$

Théorème 1.10. *Sous des hypothèses adaptées, il existe $m < 1$, tel que*

$$\hat{h} - h_{opt} = O\left(\frac{a_T^\alpha}{T}\right)^{\min\{\frac{2}{3q}-\frac{2}{3}, \frac{1}{3}\}} + O_{p.s}\left[\frac{\log T}{\left(T^{\frac{1}{m}-1}a_T^{\alpha(4-\frac{1}{m})}\right)^{\frac{1}{3}}}\right]^{\frac{1}{2}}. \quad (1.10)$$

- Pour $q < \frac{2}{3}$ et $m < \frac{1}{3}$, nous obtenons grâce au théorème 3.4

$$L_{T,\hat{h}} - L_{T,h_0} = O\left(\frac{a_T^\alpha}{T}\right)^{\frac{2}{3}} + O_{p.s}\left[\frac{\log T}{\left(T^{\frac{1}{m}+1}a_T^{\alpha(2-\frac{1}{m})}\right)^{\frac{1}{3}}}\right]^{\frac{1}{2}} + O_{p.s}\left[\frac{\log T}{\left(T^{\frac{1}{m}-1}a_T^{\alpha(4-\frac{1}{m})}\right)^{\frac{1}{3}}}\right]. \quad (1.11)$$

Chapitre 2

Procédure de sélection automatique du paramètre de lissage optimal en estimation de la densité d'un processus à temps continu

Ce chapitre a fait l'objet d'une publication dans *Statistics and Probability Letters*

Résumé

Le choix du paramètre de lissage dans l'estimation non paramétrique de la fonction de densité est d'une grande importance. La mesure de précision de cette estimation dépend fortement de la manière dont ce paramètre est choisi. Dans cet article, on construit une procédure de sélection automatique du paramètre de lissage optimal dans l'estimation à noyau de la densité d'un processus stationnaire ergodique à temps continu.

MSC[2010] 60-G10, 62-G07, 47-A35.

Mots clefs : processus stationnaire, processus à temps continu, estimation de la densité, ergodicité, estimateur à noyau, fenêtre.

2.1 Introduction

Soit $(X_t, 0 \leq t \leq T)$, $T \in \mathbb{R}^+$ un processus stationnaire ergodique à temps continu de densité marginale f . On considère l'estimateur à noyau de la densité f défini, pour tout $x \in \mathbb{R}$, par,

$$f_{T,h}(x) = \frac{1}{Th} \int_0^T K\left(\frac{x - X_t}{h}\right) dt,$$

où h est un paramètre de lissage appartenant à un ensemble $H_T := [a_T, b_T] \subset \mathbb{R}_+$, K est une fonction positive mesurable d'intégrale égale à un. Un choix inapproprié du paramètre h impacte fortement l'estimateur dans le sens où l'on peut observer des phénomènes de sur-lissage ou de sous-lissage de la fonctionnelle estimée. Dans ce travail, nous nous fixons comme objectif de déterminer le paramètre de lissage convenable qui résulte de la méthode de validation croisée adaptée à l'estimation à noyau de la densité pour un processus stochastique à temps continu. Le critère de validation croisée doit ses origines à Rudemo, (1982) et Bowman, (1984). L'historique relatif à cette méthode et les travaux qui en ont découlé a déjà été mis en avant dans la partie introduction de ce manuscrit. Le paramètre de lissage optimal pour le critère EQMI présente l'inconvénient de dépendre de quantités inconnues et son utilisation en pratique pose alors problème. Construire une procédure basée sur des critères permettant d'éviter cette issue est alors nécessaire. Dans ce qui suit, nous introduisons d'abord quelques outils permettant de résoudre le problème. L'erreur quadratique moyenne intégrée (EQMI) est définie par

$$M_h = \mathbb{E} \int [f_{T,h}(x) - f(x)]^2 dx.$$

Dans toute la suite, nous noterons h_{opt} la quantité qui minimise M_h . On note que h_{opt} dépend de quantités inconnues, ce qui rend impossible son utilisation en pratique. Pour un résultat exploitable en pratique, l'erreur quadratique intégrée (EQI), définie par

$$L_{T,h} = \int [f_{T,h}(x) - f(x)]^2 dx,$$

est plus appropriée. Comme $L_{T,h}$ peut aussi s'écrire sous la forme

$$L_{T,h} = \int f_{T,h}^2(x) dx - 2 \int f_{T,h}(x) f(x) dx + \int f^2(x) dx,$$

il ressort que minimiser $L_{T,h}$ par rapport à h revient à minimiser la quantité

$$L_{T,h} - \int f^2(x) dx = \int f_{T,h}^2(x) dx - 2 \int f_{T,h}(x) f(x) dx.$$

Cette quantité dépend encore de la densité ce qui aboutit à un paramètre h dépendant encore de quantité inconnue. Prenant appui sur le critère de validation croisée introduit par Rudémo (1982) et Bowman (1984) et utilisé par de nombreux auteurs par la suite, nous proposons le critère suivant

$$M_{T,h} = \int f_{T,h}^2(x)dx - \frac{2}{T^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int_{T_{i-1}}^{T_i} \int_{T_{j-1}}^{T_j} K_h(X_s - X_t) ds dt,$$

où, $K_h(x) = K\left(\frac{x}{h}\right)$, $T_0 = 0$ et $T_j = j\delta$ avec $\delta = \frac{T}{n}$ et $1 \leq j \leq n$.

La procédure de sélection du paramètre de lissage minimise $M_{T,h}$ pour obtenir le paramètre optimal que nous noterons dans la suite par \hat{h} .

Remarque 2.1. En notant

$$f_{T,-i}(x) := \frac{1}{(n-1)\delta} \sum_{\substack{j=1 \\ j \neq i}}^n \int_{T_{j-1}}^{T_j} K_h(x - X_t) dt,$$

Il ressort alors, un critère asymptotiquement équivalent à $M_{T,h}$, s'écrivant sous la forme

$$\frac{1}{n} \sum_{i=1}^n \int f_{T,-i}^2(x) dx - \frac{2}{T} \sum_{i=1}^n \int_{T_{i-1}}^{T_i} f_{T,-i}(X_s) ds.$$

$f_{T,-i}(x)$ s'agit de l'estimateur à noyau de f sans la partie $\{X_t : t \in [T_{i-1}, T_i]\}$ du processus.

2.2 Résultats

Introduisons tout d'abord les notations nécessaires avant de poser le hypothèses requises pour établir nos résultats. Si K est un noyau, dans tout la suite $K^{(2)}$ désigne le noyau de convolution de K avec lui même défini par

$$K^{(2)}(x) = \int K(x-y)K(y)dy.$$

Nous signalons que $K^{(2)}$ a les mêmes propriétés que le noyau K . Dans tout le chapitre 2, f_{X_s, X_t} désigne la densité conjointe du vecteur aléatoire (X_s, X_t) et f_{X_s} la densité marginale de X_s . Nous introduisons par ailleurs les notations suivantes

$$g_{s,t} := f_{X_s, X_t} - f_{X_s} f_{X_t}, \quad g_{0,u} := g_u, \quad \|\cdot\|_\infty := \sup_{(y,z) \in \mathbb{R}^2} |\cdot|.$$

Pour tout $t \geq 0$, nous notons $\mathcal{F}_t = \sigma((X_s) : 0 \leq s \leq t)$ la σ -algèbre générée par les données $(X_s)_{0 \leq s \leq t}$. Pour une variable aléatoire ξ et un entier $k \in \mathbb{N}$, le projecteur \mathcal{P}_k est défini par l'équation

$$\mathcal{P}_k \xi = \mathbb{E}[\xi | \mathcal{F}_k] - \mathbb{E}[\xi | \mathcal{F}_{k-1}],$$

où $\mathbb{E}[\xi | \mathcal{F}_k]$ est l'espérance conditionnelle de ξ sachant la σ -algèbre $\mathcal{F}_k := \mathcal{F}_{T_k}$, T_k étant le $k^{\text{ième}}$ élément de la partition de l'intervalle $[0, T]$, ie, $T_k = \delta k$.

Les hypothèses requises pour établir nos résultats sont posées ci-après.

Hypothèses

- (H1) (a) K est une fonction symétrique, bornée ($K \leq M$), de support compact et d'intégrale un.
 (b) K est Hölder-continue, ie, pour tout $(x, y) \in \mathbb{R}^2$, il existe deux constantes $\lambda > 0$ et $D > 0$ tel que $|K(x) - K(y)| \leq D|x - y|^\lambda$.
 (c) $\int z^2(K^{(2)}(z) - 2K(z))dz < 0$.
- (H2) (a) f est deux fois dérivable avec la première dérivée et la seconde dérivée bornées.
 (b) $\int f'' f < 0$.
- (H3) Il existe un ensemble $\Gamma \in \mathcal{B}_{\mathbb{R}^2}$ contenant $D = \{(s, t) \in \mathbb{R}^2, s = t\}$ tel que
 (a) $g_{s,t}$ existe pour tout $(s, t) \notin \Gamma$.
 (b) $\Delta_p(\Gamma) := \sup_{(s,t) \notin \Gamma} \|g_{s,t}\|_{L^p(\mathbb{R}^2)} < \infty$, pour $p \in]2, \infty[$.
 (c) $u \rightarrow \|g_u\|_\infty$ est intégrable sur $]0, \infty[$.
 (d) Il existe une fonction a_T et un réel $\alpha > 1$ tels que, $\iint_{|y-z| < a_T^\alpha} g_u(y, z) dy dz > 0$.
- (H4) (a) Pour tout $\delta > 0$, la densité conditionnelle $f_t^{\mathcal{F}_{t-\delta}}$ de X_t sachant la σ -algèbre $\mathcal{F}_{t-\delta}$ est dérivable et de dérivée presque sûrement bornée.
 (b) Pour tout $t \in [0, T]$, et tout $\delta > 0$, la fonction $f_t^{\mathcal{F}_{t-\delta}}(x)$ est presque sûrement bornée par une fonction déterministe $b_{t,\delta}(x)$.
 (c) Pour tout $\delta > 0$, $\frac{1}{T} \int_0^T b_{t,\delta}(x) dt \rightarrow D(x) \neq 0$, lorsque $T \rightarrow \infty$.
- (H5) Pour tout $\delta > 0$, $\sup_x \int_{\mathbb{R}^+} \|\mathcal{P}_1 f_t^{\mathcal{F}_{t-\delta}}(x)\|^2 dt < \infty$.
- (H6) Il existe un réel r , $0 < r < 1$, tel que

- (a) $Ta_T^{2r+1}/\log T \rightarrow \infty$, lorsque $T \rightarrow \infty$.
 (b) $\frac{b_T}{a_T^r} \rightarrow 0$, lorsque $T \rightarrow \infty$.

Commentaires sur les hypothèses.

Les conditions (H1) sont très standards en estimation fonctionnelle non paramétrique. La condition (H1)(c) est similaire à la condition $K^{(2)}(0) - 2K(0) < 0$ introduite par Stone (1984), il est à noter que le noyau gaussien satisfait cette hypothèse et qu'en général le noyau n'impacte pas significativement l'estimation. L'hypothèse (H2)(a) est une condition de régularité de la distribution du processus considéré. L'hypothèse (H2)(b) est satisfaite par le processus gaussien. Les deux hypothèses (H1)(d) et (H2)(b) assurent la positivité de h_{opt} . Les conditions (H3)(a) – (b) sont utilisées uniquement pour établir l'équivalence entre l'EQI et le critère de validation croisée. Ces dernières servent plus généralement dans la littérature à atteindre des vitesses de convergence optimales, voir par exemple Bosq (1996). Notons que la condition (H3)(c) est utilisée dans Bosq (1996, p 104) pour atteindre les vitesses de convergence sur-optimales de l'estimateur de la densité. Les conditions (H4) et (H5) contribuent à assurer la convergence de l'estimateur de la densité, comme établi dans Didi & Louani (2013). Les conditions (H6) assurent les contraintes du domaine d'appartenance du paramètre de lissage.

Afin de présenter le paramètre de lissage optimal théorique, on considère premièrement la décomposition de \mathbb{R}^2 en sous-espaces $\{|y - z| \leq a_T^\alpha\}$ et $\{|y - z| > a_T^\alpha\}$, où $\alpha > 1$. Il s'en suit alors, sous les conditions H1, H2 et H3(c) – (d), que la fenêtre optimale théorique par rapport à EQMI est

$$h_{opt} = \left\{ \frac{2 \int K^2}{\int z^2 (\frac{1}{2} K^{(2)} - K)(z) dz \int f'' f} \right\}^{\frac{1}{3}} \left\{ \frac{\iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T [T-u] g_u(y, z) du \right\} dy dz}{T^2} \right\}^{\frac{1}{3}}. \quad (2.1)$$

Les détails de la preuve sont donnés dans la section suivante.

Remarque 2.2. Si l'on suppose que $u \rightarrow \iint_{|y-z| \leq a_T^\alpha} |g_u(y, z)| dy dz$ est intégrable,

nous avons alors

$$\begin{aligned} \frac{\iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T [T-u] g_u(y, z) du \right\} dydz}{T^2} &= \frac{1}{T} \int_0^T \left[1 - \frac{u}{T} \right] \left\{ \iint_{|y-z| \leq a_T^\alpha} g_u(y, z) dydz \right\} du \\ &\leq \frac{1}{T} \int_0^\infty \left\{ \iint_{|y-z| \leq a_T^\alpha} |g_u(y, z)| dydz \right\} du. \end{aligned}$$

Par conséquent,

$$\frac{\iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T [T-u] g_u(y, z) du \right\} dydz}{T^2} = O\left(\frac{1}{T}\right).$$

Ci-après, un exemple de processus par le quel, h_{opt} est positif sous la condition $H1(c)$.

Exemple 2.1. Soit $(X_t, t \in \mathbb{R})$ un processus Gaussien stationnaire réel centré et de variance un. Soient la fonction d'auto corrélation $\rho(u)$ et la densité marginale f . La densité jointe est alors donnée par

$$f_u(y, z) := f_{X_0, X_u}(y, z) = \frac{1}{2\pi\sqrt{1-\rho^2(u)}} \exp \left\{ -\frac{1}{2} \left(\frac{y^2 + z^2 - 2\rho(u)yz}{1-\rho^2(u)} \right) \right\}.$$

Par un changement de variable convenable, nous obtenons

$$\begin{aligned} \iint_{|y-z| \leq a_T^\alpha} f_u(y, z) dydz &= \frac{1}{2\pi} \iint_{\substack{\frac{a_T^\alpha}{\sqrt{1-\rho^2(u)}} + \frac{1-\rho(u)}{\sqrt{1-\rho^2(u)}}z \\ -\frac{a_T^\alpha}{\sqrt{1-\rho^2(u)}} + \frac{1-\rho(u)}{\sqrt{1-\rho^2(u)}}z}} \exp \left\{ -\frac{y^2 + z^2}{2} \right\} dydz \\ &\geq \frac{1}{2\pi} \iint_{-a_T^\alpha+z}^{a_T^\alpha+z} \exp \left\{ -\frac{y^2 + z^2}{2} \right\} dydz \\ &= \iint_{|y-z| \leq a_T^\alpha} f(y)f(z) dydz. \end{aligned}$$

Il s'en suit alors que

$$\iint_{|y-z| \leq a_T^\alpha} g_u(y, z) dydz \geq 0.$$

h_{opt} étant dépendant d'une quantité inconnue, la proposition suivante permet d'établir une équivalence asymptotique entre l'EQI et le critère de la validation croisée.

Proposition 2.1. Nous supposons les hypothèses $(H1(a))$ et $(H3(a) - (b))$ satisfaites. Nous avons alors

$$\lim_{T \rightarrow \infty} \mathbb{E} \left[L_{T,h} - \int f^2(x) dx \right] = \lim_{T \rightarrow \infty} \mathbb{E} [M_{T,h}].$$

La règle de sélection automatique s'exprime ainsi par

$$\hat{h} = \operatorname{argmin}_{h \in H_T} M_{T,h}.$$

Le résultat principal de ce chapitre est donné dans le théorème suivant.

Théorème 2.1. Sous les hypothèses (H1)(a) – (b), (H2)(a), (H3)(a) – (b), (H4), (H5) et (H6), le paramètre de lissage \hat{h} résultant de la règle de sélection est asymptotiquement optimal.

2.3 Preuves

Lemme 2.1. Supposons les hypothèses (H1), (H2) et (H3)(c) – (d) satisfaites. Le paramètre de lissage optimal théorique pour le critère EQMI est donné par

$$h_{\text{opt}} = \left\{ \frac{2 \int K^2}{\int z^2 \left(\frac{1}{2} K^{(2)} - K \right)(z) dz \int f'' f} \right\}^{\frac{1}{3}} \left\{ \frac{\iint_{|y-z| \leq a_T^2} \left\{ \int_0^T [T-u] g_u(y,z) du \right\} dy dz}{T^2} \right\}^{\frac{1}{3}}. \quad (2.2)$$

Démonstration. Observons d'abord que

$$\begin{aligned} M_h &= \mathbb{E} \int f_{T,h}^2(x) dx - 2 \int \mathbb{E} f_{T,h}(x) f(x) dx + \int f^2(x) dx \\ &= \frac{1}{T^2 h^2} \int \mathbb{E} \left[\int_0^T K \left(\frac{x - X_t}{h} \right) dt \right]^2 dx - \frac{2}{T, h} \int \mathbb{E} \left[\int_0^T K \left(\frac{x - X_t}{h} \right) dt \right] f(x) dx \\ &\quad + \int f^2(x) dx \\ &= \frac{1}{T^2 h^2} \int \mathbb{E} \left[\int_0^T \int_0^T K \left(\frac{x - X_t}{h} \right) K \left(\frac{x - X_s}{h} \right) dt ds \right] dx \\ &\quad - \frac{2}{Th} \int \int_0^T \mathbb{E} K \left(\frac{x - X_t}{h} \right) dt f(x) dx + \int f^2(x) dx \\ &= \frac{1}{T^2 h^2} \iiint K \left(\frac{x - y}{h} \right) K \left(\frac{x - z}{h} \right) \left\{ \int_0^T \int_0^T f_{t,s}(y, z) dt ds \right\} dy dz dx \\ &\quad - \frac{2}{h} \iint K \left(\frac{x - y}{h} \right) f(y) f(x) dy dx + \int f^2(x) dx. \end{aligned}$$

En utilisant du théorème de Fubini, il s'en suit que

$$\begin{aligned} M_h &= \frac{1}{T^2 h^2} \iint \left\{ \int K \left(\frac{x - y}{h} \right) K \left(\frac{x - z}{h} \right) dx \right\} \left\{ \int_0^T \int_0^T g_{t,s}(y, z) dt ds \right\} dy dz \\ &\quad + \frac{1}{h} \iint K^{(2)} \left(\frac{y - z}{h} \right) f(y) f(z) dy dz \\ &\quad - \frac{2}{h} \iint K \left(\frac{x - y}{h} \right) f(y) f(x) dy dx + \int f^2(x) dx. \end{aligned}$$

En décomposant l'espace \mathbb{R}^2 en deux sous-espaces $\{|y - z| \leq a_T^\alpha\}$ et $\{|y - z| > a_T^\alpha\}$ où $\alpha > 1$, et en procédant à un changement de variables adéquat et un développement de Taylor, nous obtenons

$$\begin{aligned} M_h &= \frac{1}{T^2 h} \int K^2 \iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T \int_0^T g_{t,s}(y, z) dt ds \right\} dy dz \\ &+ \frac{1}{T^2 h^2} \iint_{|y-z| \leq a_T^\alpha} \left\{ \int K \left(\frac{x-y}{h} \right) \left[K \left(\frac{x-z}{h} \right) - K \left(\frac{x-y}{h} \right) \right] dx \right\} \\ &\times \left\{ \int_0^T \int_0^T g_{t,s}(y, z) dt ds \right\} dy dz + \frac{1}{T^2 h} \iint_{|y-z| > a_T^\alpha} K^{(2)} \left(\frac{y-z}{h} \right) \left\{ \int_0^T \int_0^T g_{t,s}(y, z) dt ds \right\} \\ &dy dz + h^2 \left\{ \int z^2 \left(\frac{1}{2} K^{(2)} - K \right)(z) dz \int f'' f \right\} + o(h^2). \end{aligned}$$

De façon similaire à Bosq (1996, p 104), par la stationnarité du processus, nous obtenons

$$\begin{aligned} \int_0^T \int_0^T g_{t,s}(y, z) dt ds &= \int_0^T \left\{ \int_{\{0 \leq s \leq t \leq T\}} g_{t,s}(y, z) ds \right\} dt + \int_0^T \left\{ \int_{\{s > t\}} g_{t,s}(y, z) dt \right\} ds \\ &= 2 \int_0^T \left\{ \int_0^s g_{t,s}(y, z) dt \right\} ds \\ &= 2 \int_0^T [T - u] g_u(y, z) du. \end{aligned} \tag{2.3}$$

Par conséquent,

$$\begin{aligned} M_h &= \frac{2}{T^2 h} \int K^2 \iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T [T - u] g_u(y, z) du \right\} dy dz \\ &+ \frac{2}{T^2 h^2} \iint_{|y-z| \leq a_T^\alpha} \left\{ \int K \left(\frac{x-y}{h} \right) \left[K \left(\frac{x-z}{h} \right) - K \left(\frac{x-y}{h} \right) \right] dx \right\} \\ &\times \left\{ \int_0^T [T - u] g_u(y, z) du \right\} dy dz + \frac{2}{T^2 h} \iint_{|y-z| > a_T^\alpha} K^{(2)} \left(\frac{y-z}{h} \right) \left\{ \int_0^T [T - u] g_u(y, z) du \right\} \\ &dy dz + h^2 \left\{ \int z^2 \left(\frac{1}{2} K^{(2)} - K \right)(z) dz \int f'' f \right\} + o(h^2) \\ &:= \frac{2}{T^2 h} \int K^2 \iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T [T - u] g_u(y, z) du \right\} dy dz + A_T + B_T \\ &+ h^2 \left\{ \int z^2 \left(\frac{1}{2} K^{(2)} - K \right)(z) dz \int f'' f \right\} + o(h^2). \end{aligned}$$

En utilisant la propriété de Hölder-continuité de K , le terme A_T est contrôlé par

$$\begin{aligned}
|A_T| &\leq 2D \left(\frac{a_T^\alpha}{h}\right)^\lambda \left| \frac{1}{T^2 h} \iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T [T-u] g_u(y, z) du \right\} dy dz \right| \\
&\leq 2D a_T^{(\alpha-1)\lambda} \left| \frac{1}{T^2 h} \iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T [T-u] g_u(y, z) du \right\} dy dz \right|.
\end{aligned}$$

Par conséquent,

$$A_T = o \left(\frac{1}{T^2 h} \iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T [T-u] g_u(y, z) du \right\} dy dz \right).$$

En supposant la fonction $u \rightarrow \|g_u\|_\infty$ intégrable et l'hypothèse (H1)(a) satisfaite, nous obtenons

$$\begin{aligned}
|B_T| &\leq \frac{4}{T^2 h} \iint_{|y-z| > a_T^\alpha} K^{(2)} \left(\frac{y-z}{h} \right) \left\{ \int_0^T T \|g_u\|_\infty du \right\} dy dz \\
&\leq \frac{4h}{T} \int_0^\infty \|g_u\|_\infty du \iint K^{(2)}(u-v) dudv \\
&= O \left(\frac{h}{T} \right).
\end{aligned}$$

Par conséquent, il ressort que

$$\begin{aligned}
M_h &= 2 \int K^2 \left\{ \frac{1}{T^2 h} \iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T [T-u] g_u(y, z) du \right\} dy dz \right\} \\
&\quad + h^2 \left\{ \int z^2 \left(\frac{1}{2} K^{(2)} - K \right)(z) dz \int f'' f \right\} \\
&\quad + o \left(\frac{1}{T^2 h} \iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T [T-u] g_u(y, z) du \right\} dy dz + h^2 \right).
\end{aligned} \tag{2.4}$$

Nous devons alors résoudre l'équation $M'_h = 0$, où M'_h est la dérivée première de M_h . Notons que sous les hypothèses (H1)(c), (H2)(b) et (H3)(c), la dérivée seconde M''_h de M_h est positive. \square

2.3.1 Preuve de la proposition 2.1

Nous estimons tout d'abord la quantité $\int f_{T,h} f$. Dans la suite, posons $\Gamma = \cup_{i=1}^n [T_{i-1}, T_i]^2$, Γ^c est le complémentaire de Γ dans $[0, T]^2$. Comme

$$\int f_{T,h}(x) f(x) dx = \int \left\{ \frac{1}{Th} \int_0^T K \left(\frac{x - X_t}{h} \right) dt \right\} f(x) dx,$$

il en découle directement alors que

$$\mathbb{E} \left[\int f_{T,h}(x) f(x) dx \right] = \frac{1}{h} \iint K \left(\frac{x-y}{h} \right) f(y) f(x) dy dx.$$

D'autre part, sous l'hypothèse (H3)(a), nous avons

$$\begin{aligned} \mathbb{E} \left[\frac{1}{T^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int_{T_{i-1}}^{T_i} \int_{T_{j-1}}^{T_j} K_h(X_t - X_s) ds dt \right] &= \mathbb{E} \left[\frac{1}{T^2} \int_{\Gamma^c} K_h(X_t - X_s) ds dt \right] \\ &= \frac{1}{T^2 h} \int_{\Gamma^c} \left\{ \iint K \left(\frac{x-y}{h} \right) f_{s,t}(x, y) dx dy \right\} ds dt \\ &= \frac{1}{T^2 h} \int_{\Gamma^c} \left\{ \iint K \left(\frac{x-y}{h} \right) g_{s,t}(x, y) dx dy \right\} ds dt \\ &+ \frac{1}{T^2 h} \int_{\Gamma^c} \left\{ \iint K \left(\frac{x-y}{h} \right) f(x) f(y) dx dy \right\} ds dt \\ &:= J_1 + \frac{1}{T^2 h} \int_{\Gamma^c} \left\{ \iint K \left(\frac{x-y}{h} \right) f(x) f(y) dx dy \right\} ds dt \\ &= J_1 + \left(\frac{n-1}{n} \right) \frac{1}{h} \iint K \left(\frac{x-y}{h} \right) f(x) f(y) dx dy. \end{aligned}$$

En utilisant l'inégalité de Hölder avec $\frac{1}{p} + \frac{1}{q} = 1$ et $p \in]2, \infty[$ et en considérant la condition (H1)(a), nous obtenons

$$\begin{aligned} |J_1| &\leq \frac{1}{T^2 h} \int_{\Gamma^c} \left\{ \iint K \left(\frac{x-y}{h} \right) |g_{s,t}(x, y)| dx dy \right\} ds dt \\ &\leq \frac{1}{T^2 h} \int_{\Gamma^c} \left\{ \left[\iint K^q \left(\frac{x-y}{h} \right) dx dy \right]^{\frac{1}{q}} \|g_{s,t}\|_{L^p(\mathbb{R}^2)} \right\} ds dt \\ &\leq \frac{\Delta_p(\Gamma)}{T^2 h} \left[\iint K^q \left(\frac{x-y}{h} \right) dx dy \right]^{\frac{1}{q}} \times \int_{\Gamma^c} ds dt. \end{aligned}$$

Les changement de variable $u = \frac{x}{h}$ et $v = \frac{y}{h}$, et la condition (H3)(b) permettent d'obtenir

$$|J_1| \leq \frac{n-1}{n} \Delta_p(\Gamma) h^{\frac{2}{q}-1} \left[\iint K^q(u-v) du dv \right]^{\frac{1}{q}} \longrightarrow 0, \quad p.s., \quad \text{lorsque } h \rightarrow 0.$$

Par conséquent, nous avons

$$\begin{aligned} \mathbb{E} \left[\frac{1}{T^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int_{T_{i-1}}^{T_i} \int_{T_{j-1}}^{T_j} K_h(X_t - X_s) ds dt \right] &= \frac{1}{h} \iint K \left(\frac{x-y}{h} \right) f(x) f(y) dx dy + o(1) \\ &= \mathbb{E} \left[\int f_{T,h}(x) f(x) dx \right] + o(1). \end{aligned}$$

Ainsi,

$$\lim_{T \rightarrow \infty} \mathbb{E} \left[L_{T,h} - \int f^2(x) dx \right] = \lim_{T \rightarrow \infty} \mathbb{E} [M_{T,h}].$$

2.3.2 Preuve du théorème 2.1

La preuve est répartie en plusieurs lemmes pour faciliter sa compréhension. Pour cela, considérons f_h comme le produit de convolution de K_h avec f , ie,

$$f_h(x) = \int K_h(x-y) f(y) dy.$$

Observons que

$$\begin{aligned} \mathbb{E} [f_{T,h}(x)] &= \mathbb{E} \left[\frac{1}{Th} \int_0^T K \left(\frac{x - X_t}{h} \right) dt \right] \\ &= \frac{1}{Th} \int_0^T \mathbb{E} \left[K \left(\frac{x - X_t}{h} \right) \right] dt \\ &= \int K_h(x-y) f(y) dy \\ &= f_h(x). \end{aligned}$$

L'expression de f_h est identique au cas discret. Rappelons le Lemme 1 de *Stone* (1984).

Lemma 2.1 (Stone-1984). Pour $h > 0$ relativement petit, il existe deux constantes C et r , tel que

$$\begin{aligned} \|f_h - f\|^2 &:= \int [f_h(x) - f(x)]^2 dx \\ &\geq Ch^r + \frac{1}{Th}. \end{aligned}$$

Démonstration. Voir *Stone* (1984). □

Dans un premier temps, nous établirons le théorème pour h appartenant à un ensemble fini H'_T contenu dans H_T , défini par

$$H'_T := \{h_j\}_{1 \leq j \leq N}, \quad \text{où } h_0 := a_T, \quad h_N := b_T, \quad h_j := a_T(1 + \nu_T)^j, \quad \text{où } \nu_T > 0.$$

Par un argument de Hölder-continuité, nous étendons le résultat à l'ensemble H_T . Posons maintenant

$$J_{Th} = \|f_h - f\|^2 + \frac{1}{Th}$$

et

$$J_{Th_r} = Ch^r + \frac{1}{Th}.$$

Lemma 2.2. Supposons les hypothèses (H1)(a), (H2)(a), (H4), (H5) et (H6)(a) satisfaites. Alors, il existe r , $0 < r < 1$, tel que

$$\lim_{T \rightarrow \infty} \max_{h \in H'_T} \left| \frac{\int f_{T,h}(x) [f_{T,h}(x) - f(x)] dx}{J_{Th_r}} \right| = 0, \quad p.s.$$

Démonstration. Sous les hypothèses (H2)(a), (H4) et (H5), le théorème 1.2 de Didi & Louani (2013), nous permet d'avoir l'assertion suivante

$$f_{T,h}(x) - f(x) = O(h) + O_{p.s} \left[\frac{\log T}{Th} \right]^{\frac{1}{2}}, \quad \text{lorsque } T \rightarrow \infty.$$

Par conséquent, comme $\int f_{Th}(x) dx = 1$, nous avons

$$\int f_{T,h}(x) [f_{T,h}(x) - f(x)] dx = O(h) + O_{p.s} \left[\frac{\log T}{Th} \right]^{\frac{1}{2}}.$$

Comme H'_T est un ensemble fini, le maximum est atteint. Alors, il existe $\tilde{h} \in H'_T$ tel que

$$\begin{aligned} \max_{h \in H'_T} \left| \frac{\int f_{T,h}(x) [f_{T,h}(x) - f(x)] dx}{J_{Th_r}} \right| &= \left| \frac{\int f_{T,\tilde{h}}(x) [f_{T,\tilde{h}}(x) - f(x)] dx}{J_{T\tilde{h}_r}} \right| \\ &= \frac{O(\tilde{h}) + O_{p.s} \left[\frac{\log T}{T\tilde{h}} \right]^{\frac{1}{2}}}{C\tilde{h}^r + \frac{1}{T\tilde{h}}} \\ &= \frac{T\tilde{h} \times \left[O(\tilde{h}) + O_{p.s} \left[\frac{\log T}{T\tilde{h}} \right]^{\frac{1}{2}} \right]}{CT\tilde{h}^{r+1} + 1}, \end{aligned}$$

qui, au vu de la condition (H6)(a), tend presque sûrement vers zéro lorsque $T \rightarrow \infty$. \square

Lemma 2.3. Supposons les hypothèses (H1)(a), (H2)(a), (H4), (H5) et (H6)(a) satisfaites. Alors, il existe r , $0 < r < 1$, tel que

$$\lim_{T \rightarrow \infty} \max_{h \in H'_T} \left| \frac{\int [f_{T,h}(x) - f_h(x)][f_h(x) - f(x)] dx}{J_{Th_r}} \right| = 0, \quad p.s.$$

Démonstration. Considérons la quantité $f_h - f$. Par un développement de Taylor et la condition H2(a), il ressort que

$$\begin{aligned} f_h(x) - f(x) &= \int K(z)f(x - hz)dz - f(x) \\ &= \int K(z) \{f(x - hz) - f(x)\} dz \\ &= O(h). \end{aligned}$$

Par les conditions (H2)(a), (H4) et (H5), nous avons

$$\begin{aligned} \left| \int [f_{T,h}(x) - f_h(x)][f_h(x) - f(x)] dx \right| &\leq \int \{|f_{T,h}(x) - f(x)| + |f(x) - f_h(x)|\} \\ &\quad \times |f_h(x) - f(x)| dx \\ &= \left\{ O(h) + O_{p.s.} \left[\frac{\log T}{Th} \right]^{\frac{1}{2}} \right\} \times \int |f_h(x) - f(x)| dx. \end{aligned}$$

Le résultat s'en suit par le théorème 1 de Devroye & Györfi (1985), page 6. \square

Afin de vérifier l'optimalité asymptotique de la fenêtre \hat{h} , il suffit de montrer que

$$\lim_{T \rightarrow \infty} \max_{h, h' \in H'_T} \left| \frac{L_{T,h} - L_{T,h'} - (M_{T,h} - M_{T,h'})}{L_{T,h} + L_{T,h'}} \right| = 0, \quad p.s.$$

Dans ce but, nous devons montrer simultanément

$$(A) \quad \liminf_T \min_{h \in H'_T} \left(\frac{L_{Th}}{J_{Th}} \right) > 0, \quad p.s$$

et

$$(B) \quad \lim_{T \rightarrow \infty} \max_{h, h' \in H'_T} \left| \frac{L_{T,h} - L_{T,h'} - (M_{T,h} - M_{T,h'})}{J_{Th} + J_{Th'}} \right| = 0, \quad p.s.$$

Considérons tout d'abord l'assertion **(A)** et observons que

$$\begin{aligned} L_{T,h} &= \int [f_{T,h}(x) - f(x)]^2 dx \\ &= \int [f_{T,h}(x) - f_h(x)]^2 dx + \|f_h - f\|^2 - 2 \int [f_{T,h}(x) - f_h(x)] [f_h(x) - f(x)] dx \\ &\geq \int [f_{T,h}(x) - f_h(x)]^2 dx + \|f_h - f\|^2 - 2 \max_{h \in H'_T} \left| \int [f_{T,h}(x) - f_h(x)] [f_h(x) - f(x)] dx \right|. \end{aligned}$$

Par conséquent,

$$\begin{aligned} \frac{L_{T,h}}{J_{Th}} &\geq \frac{\int [f_{T,h}(x) - f_h(x)]^2 dx + \|f_h - f\|^2 - 2 \max_{h \in H'_T} \left| \int [f_{T,h}(x) - f_h(x)] [f_h(x) - f(x)] dx \right|}{J_{Th}} \\ &\geq \frac{\|f_h - f\|^2}{J_{Th}} - 2 \frac{\max_{h \in H'_T} \left| \int [f_{T,h}(x) - f_h(x)] [f_h(x) - f(x)] dx \right|}{J_{Thr}}. \end{aligned}$$

En utilisant le Lemme 2.3, nous obtenons (A).

Pour établir l'assertion **(B)**, observons que

$$\begin{aligned} &\lim_{T \rightarrow \infty} \max_{h, h' \in H'_T} \frac{|L_{T,h} - L_{T,h'} - (M_{T,h} - M_{T,h'})|}{J_{Th} + J_{Th'}} \\ &= \lim_{T \rightarrow \infty} \max_{h, h' \in H'_T} \frac{|(L_{T,h} - M_{T,h} - \int f^2) - (L_{T,h'} - M_{T,h'} - \int f^2)|}{J_{Th} + J_{Th'}} \\ &\leq 2 \lim_{T \rightarrow \infty} \max_{h \in H'_T} \frac{|L_{T,h} - M_{T,h} - \int f^2|}{J_{Th}}. \end{aligned}$$

Pour évaluer la quantité $|L_{T,h} - M_{T,h} - \int f^2|$, notons que

$$\begin{aligned} &\left| \int [f_{T,h}(x) - f(x)]^2 dx - \int f_{T,h}^2(x) dx + \frac{2}{T^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int_{T_{i-1}}^{T_i} \int_{T_{j-1}}^{T_j} K_h(X_t - X_s) ds dt - \int f^2(x) dx \right| \\ &= 2 \left| - \int f_{T,h}(x) f(x) dx + \frac{1}{T^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int_{T_{i-1}}^{T_i} \int_{T_{j-1}}^{T_j} K_h(X_t - X_s) ds dt \right|. \end{aligned} \quad (2.5)$$

De façon similaire aux calculs produits dans le cas discret (voir *Silverman (1986), page 50*), il s'en suit que

$$\begin{aligned}
\int f_{T,h}^2(x)dx &= \int \left\{ \frac{1}{Th} \int_0^T K\left(\frac{x-X_t}{h}\right) dt \right\}^2 dx \\
&= \int \left\{ \frac{1}{T^2 h^2} \sum_{i=1}^n \sum_{j=1}^n \int_{T_{i-1}}^{T_i} \int_{T_{j-1}}^{T_j} K\left(\frac{x-X_t}{h}\right) K\left(\frac{x-X_s}{h}\right) ds dt \right\} dx \\
&= \frac{1}{T^2 h^2} \sum_{i=1}^n \sum_{j=1}^n \int_{T_{i-1}}^{T_i} \int_{T_{j-1}}^{T_j} \left\{ \int K\left(\frac{x-X_t}{h}\right) K\left(\frac{x-X_s}{h}\right) dx \right\} ds dt \\
&= \frac{1}{T^2 h} \sum_{i=1}^n \sum_{j=1}^n \int_{T_{i-1}}^{T_i} \int_{T_{j-1}}^{T_j} K^{(2)}\left(\frac{X_t-X_s}{h}\right) ds dt. \tag{2.6}
\end{aligned}$$

En utilisant les assertions (2.5) et (2.6), nous obtenons

$$\begin{aligned}
|L_{T,h} - M_{T,h} - \int f^2(x)dx| &\leq 2 \left| \int f_{T,h}(x) [f_{T,h}(x) - f(x)] dx \right| \\
&\quad + 2 \left| \frac{1}{T^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int_{T_{i-1}}^{T_i} \int_{T_{j-1}}^{T_j} \left\{ K_h(X_t - X_s) - K_h^{(2)}(X_t - X_s) \right\} ds dt \right| \\
&\quad + \frac{2}{T^2} \sum_{i=1}^n \int_{T_{i-1}}^{T_i} \int_{T_{i-1}}^{T_i} K_h^{(2)}(X_t - X_s) ds dt, \quad p.s.
\end{aligned}$$

Pour h très petit, nous avons $K_h^{(2)}(X_t - X_s) = K_h(X_t - X_s) = 0$ pour $t \in [T_{i-1}, T_i]$, $s \in [T_{j-1}, T_j]$ où $1 \leq i < j \leq n$ à l'exception d'un événement de probabilité négligeable.

Par conséquent,

$$|L_{T,h} - M_{T,h} - \int f^2(x)dx| \leq 2 \left[\int f_{T,h}(x) |f_{T,h}(x) - f(x)| dx + \frac{\delta \|K^{(2)}\|_\infty}{Th} \right], \quad p.s.$$

Ainsi, nous obtenons presque sûrement,

$$\lim_{T \rightarrow \infty} \max_{h \in H'_T} \frac{|L_{T,h} - M_{T,h} - \int f^2(x)dx|}{J_{Th}} \leq 2 \lim_{T \rightarrow \infty} \left\{ \max_{h \in H'_T} \frac{\int f_{T,h}(x) |f_{T,h}(x) - f(x)| dx}{J_{Th}} + \frac{2\delta \|K^{(2)}\|_\infty}{CTa_T^{r+1} + 1} \right\}.$$

Le résultat s'en suit immédiatement à partir du *Lemme 2.2* et de la condition (H6)(a).

Il reste maintenant à considérer le résultat sur l'ensemble H_T . Pour tout $h_j \in H'_T$, $1 \leq j \leq N$, posons

$$B_j := [h_{j-1}, h_j].$$

Comme $H_T = \cup_{j=1}^N B_j$, nous obtenons alors

$$\begin{aligned}
& \sup_{h, h' \in H_T} \frac{|L_{T,h} - L_{T,h'} - (M_{T,h} - M_{T,h'})|}{J_{Th} + J_{Th'}} = \max_{\substack{1 \leq i \leq N \\ 1 \leq j \leq N}} \sup_{\substack{h' \in B_i \\ h \in B_j}} \frac{|L_{T,h} - L_{T,h'} - (M_{T,h} - M_{T,h'})|}{J_{Th} + J_{Th'}} \\
& \leq 2 \max_{\substack{1 \leq i \leq N \\ 1 \leq j \leq N}} \sup_{\substack{h' \in B_i \\ h \in B_j}} \left| \frac{L_{T,h} - L_{T,h_j} - (M_{T,h} - M_{T,h_j})}{(J_{Th} - J_{Th_j}) + (J_{Th'} - J_{Th_i}) + (J_{Th_i} + J_{Th_j})} \right| \\
& + \max_{\substack{1 \leq i \leq N \\ 1 \leq j \leq N}} \sup_{\substack{h' \in B_i \\ h \in B_j}} \left| \frac{L_{T,h_i} - L_{T,h_j} - (M_{T,h_i} - M_{T,h_j})}{(J_{Th} - J_{Th_j}) + (J_{Th'} - J_{Th_i}) + (J_{Th_i} + J_{Th_j})} \right| \\
& := J_3 + J_4. \tag{2.7}
\end{aligned}$$

Lemma 2.4. Supposons les hypothèses (H1)(a) et (H1)(b) satisfaites. Pour tout $1 \leq j \leq N$, nous avons

$$\lim_{T \rightarrow \infty} \sup_{h \in B_j} \int |f_h(x) - f_{h_j}(x)| dx = 0.$$

Démonstration. Pour tout $1 \leq j \leq N$, observons que

$$\begin{aligned}
\sup_{h \in B_j} \int |f_h(x) - f_{h_j}(x)| dx &= \sup_{h \in B_j} \int \left| \int K_h(x-y) f(y) dy - \int K_{h_j}(x-y) f(y) dy \right| dx \\
&= \sup_{h \in B_j} \int \left| \int [K_h(x-y) - K_{h_j}(x-y)] f(y) dy \right| dx \\
&\leq \sup_{h \in B_j} \iint |K_h(x-y) - K_{h_j}(x-y)| f(y) dy dx \\
&\leq \sup_{h \in B_j} \iint \left[\frac{1}{h} \left| K\left(\frac{x-y}{h}\right) - K\left(\frac{x-y}{h_j}\right) \right| + \frac{h_j - h}{hh_j} K\left(\frac{x-y}{h_j}\right) \right] \\
&\quad f(y) dy dx.
\end{aligned}$$

Soit $[-V, V]$ le support de K . Pour tout $h \in B_j$, l'intervalle $[y - Vh, y + Vh]$ est contenu dans $[y - Vh_j, y + Vh_j]$. En utilisant la condition de bornitude et la propriété de Hölder-continuité de K , nous obtenons

$$\begin{aligned}
\sup_{h \in B_j} \int |f_h(x) - f_{h_j}(x)| dx &\leq \sup_{h \in B_j} \left\{ \int \left\{ \int_{y-Vh_j}^{y+Vh_j} \left[\frac{D}{h} \left| \frac{x-y}{h} - \frac{x-y}{h_j} \right|^\lambda + M \frac{h_j - h}{hh_j} \right] dx \right\} f(y) dy \right\} \\
&\leq R_1 \left\{ \frac{h_j [h_j - h_{j-1}]^\lambda}{h_{j-1}^{\lambda+1}} + \frac{h_j - h_{j-1}}{h_{j-1}} \right\} \\
&= R_1 \left\{ \nu_T^\lambda (1 + \nu_T) + \nu_T \right\} \longrightarrow 0, \quad \text{lorsque } T \rightarrow 0.
\end{aligned}$$

Ici R_1 est une constante positive. \square

Lemma 2.5. Sous les hypothèses (H1)(a) – (b), (H2)(a), (H4), (H5) et (H6), nous avons

$$\lim_{T \rightarrow \infty} \max_{\substack{1 \leq i \leq N \\ 1 \leq j \leq N}} \sup_{h \in B_j} \frac{|L_{T,h} - L_{T,h_j} - (M_{T,h} - M_{T,h_j})|}{J_{Th_i} + J_{Th_j}} = 0, \quad p.s.$$

Démonstration. Observons que

$$\begin{aligned} & |L_{T,h} - L_{T,h_j} - (M_{T,h} - M_{T,h_j})| \\ & \leq 2 \int |f_{T,h_j}(x) - f_{T,h}(x)| f(x) dx + \frac{2}{T^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int_{T_{i-1}}^{T_i} \int_{T_{j-1}}^{T_j} |K_h(X_t - X_s) - K_{h_j}(X_t - X_s)| ds dt \\ & \leq 2 \int [|f_{T,h_j}(x) - f(x)| + |f_{T,h}(x) - f(x)|] f(x) dx \\ & + \frac{2}{T^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int_{T_{i-1}}^{T_i} \int_{T_{j-1}}^{T_j} K \left(\frac{X_t - X_s}{h} \right) \left| \frac{1}{h} - \frac{1}{h_j} \right| + \frac{1}{h_j} \left| K \left(\frac{X_t - X_s}{h} \right) - K \left(\frac{X_t - X_s}{h_j} \right) \right| ds dt. \end{aligned}$$

En utilisant le Théorème 1.2 de Didi & Louani (2013) avec la bornitude du noyau K et sa Hölder-continuité, nous déduisons que

$$\begin{aligned} |L_{T,h} - L_{T,h_j} - (M_{T,h} - M_{T,h_j})| & = O(h_j) + O_{a.s.} \left[\frac{\log T}{Th_{j-1}} \right]^{\frac{1}{2}} \\ & + \frac{R_1}{2V} \times \frac{n-1}{n} \left\{ \frac{h_j - h_{j-1}}{h_j h_{j-1}} + \frac{1}{h_j} \left[\frac{h_j - h_{j-1}}{h_{j-1}} \right]^\lambda \right\}, \end{aligned}$$

ceci conduit à

$$\max_{\substack{1 \leq i \leq N \\ 1 \leq j \leq N}} \sup_{h \in B_j} \frac{|L_{T,h} - L_{T,h_j} - (M_{T,h} - M_{T,h_j})|}{J_{Th_i} + J_{Th_j}} \leq \tilde{A}_1 \left\{ \frac{b_T + \left[\frac{\log T}{T a_T} \right]^{\frac{1}{2}} + \frac{\nu_T + \nu_T^\lambda}{a_T}}{a_T^r + \frac{1}{T b_T}} \right\}, \quad p.s.$$

En choisissant la quantité ν_T telle que $\nu_T = o\left(a_T^{\frac{r+1}{\lambda}}\right)$ pour $0 < \lambda < 1$ et $\nu_T = o\left(a_T^{r+1}\right)$ pour $\lambda > 1$, le résultat suit immédiatement. \square

Lemma 2.6. Supposons les hypothèses (H1)(a) et (H1)(b) satisfaites. Alors, pour tout $1 \leq j \leq N$, nous avons

$$\lim_{T \rightarrow \infty} \frac{a_T}{\nu_T^\lambda + \nu_T} \sup_{h \in B_j} |J_{Th} - J_{Th_j}| = 0.$$

Démonstration. Pour tout $1 \leq j \leq N$ et tout $h \in B_j$, observons que

$$\begin{aligned}
 |J_{Th} - J_{Th_j}| &= \left| \|f_h - f\|^2 - \|f_{h_j} - f\|^2 + \frac{1}{Th} - \frac{1}{Th_j} \right| \\
 &= \left| \int [f_h(x) - f(x)]^2 dx - \int [f_{h_j}(x) - f(x)]^2 dx + \frac{h_j - h}{Thh_j} \right| \\
 &= \left| \int \{ [f_h(x) - f_{h_j}(x)][f_h(x) - f(x) + f_{h_j}(x) - f(x)] \} dx + \frac{h_j - h}{Thh_j} \right| \\
 &\leq \int \{ |f_h(x) - f_{h_j}(x)| [|f_h(x) - f(x)| + |f_{h_j}(x) - f(x)|] \} dx + \frac{h_j - h}{Thh_j} \\
 &\leq \int \{ |f_h(x) - f_{h_j}(x)| [|f_h(x) - f_{h_j}(x)| + 2|f_{h_j}(x) - f(x)|] \} dx + \frac{h_j - h}{Thh_j} \\
 &\leq \frac{R_1}{2V} \times \frac{\nu_T^\lambda + \nu_T}{a_T} \int [|f_h(x) - f_{h_j}(x)| + 2|f_{h_j}(x) - f(x)|] dx + \frac{\nu_T}{Ta_T}.
 \end{aligned}$$

De plus,

$$\sup_{h \in B_j} |J_{Th} - J_{Th_j}| \leq \frac{R_1}{2V} \times \frac{\nu_T^\lambda + \nu_T}{a_T} \left[\sup_{h \in B_j} \int |f_h(x) - f_{h_j}(x)| dx + 2 \int |f_{h_j}(x) - f(x)| dx \right] + \frac{\nu_T}{Ta_T}.$$

Ainsi,

$$\begin{aligned}
 \frac{a_T}{\nu_T^\lambda + \nu_T} \sup_{h \in B_j} |J_{Th} - J_{Th_j}| &\leq \frac{R_1}{2V} \left[\sup_{h \in B_j} \int |f_h(x) - f_{h_j}(x)| dx + 2 \int |f_{h_j}(x) - f(x)| dx \right] \\
 &\quad + \frac{\nu_T}{T[\nu_T + \nu_T^\lambda]} \\
 &\leq \frac{R_1}{2V} \left[\sup_{h \in B_j} \int |f_h(x) - f_{h_j}(x)| dx + 2 \int |f_{h_j}(x) - f(x)| dx \right] + \frac{1}{T}.
 \end{aligned}$$

Il suffit alors d'utiliser le Lemme 5 et le théorème 1 de *Devroye & Györfi* (1985), pour conclure. \square

En combinant les Lemmes 6 et 7, il est facile de constater que la quantité J_3 dans l'assertion (2.7) tend vers zéro lorsque T tend vers l'infini. En procédant de la même manière que l'assertion (B), on peut voir que J_4 dans (2.7) est asymptotiquement négligeable. Ceci conclut la preuve du théorème 1. \square

Chapitre 3

Vitesse de convergence du paramètre de lissage optimal en estimation de la densité d'un processus à temps continu

3.1 Introduction

Soit $(X_t)_{t \in [0, T]}$, $t \in \mathbb{R}$, un processus stationnaire ergodique de densité marginale f . L'estimateur à noyau de la fonction f est donné par

$$f_{T,h}(x) = \frac{1}{Th} \int_0^T K\left(\frac{x - X_t}{h}\right) dt,$$

où h est le paramètre de lissage appartenant à $H_T := [a_T, b_T] \subset \mathbb{R}^+$ et K est un noyau vérifiant les propriétés indiquées plus loin. L'EQI et l'EQMI sont définies respectivement par

$$L_{T,h} = \int [f_{T,h}(x) - f(x)]^2 dx$$

et

$$M_h = \mathbb{E} \int [f_{T,h}(x) - f(x)]^2 dx.$$

Le choix du paramètre de lissage est d'une importance majeure dans les problèmes d'estimation non paramétrique. Le choix approprié pour les besoins d'ordre pratique fait apparaître un paramètre aléatoire dépendant des données. De ce fait, l'EQI répond de façon plus adéquate à la problématique de choix pour laquelle la

méthode de validation croisée proposée par Rudemo (1982) et Bowman (1984) est utilisée. Rappelons que cette méthode repose sur la quantité $L_{T,h} - \int f^2(x)dx$ et permet une sélection automatique du paramètre de lissage. Elle est largement étudiée dans la littérature. L'optimalité asymptotique du paramètre de lissage issu de la méthode de la validation croisée a été établie dans le cadre de données indépendantes et identiquement distribuées par Stone (1984) sous des conditions très peu contraignantes. En utilisant l'équivalence asymptotique entre l'EQI et l'EQMI établie par Hall, P. (1982), il ressort que ce choix du paramètre de lissage est aussi asymptotiquement optimal pour l'EQMI, voir Hall, P. (1983). Dans le cadre de données dépendantes α -mélangeantes, l'optimalité du paramètre de lissage obtenu via la méthode de validation croisée a été établie par Hart & Vieu (1990). Le cadre associé aux processus stationnaires ergodiques à temps continu a été considéré dans le travail de El heda & Louani (2018) dans le quel ils proposent une procédure de sélection du paramètre de lissage optimal. Le critère de validation approprié au cas de processus à temps continus proposé dans le travail de El heda & Louani (2018) se présente sous la forme,

$$M_{T,h} = \int f_{T,h}^2(x)dx - \frac{2}{T^2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int_{T_{i-1}}^{T_i} \int_{T_{j-1}}^{T_j} K_h(X_s - X_t) ds dt.$$

Dans la suite, nous noterons h_0 , h_{opt} et \hat{h} les paramètres de lissage minimisant dans l'ensemble $H_T := [a_T, b_T]$ l'EQI, l'EQMI et $M_{T,h}$ respectivement. L'ordre de grandeur du paramètre de lissage optimal pour le critère EQMI dans le cas de données indépendantes et identiquement distribuées est $n^{-\frac{1}{5}}$. Il est établi dans El heda & Louani (2018) que cet ordre de grandeur dans le cas de processus à temps continus est $T^{-\frac{1}{3}}$. L'objet de ce chapitre est d'évaluer les distances $L_{T,h_{opt}} - L_{T,h_0}$, $L_{T,\hat{h}} - L_{T,h_0}$ et $L_{T,h_{opt}} - L_{T,\hat{h}}$ ainsi que les vitesses de convergence presque sûres de h_0 vers h_{opt} et \hat{h} vers h_{opt} . Notons que des résultats similaires ont été présentés pour des données indépendantes et identiquement distribuées par Hall, P. & Marron, J. (1987) dans le cas de l'estimation de la densité et par Rice, J. (1984) dans le cas de l'estimation de la régression. Dans le cas de l'estimation de la fonction taux de hasard, Esévez-pérez, quintela-del-río et Vieu (2002) ont présenté des résultats similaires pour des données α -mélangeantes.

3.2 Résultats

Afin de poser les hypothèses nécessaires à l'établissement de nos résultats, notons $K^{(2)}$ le produit de convolution de K avec lui même. Notons que $K^{(2)}$ a les mêmes propriétés que K . f_{X_s, X_t} désigne la densité conjointe du vecteur aléatoire (X_s, X_t) et f_{X_s} la densité marginale de X_s . Nous introduisons les notations suivantes

$$g_{X_s, X_t} := f_{X_s, X_t} - f_{X_s} f_{X_t}, \quad g_{0, u} := g_u, \quad \text{et} \quad \|\cdot\|_\infty := \sup_{(y, z) \in \mathbb{R}^2} |\cdot|.$$

Par ailleurs, pour tout $t \geq 0$, notons $\mathcal{F}_t := \sigma((X_s) : 0 \leq s \leq t)$ la σ -algèbre générée par les données $\{X_s : 0 \leq s \leq t\}$. On note \mathcal{F}_k la tribu engendrée par $\{X_s : 0 \leq s \leq T_k\}$, où $(T_k)_{0 \leq k \leq T}$ désigne une partition de l'intervalle $[0, T]$. Pour une variable aléatoire ξ et tout $k \in \mathbb{N}$, le projecteur \mathcal{P}_k est défini par $\mathcal{P}_k \xi = \mathbb{E}[\xi | \mathcal{F}_k] - \mathbb{E}[\xi | \mathcal{F}_{k-1}]$, où $\mathbb{E}[\xi | \mathcal{F}_k]$ est l'espérance conditionnelle de ξ sachant la σ -algèbre \mathcal{F}_k .

Les hypothèses nécessaires pour établir nos résultats sont posées ci-après

Hypothèses

(H1) (a) K est une fonction symétrique, bornée ($K \leq M$), de support compact et d'intégrale égale à un.

(b) K est Lipschitzienne, ie, pour tout $(x, y) \in \mathbb{R}^2$, il existe une constante $d > 0$, tel que $|K(x) - K(y)| \leq d|x - y|$.

(c) $\int z^2(K^{(2)}(z) - 2K(z))dz < 0$.

(H2) (a) f est deux fois dérivable avec la première dérivée et la seconde dérivée bornées.

(b) $\int f'' f < 0$.

(H3) Il existe un ensemble $\Gamma \in \mathcal{B}_{\mathbb{R}^2}$ contenant $D = \{(s, t) \in \mathbb{R}^2, s = t\}$ tel que

(a) $g_{s, t}$ existe pour tout $(s, t) \notin \Gamma$.

(b) $\Delta_p(\Gamma) := \sup_{(s, t) \notin \Gamma} \|g_{s, t}\|_{L^p(\mathbb{R}^2)} < \infty$, pour $p \in]2, \infty[$.

(c) $u \rightarrow \|g_u\|_\infty$ et $u \rightarrow \int \sup_y \left| \frac{\partial g_u(y, z)}{\partial y} \right| dz$ sont intégrables sur $]0, \infty[$.

(d) Il existe une fonction a_T et un réel $1 < \alpha < 2$, tels que $\iint_{|y-z| < a_T^\alpha} g_u(y, z) dy dz > 0$.

- (H4) (a) Pour tout $\delta > 0$, la densité conditionnelle $f_t^{\mathcal{F}_{t-\delta}}$ de X_t sachant la σ -algèbre $\mathcal{F}_{t-\delta}$ est différentiable de dérivée presque sûrement bornée.
- (b) Pour tout $t \in [0, T]$, et tout $\delta > 0$, la fonction $f_t^{\mathcal{F}_{t-\delta}}(x)$ est presque sûrement bornée par une fonction déterministe $b_{t,\delta}(x)$.
- (c) Pour tout $\delta > 0$, il existe un réel $m' > 1$, tel que $\frac{1}{T} \int_0^T \|b_{t,\delta}\|_{L^{m'}} dt \rightarrow D \neq 0$, lorsque $T \rightarrow \infty$.
- (H5) Pour tout $\delta > 0$, $\sup_x \int_{\mathbb{R}^+} \|\mathcal{P}_1 f_t^{\mathcal{F}_{t-\delta}}(x)\|^2 dt < \infty$.
- (H6) (a) Pour tout réel $0 < l < 1$, $Ta_T^l / \log T \rightarrow \infty$, lorsque $T \rightarrow \infty$.
- (b) Il existe un réel $0 < m < 1$, tel que $\frac{|h'| T^{\frac{1}{2}}}{h^{\frac{4-m}{2}} (\log T)^{\frac{1}{2}}} < \infty$, où h' est la dérivée de h par rapport à T .

Commentaires sur les Hypothèses.

Les conditions (H1) sont très standards en statistique non paramétrique. La condition (H1)(c) est similaire à la condition $K^{(2)}(0) - 2K(0) < 0$ introduite par Stone (1984), notons que le noyau gaussien satisfait cette hypothèse et qu'en général le noyau n'impacte pas significativement l'estimation. La condition (H2)(a) est une condition de régularité de la distribution du processus considéré. L'hypothèse (H2)(b) est satisfaite par le processus gaussien. Les deux hypothèses (H1)(d) et (H2)(b) assurent la positivité de h_{opt} . Les conditions (H3)(a) et (H3)(b) sont considérées dans la littérature pour atteindre des vitesses de convergence optimales, voir Bosq (1996) tandis que la condition (H3)(c) est similaire à celle de Bosq (1996, p 104) pour atteindre des vitesses de convergence super optimales de l'estimateur de la densité. Les hypothèses (H4) et (H5) sont nécessaires pour établir la vitesse de convergence de l'estimateur f_{Th} de la densité. Notons que la condition (H5) est similaire à celle de Wu (2003) et elle est satisfaite par aussitôt les processus linéaires que plusieurs processus non linéaires. La condition (H6)(c) est similaire à celle de Didi & Louani (2013). Notons que la condition (H6)(a) sur le paramètre de lissage est moins forte que celle de Didi & Louani (2013).

Après avoir donné la forme de h_{opt} dans le chapitre 2, la proposition suivante donne son l'ordre de grandeur.

Proposition 3.1. Sous les hypothèses (H1)(a)–(b), (H2), (H3)(a) et (H3)(c)–(d),

$$h_{opt} = \left\{ \frac{4 \int K^2 \int_0^\infty \int g_u(z, z) dz du}{\int z^2 (\frac{1}{2} K^{(2)} - K)(z) dz \int f'' f} \right\}^{\frac{1}{3}} \left\{ \frac{a_T^\alpha}{T} \right\}^{\frac{1}{3}} + o \left(\frac{a_T^\alpha}{T} \right)^{\frac{1}{3}}.$$

Afin d'établir une équivalence asymptotique entre l'EQI et l'EQMI, nous énonçons le théorème suivant.

Théorème 3.1. Sous les hypothèses (H1)(a), (H2)(a), (H3)(a)–(b), (H4), (H5) et (H6)(a), il existe $m < 1$ et $q \leq 2$ tel que,

$$L_{T,h} = M_h + O_{p.s.}(h)^{\min(\frac{2}{q}-1, 2)} + O_{p.s.} \left[\frac{\log T}{T h^{2-\frac{1}{m}}} \right]^{\frac{1}{2}}, \quad (3.1)$$

Remarque 3.1. Le choix de $h \in \left[\epsilon \left(\frac{a_T^\alpha}{T} \right)^{\frac{1}{3}}, \beta \left(\frac{a_T^\alpha}{T} \right)^{\frac{1}{3}} \right]$ vient du fait que dans la proposition 3.1, h_{opt} est de l'ordre de grandeur de $\left(\frac{a_T^\alpha}{T} \right)^{\frac{1}{3}}$.

Nous présentons maintenant le théorème suivant permettant de montrer que, choisir \hat{h} de l'ordre de grandeur de $\left(\frac{a_T^\alpha}{T} \right)^{\frac{1}{3}}$ pour minimiser le critère de la validation croisée est asymptotiquement équivalent à choisir h_0 du même ordre de grandeur pour minimiser l'EQI.

Théorème 3.2. Sous les hypothèses (H1), (H2), (H3)(a), (H3)(c)–(d) (H4), (H5) et (H6)(a), pour tout $h \in \left[\epsilon \left(\frac{a_T^\alpha}{T} \right)^{\frac{1}{3}}, \beta \left(\frac{a_T^\alpha}{T} \right)^{\frac{1}{3}} \right]$, nous avons

$$M_{T,h} = L_{T,h} - \frac{2}{T} \int_0^T f(X_t) dt + \int f^2(x) dx + O \left(\frac{a_T^\alpha}{T} \right)^{\frac{2}{3}} + O_{p.s.} \left[\frac{\log T}{\left(T^{1+\frac{1}{m}} a_T^{\alpha(2-\frac{1}{m})} \right)^{\frac{1}{3}}} \right]^{\frac{1}{2}},$$

pour tout $0 < \epsilon < \beta < \infty$.

Un estimateur \tilde{h} est dit asymptotiquement aussi "bon" que le paramètre de lissage optimal h_{opt} , si

$$\frac{L_{T,\tilde{h}}}{M_{h_{opt}}} \longrightarrow 1, \quad p.s. \quad (3.2)$$

Cette définition a été présentée par Hall, (1983) pour le critère d'optimalité asymptotique en probabilité de l'estimateur du paramètre de lissage optimal. Le corollaire suivant compare h_0 à h_{opt} et \hat{h} à h_{opt} .

Corollaire 3.1. Sous les hypothèses (H1), (H2), (H3), (H4), (H5) et (H6)(a), pour tout $h \in \left[\epsilon \left(\frac{a_T^\alpha}{T} \right)^{\frac{1}{3}}, \beta \left(\frac{a_T^\alpha}{T} \right)^{\frac{1}{3}} \right]$, nous obtenons

$$\frac{L_{T,h_0}}{M_{h_{opt}}} \rightarrow 1, \quad p.s. \quad (3.3)$$

et

$$\frac{L_{T,\hat{h}}}{M_{h_{opt}}} \rightarrow 1, \quad p.s. \quad (3.4)$$

Corollaire 3.2. Soit $\tilde{h} \in \{h_0, \hat{h}\}$ un estimateur aussi "bon" que h_{opt} . Sous les conditions du théorème 3.1 et théorème 3.2, nous obtenons

$$\frac{\tilde{h}}{h_{opt}} \rightarrow 1, \quad p.s. \quad (3.5)$$

Les deux théorèmes suivants permettent de donner des vitesses de convergence respectives de h_0 et \hat{h} .

Théorème 3.3. Sous les hypothèses (H1), (H2), (H3), (H4), (H5), (H6)(a) et H6(c), il existe $q < 1$ et $m < 1$, tels que

$$h_0 - h_{opt} = O\left(\frac{a_T^\alpha}{T}\right)^{\min\{\frac{2}{3q} - \frac{2}{3}, \frac{1}{3}\}} + O_{p.s.} \left[\frac{\log T}{\left(T^{\frac{1}{m}-1} a_T^{\alpha(4-\frac{1}{m})}\right)^{\frac{1}{3}}} \right]^{\frac{1}{2}}. \quad (3.6)$$

- Pour $q < \frac{2}{3}$ et $m < \frac{1}{3}$ et par un développement de Taylor de $L_{T,h_{opt}}$, nous avons

$$\begin{aligned} L_{T,h_{opt}} &= L_{T,h_0} + \frac{1}{2} (h_{opt} - h_0)^2 L''_{T,h^{**}} \\ &= L_{T,h_0} + \frac{1}{2} (h_{opt} - h_0)^2 \{M''_{h^{**}} + O(1)\}, \end{aligned} \quad (3.7)$$

où h^{**} est compris entre h_{opt} et h_0 . Par conséquent,

$$L_{T,h_{opt}} - L_{T,h_0} = O\left(\frac{a_T^\alpha}{T}\right)^{\frac{2}{3}} + O_{p.s} \left[\frac{\log T}{\left(T^{\frac{1}{m}+1} a_T^{\alpha(2-\frac{1}{m})}\right)^{\frac{1}{3}}} \right]^{\frac{1}{2}} + O_{p.s} \left[\frac{\log T}{\left(T^{\frac{1}{m}-1} a_T^{\alpha(4-\frac{1}{m})}\right)^{\frac{1}{3}}} \right]. \quad (3.8)$$

Théorème 3.4. Sous les hypothèses (H1), (H2), (H3)(a) – (b), (H4), (H5) et (H6)(a), il existe $q < 1$ et $m < 1$, tels que

$$\hat{h} - h_{opt} = O\left(\frac{a_T^\alpha}{T}\right)^{\min\{\frac{2}{3q}-\frac{2}{3}, \frac{1}{3}\}} + O_{p.s} \left[\frac{\log T}{\left(T^{\frac{1}{m}-1} a_T^{\alpha(4-\frac{1}{m})}\right)^{\frac{1}{3}}} \right]^{\frac{1}{2}}. \quad (3.9)$$

• Pour $q < \frac{2}{3}$ et $m < \frac{1}{3}$ et par un développement de Taylor de $L_{T,\hat{h}}$ et $L_{T,h_{opt}}$, nous obtenons

$$\begin{aligned} L_{T,\hat{h}} - L_{T,h_0} &= L_{T,\hat{h}} - L_{T,h_{opt}} + L_{T,h_{opt}} - L_{T,h_0} \\ &= \frac{1}{2} (\hat{h} - h_{opt})^2 L''_{T,h_1} + \frac{1}{2} (h_{opt} - h_0)^2 L''_{T,h_2}. \end{aligned} \quad (3.10)$$

Où h_1 est compris entre \hat{h} et h_{opt} et h_2 est compris entre h_{opt} et h_0 .

$$L_{T,\hat{h}} - L_{T,h_0} = \frac{1}{2} (\hat{h} - h_{opt})^2 \{M''_{h_1} + O(1)\} + \frac{1}{2} (h_{opt} - h_0)^2 \{M''_{h_2} + O(1)\}.$$

En combinant les assertions (3.6) et (3.9), nous obtenons

$$L_{T,\hat{h}} - L_{T,h_0} = O\left(\frac{a_T^\alpha}{T}\right)^{\frac{2}{3}} + O_{p.s} \left[\frac{\log T}{\left(T^{\frac{1}{m}+1} a_T^{\alpha(2-\frac{1}{m})}\right)^{\frac{1}{3}}} \right]^{\frac{1}{2}} + O_{p.s} \left[\frac{\log T}{\left(T^{\frac{1}{m}-1} a_T^{\alpha(4-\frac{1}{m})}\right)^{\frac{1}{3}}} \right]. \quad (3.11)$$

3.3 Preuves

3.3.1 Preuve de la Proposition 3.1

En utilisant un résultat de EL HEDA & LOUANI (2018),

$$h_{opt} = C(f, K) \left\{ \frac{\iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T [T-u] g_u(y, z) du \right\} dy dz}{T^2} \right\}^{\frac{1}{3}},$$

$$\text{où } C(f, K) := \left\{ \frac{2 \int K^2}{\int z^2 (\frac{1}{2} K^{(2)} - K)(z) dz \int f'' f} \right\}^{\frac{1}{3}}.$$

Par un développement de Taylor de $g_u(\cdot, z)$, nous obtenons

$$\begin{aligned} h_{opt} &= C(f, K) \left\{ \int_0^T \frac{[T-u]}{T^2} \int \left\{ \int_{-a_T^\alpha+z}^{a_T^\alpha+z} g_u(z, z) + (y-z) \frac{\partial g_u(y, z)}{\partial y} \Big|_{y=y_T^*(z)} dy \right\} dz du \right\}^{\frac{1}{3}} \\ &= C(f, K) \left\{ \int_0^T \frac{[T-u]}{T^2} \int \left\{ 2a_T^\alpha g_u(z, z) + \int_{-a_T^\alpha+z}^{a_T^\alpha+z} (y-z) \frac{\partial g_u(y, z)}{\partial y} \Big|_{y=y_T^*(z)} dy \right\} dz du \right\}^{\frac{1}{3}} \\ &= C(f, K) \left\{ \frac{2a_T^\alpha}{T} \int_0^T \frac{[T-u]}{T} \int g_u(z, z) dz du \right. \\ &\quad \left. + \int_0^T \frac{[T-u]}{T^2} \int \left\{ \int_{-a_T^\alpha+z}^{a_T^\alpha+z} (y-z) \frac{\partial g_u(y, z)}{\partial y} \Big|_{y=y_T^*(z)} dy \right\} dz du \right\}^{\frac{1}{3}}, \end{aligned}$$

où $y_T^*(z)$ est compris entre y et z . Observons que

$$\begin{aligned} &\int_0^T \frac{[T-u]}{T^2} \int \left\{ \int_{-a_T^\alpha+z}^{a_T^\alpha+z} (y-z) \frac{\partial g_u(y, z)}{\partial y} \Big|_{y=y_T^*(z)} dy \right\} dz du \\ &\leq \int_0^T \frac{[T-u]}{T^2} \int \left\{ \int_{-a_T^\alpha+z}^{a_T^\alpha+z} |y-z| \sup_y \left| \frac{\partial g_u(y, z)}{\partial y} \right| dy \right\} dz du \\ &\leq a_T^\alpha \int_0^T \frac{[T-u]}{T^2} \int \left\{ \sup_y \left| \frac{\partial g_u(y, z)}{\partial y} \right| \left\{ \int_{-a_T^\alpha+z}^{a_T^\alpha+z} dy \right\} dz \right\} du \\ &= 2a_T^{2\alpha} \int_0^T \frac{[T-u]}{T^2} \int \sup_y \left| \frac{\partial g_u(y, z)}{\partial y} \right| dz du \\ &= \frac{2a_T^{2\alpha}}{T} \int_0^T \frac{[T-u]}{T} \int \sup_y \left| \frac{\partial g_u(y, z)}{\partial y} \right| dz du \\ &\leq \frac{2a_T^{2\alpha}}{T} \int_0^\infty \int \sup_y \left| \frac{\partial g_u(y, z)}{\partial y} \right| dz du. \end{aligned}$$

En utilisant l'intégrabilité de $u \rightarrow \int \sup_y \left| \frac{\partial g_u(y, z)}{\partial y} \right| dz$ sur $]0, \infty[$,

$$\int_0^T \frac{[T-u]}{T^2} \int \left\{ \int_{-a_T^\alpha+z}^{a_T^\alpha+z} (y-z) \frac{\partial g_u(y, z)}{\partial y} \Big|_{y=y_T^*(z)} dy \right\} dz du = O\left(\frac{a_T^{2\alpha}}{T}\right). \quad (3.12)$$

Par ailleurs, nous avons

$$\begin{aligned} \left| \int_0^\infty \int g_u(z, z) dz du - \int_0^T \frac{[T-u]}{T} \int g_u(z, z) dz du \right| &= \left| \int_T^\infty \int g_u(z, z) dz du + \int_0^T \frac{u}{T} \int g_u(z, z) dz du \right| \\ &\leq \int_T^\infty \left| \int g_u(z, z) dz \right| du + \int_0^T \frac{u}{T} \left| \int g_u(z, z) dz \right| du. \end{aligned} \quad (3.13)$$

Par l'intégrabilité de $u \rightarrow \left| \int g_u(z, z) dz \right|$ sur $]0, \infty[$ et l'usage du théorème de convergence dominée de Lebesgue, la borne (3.13) tend vers zéro. Il s'en suit directement des assertions (3.12) et (3.13), que

$$h_{opt} = C(f, K) \left\{ \int_0^\infty \int g_u(z, z) dz du \right\}^{\frac{1}{3}} \left\{ \frac{2a_T^\alpha}{T} \right\}^{\frac{1}{3}} + o\left(\frac{a_T^\alpha}{T}\right)^{\frac{1}{3}}.$$

Afin de démontrer nos résultats, nous avons besoin du lemme suivant établissant des résultats de vitesses de convergence ponctuelles de $f_{T,h}$ plus rapides que celles présentées par Didi & Louani.

Lemma 3.1. Supposons les hypothèses (H1)-(a), (H2)-(a), (H4), (H5) et (H6)(a) satisfaites, il existe $m < 1$, tel que, pour tout $x \in \mathbb{R}$, nous avons

$$f_{T,h}(x) - f(x) = O(h^2) + O_{p.s} \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}}. \quad \text{lorsque } T \rightarrow \infty. \quad (3.14)$$

Démonstration. La preuve de ce Lemme suit le même raisonnement que Didi & Louani.

Pour tout $x \in \mathbb{R}$,

$$\begin{aligned} f_{T,h}(x) - f(x) &= \frac{1}{Th} \int_0^T K\left(\frac{x - X_t}{h}\right) - \mathbb{E} \left[K\left(\frac{x - X_t}{h}\right) \middle| \mathcal{F}_{t-\delta} \right] dt \\ &+ \frac{1}{Th} \int_0^T \mathbb{E} \left[K\left(\frac{x - X_t}{h}\right) \middle| \mathcal{F}_{t-\delta} \right] dt - f(x) \\ &:= S_1^T + S_2^T. \end{aligned}$$

Pour traiter le terme S_1^T , nous faisons appel au lemme de De la peña & Giné[8] sur les inégalités exponentielles.

Lemma 3.2. Soit $(W_n)_{n \geq 1}$ une différence de martingales par rapport à la σ -algèbre $(\mathcal{F}_n)_{n \geq 1}$ générée par les variables aléatoires W_1, \dots, W_n . Pour tout $l \geq 2$ et tout $n \geq 1$, tel qu'il existe deux constantes positives C et d_n vérifiant

$$\mathbb{E} \left[W_n^l \middle| \mathcal{F}_{n-1} \right] \leq C^{l-2} l! d_n^2 \quad p.s,$$

alors, pour tout $\epsilon > 0$, nous avons

$$\mathbb{P} \left(\left| \sum_{i=1}^n W_i \right| > \epsilon \right) \leq 2 \exp \left\{ -\frac{\epsilon^2}{2(D_n + C\epsilon)} \right\}, \quad (3.15)$$

Où $D_n = \sum_{i=1}^n d_i^2$.

Soit Y_i la différence de martingales donnée par

$$Y_i := \int_{T_{i-1}}^{T_i} K \left(\frac{x - X_t}{h} \right) - \mathbb{E} \left[K \left(\frac{x - X_t}{h} \right) \middle| \mathcal{F}_{t-\delta} \right] dt,$$

notons que pour tout $\delta > 0$, Y_i est une différence de martingales par rapport aux sigma-algèbres $(\mathcal{G}_{i-1})_{1 \leq i \leq n}$ tel que, pour tout $t \in [T_{i-1}, T_i]$, $\mathcal{G}_{i-2} \subset \mathcal{F}_{t-\delta} \subset \mathcal{G}_{i-1}$. En utilisant les inégalités de Jensen et Minkovski, voir Didi & Louani(2013), nous obtenons

$$\mathbb{E} [Y_i^l | \mathcal{G}_{i-2}] \leq 2^l \int_{T_{i-1}}^{T_i} \mathbb{E} \left[K^l \left(\frac{x - X_t}{h} \right) \middle| \mathcal{G}_{i-2} \right] dt. \quad (3.16)$$

En utilisant l'inégalité de Hölder avec $\frac{1}{m} + \frac{1}{m'} = 1$ et un changement de variable $z = \frac{x - y}{h}$, nous obtenons,

$$\begin{aligned} \mathbb{E} \left[K^l \left(\frac{x - X_t}{h} \right) \middle| \mathcal{G}_{i-2} \right] &= \int K^l \left(\frac{x - y}{h} \right) f^{\mathcal{G}_{i-2}}(y) dy \\ &\leq \left(\int K^{lm} \left(\frac{x - y}{h} \right) dy \right)^{\frac{1}{m}} \times \|f^{\mathcal{G}_{i-2}}\|_{L^{m'}} \\ &= h^{\frac{1}{m}} \left(\int K^{lm}(z) dz \right)^{\frac{1}{m}} \times \|f^{\mathcal{G}_{i-2}}\|_{L^{m'}} \\ &\leq h^{\frac{1}{m}} \|K\|_{\infty}^{l-1} \|K\|_{L^m} \times \|b_{i-2}\|_{L^{m'}}. \end{aligned}$$

Il s'en suit que,

$$|\mathbb{E} [Y_i^l | \mathcal{G}_{i-2}]| \leq 2^l \delta h^{\frac{1}{m}} \|K\|_{\infty}^{l-1} \|K\|_{L^m} \|b_{i-2}\|_{L^{m'}}.$$

Posons $C = 2 \|K\|_{\infty}$, $d_i^2 = 2^3 \delta h^{\frac{1}{m}} \|K\|_{L^m} \|b_{i-2}\|_{L^{m'}}$ et

$$\begin{aligned} D_n &= \sum_{i=2}^n d_i^2, \\ &= 2^3 h^{\frac{1}{m}} \|K\|_{L^m} \sum_2^n \delta \|b_{i-2}\|_{L^{m'}}, \end{aligned}$$

c'est ainsi que,

$$|\mathbb{E} [Y_i^l]| \leq l! C^{l-2} d_i^2.$$

En approchant $\int_0^T \|b_t\|_{L^{m'}} dt$ par la somme de Riemann $\delta \sum_2^n \|b_{i-2}\|_{L^{m'}}$,

$$\begin{aligned} D_n &= 2^3 \delta \|K\|_{L^m} T h^{\frac{1}{m}} \times \frac{1}{T} \int_0^T \|b_t\|_{L^{m'}} dt \\ &= O \left(T h^{\frac{1}{m}} \right). \end{aligned}$$

En utilisant le lemme 3.2, il s'en suit que, pour tout $\epsilon > 0$,

$$\begin{aligned}
\mathbb{P} \left(\left| S_1^T \right| > \epsilon \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}} \right) &= \mathbb{P} \left(\left| \sum_{i=1}^n Y_i \right| > \epsilon Th \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}} \right) \\
&= \mathbb{P} \left(\left| \sum_{i=1}^n Y_i \right| > \epsilon Th^{\frac{1}{m}} \left[\frac{\log T}{Th^{\frac{1}{m}}} \right]^{\frac{1}{2}} \right) \\
&\leq 2 \exp \left\{ - \frac{\epsilon^2 T^2 h^{\frac{2}{m}} \left(\frac{\log T}{Th^{\frac{1}{m}}} \right)}{O \left(Th^{\frac{1}{m}} \right) + 2\epsilon C Th^{\frac{1}{m}} \left[\frac{\log T}{Th^{\frac{1}{m}}} \right]^{\frac{1}{2}}} \right\} \\
&= 2 \exp \left\{ -\epsilon^2 O \left(\log T \right) \right\}. \tag{3.17}
\end{aligned}$$

Par ailleurs, par les mêmes calculs que l'assertion (3.17), nous obtenons

$$\mathbb{P} \left(\frac{1}{(n+1)h} \left| \sum_{i=1}^n Y_i \right| > \epsilon \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}} \right) \leq 2T^{-C'\epsilon^2}$$

et

$$\mathbb{P} \left(\frac{1}{nh} \left| \sum_{i=1}^n Y_i \right| > \epsilon \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}} \right) \leq 2T^{-C''\epsilon^2}.$$

En plus, nous avons que

$$\begin{aligned}
\mathbb{P} \left(\frac{1}{(n+1)h} \left| \sum_{i=1}^n Y_i \right| > \epsilon \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}} \right) &\leq \mathbb{P} \left(\frac{1}{Th} \left| \sum_{i=1}^n Y_i \right| > \epsilon \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}} \right) \\
&\leq \mathbb{P} \left(\frac{1}{nh} \left| \sum_{i=1}^n Y_i \right| > \epsilon \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}} \right).
\end{aligned}$$

En choisissant ϵ convenable et par l'usage du lemme de Borel-Cantelli,

$$S_1^T = O_{p.s} \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}}. \tag{3.18}$$

L'hypothèse (H5) et la symétrie de K voir Didi & Louani (2013), il en ressort que

$$S_2^T = O_{p.s} \left(T^{-\frac{1}{2}} \right) + O \left(h^2 \right). \tag{3.19}$$

Le lemme s'achève en combinant les résultats (3.18) et (3.19). \square

Le lemme suivant présente un résultat de convergence uniforme de f_{Th} . Ce résultat présente des vitesses de convergence plus rapide que celle établie dans Didi & Louani (2013).

Lemma 3.3. Sous les hypothèses (H1)(a), (H2)(a), (H4), (H5) et (H6)(b), il existe $m < 1$, tel que,

$$\sup_{x \in \mathbb{R}} |f_{T,h}(x) - f(x)| = O(h^2) + O_{p.s} \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}}, \quad \text{lorsque } T \rightarrow \infty. \quad (3.20)$$

Démonstration. La démonstration est assez similaire à celle de Didi & Louani [11]. Considérons l'ensemble B_T défini par

$$B_T := \{x : |x| \leq T^\tau, 0 \leq \tau \leq 1\}.$$

Soit

$$B_T^i = \left\{ x : \|x - x_i\| \leq T^\tau \gamma_T^{-1}, \quad (x_i)_{1 \leq i \leq \gamma_T} \in B_T \right\},$$

une partition de B_T avec γ_T une fonction entière croissante tendant vers l'infini lorsque $T \rightarrow \infty$. En faisant appel à un calcul de Didi & Louani [11], nous obtenons

$$\begin{aligned} \sup_{x \in B_T} |f_{T,h}(x) - f(x)| &\leq \sup_{x \in B_T} \left| \frac{1}{Th} \int_0^T K\left(\frac{x - X_t}{h}\right) - \mathbb{E} \left[K\left(\frac{x - X_t}{h}\right) \middle| \mathcal{F}_{t-\delta} \right] dt \right| \\ &+ \sup_{x \in B_T} \left| \frac{1}{Th} \int_0^T \mathbb{E} \left[K\left(\frac{x - X_t}{h}\right) \middle| \mathcal{F}_{t-\delta} \right] dt - f(x) \right| \\ &= \sup_{x \in B_T} A^T(x) + \sup_{x \in B_T} B^T(x). \end{aligned} \quad (3.21)$$

Le premier terme de l'égalité (3.21) est contrôlé par

$$\sup_{x \in B_T} A^T(x) \leq I^1 + I^2,$$

où

$$I^1 = \max_{1 \leq i \leq \gamma_T} \sup_{x \in B_T^i} |A^T(x) - A^T(x_i)| \quad (3.22)$$

et

$$I^2 = \max_{1 \leq i \leq \gamma_T} \sup_{x \in B_T^i} |A^T(x_i)|. \quad (3.23)$$

En utilisant la condition (H1)(b) et en prenant $\gamma_T = T^{\tau+2}$, il en découle que

$$\begin{aligned} I^1 &\leq \frac{2dT^\tau}{h^2\gamma_T} \\ &= o\left(\frac{\log T}{Th^{2-\frac{1}{m}}}\right)^{\frac{1}{2}}. \end{aligned} \quad (3.24)$$

Par ailleurs, en procédant comme dans l'assertion (3.17) avec $\gamma_T = T^{\tau+2}$, nous avons

$$\mathbb{P} \left(\max_{1 \leq k \leq \gamma_T} \frac{1}{Th} \left| \sum_{i=1}^n Y_i \right| > \epsilon \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}} \right) \leq 2T^{\tau+2-\epsilon^2 C}, \quad (3.25)$$

où C est une constante positive.

En choisissant ϵ_T convenable et en utilisant le lemme de Borel-Cantelli,

$$I^2 = O_{p.s} \left(\frac{\log T}{Th^{2-\frac{1}{m}}} \right)^{\frac{1}{2}}. \quad (3.26)$$

Il s'en suit, d'après les assertions (3.24) et (3.26), que

$$A^T = O_{p.s} \left(\frac{\log T}{Th^{2-\frac{1}{m}}} \right)^{\frac{1}{2}}. \quad (3.27)$$

D'autre part, par un calcul similaire à Didi & Louani [11] et la symétrie de K , nous avons

$$B^T = O_{p.s} \left(T^{-\frac{1}{2}} \right) + O \left(h^2 \right). \quad (3.28)$$

Par conséquent, en utilisant les assertions (3.27) et (3.28), nous obtenons

$$\sup_{x \in B_T} |f_{Th}(x) - f(x)| = O_{p.s} \left(h^2 \right) + O_{p.s} \left(\frac{\log T}{Th^{2-\frac{1}{m}}} \right)^{\frac{1}{2}}. \quad (3.29)$$

Il reste maintenant à évaluer le terme $\sup_{x \in B_T^c} |f_{Th}(x) - f(x)|$ où B_T^c est le complémentaire de B_T dans \mathbb{R} . Par le même procédé que dans l'assertion (3.28), nous avons que

$$\sup_{x \in B_T^c} \left| \frac{1}{Th} \int_0^T \mathbb{E} \left[K \left(\frac{x - X_t}{h} \right) | \mathcal{F}_{t-\delta} \right] dt - f(x) \right| = O_{p.s} \left(T^{-\frac{1}{2}} \right) + O \left(h^2 \right). \quad (3.30)$$

En fin, pour avoir,

$$\left(\frac{Th^{2-\frac{1}{m}}}{\log T} \right)^{\frac{1}{2}} \sup_{x \in B_T^c} \left| f_{T,h} - \frac{1}{Th} \int_0^T \mathbb{E} \left[K \left(\frac{x - X_t}{h} \right) | \mathcal{F}_{t-\delta} \right] dt \right| = 0 \quad p.s., \quad (3.31)$$

nous faisons appel au lemme 4-2 de Bosq (1996) [3] pour terminer la démonstration.

En effet, nous devons démontrer d'une part que l'application $T \rightarrow \sup_{|x| > T^{2\tau}} \left| \frac{\epsilon_T}{Th} \sum_{i=1}^n Y_i \right|$,

où $\epsilon_T = \left(\frac{\log T}{Th^{2-\frac{1}{m}}} \right)^{\frac{1}{2}}$, est uniformément continue, ie, il existe une constante positive θ , tel que

$$\sup_{|x| > T^{2\tau}} \left| \frac{\epsilon_T}{Th} \sum_{i=1}^n Y_i - \frac{\epsilon_S}{Sh} \sum_{i=1}^{n'} Y'_i \right| \leq \theta |T - S|, \quad (3.32)$$

où $S = n'\delta$, $S_j = j\delta$, $Y'_j := \frac{1}{Sh} \int_{S_{j-1}}^{S_j} K \left(\frac{x - X_t}{h} \right) - \mathbb{E} \left[K \left(\frac{x - X_t}{h} \right) | \mathcal{F}_{t-\delta} \right] dt$.

Notons que

$$\begin{aligned}
\sup_{|x|>T^{2\tau}} \left| \frac{\epsilon_T}{Th} \sum_{i=1}^n Y_i - \frac{\epsilon_S}{Sh} \sum_{i=1}^{n'} Y'_i \right| &\leq \sup_{|x|>T^{2\tau}} |\epsilon_T f_{Th}(x) - \epsilon_S f_{Sh}(x)| \\
&\quad + \sup_{|x|>T^{2\tau}} |\epsilon_T \tilde{f}_{Th}(x) - \epsilon_S \tilde{f}_{Sh}(x)| \\
&\leq \sup_{|x|>T^{2\tau}} \left| \frac{d(\epsilon_T f_{Th})}{dT} \right| |T - S| + \sup_{|x|>T^{2\tau}} \left| \frac{d(\epsilon_T \tilde{f}_{Th})}{dT} \right| |T - S| \\
&\leq 2 \max \left\{ \sup_{|x|>T^{2\tau}} \left| \frac{d(\epsilon_T f_{Th})}{dT} \right|, \sup_{|x|>T^{2\tau}} \left| \frac{d(\epsilon_T \tilde{f}_{Th})}{dT} \right| \right\} |T - S|,
\end{aligned}$$

$$\text{où } \tilde{f}_{T,h} := \frac{1}{Th} \int_0^T \mathbb{E} \left[K \left(\frac{x - X_t}{h} \right) \middle| \mathcal{F}_{t-\delta} \right] dt.$$

Par un calcul similaire à Didi & Louani et sous les conditions (H1)(a) – (b) et (H6)(b), il en résulte que

$$\begin{aligned}
\sup_{|x|>T^{2\tau}} \left| \frac{\epsilon_T}{Th} \sum_{i=1}^n Y_i - \frac{\epsilon_S}{Sh} \sum_{i=1}^{n'} Y'_i \right| &\leq \max \left\{ \frac{\|K\|_\infty}{(Th^{\frac{1}{m}} \log T)^{\frac{1}{2}}} \left[\frac{3}{2} + \frac{1}{2 \log T} \right], \right. \\
&\quad \left. \frac{|h'| (Th^{2-\frac{1}{m}})^{\frac{1}{2}}}{h^2 (\log T)^{\frac{1}{2}}} [2|K|_\infty + 2V|K'|_\infty] \right\} |T - S|.
\end{aligned}$$

Voir Didi & Louani pour plus de détails. D'autre part, en utilisant le lemme 3.2, nous avons que

$$\mathbb{P} \left(\left| \sum_{i=1}^n Y_i \right| > Th\epsilon_T\epsilon_0 \right) \leq \frac{2}{T\epsilon_0^2}. \quad (3.33)$$

Ce qui achève la démonstration du lemme. □

3.3.2 Preuve du Théorème 3.1

Observons que,

$$\begin{aligned}
L_{T,h} - M_h &= \int [f_{T,h}(x) - f(x)]^2 dx - \mathbb{E} \int [f_{T,h}(x) - f(x)]^2 dx \\
&= \int f_{T,h}^2(x) dx - \mathbb{E} \int f_{T,h}^2(x) dx + 2\mathbb{E} \int f_{T,h}(x) f(x) dx - 2 \int f_{T,h}(x) f(x) dx \\
&:= A + B.
\end{aligned}$$

Considérons tout d'abord le terme (A),

$$\begin{aligned}
A &= \int f_{T,h}^2(x) dx - \mathbb{E} \left[\int f_{T,h}^2(x) dx \right] \\
&= \int f_{T,h}^2(x) dx - \int [\mathbb{E} f_{T,h}(x)]^2 - \mathbb{E} \int f_{T,h}^2(x) dx + \int [\mathbb{E} f_{T,h}(x)]^2 \\
&= \int [f_{T,h}^2(x) - [\mathbb{E} f_{T,h}(x)]^2] dx - \int [\mathbb{E} f_{T,h}^2(x) - [\mathbb{E} f_{T,h}(x)]^2] dx \\
&= \int [f_{T,h}(x) - \mathbb{E} f_{T,h}(x)] [f_{T,h}(x) + \mathbb{E} f_{T,h}(x)] dx - \int \text{Var} f_{T,h}(x) dx. \\
&:= A_1 + A_2.
\end{aligned}$$

En décomposant l'ensemble $[0, T]^2$ en deux ensembles Γ et Γ^c , $\Gamma = \cup_{i=1}^n [T_{i-1}, T_i]^2$ et Γ^c est son complémentaire dans $[0, T]$, nous avons

$$\begin{aligned}
|A_2| &= \int \text{Cov} \left(\frac{1}{Th} \int_0^T K \left(\frac{x - X_t}{h} \right) dt, \frac{1}{Th} \int_0^T K \left(\frac{x - X_s}{h} \right) ds \right) dx \\
&= \frac{1}{T^2 h^2} \int \left\{ \int_0^T \int_0^T \text{Cov} \left(K \left(\frac{x - X_t}{h} \right), K \left(\frac{x - X_s}{h} \right) \right) ds dt \right\} dx \\
&= \frac{1}{T^2 h^2} \int \left\{ \int_{\Gamma} \text{Cov} \left(K \left(\frac{x - X_t}{h} \right), K \left(\frac{x - X_s}{h} \right) \right) ds dt \right\} dx \\
&+ \frac{1}{T^2 h^2} \int \left\{ \int_{\Gamma^c} \text{Cov} \left(K \left(\frac{x - X_t}{h} \right), K \left(\frac{x - X_s}{h} \right) \right) ds dt \right\} dx \\
&:= A_{2,1} + A_{2,2}.
\end{aligned}$$

Le terme $A_{2,1}$ est contrôlé par

$$\begin{aligned}
A_{2,1} &= \frac{1}{T^2 h^2} \int \left\{ \int_{\Gamma} \mathbb{E} \left[K \left(\frac{x - X_t}{h} \right) K \left(\frac{x - X_s}{h} \right) \right] - \mathbb{E} \left[K \left(\frac{x - X_t}{h} \right) \right] \mathbb{E} \left[K \left(\frac{x - X_s}{h} \right) \right] ds dt \right\} dx \\
&\leq \frac{1}{T^2 h^2} \int \left\{ \int_{\Gamma} \mathbb{E} \left[K \left(\frac{x - X_t}{h} \right) K \left(\frac{x - X_s}{h} \right) \right] \right\} dx.
\end{aligned}$$

En utilisant l'inégalité de Cauchy-schwartz et la stationnarité du processus, nous obtenons,

$$\begin{aligned}
A_{2,1} &\leq \frac{1}{T^2 h^2} \left\{ \int_{\Gamma} \mathbb{E} \left[K^2 \left(\frac{x - X_s}{h} \right) \right] \right\} dx. \\
&= \frac{\delta}{Th^2} \iint K^2 \left(\frac{x - y}{h} \right) f(y) dy dx.
\end{aligned}$$

Par le changement de variable $z = \frac{x - y}{h}$ et le développement de Taylor, nous avons

$$\begin{aligned}
A_{2,1} &\leq \frac{\delta}{Th} \iint K^2(z) f(x - hz) dy dx \\
&= \frac{\delta}{Th} \iint K^2(z) \left[f(x) - hz f'(x) + \frac{h^2}{2} z^2 f''(x^*) \right] dy dx,
\end{aligned}$$

où x^* est compris entre x et $x - hz$. Puis que le noyau K est à support compact et f'' intégrable,

$$\begin{aligned} A_{2,1} &\leq \frac{\delta}{Th} \left[\int K^2 + \frac{h^2}{2} \left\{ \int z^2 K^2(z) dz \right\} \int f'' + o(h^2) \right]. \\ &= \frac{\delta}{Th} \int K^2 + o\left(\frac{1}{Th}\right). \end{aligned} \quad (3.34)$$

En utilisant un calcul similaire à EL HEDA & Louani (2018), nous obtenons

$$\begin{aligned} A_{2,2} &= \frac{1}{T^2 h^2} \iint_{\Gamma^c} \text{Cov} \left(K \left(\frac{x - X_t}{h} \right), K \left(\frac{x - X_s}{h} \right) \right) ds dt dx \\ A_{2,2} &\leq \frac{n-1}{n} \Delta_p(\Gamma) h^{\frac{2}{q}-1} \\ &= O_{p.s} \left(h^{\frac{2}{q}-1} \right). \end{aligned} \quad (3.35)$$

La combinaison des assertions (3.34) et (3.35) donne

$$A_2 \leq \frac{\delta}{Th} \int K^2 + O_{p.s} \left(h^{\frac{2}{q}-1} \right) + o\left(\frac{1}{Th}\right). \quad (3.36)$$

D'une part, en utilisant le lemme 3.1,

$$f_{T,h} - f = O(h^2) + O_{p.s} \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}}. \quad (3.37)$$

D'autre part, en utilisant un changement de variable et un développement de Taylor, pour tout $x \in \mathbb{R}$, nous avons

$$\begin{aligned} f(x) - \mathbb{E}f_{T,h}(x) &= f(x) - \mathbb{E} \left[\frac{1}{Th} \int_0^T K \left(\frac{x - X_t}{h} \right) dt \right] \\ &= -\frac{h^2}{2} \left\{ \int z^2 K(z) dz \right\} f''(x) + o(h^2). \end{aligned}$$

Ceci conduit à

$$\begin{aligned} f_{T,h}(x) - \mathbb{E}f_{T,h}(x) &= f_{T,h}(x) - f(x) + f(x) - \mathbb{E}f_{T,h}(x) \\ &= O(h^2) + O_{p.s} \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}}. \end{aligned} \quad (3.38)$$

Par conséquent,

$$\begin{aligned}
A_1 &= \int [f_{T,h}(x) - \mathbb{E}f_{T,h}(x)] [f_{T,h}(x) + \mathbb{E}f_{T,h}(x)] dx \\
&= \left\{ O(h^2) + O_{p.s} \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}} \right\} \int [f_{T,h}(x) + \mathbb{E}f_{T,h}(x)] dx \\
&= O(h^2) + O_{p.s} \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}}.
\end{aligned} \tag{3.39}$$

En combinant les assertions (3.39) et (3.36), nous obtenons

$$A = O_{p.s}(h^{\min(\frac{2}{q}-1,2)}) + O_{p.s} \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}}. \tag{3.40}$$

Par ailleurs, nous avons que

$$B = O(h^2) + O_{p.s} \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}} \tag{3.41}$$

Le résultat découle directement des assertions (3.41) et (3.40).

3.3.3 Preuve du Théorème 3.2

Observons que

$$\begin{aligned}
L_{T,h} - M_{T,h} &= -2 \int f_{T,h}(x)f(x)dx + \int f^2(x)dx + \frac{2}{T^2h} \sum_{i=1}^n \sum_{j=1}^n \int_{T_{i-1}}^{T_i} \int_{T_{j-1}}^{T_j} K \left(\frac{X_s - X_t}{h} \right) dsdt \\
&\quad - \frac{2}{T^2h} \sum_{i=1}^n \int_{T_{i-1}}^{T_i} \int_{T_{i-1}}^{T_i} K \left(\frac{X_s - X_t}{h} \right) dsdt \\
&= 2 \int f(x) [f(x) - f_{T,h}(x)] dx - \int f^2(x)dx + \frac{2}{T} \int_0^T f_{T,h}(X_t) dt \\
&\quad - \frac{2}{T^2h} \sum_{i=1}^n \int_{T_{i-1}}^{T_i} \int_{T_{i-1}}^{T_i} K \left(\frac{X_s - X_t}{h} \right) dsdt \\
&= 2 \int f(x) [f(x) - f_{T,h}(x)] dx - \int f^2(x)dx + \frac{2}{T} \int_0^T [f_{T,h}(X_t) - f(X_t)] dt \\
&\quad + \frac{2}{T} \int_0^T f(X_t) dt - \frac{2}{T^2h} \sum_{i=1}^n \int_{T_{i-1}}^{T_i} \int_{T_{i-1}}^{T_i} K \left(\frac{X_s - X_t}{h} \right) dsdt,
\end{aligned}$$

En utilisant le lemme 3.3, nous obtenons

$$\begin{aligned}
\frac{2}{T} \left| \int_0^T [f_{T,h}(X_t) - f(X_t)] dt \right| &\leq \frac{2}{T} \int_0^T \sup_{t \in [0,T]} |f_{T,h}(X_t) - f(X_t)| dt \\
&= O(h^2) + O_{p.s} \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}}.
\end{aligned} \tag{3.42}$$

Par ailleurs, nous avons que

$$\frac{2}{T^2} \sum_{i=1}^n \int_{T_{i-1}}^{T_i} \int_{T_{i-1}}^{T_i} K\left(\frac{X_s - X_t}{h}\right) ds dt \leq \frac{2\delta M}{Th} = O_{p.s} \left[\frac{\log T}{Th^{2-\frac{1}{m}}} \right]^{\frac{1}{2}}. \quad (3.43)$$

Le résultat se déduit directement des résultats (3.42) et (3.43).

3.3.4 Preuve du Corollaire 3.1

En utilisant le théorème 3.1, nous avons

$$L_{T,h_0} = M_{h_{opt}} + O\left(\frac{a_T^\alpha}{T}\right)^{\min(\frac{2}{3q}-\frac{1}{3}, \frac{2}{3})} + O_{p.s} \left[\frac{\log T}{\left(T^{1+\frac{1}{m}} a_T^{\alpha(2-\frac{1}{m})}\right)^{\frac{1}{3}}} \right]^{\frac{1}{2}}. \quad (3.44)$$

Par conséquent, h_0 est asymptotiquement presque sûrement aussi "bon" que h_{opt} .

Soit \hat{h} appartenant à $\left[\epsilon \left(\frac{a_T^\alpha}{T}\right)^{\frac{1}{3}}, \beta \left(\frac{a_T^\alpha}{T}\right)^{\frac{1}{3}} \right]$, une conséquence immédiate du théorème 3.2 est d'une part,

$$M_{T,\hat{h}} = L_{T,\hat{h}} - \frac{2}{T} \int_0^T f(X_t) dt + \int f^2(x) dx + O\left(\frac{a_T^\alpha}{T}\right)^{\frac{2}{3}} + O_{p.s} \left[\frac{\log T}{\left(T^{1+\frac{1}{m}} a_T^{\alpha(2-\frac{1}{m})}\right)^{\frac{1}{3}}} \right]^{\frac{1}{2}}. \quad (3.45)$$

D'autre part,

$$M_{T,\hat{h}} = \inf_{\epsilon \left(\frac{a_T^\alpha}{T}\right)^{\frac{1}{3}} \leq h \leq \beta \left(\frac{a_T^\alpha}{T}\right)^{\frac{1}{3}}} L_{T,h} - \frac{2}{T} \int_0^T f(X_t) dt + \int f^2(x) dx + O\left(\frac{a_T^\alpha}{T}\right)^{\frac{2}{3}} + O_{p.s} \left[\frac{\log T}{\left(T^{1+\frac{1}{m}} a_T^{\alpha(2-\frac{1}{m})}\right)^{\frac{1}{3}}} \right]^{\frac{1}{2}}. \quad (3.46)$$

En utilisant les assertions (3.45) et (3.46), nous obtenons

$$L_{T,\hat{h}} = \epsilon \left(\frac{a_T^\alpha}{T}\right)^{\frac{1}{3}} \inf_{\leq h \leq \beta \left(\frac{a_T^\alpha}{T}\right)^{\frac{1}{3}}} L_{T,h} + O\left(\frac{a_T^\alpha}{T}\right)^{\frac{2}{3}} + O_{p.s.} \left[\frac{\log T}{\left(T^{1+\frac{1}{m}} a_T^{\alpha(2-\frac{1}{m})}\right)^{\frac{1}{3}}} \right]^{\frac{1}{2}}. \quad (3.47)$$

Les assertions (3.44) et (3.47) montrent que \hat{h} est asymptotiquement aussi "bon" que h_{opt} .

3.3.5 Preuve du Corollaire 3.2

Ce résultat se démontre par l'absurde de la même manière qu'en temps discret (voir Hall (1983), p 1160), en effet,

$$\frac{\tilde{h}}{h_{opt}} = \frac{\tilde{h}(T)}{h_{opt}(T)} = \frac{\tilde{h}(n\delta)}{h_{opt}(n\delta)}.$$

Supposons alors que

$$\frac{\tilde{h}}{h_{opt}} \not\rightarrow 1, \quad p.s. \quad (3.48)$$

Comme \tilde{h} et h_{opt} sont du même ordre de grandeur et que $h_{opt} = \gamma \left(\frac{a_T^\alpha}{T}\right)^{\frac{1}{3}}$, nous avons

$$\frac{\tilde{h}}{h_{opt}} \in \left[\frac{\epsilon}{\gamma}, \frac{\beta}{\gamma} \right].$$

On peut alors extraire une sous suite telle que

$$\frac{\tilde{h}(\varphi(n))}{h_{opt}(\varphi(n))} \longrightarrow l \neq 1 \quad p.s.,$$

avec $l \in \left[\frac{\epsilon}{\gamma}, \frac{\beta}{\gamma} \right]$. Considérons le cas $l \in]0, 1[$ (le cas $l > 1$ est traité de la même manière). On peut alors choisir $\tilde{h}(\varphi(n)) = Zh_{opt}(\varphi(n))$, tel que, $Z \in [l - \eta, l + \eta]$ avec $\eta = \frac{1}{2} \min(l, 1 - l)$. Comme le choix de η assure l'écart entre $\tilde{h}(\varphi(n))$ et $h_{opt}(\varphi(n))$, nous avons alors

$$\frac{L_{T,\tilde{h}(\varphi(n))}}{M_{h_{opt}(\varphi(n))}} \longrightarrow \xi > 1, \quad p.s.$$

ce qui contredit le fait que \tilde{h} est asymptotiquement aussi "bon" que h_{opt} .

3.3.6 Preuve du Théorème 3.3

Soit $L(Z) = -ZK'(Z)$. Notons que L est un noyau qui satisfait les mêmes hypothèses que K . Soit $\hat{f}_{T,h}$, $g_{T,h}$ et $\hat{g}_{T,h}$ les estimateurs à noyaux de f définis par

$$\begin{aligned}\hat{f}_{T,h}(x) &= \frac{1}{Th} \int_0^T K^{(2)}\left(\frac{x - X_t}{h}\right) dt, & g_{T,h}(x) &= \frac{1}{Th} \int_0^T L\left(\frac{x - X_t}{h}\right) dt \quad \text{et} \\ \hat{g}_{T,h}(x) &= \frac{1}{Th} \int_0^T L^{(2)}\left(\frac{x - X_t}{h}\right) dt.\end{aligned}$$

En posant

$$L_{T,h} = M_h + D_h \tag{3.49}$$

et en dérivant par rapport à h , $L_{T,h}$ donné dans l'assertion (3.49), nous avons

$$\begin{aligned}L'_{T,h} &= M'_h + D'_h \\ L'_{T,h_0} &= (M'_{h_0} - M'_{h_{opt}}) + D'_{h_0}\end{aligned} \tag{3.50}$$

$$0 = L'_{T,h_0} = (h_0 - h_{opt}) M''_{h^*} + D'_{h_0}, \tag{3.51}$$

où h^* est compris entre h_0 et h_{opt} . Il découle de l'égalité (3.51) que

$$h_{opt} - h_0 = \frac{D'_{h_0}}{M''_{h^*}}. \tag{3.52}$$

D'une part, nous avons, sous les hypothèses (H1), (H2), (H3)(a), (H3)(c) – (d), que

$$\begin{aligned}M''_h &= \left\{ 8 \int K^2 \left\{ \int \int_0^\infty g_u(z, z) dz du \right\} \right\} \frac{a_T^\alpha}{Th^3} + 2 \int z^2 \left(\frac{1}{2} K^{(2)} - K \right)(z) dz \int f'' f \\ &+ o \left(\left\{ 8 \int K^2 \left\{ \int \int_0^\infty g_u(z, z) dz du \right\} \right\} \frac{a_T^\alpha}{Th^3} + 2 \int z^2 \left(\frac{1}{2} K^{(2)} - K \right)(z) dz \int f'' f \right).\end{aligned} \tag{3.53}$$

D'autre part,

$$\begin{aligned}
D(h) &= L_{T,h} - M_h \\
&= \int f_{T,h}^2(x)dx - 2 \int f_{T,h}(x)f(x)dx - \mathbb{E} \int f_{T,h}^2(x)dx + 2\mathbb{E} \int f_{T,h}(x)f(x)dx \\
&= \int f_{T,h}^2(x)dx + 2 \int [\mathbb{E} [f_{T,h}(x)] - f_{T,h}(x)] f(x)dx - \mathbb{E} \int f_{T,h}^2(x)dx \\
&= \frac{1}{T^2h^2} \int \int_0^T \int_0^T K\left(\frac{x-X_t}{h}\right) K\left(\frac{x-X_s}{h}\right) dt ds dx + \frac{2}{h} \int \left[\int K\left(\frac{x-y}{h}\right) f(y)dy \right. \\
&\quad \left. - \frac{1}{T} \int_0^T K\left(\frac{x-X_t}{h}\right) dt dx \right] f(x)dx - \frac{1}{T^2h^2} \int \int_0^T \int_0^T \left\{ \iint K\left(\frac{x-y}{h}\right) K\left(\frac{x-z}{h}\right) \right. \\
&\quad \left. f_{t,s}(y,z)dydz \right\} dt ds dx.
\end{aligned}$$

En utilisant le théorème de Fubini et un changement de variables, nous obtenons

$$\begin{aligned}
D(h) &= \frac{1}{T^2h} \int_0^T \int_0^T K^{(2)}\left(\frac{X_t - X_s}{h}\right) dt ds - \frac{2}{h} \int \left[\frac{1}{T} \int_0^T K\left(\frac{x-X_t}{h}\right) dt \right. \\
&\quad \left. - \int K\left(\frac{x-y}{h}\right) f(y)dy \right] f(x)dx - \frac{1}{T^2h} \int_0^T \int_0^T \iint K^{(2)}\left(\frac{y-z}{h}\right) f_{t,s}(y,z)dydz dt ds.
\end{aligned}$$

En dérivant $D(h)$ par rapport à h , il en découle

$$\begin{aligned}
D'(h) &= -\frac{1}{T^2h^2} \int_0^T \int_0^T K^{(2)}\left(\frac{X_t - X_s}{h}\right) dt ds + \frac{1}{T^2h^2} \int_0^T \int_0^T L^{(2)}\left(\frac{X_t - X_s}{h}\right) dt ds \\
&\quad + \frac{2}{h^2} \int \left[\frac{1}{T} \int_0^T K\left(\frac{x-X_t}{h}\right) dt - \int K\left(\frac{x-y}{h}\right) f(y)dy \right] f(x)dx \\
&\quad - \frac{2}{h^2} \int \left[\frac{1}{T} \int_0^T L\left(\frac{x-X_t}{h}\right) dt - \int L\left(\frac{x-y}{h}\right) f(y)dy \right] f(x)dx \\
&\quad + \frac{1}{T^2h^2} \int_0^T \int_0^T \iint K^{(2)}\left(\frac{y-z}{h}\right) f_{t,s}(y,z)dydz dt ds \\
&\quad - \frac{1}{T^2h^2} \int_0^T \int_0^T \iint L^{(2)}\left(\frac{y-z}{h}\right) f_{t,s}(y,z)dydz dt ds,
\end{aligned}$$

où $L^{(2)}(Z) = -ZK^{(2)'}(Z)$.

$$\begin{aligned}
 D'(h) &= -\frac{1}{Th} \int_0^T \hat{f}_{T,h}(X_t) dt + \frac{1}{Th} \int_0^T \hat{g}_{T,h}(X_t) dt + \frac{2}{h} \int [f_{T,h} - \mathbb{E}f_{T,h}] f(x) dx \\
 &- \frac{2}{h} \int [g_{T,h}(x) - \mathbb{E}[g_{T,h}(x)]] f(x) dx + \frac{1}{T^2 h^2} \int \int \int K^{(2)}\left(\frac{y-z}{h}\right) f_{t,s}(y,z) dy dz dt ds \\
 &+ \frac{1}{T^2 h^2} \int_{\Gamma^c} \int \int K^{(2)}\left(\frac{y-z}{h}\right) g_{t,s}(y,z) dy dz dt ds \\
 &+ \frac{1}{T^2 h^2} \int_{\Gamma^c} \int \int K^{(2)}\left(\frac{y-z}{h}\right) f(y)f(z) dy dz dt ds \\
 &- \frac{1}{T^2 h^2} \int \int \int L^{(2)}\left(\frac{y-z}{h}\right) f_{t,s}(y,z) dy dz dt ds - \frac{1}{T^2 h^2} \int_{\Gamma^c} \int \int L^{(2)}\left(\frac{y-z}{h}\right) g_{t,s}(y,z) dy dz \\
 &dt ds - \frac{1}{T^2 h^2} \int_{\Gamma^c} \int \int L^{(2)}\left(\frac{y-z}{h}\right) f(y)f(z) dy dz dt ds.
 \end{aligned}$$

Le Lemme 3.1, un changement de variables $Z = \frac{y-z}{h}$ et un développement de Taylor, donnent

$$\begin{aligned}
 \frac{2}{h} \int [f_{T,h}(x) - \mathbb{E}[f_{T,h}]] f(x) dx &= \frac{2}{h} \int [f_{T,h} - f(x) + f(x) - \mathbb{E}f_{T,h}] f(x) dx \\
 &= O(h) + O\left[\frac{\log T}{Th^{4-\frac{1}{m}}}\right]^{\frac{1}{2}}. \tag{3.54}
 \end{aligned}$$

De la même manière, nous avons

$$\frac{2}{h} \int [g_{T,h}(x) - \mathbb{E}[g_{T,h}]] f(x) dx = O(h) + O_{p.s}\left[\frac{\log T}{Th^{4-\frac{1}{m}}}\right]^{\frac{1}{2}}. \tag{3.55}$$

Par ailleurs, en utilisant le théorème de Fubini, un changement de variable $Z = \frac{y-z}{h}$ et les conditions $H1(a)$ et $H2(a)$, il en ressort que

$$\begin{aligned}
 \frac{1}{T^2 h^2} \int_{\Gamma^c} \int \int K^{(2)}\left(\frac{y-z}{h}\right) f(y)f(z) dy dz dt ds &= \frac{T^2 - n\delta^2}{T^2 h^2} \int \int K^{(2)}\left(\frac{y-z}{h}\right) f(y)f(z) dy dz \\
 &= \frac{T^2 - n\delta^2}{T^2 h} \int f^2(y) dy + O(h). \tag{3.56}
 \end{aligned}$$

De la même manière,

$$\frac{1}{T^2 h^2} \int_{\Gamma^c} \int \int L^{(2)}\left(\frac{y-z}{h}\right) f(y)f(z) dy dz dt ds = \frac{T^2 - n\delta^2}{T^2 h} \int f^2(y) dy + O(h). \tag{3.57}$$

Sous les hypothèses $H3(a) - (b)$ et par un calcul de EL HEDA & Louani [13], il s'en suit que

$$\frac{1}{T^2 h^2} \int_{\Gamma^c} \iint K^{(2)} \left(\frac{y-z}{h} \right) g_{t,s}(y,z) dy dz dt ds = O \left(h^{\frac{2}{q}-2} \right). \quad (3.58)$$

De la même manière, nous avons

$$\frac{1}{T^2 h^2} \int_{\Gamma^c} \iint L^{(2)} \left(\frac{y-z}{h} \right) g_{t,s}(y,z) dy dz dt ds = O \left(h^{\frac{2}{q}-2} \right). \quad (3.59)$$

En utilisant les assertions (3.55), (3.57), (3.58) et (3.59), nous obtenons

$$\begin{aligned} D'(h) &= \frac{1}{Th} \int_0^T \{f(X_t) - \hat{f}_{T,h}(X_t)\} dt + \frac{1}{Th} \int_0^T \{\hat{g}_{T,h}(X_t) - f(X_t)\} dt + \\ &O \left(h^{\min\{\frac{2}{q}-2, 1\}} \right) + O_{p.s} \left[\frac{\log T}{Th^{4-\frac{1}{m}}} \right]^{\frac{1}{2}}. \end{aligned} \quad (3.60)$$

En utilisant le Lemme 3.3, nous avons que

$$\begin{aligned} \frac{1}{Th} \int_0^T \{f(X_t) - \hat{f}_{T,h}(X_t)\} dt &\leq \frac{1}{Th} \int_0^T \sup_{t \in [0, T]} |f(X_t) - \hat{f}_{T,h}(X_t)| dt \\ &\leq \frac{1}{h} \sup_t |f(X_t) - \hat{f}_{T,h}(X_t)| \\ &= O(h) + O_{p.s} \left[\frac{\log T}{Th^{4-\frac{1}{m}}} \right]^{\frac{1}{2}}. \end{aligned}$$

Par conséquent,

$$D'(h) = O(h)^{\min\{\frac{2}{q}-2, 1\}} + O_{p.s} \left[\frac{\log T}{Th^{4-\frac{1}{m}}} \right]^{\frac{1}{2}}. \quad (3.61)$$

Ainsi, la dérivée de D_h par rapport à h prise en h_0 est de la forme

$$D'(h_0) = O_{p.s} \left(\frac{a_T^\alpha}{T} \right)^{\min\{\frac{2}{3q}-\frac{2}{3}, \frac{1}{3}\}} + O_{p.s} \left[\frac{\log T}{\left(T^{\frac{1}{m}-1} a_T^{\alpha(4-\frac{1}{m})} \right)^{\frac{1}{3}}} \right]^{\frac{1}{2}}, \quad (3.62)$$

et la dérivée seconde de M_h par rapport à h prise en h^* est donnée par

$$\begin{aligned} M''_{h^*} &= \left\{ 8 \int K^2 \left\{ \int \int_0^\infty g_u(z, z) dz du \right\} \right\} + 2 \int z^2 \left(\frac{1}{2} K^{(2)} - K \right)(z) dz \int f'' f \\ &+ o \left(\left\{ 8 \int K^2 \left\{ \int \int_0^\infty g_u(z, z) dz du \right\} \right\} + 2 \int z^2 \left(\frac{1}{2} K^{(2)} - K \right)(z) dz \int f'' f \right). \end{aligned}$$

En combinant les assertions (3.53), (3.62) et (3.63) , il s'en suit que

$$h_{opt} - h_0 = O\left(\frac{a_T^\alpha}{T}\right)^{\min\{\frac{2}{3q}-\frac{2}{3}, \frac{1}{3}\}} + O_{p.s} \left[\frac{\log T}{\left(T^{\frac{1}{m}-1} a_T^{\alpha(4-\frac{1}{m})}\right)^{\frac{1}{3}}} \right]^{\frac{1}{2}}. \quad (3.63)$$

3.3.7 Preuve du Théorème 3.4

Soit δ_h défini par

$$\delta_h = M_{T,h} - L_{T,h} + \int f^2. \quad (**)$$

En dérivant δ_h donnée par (**) par rapport à h , nous obtenons

$$\begin{aligned} M'_{T,h} &= \delta'_h + L'_{T,h} \\ &= \delta'_h + M'_h + D'_h \\ 0 = M'_{T,\hat{h}} &= \delta'_h + (M'_{\hat{h}} - M'_{h_{opt}}) + D'_{\hat{h}} \\ &= \delta'_h + (\hat{h} - h_{opt}) M''_{h^*} + D'_{\hat{h}}, \end{aligned}$$

où h^* est compris entre \hat{h} et h_{opt} . Il s'en suit que

$$h_{opt} - \hat{h} = \frac{\delta'_h + D'_{\hat{h}}}{M''_{h^*}}. \quad (3.64)$$

Par ailleurs, le terme δ_h est donné par,

$$\begin{aligned} \delta_h &= M_{T,h} - L_{T,h} + \int f^2 \\ &= -\frac{2}{T^2 h} \int_{\Gamma^c} K\left(\frac{X_t - X_s}{h}\right) dsdt + \frac{2}{Th} \int_0^T \int_0^T K\left(\frac{x - X_t}{h}\right) f(x) dt dx. \end{aligned}$$

En dérivant δ_h par rapport à h .

$$\begin{aligned} \delta'_h &= \frac{2}{T^2 h^2} \int_{\Gamma^c} K\left(\frac{X_t - X_s}{h}\right) dsdt - \frac{2}{T^2 h^2} \int_{\Gamma^c} L\left(\frac{X_t - X_s}{h}\right) dsdt \\ &\quad - \frac{2}{Th^2} \int_0^T \int_0^T K\left(\frac{x - X_t}{h}\right) f(x) dt dx + \frac{2}{Th^2} \int_0^T \int_0^T L\left(\frac{x - X_t}{h}\right) f(x) dt dx. \end{aligned}$$

Observons que pour h très petit, nous avons $\frac{1}{h}K_h(X_t - X_s) = \frac{1}{h}L_h(X_t - X_s) = 0$ pour $t \in [T_{i-1}, T_i], s \in [T_{j-1}, T_j]$ où $1 \leq i < j \leq n$ à l'exception d'un événement de probabilité négligeable. Par conséquent,

$$\begin{aligned} \delta'_h &= -\frac{2}{Th^2} \int \int_0^T K\left(\frac{x - X_t}{h}\right) f(x) dt dx + \frac{2}{Th^2} \int \int_0^T L\left(\frac{x - X_t}{h}\right) f(x) dt dx \\ &= \frac{2}{h} \int [f(x) - f_{T,h}(x)] f(x) dx + \frac{2}{h} \int [g_{T,h}(x) - f(x)] f(x) dx. \end{aligned}$$

En utilisant le lemme 3.1,

$$\delta'_h = O(h) + O_{p.s} \left[\frac{\log T}{Th^{4-\frac{1}{m}}} \right]^{\frac{1}{2}}. \quad (3.65)$$

Ainsi,

$$\delta'_h = O\left(\frac{a_T^\alpha}{T}\right)^{\frac{1}{3}} + O_{p.s} \left[\frac{\log T}{\left(T^{\frac{1}{m}-1} a_T^{\alpha(4-\frac{1}{m})}\right)^{\frac{1}{3}}} \right]^{\frac{1}{2}}. \quad (3.66)$$

Le résultat s'obtient en combinant les résultats (3.66), (3.62) et (3.63).

Annexe : Quelques outils d'analyse et de probabilités

Notations et définitions

Un espace mesurable (Ω, \mathcal{A}) est défini par un ensemble non vide Ω muni d'une tribu \mathcal{A} , c'est-à-dire une classe d'ensembles de Ω vérifiant les propriétés suivantes

- Contenant l'ensemble vide.
- Stable par union finie.
- Stable par complémentarité.

Définition 1 (Ensemble mesurable). les éléments de la tribu ou la σ -algèbre \mathcal{A} sont des ensembles mesurables.

Définition 2 (Application mesurable). les éléments de la tribu ou la σ -algèbre \mathcal{A} sont des ensembles mesurables.

Définition 3 (Filtration). Une filtration est une suite croissante de tribus $(\mathcal{F}_t)_{t \geq 0}$, c'est-à-dire

$$\mathcal{F}_t \subset \mathcal{F}_{t+s} \quad \text{pour tout } t, s \geq 0.$$

Soit $(\Omega, \mathcal{A}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ un espace probabilisé filtré.

Définition 4 (Processus en temps continu). Soit $X = (X_t, t \in \mathbb{R}^+)$ un processus défini sur $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}^+}, \mathbb{P})$. Le processus X est dit mesurable si l'application

$$\begin{aligned} X : [0, \infty[\times \Omega &\rightarrow (E, \mathcal{E}) \\ (t, \omega) &\mapsto X_t(\omega) \end{aligned}$$

est mesurable par rapport à $\mathcal{B}([0, \infty[) \otimes \mathcal{F}_t$, ie, pour tout $B \in \mathcal{E}$, $X^{-1}(B) \in \mathcal{B}([0, \infty[) \otimes \mathcal{F}_t$. Le processus X est dit adapté si $\forall t \in \mathbb{R}^+$, X_t est \mathcal{F}_t -mesurable.

Définition 5 (Processus stationnaires en temps continu). Un processus $(X_t)_{t \in \mathbb{R}^+}$ est dit *strictement stationnaire* si les lois jointes de $(X_{t_1}, \dots, X_{t_k})$ et de $(X_{t_1+h}, \dots, X_{t_k+h})$ sont identiques pour tout $k \in \mathbb{N}$ et pour tout $t_1, \dots, t_k, h \in \mathbb{R}^+$.

Définition 6 (Ensembles δ -invariants). Soit $(X_t)_{t \in \mathbb{R}^+}$ un processus défini sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, un ensemble mesurable A est dit δ -invariant pour une transformation T (un opérateur linéaire), si $T^\delta(A) = A$.

Définition 7 (Processus δ -ergodique). Un processus $(X = X_t)_{t \in \mathbb{R}^+}$ est dit δ -ergodique, si pour tout ensemble A δ -invariant, $\mathbb{P}(A) = 1$ ou 0 .

Définition 8 (Processus ergodique). Un processus $X = (X_t)_{t \in \mathbb{R}^+}$ est dit ergodique s'il est δ -ergodique pour tout $\delta > 0$.

Convergences

Définition 9 (Convergence en Loi). La suite de variables aléatoires $(X_n)_{n \in \mathbb{Z}}$ de fonction de répartition F_n converge en Loi vers Y de fonction de répartition F , si pour tout x où F est continue, on a

$$F_n(x) \longrightarrow F(x).$$

Définition 10 (Convergence en probabilité). La suite de v.a $(X_n)_{n \in \mathbb{Z}}$ converge en probabilité, si

$$\forall \epsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbb{P} \{d(X_n, X) \geq \epsilon\} = 0.$$

Définition 11 (Convergence presque sûre). La suite $(X_n)_{n \in \mathbb{Z}}$ converge presque sûrement (p.s.) vers Y , si

$$\mathbb{P} \left\{ \omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\} = 1.$$

Remarque : La convergence presque sûre entraîne la convergence en probabilité qui entraîne la convergence en loi.

Théorème ergodique en temps continu

Théorème 1. Si $X = (X_t)_{t \in [0, T], T \in \mathbb{R}^+}$ est un processus ergodique stationnaire et si X_1 est intégrable, on a

$$\lim_{n \rightarrow \infty} \frac{1}{T} \int_{i=1}^T X_t dt = \mathbb{E}[X_1] \quad p.s.$$

Inégalités dans les espaces L_p

- Inégalité de Minkovski

$$\int |f + g| \leq \|f\|_{L_p} + \|g\|_{L_q}, \quad \text{où } \frac{1}{p} + \frac{1}{q} = 1.$$

- Inégalité de Hölder

$$\int |fg| \leq \|f\|_{L_p} \|g\|_{L_q}, \quad \text{où } \frac{1}{p} + \frac{1}{q} = 1.$$

- Inégalité de Jensen

Pour toute fonction intégrable g sur un ensemble $(\Omega, \mathcal{A}, \mu)$ et toute fonction convexe φ

$$\varphi\left(\int_{\Omega} g d\mu\right) \leq \int_{\Omega} \varphi(g) d\mu.$$

Perspectives de recherche

L'une des questions qui vient juste après la construction de la procédure de sélection optimale du paramètre de lissage et l'étude des propriétés de convergence des estimateurs qui en résultent concerne la normale asymptotique de ces derniers. Ce type de résultats permettra d'évaluer au mieux les performances des estimateurs par ailleurs. Des études d'autres fonctionnelles et autres opérateurs liés à des modèles de régression sont aussi envisagées.

Outre le volet théorique, les aspects pratiques soulèvent de nombreuses questions. Identifier l'impact de la partition de l'intervalle de temps $[0, T]$ sur les résultats est un travail qui permettra un meilleur usage en pratique du paramètre de lissage obtenu et de meilleurs résultats d'estimation.

Ce travail est de nature à demander un effort important de programmation et calculs informatique.

Bibliographie

- [1] Birkhoff, G. D. (1931) *Proof of the ergodic theorem. Proc. Nat. Aad. Sci. USA*, **17**, p. 656-660.
- [2] Blum, J. R., Hanson, D. L. and Koopmans, L. H. (1963) *On the strong law of large numbers for a class of stochastic processes*, **2**, p. 1-11.
- [3] Bosq, D. (1996) *Nonparametric statistics for stochastic processes. Springer, New York*.
- [4] Bowman, A. W. (1984) *An alternative method of cross-validation for the smoothing of density estimates. Biometrika*, **71**, p. 353-360.
- [5] Bowman, A.W. (1985) *A comparative study of some kernel-based nonparametric density estimators. J. Statistic. Comput. Simul*, **21**, p. 313-327.
- [6] Chacón, J. E., Montanero, J., Nogales, A. G. and Pérez, P. (2007) *On the existence and limit behavior of the optimal bandwidth foe kernel density estimation. Statistica Sinica*, **17**, p. 289-300.
- [7] Chacón, J. E. and Tenreiro, C. (2012) *Exact and Asymptotically Optimal Bandwidths for Kernel Estimation of Density Functionals. Methodol Comput Appl Probab*, **14**, p. 523-548.
- [8] De la penã, V.H and Giné, E. (1999) *Decoupling, from depedence to independence. Probability and its applications. Springer- veslage, New York*.
- [9] Delecroix, M. (1987) *Sur l'estimation et la prévision nonparamétrique des processus ergodiques. Université des sciences et techniques de Lille Flandres-Artois*.
- [10] Devroye, L. and Györfi, L. (1996) *Nonparametric density estimation. The L_1 view. Ohon Wiley and sons .*
- [11] Didi, S. and Louani, D. (2013). *Consistency results for the kernel density estimate on continuous time stationary and dependent data. Statistics and Probability Letters* **83**, p. 1262-1270.

- [12] Duin, R. P. W. (1976) *On the choice of smoothing parameter for Parzen estimators of probability density function. IEEE Trans. Comput*, **25**, p. 1175-1179.
- [13] El heda, K. and Louani, D. (2018) *Optimal bandwidth selection in kernel density estimation for continuous time dependent process. Statistics and Probability Letters*, **138**, p. 9-19.
- [14] Estévez-Pérez, G. and quintela-del-Río, A. and Vieu, P. (2002) *convergence rate for cross-validatory bandwidth in kernel hazard estimation from dependent samples. Journal of Statistical Planning and inference*, **104**, p. 1-30.
- [15] Habbema, J. D. F., Hermans, J. and Van der Broek, K. (1974) *A stepwise discrimination program using density estimation.. Physica Verlag*, p. 100-110.
- [16] Hall, P. (1982) *Limit theorems for stochastic measures of the accuracy of density estimators. Stochastic process appl*, **13** , p. 11-25.
- [17] Hall, P. (1983) *Large Sample Optimality of Least-Squares Cross-Validation in Density Estimation. The Annals of Statistics*, **11** , p. 1156-1174.
- [18] Hall, P. and Marron, J. (1987) *Extent to which Least-Squares Cross-Validation Minimises Integrated Square Error in Nonparametric Density Estimation. Probab. Th. Rel. Fields*, **74** , p. 567-581.
- [19] Härdle, W. and Marron, J. S. (1985) *Optimal bandwidth selection in nonparametric regression function estimation. Ann. Statist.* **13(4)**. p. 1465–1481.
- [20] Marron, J. S. and Härdle, W. (1986) *Random approximations to some measures of accuracy in nonparametric curve estimation. Journal of multivariate anamysis.* **20**. p. 91-113.
- [21] Härdle, W. (1990) *Applied nonparametric regression. TCampbridge university press*.
- [22] Hart, D. and Vieu, P. (1990) *Data-driven Bandwidth Choice for Density Estimation Based on Dependent Data. The annals of statistics*, **18**, p. 873-890.
- [23] Heidenreich, N. B., Schindler, A. and Sperlich, S. (2013) *Bandwidth selection for kernel density estimation : a review of fully automatic selectors. AStA Adv Stat Anal*, **97**, p. 403-433.
- [24] Kim, T. Y. and Denis, D. C. (1997) *A Study on Bandwidth Selection in Density Estimation under Dependence. journal of multivariate analysis*, **60**, p. 190–203.

- [25] Kolmogorov, A. N and Rosanov, Yu. A. (1960) *On strong mixing conditions for stationary Gaussian processes. Teor. Veroyatnost. i. Primenen*, **5**, p. 222-227.
- [26] Krengel, U. (1985) *Ergodic theorems. Walter de Gruyter Berlin. Newyork.*
- [27] Laïb, N. and Louani, D. (2016) *Uniform in bandwidth rate of convergence of the conditional mode estimate on functional stationary ergodic data. Journal of the Korean Statistical Society*, **45**, p. 1-13.
- [28] Masry, E. (1983) *Probability density estimation from sampled dataIEEE transf. inf. Th* , **29**, p. 696-709.
- [29] Marron, J. S. (1987) *A comparison of cross-validation techniques in density estimation. University of North Carolina, Chapel Hill*, **15**, p. 152-162.
- [30] Nadaraya, E. N. (1965) *On nonparametric estimates of density functions and regression curves . Thor. Veroyatnost. i primenen*, **10**, p. 199-203.
- [31] Parzen, E. (1962). *On estimation of a probability density function and mode, Ann. Math. Statist*, p. 1065–1076,
- [32] Peskir, G. (2000) *From uniform laws of large numbers to uniform ergodic theorem for wide sense stationary process, . Stochastic Annal.Appl*, **16** , p. 697-720.
- [33] Perasaka Rao (1983) *Nonparametric density estimation. Academic press.*
- [34] Rachdi, M. and Vieu, P. (2007) *Nonparametric regression for functional data : Automatic smoothing parameter selection. Journal of Statistical Planning and Inference*, **137**, p. 2784-2801.
- [35] Rice, J. (1984) *bandwidth choice for nonparametric regression. Ann. stat.*, **12**, p. 1215-1230.
- [36] Rosenblatt, M. (1956) *Remarks on some nonparametric estimates of a density function. Ann. Math. Statisti.* , **27**, p. 832-837.
- [37] Rosenblatt, M. (1975) *A quadratic measure of deviation of two dimensions on a density estimates and a test of independence . University of California*, **1**, p. 1-14.
- [38] Rosenblatt, M. (1971) *Curve estimates. Ann. Math. Statist*, **42**, p. 1815-1842.
- [39] Roussas. G, (1990) *Nonparametric regression estimation under mixig coditions. Stoch. Processes. Appl* . **36**, p. 107-116.

- [40] Rudemo, M. (1982) *Empirical Choice of Histograms and Kernel Density Estimators. Scandinavian Journal of Statistics*, **9**, p. 65-78.
- [41] Sarda, P. and Vieu, P. (1991). *Smoothing parameter selection in hazard estimation. Statist. Probab. Lett.* **11(5)**. p. 429-434.
- [42] Scott, D.W. and Factor, L.E. (1981) *Montecarlo study of three data-based nonparametric density estimators . J. Amer. Statist. Assoc* , **76**, p. 9-15.
- [43] Scott, D.W., Tapia, R.A and Thompson, J.R. (1977) *Kernel density estimation revisited . Nonlinear Analysis* , **1** , p. 339-372.
- [44] Silverman, B.W. (1978a) *Choosing the window width when estimating a density. Biometrika*, **65**, p. 1-11.
- [45] Silverman, B. W. (1986) *Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability, London : Chapman and Hall.*
- [46] Stone, C. J. (1984) *An asymptotically optimal window selection rule for kernels density estimates. The anals of statistics*, **12**, p. 1285-1297.
- [47] Tenreiro, C. (2017) *A weighted least-squares cross-validation bandwidth selector for kernel density estimation. Communications in Statistics-Theory and Methods*, **46**, p. 3438-3458.
- [48] Tran, L. T. (2017) *kernel density and regression estimation for dependent random variables and times series. Techn. report. univ. Indiana.*
- [49] Vieu, P. (1989) *Quadratic error for nonparametric estimate under dependence . Unpublished manuscript.*
- [50] Von Neumann, J. (1932) *Proof of the quasi-ergodic hypothesis. Proc. Nat. Acad. Sci. USA*, **18**, p. 70-82.
- [51] Watson, G.S. (1965) *Smooth regression analysis . The Indian Journal of Statistics, Series A*, **26**, p. 359-372.
- [52] Woodroof, M. (1970) *On choosing a delta sequence. Ann. Math. Statist*, **41**, p. 1665-1671.
- [53] Wu, W.B. (2003) *Nonparametric Estimation For stationary Processes*. Technical Report 536 *University of Chicago*.
- [54] Youndjé, É., Sarda, P. and Vieu, P. (1996) *Optimal Smooth Hazard Estimates. Test*, **5**, p. 379-394.