

République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique



N° Attribué par la bibliothèque



Année univ.: 2019/2020

Sur le choix du paramètre de lissage dans l'estimation par noyau de la densité

Mémoire présenté en vue de l'obtention du diplôme de

Master Académique

Université de Saida - Dr Moulay Tahar

Discipline : MATHEMATIQUES

Spécialité : Analyse Stochastiques, Statistique des Processus et
Applications

par

Souar Moussa¹

Sous la direction de

Dr/Mme. F.Benziadi

Soutenue le 14/09/2020 devant le jury composé de

Mlle. S. Idrissi Université Dr Tahar Moulay - Saïda Présidente

Mme. F. Benziadi Université Dr Tahar Moulay - Saïda Encadreur

Mr. T. Djebbouri Université Dr Tahar Moulay - Saïda Examinateur

Mme. W. Benzatout Université Dr Tahar Moulay - Saïda Examinatrice

1. e-mail : moussasouar22@gmail.com

Remerciements

Je remercie tout d'abord **ALLAH** qui m'a guidé tout au long de ma vie et m'avoir donné la santé, la force et le courage pour pouvoir surmonter les moments difficiles et atteindre mon objectif.

Je tiens à remercier tout particulièrement mon encadreur, **Dr. Fatima Benziadi** pour ses conseils, ses critiques et sa disponibilité. La pertinence de ses remarques ont toujours su me motiver et me diriger.

Je voudrais également remercier tous les membres de jury, **Dr. Idrissi Soumia** de m'avoir honorée en acceptant de présider le jury, **Dr. Benzatout Wahiba** et **Dr. Djebbouri Tayeb** d'avoir accepté d'examiner ce travail, merci pour toutes leurs remarques et critiques.

Je remercie chaleureusement ma famille, qui m'a soutenu, encouragé et poussé durant toutes mes années d'études.

Je remercie mes amis pour leur soutien et leurs encouragements.
J'exprime mes remerciements à les responsables du laboratoire des Modèles Stochastique, Statistique et Application pour leur accueil et leurs conseils.

Merci à tous !

Table des matières

Introduction	7
1 Introduction générale	9
1.1 Méthodes du choix de paramètre de lissage	10
1.1.1 Méthode Rule of Thumb	12
1.1.2 Méthode plug-in	13
1.1.3 Méthode de validation croisée	14
1.1.4 Méthode de vraisemblance de la validation croisée	15
1.1.5 Méthode de validation croisée biaisée	15
1.2 Ergodicité des processus stationnaires	17
2 Choix optimal du paramètre de lissage dans l'estimation de la densité à temps discret	19
2.1 Estimation de la fonction de densité par les histogrammes	19
2.1.1 Estimation par histogramme mobile	20
2.1.2 Risque de l'estimateur par histogramme mobile	20
2.1.3 Choix de la fenêtre par la méthode de validation croisée	22
2.2 Estimation de la fonction de densité par la méthode du noyau	24
2.2.1 Propriétés de l'estimateur à noyau	25
2.2.2 Résultats asymptotiques	28
2.3 Choix du paramètre de lissage par la méthode de validation croisée	29
2.3.1 Choix pour des données indépendantes	29
2.3.2 Choix pour des données dépendantes	32
3 Choix optimal du paramètre de lissage dans l'estimation de la densité pour des processus à temps continu	35

3.1 Modèle	35
3.2 Hypothèses	36
3.3 Choix du paramètre de lissage	37
3.4 Résultats asymptotiques	38
3.4.1 Convergence des paramètres de lissage	40
Conclusion	57
bibliographie	59

Introduction

La théorie de l'estimation est une des préoccupations majeures des statisticiens. On trouve deux approches d'estimation dans la littérature, la première est l'approche paramétrique qui se résume en l'estimation d'un nombre fini de paramètres réels associés à la loi de l'échantillon. Dans ce cas, on construit les estimateurs en utilisant soit la méthode des moments soit celle du maximum de vraisemblance ou encore celle des moindres carrés. Cependant, l'approche non-paramétrique consiste, généralement à estimer à partir des observations une fonction inconnue appartenant à une certaine classe de fonctions. De ce fait, l'estimation non-paramétrique offre une très grande flexibilité de modélisation pour les applications réelles. Dans ce travail, nous nous intéressons à le choix du paramètre de lissage dans l'estimation de la fonction de densité.

La sélection du paramètre de lissage a fait l'objet de nombreuses contributions. 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)[17]. 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)[24]. Il décrit une règle de sélection de la fenêtre en utilisant la méthode de la validation croisée. Sous l'hypothèse que la densité est bornée, ce choix est ainsi asymptotiquement optimal. Notons que ce critère a été introduit par Rudemo (1982)[19] et Bowman (1984)[5]. Habbema, Hermans et Vandenbroek (1974)[11] et Duin (1976)[9] ont proposé une méthode fondée sur un critère du maximum de vraisemblance. Une autre méthode dite, validation croisée biaisée introduite par Scott et Terrell (1987)[21], Il s'agit d'introduire un biais dans le critère de validation croisée afin de réduire sa variance. Les méthodes de Rule of Thumb et

plug-in qui repose sur l'estimation d'une quantité qui dépend de la dérivée seconde de la densité de probabilité inconnue f . Dans le cas des données α -mélangeantes, Hart & Vieu (1990)[12] 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 présent manuscrit est composé de trois chapitres :

Dans le premier chapitre, nous citons les différentes méthodes de sélection du paramètre de lissage et de donner une présentation sur la théorie ergodique.

Dans le deuxième chapitre, nous nous intéressons à la question du choix du paramètre de lissage dans le cas discret, nous utilisons la méthode de validation croisée afin d'avoir un paramètre de lissage asymptotiquement optimale.

Le troisième chapitre présente une procédure de choix du paramètre de lissage dans l'estimation de la densité pour un processus stationnaire ergodique à temps continu basée sur le critère de validation croisée et nous établissons des résultats asymptotiques sur la fenêtre obtenue.

Chapitre 1

Introduction générale

L'estimation de la densité de probabilité sous-jacente à un ensemble fini d'observations est un problème fondamental en statistique. Dans la littérature, plusieurs méthodes ont été dédiées à l'estimation de la densité de probabilité. On s'intéresse dans ce qui suit par deux méthodes majeures : l'estimation par histogramme et l'estimation par noyau.

La méthode d'histogramme est historiquement la première méthode pour estimer une fonction de densité, dont l'origine est attribuée à John Grant au XVII^{ème} siècle. Pour un échantillon de variables aléatoires réelles X_1, \dots, X_n , indépendantes, identiquement distribuées et de densité f . L'estimateur d'histogramme est défini, pour tout $x \in \mathbb{R}$ par :

$$\hat{f}_{nh}(x) = \frac{1}{nh} \sum_{i=1}^n \mathbf{1}_{[x-\frac{h}{2}, x+\frac{h}{2}]}(X_i),$$

où $\mathbf{1}$ désigne la fonction indicatrice et $h = h_n$ est la fenêtre vérifiant $\lim_{n \rightarrow \infty} h_n = 0$.

Par ailleurs la méthode du noyau a d'abord été décrite en 1951 dans un rapport non publié par Fix et Hodges (voir Silverman and Jones (1989)[23]). La première forme de l'estimateur à noyau a été introduite par Rosenblatt (1956)[18] suivi par Parzen (1962)[17]. Ils ont proposé une classe d'estimateurs à noyau d'une densité univariée, l'estimateur à noyau est défini par :

$$\hat{f}_{nh}(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right), \quad \forall x \in \mathbb{R},$$

où $K : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction positive d'intégrale 1, cette méthode est une généralisation de la méthode d'estimation par histogramme.

Les propriétés de convergence de l'estimateur à noyau ont été établies par Parzen [17], Silverman [22] et Nadaraya [14]. Les théorèmes relatifs à l'erreur quadratique moyenne et l'erreur quadratique intégrée moyenne ont été obtenus sous forme élémentaire par Parzen [17].

Banon (1978)[1], fut le premier à s'intéresser à l'estimation de la densité en temps continu à partir de l'observation d'une partie $(X_t)_{0 \leq t \leq T}$ d'un processus stationnaire $(X_t)_{t \geq 0}$, cet estimateur est donné, pour tout $x \in \mathbb{R}$, par :

$$\widehat{f}_{Th}(x) = \frac{1}{Th_T} \int_0^T K\left(\frac{x - X_t}{h_T}\right) dt.$$

Des résultats ont été traités pour l'estimation de la densité relative à des processus de diffusion par Banon (1978)[1], Banon et Nguyen ((1978)[2],(1981)[3]), Nguyen (1979)[15] et Nguyen et Pham (1980)[16].

La convergence presque sûre et la convergence en moyenne quadratique des estimateurs de la densité et de la densité conditionnelle ont été étudiés par Delecroix (1979)[7] pour des processus strictement stationnaires et fortement mélangeants, récemment Didi & Louani (2013)[8] ont obtenu des résultats de convergence presque sûre, ponctuelle et uniforme, avec des vitesses de convergence sous des conditions de dépendance assez générales où des méthodes de preuve basées sur des différences de martingale et des projections successives sur une famille de σ -algèbre emboîtées, comparables à celle définies dans Wu & al (2010)[27] dans le cas discret.

1.1 Méthodes du choix de paramètre de lissage

Le paramètre de lissage est un élément très important dans les différentes méthodes d'estimation de la fonction de densité. Il dépend de l'échantillon ainsi que de sa taille. Dans le cas de l'estimation par la méthode du noyau, ce paramètre est indispensable pour la convergence de l'estimateur et l'efficacité du lissage, et donc la qualité de cette estimation. Pour cela le choix du paramètre de lissage a fait l'objet de nombreuses contributions dans la littérature, plusieurs procédures automatiques ont été développées et des études comparatives ont été effectuées.

Dans certaines situations, il suffit de choisir subjectivement le paramètre de lissage en observant les estimations de la densité produites par plusieurs paramètres, on peut commencer avec une petite valeur du paramètre et diminuer la qualité de lissage en maximisant successivement cette valeur jusqu'à atteindre l'estimateur le plus proche graphiquement à la fonction de densité (voir figure (1.1)).

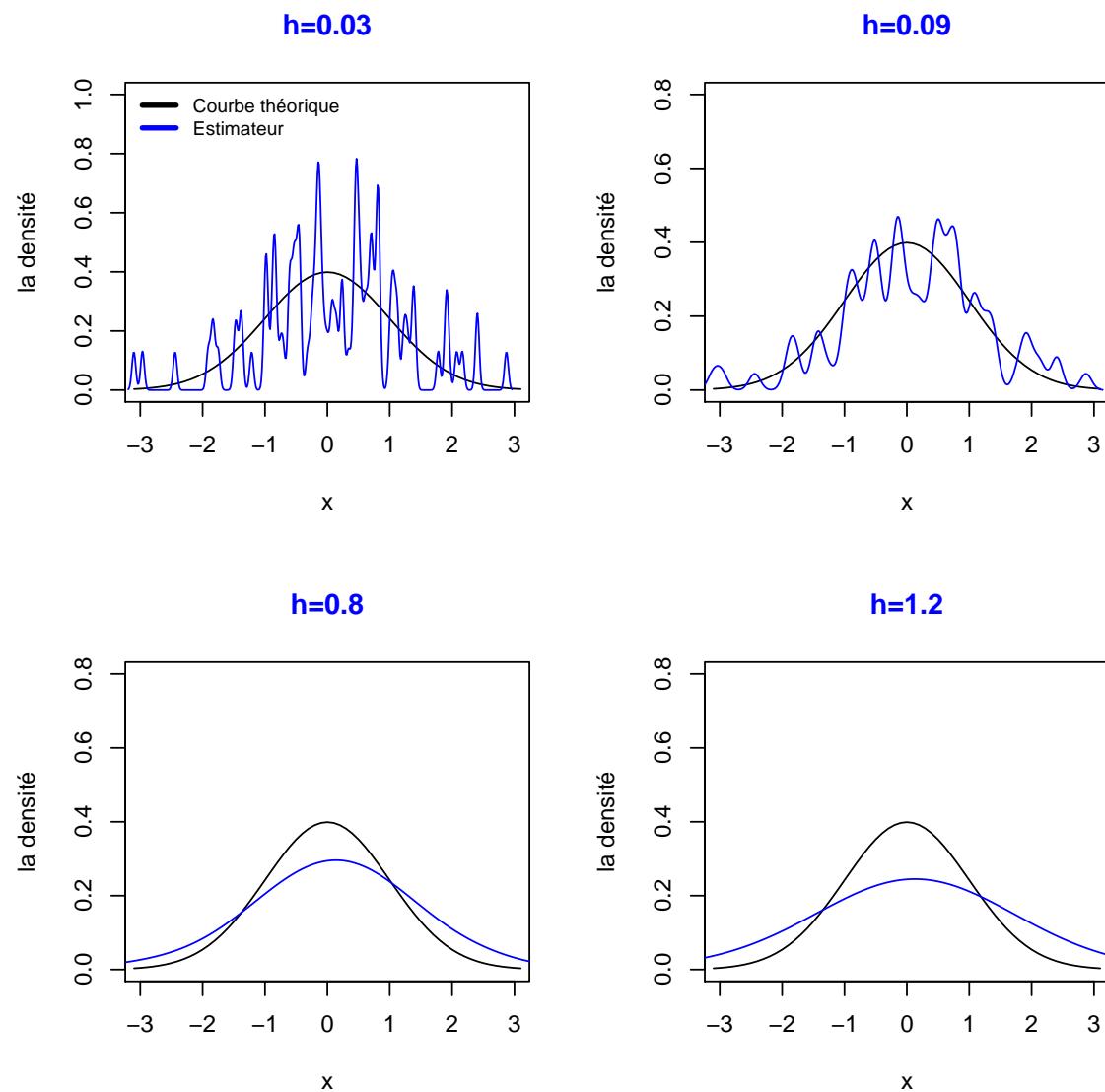


FIGURE 1.1 – L'estimation par noyau avec différents paramètres de lissage.

La décision d'un choix optimal pour le paramètre de lissage consiste à minimiser les critères d'erreurs. L'erreur quadratique moyenne intégrée (*MISE*)¹ relative à l'estimateur à noyau \hat{f}_{nh} de la densité f est donné par :

$$MISE(\hat{f}_{nh}) = \frac{h_n^4}{4} \mu_2^2(K) R(f'') + \frac{R(K)}{nh_n}, \quad (1.1)$$

avec

$$\mu_2(K) = \int_{\mathbb{R}} t^2 K(t) dt,$$

et

$$R(g) = \int_{\mathbb{R}} g^2(t) dt,$$

pour toute fonction g de carré intégrable.

La fenêtre optimale qui minimise l'erreur quadratique moyenne intégrée est donnée par :

$$h_{opt} = \left(\frac{R(K)}{\mu_2^2(K) R(f'')} \right)^{\frac{1}{5}} n^{-\frac{1}{5}}. \quad (1.2)$$

Cette expression ne peut être utilisée en pratique car elle dépend de la densité inconnue f à travers $R(f'')$. De ce fait, nous devons construire des procédures qui nous permettra d'éviter cette quantité et de donner un paramètre de lissage approprié.

1.1.1 Méthode Rule of Thumb

Le choix du paramètre de lissage par cette méthode consiste à remplacer la partie inconnue $R(f'')$ dans l'expression (1.2) par une distribution classique afin d'obtenir un estimateur pour h . Si les donner suivent la loi normale centrée et de variance σ^2 donc on a :

$$R(f'') = \int (f''(x))^2 dx = \frac{3}{8\sqrt{\pi}} \sigma^{-5}.$$

1. L'erreur quadratique moyenne intégrée (mean integrated squared error MISE) est défini par :
 $MISE(\hat{f}_n(x)) = \int \mathbb{E} [\{\hat{f}_n(x) - f(x)\}^2] dx.$

De plus, si on utilise un noyau gaussien, on aura :

$$\mu_2(K) = \int_{\mathbb{R}} u^2 K(u) du = \int u^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = 1,$$

et

$$R(K) = \int K^2(u) du = \int \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \right)^2 du = \frac{1}{2\pi} \int e^{-u^2} du = \frac{1}{2\sqrt{\pi}}.$$

Donc le paramètre de lissage défini par (1.2) et que l'on note dans ce cas par h_{rot} est donné par :

$$\begin{aligned} h_{rot} &= \left(\frac{1}{2\sqrt{\pi}} \right)^{\frac{1}{5}} \left(\frac{3}{8\sigma^5 \sqrt{\pi}} \right)^{-\frac{1}{5}} n^{-\frac{1}{5}} \\ &= \left(\frac{4}{3} \right)^{\frac{1}{5}} \sigma n^{-\frac{1}{5}} \\ &\approx 1.06 \sigma n^{-\frac{1}{5}}. \end{aligned}$$

Il suffit donc d'estimer σ à partir des données ainsi nous obtenons le paramètre h_{rot} . Une version plus robuste contre les valeurs aberrantes, si en utilisant une mesure de l'étendue de l'échantillon. L'écart inter-quartile R^2 permet d'obtenir des meilleurs résultats que l'écart-type. Le paramètre de lissage h_{rot} devient :

$$h_{rot} \approx 1.06 \min \left(\sigma, \frac{R}{1.34} \right) n^{-\frac{1}{5}}.$$

1.1.2 Méthode plug-in

L'idée de cette méthode est basée sur l'estimation de la partie inconnue de l'équation (1.2), Woodroof (1970)[26] a proposé un choix initiale h_0 de la fenêtre lié à un estimateur de $R(f'')$ qui est défini par :

$$\hat{R}_{nh}(f'') = R(\hat{f}_{nh_0}''),$$

2. Pour une série statistique des N données est rangée par ordre croissant. Le premier quartile Q_1 est la plus petite donnée de la série telle qu'au moins 25% des données soient inférieures ou égales à Q_1 . Le troisième quartile Q_3 est la plus petite donnée de la série telle qu'au moins 75% des données soient inférieures ou égales à Q_3 .

L'écart inter-quartile R est : $R = Q_3 - Q_1$.

où \widehat{f}_{nh_0}'' désigne la dérivée seconde de l'estimateur à noyau \widehat{f}_{nh_0} . Avec un noyau K deux fois dérivable, donc l'estimateur $\hat{R}_{nh}(f'')$ est donné par :

$$\hat{R}_{nh}(f'') = \frac{1}{n^2 h_0^6} \int \left(\sum_{i=1}^n K'' \left(\frac{x - X_i}{h_0} \right) \right)^2 dx.$$

Il suffit donc de mettre l'estimateur $\hat{R}_{nh}(f'')$ dans l'équation (1.2) pour obtenir un estimateur h_1 de la fenêtre h_{opt} défini par :

$$h_1 = \left(\frac{R(K)}{\mu_2^2(K) \hat{R}_{nh}(f'')} \right)^{\frac{1}{5}} n^{-\frac{1}{5}}.$$

Remarque : La fenêtre h_1 est sensible au choix initial de h_0 , Voir Scott et Factor (1981)[20] pour plus de détails.

1.1.3 Méthode de validation croisée

La méthode de validation croisée a été proposée par Rudemo (1982)[19] et Bowman (1984)[5], le principe de cette méthode est de minimiser l'erreur quadratique intégrée ISE pour un estimateur à noyau \widehat{f}_{nh} de la densité f défini par :

$$\begin{aligned} ISE(\widehat{f}_{nh}(x)) &= \int [\widehat{f}_{nh}(x) - f(x)]^2 dx \\ &= \int \widehat{f}_{nh}^2(x) dx - 2 \int \widehat{f}_{nh}(x) f(x) dx + \int f^2(x) dx. \end{aligned}$$

Minimiser $ISE(\widehat{f}_{nh})$ par rapport à h est équivalent à minimiser :

$$\int \widehat{f}_{nh}^2(x) dx - 2 \int \widehat{f}_{nh}(x) f(x) dx. \quad (1.3)$$

Un estimateur de la partie inconnue de (1.3) est donné par $\frac{1}{n} \sum_{i=1}^n \widehat{f}_{n,-i}(x)$ où $\widehat{f}_{n,-i}$ est

l'estimateur à noyau basé sur $(n-1)$ observations différentes du point d'ordre i . Donc le critère de validation croisée est donné par :

$$CV(h) = \int \widehat{f}_{nh}^2(x) dx - \frac{2}{n} \sum_{i=1}^n \widehat{f}_{n,-i}(X_i).$$

Le paramètre de lissage sélectionné par la méthode de validation croisée est le minimum de cette expression par rapport à h .

1.1.4 Méthode de vraisemblance de la validation croisée

Les paramètres d'un modèle statistique peuvent être estimés par la méthode du maximum de vraisemblance. Une méthode appelée la vraisemblance de la validation croisée a été présentée par Habbema, Hermans et Vandenbroek (1974)[11] et Duin (1976)[9], elle consiste à maximiser la fonction de vraisemblance ou, plus souvent le logarithme de cette fonction par rapport à h . L'estimateur $\log \hat{f}_{n,-i}$ de $\log f$ est construit à partir des données $\{Y, X_j, j = 1, \dots, n, j \neq i\}$ tel que Y est indépendant à X_j . Le critère de vraisemblance de validation croisée est alors donné par

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

L'estimation du paramètre de lissage par cette méthode est obtenue en maximisant cette expression.

1.1.5 Méthode de validation croisée biaisée

Cette méthode a été introduite par Scott et Terrell en 1987[21]. Rappelant qu'à partir du critère de *MISE* relative à l'estimateur à noyau \hat{f}_{nh} donné par l'expression (1.1), on peut donner le paramètre de lissage si l'on connaît $R(f'')$, Scott et Terrell [21] montraient :

$$\mathbb{E}[R(\hat{f}_{nh}'')] = R(f'') + \frac{R(K'')}{nh_n^5} + o(h^2).$$

Ils proposent alors d'estimer $R(f'')$ par :

$$\hat{R}(f'') = R(\hat{f}_{nh}'') - \frac{R(K'')}{nh_n^5}.$$

L'estimateur $\hat{R}(f'')$ est donné par :

$$\hat{R}(f'') = \frac{1}{n^2 h_n^5} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n K'' * K'' \left(\frac{X_i - X_j}{h_n} \right), \quad (1.4)$$

où $*$ représente le produit de convolution défini par :

$$K'' * K''(x) = \int K''(x - y)K''(y)dy.$$

La formule (1.4) est obtenue de la façon suivante :

$$\begin{aligned} R(\widehat{f}_{nh}'') &= \frac{1}{n^2 h_n^6} \int \left(\sum_{i=1}^n K''\left(\frac{x - X_i}{h_n}\right) \right)^2 dx \\ &= \frac{1}{n^2 h_n^6} \sum_{i=1}^n \sum_{j=1}^n \int K''\left(\frac{x - X_i}{h_n}\right) K''\left(\frac{x - X_j}{h_n}\right) dx \\ &= \frac{1}{n^2 h_n^6} \sum_{i=1}^n \int K''^2\left(\frac{x - X_i}{h_n}\right) dx + \frac{1}{n^2 h_n^6} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int K''\left(\frac{x - X_i}{h_n}\right) K''\left(\frac{x - X_j}{h_n}\right) dx, \end{aligned}$$

en effectuant le changement de variable $u = \frac{x - X_i}{h_n}$ on obtient :

$$R(\widehat{f}_{nh}'') = \frac{1}{nh_n^5} \int K''^2(u) du + \frac{1}{n^2 h_n^5} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int K''(u) K''\left(\frac{X_i - X_j}{h_n} + u\right) du,$$

comme le noyau K est symétrique alors on a :

$$\begin{aligned} R(\widehat{f}_{nh}'') &= \frac{1}{nh_n^5} \int K''^2(u) du + \frac{1}{n^2 h_n^5} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int K''(u) K''\left(\frac{X_j - X_i}{h_n} - u\right) du \\ &= \frac{1}{nh_n^5} \int K''^2(u) du + \frac{1}{n^2 h_n^5} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n K'' * K''\left(\frac{X_j - X_i}{h_n}\right) \\ &= \frac{1}{nh_n^5} \int K''^2(u) du + \frac{1}{n^2 h_n^5} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n K'' * K''\left(\frac{X_i - X_j}{h_n}\right) \\ &= \frac{R(K'')}{nh_n^5} + \frac{1}{n^2 h_n^5} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n K'' * K''\left(\frac{X_i - X_j}{h_n}\right). \end{aligned}$$

On conclut donc :

$$\hat{R}(f'') = \frac{1}{n^2 h_n^5} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n K'' * K''\left(\frac{X_i - X_j}{h_n}\right).$$

Donc on remplacer $R(f'')$ dans l'expression (1.1) par son estimateur et nous obtenons le critère de validation croisée baisée suivant :

$$BCV(h) = \frac{R(K)}{nh_n} + \frac{\mu_2^2(K)}{4n^2h_n} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n K'' * K'' \left(\frac{X_i - X_j}{h_n} \right).$$

Le paramètre de lissage h choisi par cette méthode est la valeur de h qui minimise le $BCV(h)$.

L'estimation par la méthode du noyau relative à des processus ergodiques à temps continu a reçu un intérêt particulier ces dernières années. De nombreux résultats traitant la vitesse de convergence et la normalité asymptotique.

Dans ce qui suit nous énonçons la théorie ergodique pour les processus stationnaires.

1.2 Ergodicité des processus stationnaires

La théorie ergodique s'intéresse à l'étude des systèmes dynamiques. Elle fait état de l'équivalence entre le comportement moyen de l'ensemble des systèmes dynamiques et la moyenne temporelle des comportements d'un système dynamique. Les origines de l'ergodicité remontent à la mécanique statistique, Birkoff (1931)[4] et Von Neumann (1932)[25] sont les pionniers dans ce domaine. Dans le cas des processus à temps continu le cadre ergodique est plus générale que la dépendance faible des données.

Définition 1.2.1. (*Ensemble invariant*). Soit $\{Y_t\}_{t \in \mathbb{R}^+}$ un processus à temps continu défini sur un espace mesurable $(\Omega, \mathcal{X}, \mu)$. Pour $\delta > 0$, soit T^δ une transformation δ -shift (i.e. $(T^\delta(Y))_s = Y_{s+\delta}$). Un ensemble mesurable A est dit δ -invariant, si il ne change pas sous une transformation δ -shift (i.e. $(T^\delta(A)) = A$)).

Définition 1.2.2. (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 1.2.3. (Ergodicité). $Y = \{Y_t\}_{t \in \mathbb{R}^+}$ est dit ergodique s'il est δ -ergodique pour tout $\delta > 0$.

Définition 1.2.4. (Processus stationnaire en temps discret). Soit $Y = (Y_0, Y_1, \dots)$ un processus défini sur un espace de probabilité $(E, \mathcal{F}, \mathbb{P})$. 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 $n \in \mathbb{N}$, $F_1, \dots, F_n \in \mathcal{F}$, $t_1, \dots, t_n, s \in \mathbb{Z}$, on a :

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

Définition 1.2.5. (*Processus stationnaire en temps continu*). Un processus $(Y_t)_{t \in \mathbb{R}^+}$ est dit strictement stationnaire si les lois jointes de $(Y_{t_1}, \dots, Y_{t_k})$ et de $(Y_{t_1+h}, \dots, Y_{t_k+h})$ sont identiques pour tout $k \in \mathbb{N}$ et pour tout $t_1, \dots, t_n, h \in \mathbb{R}^+$.

Donnons maintenant le théorème ponctuel ergodique de Birkhoff lié aux processus stationnaires à temps discret et continu, (voir Krengel (1985)[13], théorème 4.4 p.26).

Théorème 1.2.1. [13] (*Birkoff en temps discrète*). Soit Y_0, Y_1, \dots est un processus réel stationnaire, Y_0 est intégrable et \mathcal{F} une σ -algébre 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, alors :

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

Théorème 1.2.2. [13] (*Birkoff en temps continu*). Si $(Y_t)_{t \in [0, T]}$ avec $T \in \mathbb{R}^+$ est un processus réel stationnaire ergodique, alors :

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

Chapitre 2

Choix optimal du paramètre de lissage dans l'estimation de la densité à temps discret

On suppose que les X_i sont à valeurs réelles et que f est une densité deux fois continûment différentiable par rapport à la mesure de Lebesgue sur \mathbb{R} .

2.1 Estimation de la fonction de densité par les histogrammes

La méthode d'estimation par histogramme est la plus naturelle, elle est répondue à l'objectif d'une représentation de la distribution de données. Supposons que l'on ait n observations x_1, x_2, \dots, x_n issues d'une même loi de probabilité inconnue de densité f , où f est à support borné $[a_0, a_m]$. Estimer cette densité f par la méthode d'histogramme revient à approcher f par une fonction en escaliers. Pour cela, on partitionne l'intervalle de référence $[a_0, a_m]$ en $m \in \mathbb{N}$ classes C_j de la forme $[a_{j-1}, a_j[$, $j \in \{1, \dots, m\}$. La largeur de la classe C_j est alors $h_j = a_j - a_{j-1}$.

L'estimateur par histogramme s'écrit alors : $\exists j \in \{1, \dots, m\}$, tel que $\forall x \in [a_{j-1}, a_j[$

$$\widehat{f}_h(x) = \sum_{j=1}^m \frac{f_j}{h_j} \mathbf{1}_{[a_{j-1}, a_j[}(x),$$

où f_j est la fréquence empirique du nombre d'observations appartenant à la classe correspondante, tel que

$$f_j = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[a_{j-1}, a_j]}(X_i).$$

2.1.1 Estimation par histogramme mobile

L'estimateur par histogramme précédent \hat{f}_h n'est pas un bon estimateur. Considérons la classe $C_j = [a_{j-1}, a_j]$ et imaginons que le point $x \in C_j$ où l'on veut estimer $f(x)$ par $\hat{f}_h(x)$ se situe près de l'extrémité. Alors, toutes les observations de la classe C_j interviennent dans le calcul de \hat{f}_h , mais on se rend compte qu'une observation située près de a_j sera prise en compte, alors qu'elle est assez éloignée de x , et qu'une observation située tout près de x dans la classe C_{j-1} n'entre pas dans le calcul de \hat{f}_h . Pour remédier cet estimateur, on peut alors utiliser l'histogramme mobile, qui est un translaté de l'histogramme de manière à ce que l'observation x où l'on estime, se retrouve au centre d'une classe, plus précisément au centre de la classe $[x - \frac{h}{2}, x + \frac{h}{2}]$ où h désigne la largeur d'une classe (ou la fenêtre de l'histogramme). L'estimateur par histogramme mobile s'écrit alors :

$$\hat{f}_{nh}(x) = \frac{1}{nh} \sum_{i=1}^n \mathbf{1}_{[x - \frac{h}{2}, x + \frac{h}{2}]}(X_i).$$

2.1.2 Risque de l'estimateur par histogramme mobile

L'évolution de la similitude entre l'estimateur \hat{f}_{nh} et la vraie densité f à estimer, nécessite des critères d'erreurs. L'un des critères universelles pour mesurer la qualité de cette estimateur est représenté par l'erreur quadratique moyenne (mean squared error MSE) défini par :

$$\begin{aligned} MSE(\hat{f}_{nh}(x)) &= \mathbb{E} \left[(\hat{f}_{nh}(x) - f(x))^2 \right] \\ &= \mathbb{E} \left[\hat{f}_{nh}^2(x) \right] - \left[\mathbb{E}\{\hat{f}_{nh}(x)\} \right]^2 + \left[\mathbb{E}\{\hat{f}_{nh}(x)\} - f(x) \right]^2 \\ &= Bias^2\{\hat{f}_{nh}\} + Var\{\hat{f}_{nh}\}. \end{aligned}$$

- Etude du biais :

$$\begin{aligned}
\mathbb{E} \left[\widehat{f}_{nh}(x) \right] - f(x) &= \mathbb{E} \left[\frac{1}{nh} \sum_{i=1}^n \mathbf{1}_{[x-\frac{h}{2}, x+\frac{h}{2}[} (X_i) \right] - f(x) \\
&= \frac{1}{h} \mathbb{E} \left[\mathbf{1}_{[x-\frac{h}{2}, x+\frac{h}{2}[} (X_i) \right] - f(x) \\
&= \frac{1}{h} \int \mathbf{1}_{[x-\frac{h}{2}, x+\frac{h}{2}[} (X_i) f(x) dx - f(x) \\
&= \frac{1}{h} \int_{[x-\frac{h}{2}, x+\frac{h}{2}[} f(x) dx - f(x) \\
&= \frac{1}{h} \mathbb{P} \left\{ X_i \in \left[x - \frac{h}{2}, x + \frac{h}{2} \right] \right\} - f(x) \\
&= \frac{p_i}{h} - f(x),
\end{aligned}$$

où $p_i = \mathbb{P} \left\{ X_i \in \left[x - \frac{h}{2}, x + \frac{h}{2} \right] \right\}$ la probabilité des X_i trouvant dans l'intervalle $\left[x - \frac{h}{2}, x + \frac{h}{2} \right]$.

- Etude de la variance :

$$\begin{aligned}
Var \left[\widehat{f}_{nh}(x) \right] &= Var \left[\frac{1}{nh} \sum_{i=1}^n \mathbf{1}_{[x-\frac{h}{2}, x+\frac{h}{2}[} (X_i) \right] \\
&= \frac{1}{nh^2} Var \left[\mathbf{1}_{[x-\frac{h}{2}, x+\frac{h}{2}[} (X_i) \right] \\
&= \frac{1}{nh^2} \left\{ \mathbb{E} \left[\left(\mathbf{1}_{[x-\frac{h}{2}, x+\frac{h}{2}[} (X_i) \right)^2 \right] - \left[\mathbb{E} \left[\mathbf{1}_{[x-\frac{h}{2}, x+\frac{h}{2}[} (X_i) \right] \right]^2 \right\} \\
&= \frac{1}{nh^2} [p_i - p_i^2] \\
&= \frac{p_i(1-p_i)}{nh^2}.
\end{aligned}$$

- L'erreur quadratique moyenne est donné par :

$$MSE(\widehat{f}_{nh}(x)) = \left(\frac{p_i}{h} - f(x) \right)^2 + \frac{p_i(1-p_i)}{nh^2}.$$

Lemme 2.1.1. Si X_1, \dots, X_n sont indépendantes de même loi de densité f supportée par $[0,1]$ et \widehat{f}_{nh} est l'estimateur par histogramme mobile avec $m = 1/h$ classes, alors l'erreur quadratique moyenne intégrée (mean integrated squared error MISE) de \widehat{f}_{nh}

est donné par :

$$\begin{aligned}
 MISE(\widehat{f}_{nh}(x)) &= \int_0^1 MSE(\widehat{f}_{hh}(x))dx \\
 &= \int_0^1 \mathbb{E} [\{\widehat{f}_{nh}(x) - f(x)\}^2] dx \\
 &= \int_0^1 f^2(x)dx + \frac{1}{nh} - \frac{n+1}{nh} \sum_{j=1}^m p_j^2.
 \end{aligned}$$

2.1.3 Choix de la fenêtre par la méthode de validation croisée

Afin d'établir une méthode de choix de h indépendant de f , nous commençons par estimer le risque de l'estimateur \widehat{f}_{nh} en utilisant uniquement les observations x_1, \dots, x_n . Soit $\hat{J}(h, x_1, \dots, x_n)$ un estimateur de $MISE(\widehat{f}_{nh}(x)) - \int_0^1 f^2(x)dx$. Pour que la méthode de sélection de h conduise vers des résultats raisonnables, on demande de l'estimateur \hat{J} être sans biais, c'est-à-dire :

$$\mathbb{E} [\hat{J}(h, x_1, \dots, x_n)] = MISE(\widehat{f}_{nh}(x)) - \int_0^1 f^2(x)dx.$$

On détermine la valeur de h en minimisant $\hat{J}(h, x_1, \dots, x_n)$ par rapport à h , ce minimum est atteint et sélectionnée comme fenêtre pour l'estimateur par histogramme mobile.

Principe de la méthode : Avant de commencer, on choisit une partition uniforme C_1, \dots, C_m de l'intervalle $[0, 1[$ tel que :

$$C_j = \left[\frac{j-1}{m}, \frac{j}{m} \right] \quad j = 1, \dots, m.$$

Pour toute densité f et pour tout histogramme mobile \widehat{f}_{nh} , d'après le lemme 2.1.1, on a :

$$\begin{aligned}
 J(h, x_1, \dots, x_n) &= MISE(\widehat{f}_{nh}(x)) - \int_0^1 f^2(x)dx \\
 &= \frac{1}{nh} - \frac{n+1}{nh} \sum_{j=1}^m p_j^2. \tag{2.1}
 \end{aligned}$$

Pour déterminer un estimateur sans biais de J , il suffit de déterminer un estimateur sans biais de p_j^2 , pour tout $j = 1, \dots, m$. Une approche simple consisterait à estimer p_j^2 par \hat{p}_j^2 , où \hat{p}_j est la probabilité empirique :

$$\hat{p}_j = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{C_j}(X_i).$$

Comme $\sum_{i=1}^n \mathbb{1}_{C_j}(X_i)$ suit la loi binomiale de paramètres (n, p_j) , on a :

$$Var(\hat{p}_j) = \frac{p_j(1 - p_j)}{n},$$

par conséquent,

$$\begin{aligned} \mathbb{E}(\hat{p}_j^2) &= Var(\hat{p}_j) + (\mathbb{E}(\hat{p}_j))^2 \\ &= p_j^2(1 - \frac{1}{n}) + \frac{p_j}{n}. \end{aligned} \quad (2.2)$$

Cette égalité nous montre d'une part que l'idée naïve d'estimer p_j^2 par \hat{p}_j^2 ne conduit pas vers un estimateur sans biais. Mais, d'autre part, ce petit calcul que nous venons d'effectuer prépare le terrain pour déterminer l'estimateur utilisé par la méthode de validation croisée. En effet, comme \hat{p}_j est un estimateur sans biais de p_j , il résulte de (2.2) que $\hat{p}_j^2 - \frac{\hat{p}_j}{n}$ est un estimateur sans biais de $p_j^2(1 - \frac{1}{n})$. Par conséquent, pour tout $j = 1, \dots, m$,

$$\frac{\hat{p}_j^2 - \hat{p}_j/n}{1 - 1/n} = \frac{n}{n-1} \hat{p}_j^2 - \frac{1}{n-1} \hat{p}_j$$

est un estimateur sans biais de p_j^2 . En injectant cet estimateur dans le membre droit de l'égalité (2.1) et en utilisant le fait que $\sum_{j=1}^m \hat{p}_j = 1$, nous obtenons le résultat suivant.

Proposition 2.1.1. *Si f est une densité de carré intégrable et si \widehat{f}_{nh} est l'estimateur de l'histogramme mobile à $m = 1/h$ classes basé sur l'échantillon x_1, \dots, x_n ayant f*

pour densité de probabilité, alors :

$$\hat{J}(h, x_1, \dots, x_n) = \frac{2}{(n-1)h} - \frac{n+1}{(n-1)h} \sum_{j=1}^m \hat{p}_j^2$$

est un estimateur sans biais de $MISE(\hat{f}_{nh}) - \int_0^1 f^2$.

Nous pouvons à présent énoncer la méthode de validation croisée. Nous allons le faire dans le cadre général, sans supposer que les observations sont incluses dans $[0, 1]$. Dans ce cas, on peut poser $a = x_{\min}$ et $b = x_{\max}$ et pour tout $m \in \mathbb{N}$ choisir la fenêtre $h = (b-a)/m$. On définit alors les classes $C_j = [a + (j-1)h, a + jh[$ pour $j = 1, \dots, m-1$ et $C_m = [b-h, b]$.

2.2 Estimation de la fonction de densité par la méthode du noyau

Soit $(X_i)_{i \geq 1}$ une suite des variables aléatoires indépendantes et de même loi, de densité de probabilité f . L'estimation par la méthode du noyau est l'une des méthodes la plus utilisée qui permet d'obtenir un estimateur continu de la densité f . Cette méthode a été introduit par Rosenblatt (1956)[18] et développé par Parzen (1962)[17].

L'estimateur à noyau \hat{f}_{nh} est défini pour tout $x \in \mathbb{R}$ par :

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

où $\{h_n\}_{n \geq 1}$ est une suite de réels positifs appelés le paramètre de lissage ou la largeur de la fenêtre, qui tend vers 0 quand n tend vers l'infini et K est la fonction noyau continue symétrique (i.e. $K(t) = K(-t)$), positive (i.e. $K(t) > 0, \forall t \in \mathbb{R}$) et d'intégrale égale à 1.

Noyaux usuels :

- **Le noyau uniforme :**

$$K(u) = \frac{1}{2} \mathbf{1}_{[-1,1]}(u).$$

- Le noyau triangulaire :

$$K(u) = (1 - |u|)\mathbb{1}_{[-1,1]}(u).$$

- Le noyau d'Epanechnikov :

$$K(u) = \frac{3}{4}(1 - u^2)\mathbb{1}_{[-1,1]}(u).$$

- Le noyau gaussien :

$$K(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}, \quad u \in \mathbb{R}.$$

Lemme 2.2.1. *Un estimateur à noyau est une densité de probabilité.*

Démonstration.

$$\begin{aligned} \int_{\mathbb{R}} \widehat{f}_{nh_n}(x)dx &= \frac{1}{nh_n} \sum_{i=1}^n \int_{\mathbb{R}} K\left(\frac{x - X_i}{h_n}\right) dx \\ &= \frac{1}{nh_n} \sum_{i=1}^n \int_{\mathbb{R}} K(u)h_n du \text{ (changement de variable } u = \left(\frac{x - X_i}{h_n}\right)) \\ &= \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}} K(u)du = \frac{1}{n}n = 1. \end{aligned}$$

□

2.2.1 Propriétés de l'estimateur à noyau

Nous allons maintenant donné quelques propriétés statistiques de l'estimateur \widehat{f}_{nh} où la densité f est bornée dont la dérivée seconde est bornée et le noyau K vérifier les conditions suivantes :

$$\int_{\mathbb{R}} tK(t)dt = 0, \int_{\mathbb{R}} t^2K(t)dt = \mu_2(K) > 0, \int_{\mathbb{R}} K^2(t)dt = R(K),$$

où $\mu_2(K)$ et $R(K)$ sont deux constants.

• **Etude du biais :**

Le biais de l'estimateur \widehat{f}_{nh} est :

$$|Bias(\widehat{f}_{nh}(x))| \leq C_1 h_n^2, \quad (2.3)$$

où $C_1 = \max_x |f''(x)| \frac{\mu_2(K)}{2}$.

Démonstration.

$$\begin{aligned} \mathbb{E}[\widehat{f}_{nh}(x)] &= \mathbb{E}\left[\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)\right] \\ &= \frac{1}{h_n} \mathbb{E}\left[K\left(\frac{x-t}{h_n}\right)\right] \\ &= \frac{1}{h_n} \int_{\mathbb{R}} K\left(\frac{x-t}{h_n}\right) f(t) dt, \end{aligned}$$

on pose : $u = \frac{x-t}{h_n} \Rightarrow t = x - h_n u$

$$\mathbb{E}[\widehat{f}_{nh}(x)] = \int_{\mathbb{R}} K(u) f(x - h_n u) du,$$

Puisque f est deux fois continûment différentiable, le développement de Taylor de $f(x - hu)$ nous donne :

$$f(x - h_n u) = f(x) - h_n u f'(x) + \frac{1}{2!} (-h_n u)^2 f''(\theta_u) \quad (\text{avec } \theta_u \in]x - h_n u, x]).$$

Donc on obtient :

$$\mathbb{E}[\widehat{f}_{nh}(x)] = \int_{\mathbb{R}} K(u) \left[f(x) - h_n u f'(x) + \frac{h_n^2 u^2}{2} f''(\theta_u) \right] du.$$

Comme f'' est continue alors $f''(\theta_u) \rightarrow f''(x)$ quand $n \rightarrow \infty$, alors

$$\mathbb{E}[\widehat{f}_{nh}(x)] = f(x) \int_{\mathbb{R}} K(u) du - h_n f'(x) \int_{\mathbb{R}} u K(u) du + \frac{h_n^2}{2} f''(x) \int_{\mathbb{R}} u^2 K(u) du.$$

Puisque :

$$\int_{\mathbb{R}} K(t) dt = 1, \int_{\mathbb{R}} t K(t) dt = 0, \int_{\mathbb{R}} t^2 K(t) dt = \mu_2(K),$$

donc :

$$\mathbb{E}[\widehat{f}_{nh}(x)] = f(x) + \frac{h_n^2}{2} f''(x) \mu_2(K),$$

il en résulte que :

$$\begin{aligned} |Bias(\widehat{f}_{nh}(x))| &= |\mathbb{E}[\widehat{f}_{nh}(x)] - f(x)| \\ &\leq h_n^2 \max_x |f''(x)| \frac{\mu_2(K)}{2} \\ &\leq C_1 h_n^2. \end{aligned}$$

□

• **Etude de la variance :**

La variance de l'estimateur \widehat{f}_{nh} est :

$$Var(\widehat{f}_{nh}(x)) \leq \frac{C_2}{nh_n}, \quad (2.4)$$

où $C_2 = \max_x |f(x)|R(K)$.

Démonstration.

$$\begin{aligned} Var(\widehat{f}_{nh}(x)) &= Var\left[\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)\right] \\ &= \frac{1}{nh_n^2} Var\left[K\left(\frac{x-t}{h_n}\right)\right] \\ &\leq \frac{1}{nh_n^2} \mathbb{E}\left[\left(K\left(\frac{x-t}{h_n}\right)\right)^2\right] \\ &\leq \frac{1}{nh_n^2} \int_{\mathbb{R}} \left[K\left(\frac{x-t}{h_n}\right)\right]^2 f(t) dt \\ &\leq \frac{1}{nh_n} \int_{\mathbb{R}} K^2(u) f(x-h_n u) du \quad (\text{par changement de variable}). \end{aligned}$$

Comme f est continue donc $f(x-h_n u) \simeq f(x)$ quand n est assez grand.

$$\begin{aligned} Var(\widehat{f}_{nh}(x)) &\leq \frac{1}{nh_n} f(x) \int_{\mathbb{R}} K^2(u) du \\ &\leq \max_x |f(x)| R(K) \frac{1}{nh_n} \\ &\leq \frac{C_2}{nh_n}. \end{aligned}$$

□

- MSE de l'estimateur :

On déduit de (2.3) et (2.4) que l'erreur quadratique moyenne de \hat{f}_{nh} admet la majoration suivante :

$$MSE(\hat{f}_{nh}(x)) \leq C_1^2 h_n^4 + \frac{C_2}{nh_n}. \quad (2.5)$$

2.2.2 Résultats asymptotiques

Théorème 2.2.1. (*Parzen [17]*) (*Convergence en moyenne quadratique*). Soit f une densité continue et \hat{f}_{nh} son estimateur à noyau. Si les hypothèses suivantes sont réalisées :

$$(A1) : \int_{\mathbb{R}} K(u)du = 1, \int_{\mathbb{R}} |K(u)|du < \infty, \sup_u |K(u)| < \infty, \lim_{n \rightarrow \infty} |uK(u)| = 0$$

$$(A2) : \lim_{n \rightarrow \infty} h_n = 0, \lim_{n \rightarrow \infty} nh_n = \infty,$$

alors \hat{f}_{nh} est un estimateur convergent en moyenne quadratique c'est à dire :

$$\lim_{n \rightarrow \infty} MSE(\hat{f}_{nh}(x)) = 0.$$

Théorème 2.2.2. (*La vitesse de convergence presque complète¹*). Supposons que les hypothèses suivantes soient réalisées :

$$(A3) : f \in C^2$$

$$(A4) : \lim_{n \rightarrow \infty} h_n = 0, \lim_{n \rightarrow \infty} \left(\frac{nh_n}{\log n} \right) = \infty$$

$$(A5) : K \text{ est borné, de carré intégrable et } \int_{\mathbb{R}} tK(t)dt = 0, \int_{\mathbb{R}} t^2 K(t)dt < \infty,$$

alors on a :

$$\hat{f}_{nh}(x) - f(x) = O(h_n^2) + O_{p.co} \left(\sqrt{\frac{\log n}{nh_n}} \right).$$

1. La convergence presque complète (*p.co*). Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires, X une variable aléatoire et u_n est une suite numérique.

- On dit que $\lim_{n \rightarrow \infty} X_n = X$ *p.co*ssi $\forall \epsilon > 0, \sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| \geq \epsilon) < \infty$.
- On dit que $X_n - X = O(u_n)$ *p.co*ssi $\forall \epsilon > 0, \sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| \geq \epsilon u_n) < \infty$.

Le succès rencontré par l'estimateur à noyau s'explique par sa simplicité, sa flexibilité et aussi ses propriétés de convergence. Il laisse à l'utilisateur une grande latitude non seulement dans le choix du noyau K , mais aussi dans le choix du paramètre de lissage h .

2.3 Choix du paramètre de lissage par la méthode de validation croisée

L'estimateur à noyau et ses propriétés sont principalement conditionné par le paramètre de lissage h , ce paramètre représente en quelque sorte une fenêtre qui permet de déterminer le degré de lissage de l'estimateur \hat{f}_{nh} . D'après l'expression (2.5), si h est trop petit, le biais de l'estimateur devient petit devant sa variance et l'estimateur est très volatile et on parle de sous-lissage (under-smoothing). Dans le cas contraire, lorsque h est trop grand, la variance devient petite et c'est le biais qui devient dominant. L'estimateur est alors trop lisse et est de moins à moins influencé par les données, on parle alors d'un effet de sur-lissage (over-smoothing).

En pratique, il est nécessaire de trouver la méthode qui permet d'éviter le sous-lissage et le sur-lissage pour la sélection de la fenêtre h .

2.3.1 Choix pour des données indépendantes

Parmi les mesures de divergence entre la densité f et son estimateur \hat{f}_{nh} on trouve l'erreur quadratique intégrée (integrated squared error ISE) définie par :

$$\begin{aligned} ISE(\hat{f}_{nh}(x)) &= \int \left[\hat{f}_{nh}(x) - f(x) \right]^2 dx \\ &= \int \hat{f}_{nh}^2(x) dx - 2 \int \hat{f}_{nh}(x) f(x) dx + \int f^2(x) dx. \end{aligned}$$

Le paramètre de lissage choisi par la méthode de la validation croisée (cross validation CV) est la valeur de h qui minimise un estimateur de ISE . puisque $\int f^2(x) dx$ ne dépend pas du paramètre de lissage h . On peut choisir le paramètre de lissage de façon à ce qu'il minimise un estimateur de :

$$ISE(\hat{f}_{nh}(x)) - \int f^2(x) dx = \int \hat{f}_{nh}^2(x) dx - 2 \int \hat{f}_{nh}(x) f(x) dx.$$

On doit donc trouver un estimateur pour $\int \hat{f}_{nh}(x)f(x)dx$. Remarquons que :

$$\int \hat{f}_{nh}(x)f(x)dx = \mathbb{E}(\hat{f}_{nh}(x)),$$

son estimateur empirique est alors : $\frac{1}{n} \sum_{i=1}^n \hat{f}_{n,-i}(x)$, où $\hat{f}_{n,-i}$ est l'estimateur "one-leave-out" de la densité f construit à partir des données indépendantes $(X_j)_{1 \leq j \leq n, j \neq i}$ défini par :

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

Le critère de validation croisée est alors donné par :

$$\begin{aligned} CV(h) &= \int \hat{f}_{nh}^2(x)dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_{n,-i}(X_i) \\ &= \int \hat{f}_{nh}^2(x)dx - \frac{2}{n(n-1)h_n} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{X_i - X_j}{h_n}\right). \end{aligned}$$

- La règle de sélection du paramètre de lissage est donnée par :

$$\hat{h} = \underset{h}{\operatorname{argmin}} CV(h).$$

Proposition 2.3.1. *La statistique :*

$$CV(h) = \int \hat{f}_{nh}^2(x)dx - \frac{2}{n(n-1)h_n} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{X_i - X_j}{h_n}\right)$$

est un estimateur sans biais de $MISE(\hat{f}_{nh}) - \int f^2$.

Démonstration. D'une part, comme la densité jointe du couple (X_i, X_j) est $f(x)f(y)$

2.3 Choix du paramètre de lissage par la méthode de validation croisée 31

(car les variables sont indépendantes), on a

$$\begin{aligned}
\mathbb{E}[CV(h)] &= \mathbb{E}\left[\int \widehat{f}_{nh}^2(x)dx\right] - \frac{2}{n(n-1)h_n} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathbb{E}\left[K\left(\frac{X_i - X_j}{h_n}\right)\right] \\
&= \mathbb{E}\left[\int \widehat{f}_{nh}^2(x)dx\right] - \frac{2}{n(n-1)h_n} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \iint K\left(\frac{x-y}{h_n}\right) f(x)f(y) dx dy \\
&= \mathbb{E}\left[\int \widehat{f}_{nh}^2(x)dx\right] - \frac{2}{n(n-1)h_n} n(n-1) \iint K\left(\frac{x-y}{h_n}\right) f(x)f(y) dx dy \\
&= \mathbb{E}\left[\int \widehat{f}_{nh}^2(x)dx\right] - \frac{2}{h_n} \iint K\left(\frac{x-y}{h_n}\right) f(x)f(y) dx dy.
\end{aligned}$$

D'autre part,

$$\begin{aligned}
MISE(\widehat{f}_{nh}(x)) - \int f^2(x)dx &= \int \mathbb{E}\left[\widehat{f}_{nh}(x) - f(x)\right]^2 dx - \int f^2(x)dx \\
&= \int \mathbb{E}\left[\widehat{f}_{nh}^2(x)\right] dx - 2 \int \mathbb{E}\left[\widehat{f}_{nh}(x)\right] f(x)dx \\
&= \int \mathbb{E}\left[\widehat{f}_{nh}^2(x)\right] dx - 2 \int \mathbb{E}\left[\frac{1}{nh_n} \sum_{j=1}^n K\left(\frac{x-X_j}{h_n}\right)\right] f(x)dx \\
&= \mathbb{E}\left[\int \widehat{f}_{nh}^2(x)dx\right] - \frac{2}{h_n} \iint K\left(\frac{x-y}{h_n}\right) f(y)f(x) dy dx \\
&= \mathbb{E}[CV(h)],
\end{aligned}$$

ce qui équivaut à dire que $CV(h)$ est un estimateur sans biais de $MISE(\widehat{f}_{nh}) - \int f^2$.

□

• Optimalité asymptotique

La fenêtre \widehat{h} est dite asymptotiquement optimale si :

$$\lim_{n \rightarrow \infty} \frac{ISE(\widehat{f}_{n\widehat{h}}(x))}{\inf_h ISE(\widehat{f}_{nh}(x))} = 1 \quad p.s.$$

Théorème 2.3.1. (*Stone [24]*). *Sous les conditions suivantes :*

- ▷ K est le noyau à support compact et Hölder-continue, i.e. Pour tout $(x, y) \in \mathbb{R}^2$, il existe deux constantes $\lambda > 0$ et $D > 0$ tel que $|K(x) - K(y)| \leq D|x - y|^\lambda$.
- ▷ $h_n \in H_N := \{h_1, \dots, h_N\}$. $H_N \subset H_n$ et satisfaisant la condition suivante :

$$\#H_N \leq An^a, \quad n \geq 1, \quad A, a > 0,$$

et avec l'hypothèse que f soit bornée, \widehat{h} est asymptotiquement optimal.

2.3.2 Choix pour des données dépendantes

Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires alpha-mélangeante², pour la sélection de la fenêtre de lissage, considérons ici l'erreur quadratique intégrée pondérée relative à l'estimateur à noyau \widehat{f}_{nh} de la densité f défini par :

$$ISE(\widehat{f}_{nh}(x)) = \int \left[\widehat{f}_{nh}(x) - f(x) \right]^2 w(x) dx$$

où w est une fonction positive à support compact. Le paramètre h_n est pris ici dans un ensemble H_n . Pour minimiser l'erreur quadratique intégrée, Hart & Vieu (1990)[12] ont défini le critère de la validation croisée en introduisant, un estimateur \widehat{f}_{n,l_n} de f , appelé l'estimateur de "sequence-leave-out", basé sur les données $\{X_j\}$, tel que $|j - i| > l_n$ défini par :

$$\widehat{f}_{n,l_n}(x) = \frac{1}{n_{l_n} h_n} \sum_{|j-i|>l_n} K\left(\frac{x - X_j}{h_n}\right),$$

où l_n est une suite d'entiers positifs appelée "sequence-leave-out" tel que :

$$nn_{l_n} = \#\{(i, j) : |j - i| > l_n\}.$$

Le critère de la validation croisée est défini par :

$$CV_{l_n}(h) = \int \widehat{f}_{nh}^2(x) w(x) dx - \frac{2}{n} \sum_{i=1}^n \widehat{f}_{n,l_n}(X_i) w(X_i),$$

et donc le paramètre de lissage sélectionné est donné par :

$$\widehat{h}_{l_n} = \operatorname{argmin}_{h \in H_n} CV_{l_n}(h).$$

2. Soit $\sigma(X_1, \dots, X_n)$ la tribu engendrée par les variables aléatoires X_1, \dots, X_n . Une suite $(X_i)_{i \geq 1}$ est dite alpha-mélangeante 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 :

- ▷ $|P(A \cap B) - P(A)P(B)| \leq \alpha(m)$,
- ▷ $\lim_{m \rightarrow \infty} \alpha(m) = 0$.

- Optimalité asymptotique

Pour établir leur résultat, Hart & Vieu (1990)[12] ont considéré les hypothèses suivantes :

(H1) : Le noyau K est borné, symétrique, à support compact, $\int x^v K(x) < \infty$, $\int x^k K(x) dx =$

$0, k = 1, \dots, v - 1$ et Lipschitz continue, i.e. Hölder continue avec $\lambda = 1$.

(H2) : $H_n := [An^{-a}, Bn^{-b}]$, $0 < b \leq \frac{1}{2v+1} \leq a < \frac{2}{1+4v}$, où A et B sont deux constantes positives.

(H3) : "sequence-leave-out" $\{l_n\}_{\mathbb{N}^*}$ vérifie, $l_n \leq n^{\theta_1}$ où $\theta_1 < 1 - \frac{a(1+4v)}{2}$.

(H4) : Le coefficient de mélange vérifie, $\sup_{j>n^{\theta_1}} \alpha(j) = o(n^{-\theta_2})$, où $\theta_2 := \theta_2(a, b, v, \theta_1)$

(voir Hart & Vieu (1990)[12]).

(H5) : La densité f est bornée, admet v dérivées continues pour $v \in \mathbb{N}^*$ et $\max(f(x), f(-x)) \rightarrow 0$, quand $x \rightarrow \infty$.

Théorème 2.3.2. (*Hart & Vieu(1990)[12]*). *Sous les hypothèses (H1), (H2), (H3), (H4) et (H5), \hat{h}_{l_n} est asymptotiquement optimal.*

**CHAPITRE 2. CHOIX OPTIMAL DU PARAMÈTRE DE LISSAGE
34 DANS L'ESTIMATION DE LA DENSITÉ À TEMPS DISCRET**

Chapitre 3

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

3.1 Modèle

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

$$\widehat{f}_{Th}(x) = \frac{1}{Th_T} \int_0^T K\left(\frac{x - X_t}{h_T}\right) dt,$$

où h_T est le paramètre de lissage appartenant à un ensemble $H_T := [a_T, b_T] \subset \mathbb{R}^+$ et K est un noyau.

Avant d'établir nos résultats, notons que f_{X_s, X_t} la densité conjointe du vecteur aléatoire (X_s, X_t) et f_{X_s} la densité marginale de X_s , définissons la fonction $g_{s,t}$ par :

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

Comme le processus est stationnaire, on a :

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

Par ailleurs, pour tout $t \geq 0$, $\mathcal{F}_t = \sigma((X_s) : 0 \leq s \leq t)$ est la σ -algèbre générée par les données $(X_s)_{0 \leq s \leq t}$. 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 $[0, T]$. La projection \mathcal{P}_k est définie pour une variable aléatoire 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 $\mathcal{F}_k := \mathcal{F}_{T_k}$, T_k étant le $k^{ième}$ élément de la partition de l'intervalle $[0, T]$, i.e. $T_k = \delta k$.

3.2 Hypothèses

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

(b) K est Hölder-continue, i.e. Pour tout $(x, y) \in \mathbb{R}^2$, il existe deux constantes $\lambda > 0$ et $D > 0$ tel que $|K(x) - K(y)| \leq D|x - y|^\lambda$.

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

(H2) (a) f est deux fois dérивables 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 $]0, \infty[$, où

$$\|.\|_\infty := \sup_{x,y \in \mathbb{R}^2} |.|.$$

(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 différentiable de dérivée presque sûrement bornée.

(b) Pour tout $t \in [0, T]$, tout $\delta > 0$ et tout $x \in \mathbb{R}$, 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

$$(a) \frac{T a_T^{2r+1}}{\log T} \rightarrow \infty, \text{ lorsque } T \rightarrow \infty.$$

$$(b) \frac{b_T}{a_T^r} \rightarrow 0, \text{ lorsque } T \rightarrow \infty.$$

Pour donner l'expression de l'erreur quadratique moyenne intégrée (*MISE*) de l'estimateur \hat{f}_{Th} , considérons la décomposition de \mathbb{R}^2 en sous-espaces $\{|y - z| \leq a_T^\alpha\}$ et $\{|y - z| > a_T^\alpha\}$, où $\alpha > 1$.

Lemme 3.2.1. *Sous les hypothèses (H1), (H2) et (H3)(c,d), l'erreur quadratique moyenne intégrée (*MISE*) de l'estimateur \hat{f}_{Th} est donné par :*

$$\begin{aligned} MISE(\hat{f}_{Th}(x)) &= 2 \int K^2 \left\{ \int \frac{1}{T^2 h_T} \iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T [T-u] g_u(y, z) du \right\} dy dz \right\} \\ &\quad + h_T^2 \left\{ \int z^2 \left(\frac{1}{2} K * K - K \right) (z) dz \int f'' f \right\} \\ &\quad + o \left(\frac{1}{T^2 h_T} \iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T [T-u] g_u(y, z) du \right\} dy dz + h_T^2 \right). \end{aligned} \tag{3.1}$$

3.3 Choix du paramètre de lissage

Le paramètre de lissage optimal pour le critère *MISE* donné dans (3.1) est de la forme :

$$h_{opt} = \left\{ \frac{\int K^2}{\int z^2 \left(\frac{1}{2} K * K - K \right) (z) dz \int f'' f} \right\}^{\frac{1}{3}} \left\{ \frac{\iint_{|y-z| \leq a_T^\alpha} \left\{ \int_0^T [T-u] g_u(y, z) du \right\} dy dz}{T^2} \right\}^{\frac{1}{3}}. \tag{3.2}$$

Ce paramètre dépend des quantités inconnues f et g_u . Il est donc impossible de le calculer en pratique. Pour un résultat réalisable, considérant l'erreur quadratique

intégrée (*ISE*), définie par :

$$\begin{aligned} ISE(\hat{f}_{Th}(x)) &= \int \left[\hat{f}_{Th}(x) - f(x) \right]^2 dx \\ &= \int \hat{f}_{Th}^2(x) dx - 2 \int \hat{f}_{Th}(x) f(x) dx + \int f^2(x) dx. \end{aligned}$$

Minimisant $ISE(\hat{f}_{Th})$ par rapport à h revient à minimiser :

$$ISE(\hat{f}_{Th}(x)) - \int f^2(x) dx = \int \hat{f}_{Th}^2(x) dx - 2 \int \hat{f}_{Th}(x) f(x) dx.$$

À l'aide de l'estimateur de "one-leave-out", le critère de validation croisée approprié au cas de l'estimation de la densité pour des processus à temps continu construit à partir des données $(X_t)_{t \in \cup_{\substack{j=1 \\ j \neq i}}^{n-1} [T_{j-1}, T_j]}$ est donné par :

$$CV(h) = \int \hat{f}_{Th}^2(x) dx - \frac{2}{T^2 h_T} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int_{T_{i-1}}^{T_i} \int_{T_{j-1}}^{T_j} K \left(\frac{X_s - X_t}{h_T} \right) ds dt,$$

où, $T_0 = 0$ et $T_j = j\delta$ avec $\delta = \frac{T}{n}$ et $1 \leq j \leq n$.

• Le paramètre de lissage sélectionné par cette méthode est donné par :

$$\hat{h} = \underset{h \in H_T}{\operatorname{argmin}} CV(h).$$

3.4 Résultats asymptotiques

Proposition 3.4.1. *Supposons que les hypothèses (H1)(a) et (H3)(a,b) soient satisfaites, alors la statistique :*

$$CV(h) = \int \hat{f}_{Th}^2(x) dx - \frac{2}{T^2 h_T} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \int_{T_{i-1}}^{T_i} \int_{T_{j-1}}^{T_j} K \left(\frac{X_s - X_t}{h_T} \right) ds dt$$

est un estimateur asymptotiquement sans biais de $MISE(\hat{f}_{Th}) - \int f^2$.

Démonstration. Dans ce qui suit posons $\Gamma = \cup_{i=1}^n [T_{i-1}, T_i]$, Γ^c est le complémentaire de Γ dans $[0, T]^2$. D'une part, sous l'hypothèse (H3)(a), on a :

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

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

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

En effectuant les changement de variable $u = \frac{x}{h_T}$ et $v = \frac{y}{h_T}$, et sous l'hypothèse (H3)(b) on a :

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

Donc

$$\mathbb{E}[CV(h)] = \mathbb{E}\left[\int \widehat{f}_{Th}^2(x)dx\right] - \frac{2}{h_T} \iint K\left(\frac{x-y}{h_T}\right) f(x)f(y) dx dy - o(1).$$

D'autre part,

$$\begin{aligned}
 MISE(\widehat{f}_{Th}(x)) - \int f^2(x)dx &= \int \mathbb{E} [\widehat{f}_{Th}(x) - f(x)]^2 dx - \int f^2(x)dx \\
 &= \int \mathbb{E} [\widehat{f}_{Th}^2(x)] dx - 2 \int \mathbb{E} [\widehat{f}_{Th}(x)] f(x)dx \\
 &= \int \mathbb{E} [\widehat{f}_{Th}^2(x)] dx - 2 \int \mathbb{E} \left[\frac{1}{Th_T} \int_0^T K\left(\frac{x-X_t}{h_T}\right) dt \right] f(x)dx \\
 &= \mathbb{E} \left[\int \widehat{f}_{Th}^2(x)dx \right] - \frac{2}{h_T} \iint K\left(\frac{x-y}{h_T}\right) f(y)f(x) dydx.
 \end{aligned}$$

Ainsi,

$$\lim_{T \rightarrow \infty} \mathbb{E}[CV(h)] = MISE(\widehat{f}_{Th}(x)) - \int f^2(x)dx.$$

□

Théorème 3.4.1. *Sous les hypothèses (H1)(a,b), (H2)(a), (H3)(a,b), (H4), (H5) et (H6), le paramètre de lissage \widehat{h} est asymptotiquement optimal.*

3.4.1 Convergence des paramètres de lissage

Notons que h_0 , h_{opt} et \widehat{h} les paramètres de lissage minimisant dans l'ensemble $H_T := [a_T, b_T]$, $ISE(\widehat{f}_{Th})$, $MISE(\widehat{f}_{Th})$ et $CV(h)$ respectivement. On se présente dans ce qui suit l'ordre de grandeur de h_{opt} et les vitesses de convergence presque sûres de h_0 vers h_{opt} et \widehat{h} vers h_{opt} sous les hypothèses (H1)(a,c), (H2), (H3)(a,b,c), (H4)(a,b), (H5) et les hypothèses suivantes :

(H1') K est Lipschitzienne, i.e. Pour tout $(x, y) \in \mathbb{R}^2$, il existe une constante $d > 0$, tel que $|K(x) - K(y)| \leq d|x - y|$.

(H2') Il existe une fonction a_T et un réel $1 < \alpha < 2$ tels que, $\iint_{|y-z| < a_T^\alpha} g_u(y, z) dy dz > 0$.

(H3') 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 $T \rightarrow \infty$.

(H4') (a) Pour tout réel $0 < l < 1$, $\frac{T a_T^l}{\log T} \rightarrow \infty$, lorsque $T \rightarrow \infty$.

- (b) Il existe un réel $0 < m < 1$, tel que $\frac{|h'_T|T^{\frac{1}{2}}}{h_T^{\frac{4-m}{2}}(\log T)^{\frac{1}{2}}} < \infty$, où h'_T est la dérivée de h_T .

Après avoir donné la forme de h_{opt} dans l'expression (3.2), la proposition suivante donne son ordre de grandeur.

Proposition 3.4.2. *Supposons que les hypothèses (H1)(a), (H1'), (H2), (H3)(a,c) et (H2'), soient satisfaites, alors :*

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

Démonstration. D'après l'expression (3.2), on a :

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

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

Par un développement de Taylor de $g_u(., z)$, nous obtenons :

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

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

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

Observons que :

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

Comme $u \rightarrow \int \sup_y \left| \frac{\partial g_u(y, z)}{\partial y} \right| dz$ est intégrable sur $]0, \infty[$, donc

$$\int_0^T \frac{[T-u]}{T^2} \int \left\{ \int_{-a_T^\alpha+z}^{a_T^\alpha+z} (y-z) \frac{\partial g_u(y, z)}{\partial y} \Big|_{y=y_T^*(z)} dy \right\} dz du = O\left(\frac{a_T^{2\alpha}}{T}\right). \quad (3.3)$$

Par ailleurs, nous avons :

$$\begin{aligned}
 & \left| \int_0^\infty \int g_u(z, z) dz du - \int_0^T \frac{[T-u]}{T} \int g_u(z, z) dz du \right| \\
 & = \left| \int_T^\infty \int g_u(z, z) dz du + \int_0^T \frac{u}{T} \int g_u(z, z) dz du \right| \\
 & \leq \int_T^\infty \left| \int g_u(z, z) dz \right| du + \int_0^T \frac{u}{T} \left| \int g_u(z, z) dz \right| du. \quad (3.4)
 \end{aligned}$$

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 (3.4) tends vers zéro. Il s'en suit directement des assertions (3.3) et (3.4), que :

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

□

Nous avons besoin des deux lemmes suivants qui présentent des résultats des vitesses de convergence ponctuelle et uniforme de \hat{f}_{Th} pour démontrer nos résultats.

Lemme 3.4.1. *Supposons que les hypothèses (H1)(a), (H2)(a), (H4)(a,b), (H3'), (H5) et (H4')(a) soient satisfaites, il existe $m < 1$, tel que, pour tout $x \in \mathbb{R}$, on a :*

$$\hat{f}_{Th}(x) - f(x) = O(h_T^2) + O_{p.s} \left(\sqrt{\frac{\log T}{Th_T^{2-\frac{1}{m}}}} \right), \quad \text{lorsque } T \rightarrow \infty.$$

Lemme 3.4.2. *Supposons que les hypothèses (H1)(a), (H2)(a), (H4)(a,b), (H3'), (H5) et (H4')(b) soient satisfaites, il existe $m < 1$, alors :*

$$\sup_{x \in \mathbb{R}} |\hat{f}_{Th}(x) - f(x)| = O(h_T^2) + O_{p.s} \left(\sqrt{\frac{\log T}{Th_T^{2-\frac{1}{m}}}} \right), \quad \text{lorsque } T \rightarrow \infty.$$

Démonstration du lemme 3.4.1

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

$$\begin{aligned} \hat{f}_{Th}(x) - f(x) &= \underbrace{\frac{1}{Th_T} \int_0^T K \left(\frac{x - X_t}{h_T} \right) - \mathbb{E} \left[K \left(\frac{x - X_t}{h_T} \right) | \mathcal{F}_{t-\delta} \right] dt}_{S_1} \\ &\quad + \underbrace{\frac{1}{Th_T} \int_0^T \mathbb{E} \left[K \left(\frac{x - X_t}{h_T} \right) | \mathcal{F}_{t-\delta} \right] dt - f(x)}. \end{aligned}$$

Pour le terme S_1 en appliquant le lemme suivant sur les inégalités exponentielles :

Lemme 3.4.3. (De la Peña, V.H. and Giné, E. [6]). 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 :

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

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

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

où $D_n = \sum_{i=1}^n d_i^2$.

Soit Y_i la différence de martingales donnée par :

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

notons que pour tout $\delta > 0$, Y_i est une différence de martingales par rapport aux σ -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, on a :

$$\begin{aligned} |\mathbb{E}[Y_i^l | \mathcal{G}_{i-2}]| &\leq \mathbb{E}\left[\left|\int_{T_{i-1}}^{T_i} K\left(\frac{x - X_t}{h_T}\right) - \mathbb{E}\left[K\left(\frac{x - X_t}{h_T}\right) | \mathcal{F}_{t-\delta}\right] dt\right|^l | \mathcal{G}_{i-2}\right] \\ &\leq \int_{T_{i-1}}^{T_i} \mathbb{E}\left[\left|K\left(\frac{x - X_t}{h_T}\right) - \mathbb{E}\left[K\left(\frac{x - X_t}{h_T}\right) | \mathcal{F}_{t-\delta}\right]\right|^l | \mathcal{G}_{i-2}\right] dt \\ &\leq \int_{T_{i-1}}^{T_i} \left(\mathbb{E}\left[K^l\left(\frac{x - X_t}{h_T}\right) | \mathcal{G}_{i-2}\right]^{\frac{1}{l}} + \mathbb{E}\left[\mathbb{E}\left[K\left(\frac{x - X_t}{h_T}\right) | \mathcal{F}_{t-\delta}\right]^l | \mathcal{G}_{i-2}\right]^{\frac{1}{l}} \right)^l dt \\ &\leq \int_{T_{i-1}}^{T_i} \left(\mathbb{E}\left[K^l\left(\frac{x - X_t}{h_T}\right) | \mathcal{G}_{i-2}\right]^{\frac{1}{l}} + \mathbb{E}\left[\mathbb{E}\left[K^l\left(\frac{x - X_t}{h_T}\right) | \mathcal{F}_{t-\delta}\right] | \mathcal{G}_{i-2}\right]^{\frac{1}{l}} \right)^l dt \\ &\leq \int_{T_{i-1}}^{T_i} \left(2\mathbb{E}\left[K^l\left(\frac{x - X_t}{h_T}\right) | \mathcal{G}_{i-2}\right]^{\frac{1}{l}} \right)^l dt \\ &\leq 2^l \int_{T_{i-1}}^{T_i} \mathbb{E}\left[K^l\left(\frac{x - X_t}{h_T}\right) | \mathcal{G}_{i-2}\right] dt. \end{aligned}$$

En utilisant l'inégalité de Hölder avec $\frac{1}{m} + \frac{1}{m'} = 1$ et un changement de variable

$z = \frac{x - y}{h_T}$, nous obtenons,

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

Donc,

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

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

$$|\mathbb{E} [Y_i^l | \mathcal{G}_{i-2}]| \leq l! C^{l-2} d_i^2,$$

et

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

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

$$D_n = 2^3 \|K\|_{\infty} \|K\|_{L^m} T h_T^{\frac{1}{m}} \frac{1}{T} \int_0^T \|b_t\|_{L^{m'}} dt.$$

Sous l'hypothèse (H3'), on a

$$D_n = O(T h_T^{\frac{1}{m}}).$$

En utilisant le lemme 3.4.3, pour tout $\epsilon > 0$, on a :

$$\begin{aligned} \mathbb{P} \left(|S_1| > \epsilon \left(\frac{\log T}{T h_T^{2-\frac{1}{m}}} \right)^{\frac{1}{2}} \right) &= \mathbb{P} \left(\left| \sum_{i=1}^n Y_i \right| > \epsilon T h_T \left(\frac{\log T}{T h_T^{2-\frac{1}{m}}} \right)^{\frac{1}{2}} \right) \\ &= \mathbb{P} \left(\left| \sum_{i=1}^n Y_i \right| > \epsilon T h_T^{\frac{1}{m}} \left(\frac{\log T}{T h_T^{\frac{1}{m}}} \right)^{\frac{1}{2}} \right) \\ &\leq 2 \exp \left\{ - \frac{\epsilon^2 T^2 h_T^{\frac{2}{m}} \left(\frac{\log T}{T h_T^{\frac{1}{m}}} \right)}{O(T h_T^{\frac{1}{m}}) + 2\epsilon C T h_T^{\frac{1}{m}} \left(\frac{\log T}{T h_T^{\frac{1}{m}}} \right)^{\frac{1}{2}}} \right\} \\ &\leq 2 \exp \{-\epsilon^2 O(\log T)\} \\ &\leq 2 T^{-C' \epsilon^2}. \end{aligned} \tag{3.5}$$

Pour un choix approprié de ϵ et l'usage du lemme de Borel-Cantelli, on a :

$$S_1 = O_{p.s} \left(\sqrt{\frac{\log T}{Th_T^{2-\frac{1}{m}}}} \right). \quad (3.6)$$

Sous l'hypothèse (H5) et la symétrie de K , voir Didi & Louani (2013)[8], on a :

$$S_2 = O_{p.s} \left(T^{-\frac{1}{2}} \right) + O(h_T^2). \quad (3.7)$$

Le lemme s'achève en combinant les résultats (3.6) et (3.7). \square

Démonstration du lemme 3.4.2

Considérons l'ensemble B_T défini par :

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

Soit γ_T une fonction entière non décroissante tendant vers l'infini lorsque $T \rightarrow \infty$, considérons $\{B_{T,i}\}_{1 \leq i \leq \gamma_T}$ une partition de l'ensemble B_T définie par :

$$B_{T,i} = \{x : |x - x_i| \leq T^r \gamma_T^{-1}\}.$$

Où $(x_i)_{1 \leq i \leq \gamma_T}$ est une suite d'éléments de B_T . En faisant appel à un calcul de Didi & Louani [8], nous obtenons

$$\begin{aligned} \sup_{x \in B_T} |\hat{f}_{Th}(x) - f(x)| &\leq \sup_{x \in B_T} \left| \frac{1}{Th_T} \int_0^T K \left(\frac{x - X_t}{h_T} \right) - \mathbb{E} \left[K \left(\frac{x - X_t}{h_T} \right) \mid \mathcal{F}_{t-\delta} \right] dt \right| \\ &\quad + \sup_{x \in B_T} \left| \frac{1}{Th_T} \int_0^T \mathbb{E} \left[K \left(\frac{x - X_t}{h_T} \right) \mid \mathcal{F}_{t-\delta} \right] dt - f(x) \right| \\ &\leq A_1 + A_2. \end{aligned}$$

Où

$$A_1 = \sup_{x \in B_T} |A_1^T(x)| = \sup_{x \in B_T} \left| \frac{1}{Th_T} \int_0^T K \left(\frac{x - X_t}{h_T} \right) - \mathbb{E} \left[K \left(\frac{x - X_t}{h_T} \right) \mid \mathcal{F}_{t-\delta} \right] dt \right|,$$

et

$$A_2 = \sup_{x \in B_T} |A_2^T(x)| = \sup_{x \in B_T} \left| \frac{1}{Th_T} \int_0^T \mathbb{E} \left[K \left(\frac{x - X_t}{h_T} \right) \mid \mathcal{F}_{t-\delta} \right] dt - f(x) \right|.$$

Nous avons :

$$A_1 \leq \max_{1 \leq i \leq \gamma_T} \sup_{x \in B_{T,i}} |A_1^T(x) - A_1^T(x_i)| + \max_{1 \leq i \leq \gamma_T} |A_1^T(x_i)| = I_1 + I_2.$$

En utilisant l'hypothèse (H1'), il existe une constante positive d telle que :

$$\begin{aligned} \left| K\left(\frac{x - X_t}{h_T}\right) - K\left(\frac{x_i - X_t}{h_T}\right) \right| &\leq d \left| \frac{x - X_t}{h_T} - \frac{x_i - X_t}{h_T} \right| \\ &\leq \frac{d}{h_T} |x - x_i| \\ &\leq \frac{dT^r}{h_T \gamma_T}. \end{aligned}$$

Donc, nous avons :

$$\begin{aligned} I_1 &\leq \frac{1}{Th_T} \sum_{j=1}^n \int_{T_{j-1}}^{T_j} \left\{ \max_{1 \leq i \leq \gamma_T} \sup_{x \in B_T \cap B_{T,i}} \left| K\left(\frac{x - X_t}{h_T}\right) - K\left(\frac{x_i - X_t}{h_T}\right) \right| \right. \\ &\quad \left. + \mathbb{E} \left[\max_{1 \leq i \leq \gamma_T} \sup_{x \in B_T \cap B_{T,i}} \left| K\left(\frac{x - X_t}{h_T}\right) - K\left(\frac{x_i - X_t}{h_T}\right) \right| \middle| \mathcal{G}_{j-1} \right] \right\} dt \\ &\leq \frac{2\delta n}{Th_T} \times \frac{dT^r}{h_T \gamma_T} = \frac{2dT^r}{h_T^2 \gamma_T}. \end{aligned}$$

Prenant $\gamma_T = T^{r+2}$, il en découle que :

$$I_1 = o_{p.s.} \left(\sqrt{\frac{\log T}{Th_T^{2-\frac{1}{m}}}} \right). \quad (3.8)$$

Pour I_2 , nous observons que :

$$\max_{1 \leq i \leq \gamma_T} |A_1^T(x_i)| = \max_{1 \leq i \leq \gamma_T} \left| \sum_{j=1}^n Y_j(x_i) \right|.$$

Donc par ailleurs, comme dans (3.5) avec $\gamma_T = T^{r+2}$, nous avons :

$$\mathbb{P} \left(\max_{1 \leq i \leq \gamma_T} \frac{1}{Th_T} \left| \sum_{j=1}^n Y_j(x_i) \right| > \epsilon \sqrt{\frac{\log T}{Th_T^{2-\frac{1}{m}}}} \right) \leq 2T^{r+2-\epsilon^2 C},$$

où C est une constante positive. En choisissant un ϵ approprié et en utilisant le lemme de Borel-Cantelli, on obtient :

$$I_2 = O_{p.s} \left(\sqrt{\frac{\log T}{Th_T^{2-\frac{1}{m}}}} \right). \quad (3.9)$$

Il résulte alors les assertions (3.8) et (3.9), que :

$$A_1 = O_{p.s} \left(\sqrt{\frac{\log T}{Th_T^{2-\frac{1}{m}}}} \right). \quad (3.10)$$

D'autre part, par un calcul similaire à Didi & Louani [8] et la symétrie de K , nous avons :

$$A_2 = O_{p.s} \left(T^{-\frac{1}{2}} \right) + O(h_T^2). \quad (3.11)$$

Par conséquent, d'après (3.10) et (3.11) nous obtenons :

$$\sup_{x \in B_T} |\hat{f}_{Th}(x) - f(x)| = O(h_T^2) + O_{p.s} \left(\sqrt{\frac{\log T}{Th_T^{2-\frac{1}{m}}}} \right).$$

Il reste maintenant à évaluer le terme $\sup_{x \in B_T^c} |\hat{f}_{Th}(x) - f(x)|$, où B_T^c est le complémentaire de B_T dans \mathbb{R} . Comme pour l'équation (3.11), il est clair que :

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

Il reste à prouver que :

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

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

où $\epsilon_T = \sqrt{\frac{\log T}{Th_T^{2-\frac{1}{m}}}}$, est uniformément continue, i.e., il existe une constante positive θ ,

tel que :

$$\sup_{|x|>T^{2r}} \left| \frac{\epsilon_T}{Th_T} \sum_{i=1}^n Y_i - \frac{\epsilon_S}{Sh_T} \sum_{i=1}^{n'} Y'_i \right| \leq \theta |T - S|,$$

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

Observons que :

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

où $\bar{f}_{Th} := \frac{1}{Th_T} \int_0^T \mathbb{E}\left[K\left(\frac{x - X_t}{h_T}\right) | \mathcal{F}_{t-\delta}\right] dt$. Par un calcul similaire à Didi & Louani et sous les conditions (H1)(a), (H1') et (H4')(b), il en résulte que :

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

Voir Didi & Louani [8] pour plus de détails.

D'autre part, en utilisant le lemme 3.4.3, nous avons que :

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

Ce qui achève la démonstration du lemme. \square

Théorème 3.4.2. *Sous les hypothèses (H1)(a,c), (H1'), (H2), (H3)(a,b,c), (H2'),*

(H4)(a,b), (H3'), (H5) et (H4'), il existe $q < 1$ et $m < 1$, tels que :

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

Preuve

Soit $L(Z) = -ZK'(Z)$. Notons que L est un noyau qui satisfait les mêmes hypothèses que K . Soit \tilde{f}_{Th} , \hat{g}_{Th} et \tilde{g}_{Th} les estimateurs à noyaux de f définis par :

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

En posant :

$$ISE(\hat{f}_{Th}(x)) = MISE(\hat{f}_{Th}(x)) + D_h(x),$$

et en dérivant par rapport à h_T ,

$$\begin{aligned}ISE'(\hat{f}_{Th}(x)) &= MISE'(\hat{f}_{Th}(x)) + D'_h(x) \\ ISE'(\hat{f}_{Th_0}(x)) &= \left\{MISE'(\hat{f}_{Th_0}(x)) - MISE'(\hat{f}_{Th_{opt}}(x))\right\} + D'_{h_0}(x) \\ ISE'(\hat{f}_{Th_0}(x)) &= (h_0 - h_{opt})MISE''(\hat{f}_{Th^*}(x)) + D'_{h_0}(x),\end{aligned}$$

où h^* est compris entre h_0 et h_{opt} . Comme h_0 est le paramètre de lissage qui minimise le critère $ISE(\hat{f}_{Th})$, donc :

$$ISE'(\hat{f}_{Th_0}(x)) = 0,$$

par conséquent,

$$h_{opt} - h_0 = \frac{D'_{h_0}(x)}{MISE''(\hat{f}_{Th^*}(x))}.$$

D'une part, nous avons, sous les hypothèses (H1)(a,b), (H1'), (H2), (H3)(a,c) et (H2'), que :

$$\begin{aligned}
 MISE''(\hat{f}_{Th}(x)) &= \left\{ 8 \int K^2 \left\{ \int \int_0^\infty g_u(z, z) dz du \right\} \right\} \frac{a_T^\alpha}{Th_T^3} + 2 \int z^2 \left(\frac{1}{2} K * K - K \right)(z) dz \\
 &\quad \int f'' f + o \left(\left\{ 8 \int K^2 \left\{ \int \int_0^\infty g_u(z, z) dz du \right\} \right\} \frac{a_T^\alpha}{Th_T^3} \right. \\
 &\quad \left. + 2 \int z^2 \left(\frac{1}{2} K * K - K \right)(z) dz \int f'' f \right).
 \end{aligned} \tag{3.12}$$

D'autre part,

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

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

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

En dérivant D_h par rapport à h_T , il en découle,

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

où $L * L(Z) = -Z K' * K'(Z)$.

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

Par le lemme 3.4.1, un changement de variables $Z = \frac{y - z}{h_T}$ et un développement de Taylor, donnent

$$\begin{aligned} \frac{2}{h_T} \int \left[\hat{f}_{Th}(x) - \mathbb{E}[\hat{f}_{Th}(x)] \right] f(x) dx &= \frac{2}{h_T} \int \left[\hat{f}_{Th}(x) - f(x) + f(x) - \mathbb{E}[\hat{f}_{Th}(x)] \right] f(x) dx \\ &= O(h_T) + O_{p.s} \left(\sqrt{\frac{\log T}{Th_T^{4-\frac{1}{m}}}} \right). \end{aligned}$$

De la même manière, nous avons

$$\frac{2}{h_T} \int [\hat{g}_{Th}(x) - \mathbb{E}[\hat{g}_{Th}(x)]] f(x) dx = O(h_T) + O_{p.s} \left(\sqrt{\frac{\log T}{Th_T^{4-\frac{1}{m}}}} \right). \quad (3.13)$$

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

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

De la même manière,

$$\frac{1}{T^2 h_T^2} \int_{\Gamma^c} \iint L * L \left(\frac{y-z}{h_T} \right) f(y) f(z) dy dz dt ds = \frac{T^2 - n\delta^2}{T^2 h_T} \int f^2(y) dy + O(h_T). \quad (3.14)$$

Sous les hypothèses (H3)(a,b) et par un calcul de EL HEDA & Louani [10], il s'en suit que :

$$\frac{1}{T^2 h_T^2} \int_{\Gamma^c} \iint K * K \left(\frac{y-z}{h_T} \right) g_{t,s}(y, z) dy dz dt ds = O(h_T^{2/q-2}). \quad (3.15)$$

De la même manière, nous avons :

$$\frac{1}{T^2 h_T^2} \int_{\Gamma^c} \iint L * L \left(\frac{y-z}{h_T} \right) g_{t,s}(y, z) dy dz dt ds = O(h_T^{2/q-2}). \quad (3.16)$$

En utilisant les équations (3.13), (3.14), (3.15) et (3.16), nous obtenons :

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

En utilisant le lemme 3.4.2, on a :

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

Par conséquent,

$$D'_h(x) = O(h_T)^{\min\{\frac{2}{q}-2, 1\}} + O_{p.s} \left(\sqrt{\frac{\log T}{Th_T^{4-\frac{1}{m}}}} \right).$$

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

$$D'_{h_0}(x) = O \left(\frac{a_T^\alpha}{T} \right)^{\min\{\frac{2}{3q}-\frac{2}{3}, \frac{1}{3}\}} + O_{p.s} \left(\sqrt{\frac{\log T}{\left(T^{\frac{1}{m}-1} a_T^{\alpha(4-\frac{1}{m})}\right)^{\frac{1}{3}}}} \right), \quad (3.17)$$

et la dérivée seconde de $MISE(\hat{f}_{Th})$ par rapport à h_T prise en h^* est donnée par :

$$\begin{aligned} MISE''(\hat{f}_{Th^*}(x)) &= \left\{ 8 \int K^2 \left\{ \int \int_0^\infty g_u(z, z) dz du \right\} \right\} + 2 \int z^2 \left(\frac{1}{2} K * K - K \right)(z) dz \int f'' f \\ &\quad + o \left(\left\{ 8 \int K^2 \left\{ \int \int_0^\infty g_u(z, z) dz du \right\} \right\} + 2 \int z^2 \left(\frac{1}{2} K * K - K \right)(z) dz \right. \\ &\quad \left. \int f'' f \right). \end{aligned} \quad (3.18)$$

En combinant les assertions (3.12), (3.17) et (3.18) il s'en suit que :

$$h_{opt} - h_0 = O \left(\frac{a_T^\alpha}{T} \right)^{\min\{\frac{2}{3q}-\frac{2}{3}, \frac{1}{3}\}} + O_{p.s} \left(\sqrt{\frac{\log T}{\left(T^{\frac{1}{m}-1} a_T^{\alpha(4-\frac{1}{m})}\right)^{\frac{1}{3}}}} \right). \quad (3.19)$$

■

Théorème 3.4.3. *Sous les hypothèses (H1)(a,c), (H1'), (H2), (H3)(a,b), (H4)(a,b), (H3'), (H5) et (H4')(a), il existe $q < 1$ et $m < 1$, tels que :*

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

Preuve

Soit λ_h défini par :

$$\lambda_h(x) = CV(h) - ISE(\hat{f}_{Th}(x)) + \int f^2(x)dx.$$

En dérivant λ_h par rapport à h_T , nous obtenons :

$$\begin{aligned} CV'(h) &= \lambda'_h(x) + ISE'(\hat{f}_{Th}(x)) \\ &= \lambda'_h(x) + MISE'(\hat{f}_{Th}(x)) + D'_h(x) \\ CV'(\hat{h}) &= \lambda'_{\hat{h}}(x) + \left\{ MISE'(\hat{f}_{T\hat{h}}(x)) - MISE'(\hat{f}_{Th_{opt}}(x)) \right\} + D'_{\hat{h}}(x) \\ &= \lambda'_{\hat{h}}(x) + (\hat{h} - h_{opt}) MISE''(\hat{f}_{Th^*}(x)) + D'_{\hat{h}}(x), \end{aligned}$$

où h^* est compris entre \hat{h} et h_{opt} . Comme \hat{h} est le paramètre de lissage qui minimise le critère de validation croisée, donc :

$$CV'(\hat{h}) = 0,$$

il s'en suit que :

$$h_{opt} - \hat{h} = \frac{\lambda'_{\hat{h}}(x) + D'_{\hat{h}}(x)}{MISE''(\hat{f}_{Th^*}(x))}.$$

Par ailleurs, le terme λ_h est donné par :

$$\begin{aligned} \lambda_h(x) &= CV(h) - ISE(\hat{f}_{Th}(x)) + \int f(x)^2 dx \\ &= -\frac{2}{T^2 h_T} \int_{\Gamma^c} K\left(\frac{X_t - X_s}{h_T}\right) dsdt + \frac{2}{Th_T} \int \int_0^T K\left(\frac{x - X_t}{h_T}\right) f(x) dt dx. \end{aligned}$$

En dérivant λ_h par rapport à h_T .

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

Observons que pour h_T très petit, nous avons :

$$\frac{1}{h_T^2} K\left(\frac{X_t - X_s}{h_T}\right) = \frac{1}{h_T^2} L\left(\frac{X_t - X_s}{h_T}\right) = 0,$$

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,

$$\begin{aligned}\lambda'_h(x) &= -\frac{2}{Th_T^2} \int \int_0^T K\left(\frac{x-X_t}{h_T}\right) f(x) dt dx + \frac{2}{Th_T^2} \int \int_0^T L\left(\frac{x-X_t}{h_T}\right) f(x) dt dx \\ &= \frac{2}{h_T} \int [f(x) - \hat{f}_{Th}(x)] f(x) dx + \frac{2}{h_T} \int [\hat{g}_{Th}(x) - f(x)] f(x) dx.\end{aligned}$$

En utilisant le lemme 3.4.1,

$$\lambda'_h(x) = O(h_T) + O_{p.s} \left(\sqrt{\frac{\log T}{Th_T^{4-\frac{1}{m}}}} \right).$$

Ainsi,

$$\lambda'_h(x) = O\left(\frac{a_T^\alpha}{T}\right)^{\frac{1}{3}} + O_{p.s} \left(\sqrt{\frac{\log T}{\left(T^{\frac{1}{m}-1} a_T^{\alpha(4-\frac{1}{m})}\right)^{\frac{1}{3}}}} \right). \quad (3.20)$$

Le résultat s'obtient en combinant les résultats (3.20), (3.17) et (3.19). ■

Conclusion

L'objectif de ce mémoire porte sur le problème de choix du paramètre de lissage dans l'estimation de la densité.

Dans la première partie nous avons présenté les différentes méthodes de sélection du paramètre de lissage, les méthodes basée sur l'estimation de la quantité inconnue $R(f'')$ donnée dans l'expression du paramètre de lissage théorique optimal et les méthodes reposant sur la validation croisée qui sont des méthodes direct, consistent à optimiser le critère d'erreur. Ensuite, nous avons présenté le cadre stationnaire ergodique en temps discret et continu.

Le choix de la fenêtre de lissage dans l'estimation de la densité dans le cas discret est présenté dans la deuxième partie, nous avons commencé par la méthode d'histogramme, nous avons donné sa forme améliorée, son risque quadratique moyen et nous avons fondé le choix approprié du la fenêtre h . Par la suite, nous nous sommes intéressés à la méthode du noyau, en examinant les propriétés de l'estimateur plus précisément le biais, la variance et l'erreur quadratique moyenne. Nous avons établi des résultats de convergences en moyenne quadratique et presque complète. Tous ces résultats mettent en évidence le rôle du paramètre de lissage h , en regardant par exemple l'expression de l'erreur quadratique moyenne, le terme de biais est proportionnelle à h tandis que le terme de variance est inversement proportionnelle à h , ce qui nous indique que le but est de minimiser le critère d'erreur. Nous avons cherché à sélectionner ce paramètre dans deux cas (indépendant et alpha-mélange) utilisons la technique de la validation croisée. Ce choix issu de la méthode de la validation croisée est asymptotiquement optimal.

La dernière partie est consacrée sur le choix du paramètre de lissage dans l'estimation de la densité pour un processus stationnaire ergodique à temps continu. Le paramètre de lissage optimal pour le critère $MISE$ présente l'inconvénient de dépendance

des quantités inconnues et son utilisation en pratique qui pose alors problème. De ce fait, nous avons donné un critère de validation croisée adaptée inspirée de l'estimateur de "one-leave-out". Enfin nous avons établis quelques propriétés asymptotiques et des vitesses de convergence presque sur pour le paramètre de lissage.

Bibliographie

- [1] Banon, G. (1978). Nonparametric identification for diffusion processes. SIAM J. Control Optim. 16(3), p. 380-395.
- [2] Banon, G. Nguyen, H.T. (1978). Sur l'estimation récurrente de la densité et de sa dérivée pour un processus de Markov. C. R. Acad. Sci. Paris Sér. A Math. 286(16), p. 691-4.
- [3] Banon, G. Nguyen, H.T. (1981). Recursive estimation in diffusion model.SIAM J. Control Optim. 19(5), p. 676-85.
- [4] Birkhoff, G. D. (1931). Proof of the ergodic theorem. Proc. Nat. Aad. Sci. USA, 17, p. 656-660.
- [5] Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. Biometrika, 71, p. 353-360.
- [6] De la Peña, V.H. and Giné, E. (1999). Decoupling, from dependence to independence. Probability and its applications. Springer-Verlag, New York.
- [7] Delecroix, M. (1979). Sur l'estimation des densités d'un processus stationnaire et mélangeant. Publ. U.E.R. Math. Pures Appl. IRMA, 1(4), exp. no. I, 24. Seminar on Mathematical Statistics.
- [8] Didi, S. Louani, D. (2013). Consistency results for the kernel density estimate on continuous time stationary and dependent data. Journal of Statistics & Probability Letters, 83, 4, p. 1262-1270.
- [9] 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.
- [10] 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.

- [11] Habbema, J. D. F. Hermans, J. and Van der Broek, K. (1974). A stepwise discrimination program using density estimation. Physica Verlag, p. 100-110.
- [12] 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.
- [13] Krengel, U. (1985). Ergodic Theorems, Walter degruyter Co.Berlin.
- [14] Nadaraya, E. (1965). On non parametric estimation density function and regression. *Theory Probab P.P.L.*
- [15] Nguyen, H.T. (1979). Density estimation in a continuous-time stationary Markov process.*Ann. Statist.* 7(2), p. 341-348.
- [16] Nguyen, H.T. Pham, T.D. (1980). Sur l'utilisation du temps local en statistique des processus. *C. R. Acad. Sci. Paris SeÃr. A Math.* 290(3), p. 165-168.
- [17] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33, 1065-1076.
- [18] Rosenblatt, M. (Sept.1956). Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics.* 27, p. 832-837.
- [19] Rudemo, M. (1982). Empirical Choice of Histograms and Kernel Density Estimators. *Scandinavian Journal of Statistics*, 9, p. 65-78.
- [20] 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.
- [21] Scott, D. W., and Terrell, G. R. (1987). Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association* 82, 400, 1131-1146.
- [22] Silverman, B. W. (1986). Density estimation for statistics and data analysis, vol. 26. CRC press.
- [23] Silverman, B.W. and Jones, MC. 1989. E. Fix and J. L. Hodges. (1951). An important contribution to nonparametric discriminant analysis and density estimation. *Int. Stat. Rev.* 57(3) 233-247.
- [24] Stone, C. J. (1984). An asymptotically optimal window selection rule for kernels density estimates. *The analas of statistics*, 12, p. 1285-1297.
- [25] Von Neumann, J. (1932). Proof of the quasi-ergodic hypothesis. *Proc. Nat. Acad. Sci. USA*, 18, p. 70-82.

- [26] Woodroof, M. (1970). On choosing a delta sequence. *Ann. Math. Statist.*, 41, p. 1665-1671.
- [27] Wu, W. B., Huang, Y. and Huang, Y. (2010). Kernel estimation for time series : An asymptotic theory. *Stochastic Processes and their Applications*, 120, p. 2412-2431.