BEATRICE LAURENT-BONNEAU

Estimation de fonctionnelles intégrales non linéaires d'une densité et de ses dérivées, 1993

UNIVERSITE PARIS SUD

CENTRE d'ORSAY

THESE

présentée pour obtenir

Le TITRE de DOCTEUR EN SCIENCES

SPECIALITE : MATHEMATIQUES

par

Béatrice LAURENT-BONNEAU

Sujet : **ESTIMATION DE FONCTIONNELLES INTEGRALES NON LINEAIRES D'UNE DENSITE ET DE SES DERIVEES.**

Soutenue le 4 février 1993 devant le jury composé de :

Monsieur Lucien BIRGÉ Président
Monsieur Jean BRETAGNOLLE
Monsieur Paul DOUKHAN
Monsieur Yannick HEURTEAUX
Monsieur Pascal MASSART
Madame Dominique PICARD

Remerciements

Je tiens en premier lieu à remercier chaleureusement mon directeur de thèse, Pascal Massart. Je lui suis très reconnaissante de m'avoir fait profiter de sa vaste culture mathématique toujours avec beaucoup de gentillesse, de patience et un immense enthousiasme. Le travail en sa compagnie a été pour moi très agréable et très enrichissant.

Je souhaite remercier tout particulièrement Lucien Birgé pour l'attention qu'il a porté à mon travail. Il m'a notamment donné de précieux conseils lors de la rédaction de mon premier article. Je le remercie également d'avoir accepté de rapporter cette thèse et d'assurer la présidence de mon jury.

Je tiens à exprimer ma sincère gratitude à Boris Levit, rapporteur de cette thèse.

Paul Doukhan m'a fait profiter de ses connaissances en matière d'estimation fonctionnelle, je le remercie également d'avoir accepté de faire partie de mon jury, ainsi que Jean Bretagnolle et Dominique Picard.

Je suis très reconnaissante envers Yannick Heurteaux qui a accepté de répondre aux questions d'analyse que je lui ai posées au cours de ma thèse, je suis particulièrement sensible à sa présence au sein de mon jury.

Je remercie mes collègues et amis du laboratoire de statistique de l'Université d'Orsay pour l'ambiance sympathique et stimulante qui y règne et pour les échanges enrichissants que nous avons eus.

Je remercie Sabine Hoarau pour sa gentillesse et son efficacité ainsi que madame Baillet, les ingénieurs informaticiens du laboratoire et le personnel de la bibliothèque de mathématiques. Je remercie également Sandrine Lecouteux qui a veillé à l'édition de cette thèse.

Je n'oublie pas mes proches, à qui j'exprime ma profonde gratitude. Nicolas a sû m'encourager aux moments opportuns, je le remercie pour sa patience exemplaire et son soutien affectif. Je remercie du fond du coeur mes parents pour l'affection qu'ils m'ont toujours prodiguée et pour leurs encouragements.

Table des matières

Chapitre 1

Introduction

1.1 Etude bibliographique

L'étude de l'estimation des fonctionnelles intégrales non linéaires d'une densité et de ses dérivées se développe à partir des années 70. On dispose alors d'un certain nombre de résultats concernant l'estimation de densités, notamment par la méthode des noyaux ; résultats qui sont utilisés pour estimer les fonctionnelles du type $I(f) = \int \phi(f, f', \dots f^{(k)})$. Ainsi, Dmitriev et Tarasenko (1973) proposent d'estimer $I(f)$ dans un modèle i.i.d. par

$$I(\hat{f}_n) = \int \phi(\hat{f}_n, \hat{f}_n{}', \dots \hat{f}_n{}^{(k)})$$

où $\hat{f}_n$ est un estimateur à noyau de f :

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{j=1}^{n} K(\frac{x - X_j}{h_n}).$$

Les estimateurs à noyau ont été proposés pour la première fois par Rosenblatt (1956); leurs propriétés asymptotiques ont tout d'abord été étudiées par Parzen (1962). Un peu plus tard Bhattacharya (1967) et Schuster (1969) étudient les propriétés de $\hat{f}_n{}^{(k)}$ pour l'estimation de $f^{(k)}$. A l'aide de tous ces résultats, Dmitriev et Tarasenko démontrent la convergence presque sûre de leur estimateur dans les cas particuliers suivants :

- l'estimation de l'information de Fisher $\int \frac{f'^2}{f}$;
- l'estimation de l'entropie $-\int f \log(f)$;
- l'estimation de $\int f^2$.

L'estimation de l'entropie avait déjà été étudiée par Basharin (1956) dans le cas de variables aléatoires discrètes.

Levit (1978) s'intéresse à l'estimation de fonctionnelles du type

$$\Phi = \int \phi(F(x), F'(x), \ldots, F^{(r)}(x), x) dF(x),$$

F étant la fonction de répartition des X_i. Il propose d'estimer Φ par la quantité

$$\widehat{\Phi_n} = \frac{1}{n} \sum_{i=1}^{n} \phi(F_n(X_i), \hat{f}_n(X_i), \ldots \hat{f}_n^{(r-1)}(X_i), X_i)$$

où F_n est la fonction de répartition empirique des X_i et $\hat{f}_n$ un estimateur à noyau de la densité f. Sous des conditions de régularité pour la densité f qui ne sont pas les conditions optimales, Levit obtient des estimateurs efficaces et asymptotiquement gaussiens :

$$\begin{aligned} \sqrt{n}(\widehat{\Phi_n} - \Phi) &\xrightarrow{\mathcal{L}} \mathcal{N}(0, C(f, \phi)) \\ nE(\widehat{\Phi_n} - \Phi)^2 &\rightarrow C(f, \phi) \end{aligned}$$

où $C(f, \phi)$ est la variance asymptotique minimale pour ce problème d'estimation. Pour le démontrer, Levit s'appuie sur les résultats de Koshevnik et Levit (1976).

Les premiers résultats de minoration de la vitesse de convergence des estimateurs pour ce type de problème sont donnés par Ibragimov, Nemirovskii et Hasminskii (1987) dans le cadre du modèle du bruit blanc. On observe

$$Y(t) = \int_0^t f(u) du + \epsilon w(t), \quad t \in [0, 1].$$

w est un processus de Wiener standard. On s'intéresse à l'estimation d'une fonctionnelle du type $\Phi(f)$. Les trois auteurs proposent un estimateur de $\Phi(f)$ qui est efficace en particulier lorsque Φ'' est lipschitzienne et lorsque f est supposée à priori appartenir à la classe de fonctions

$$\Lambda_\alpha = \{f \in \mathbb{L}^2([0, 1]), |f(x) - f(y)| \leq C|x - y|^\alpha\} \quad \text{avec } \alpha > \frac{1}{2}.$$

Lorsque $\alpha < \frac{1}{4}$, ils démontrent le résultat suivant de minoration :

$$\inf_{T_\epsilon} \sup_{f \in \Lambda_\alpha} E_f(T_\epsilon - \Phi(f))^2 \geq \epsilon^{\frac{16\alpha}{1+4\alpha}}.$$

En 1987 Hall et Marron proposent des estimateurs de $\int (f^{(k)})^2$ dans le cadre d'un modèle i.i.d. . Pour simplifier la présentation de leurs résultats, intéressons-nous seulement à l'estimation de $\theta = \int f^2$. Soit $\hat{f}$ un estimateur à noyau de la densité f.

$$\hat{f} = \frac{1}{n}\sum_{j=1}^{n} K_h(x - X_j), \quad \text{avec } K_h = \frac{1}{h}K(\frac{\cdot}{h}).$$

Il est assez naturel d'estimer $\int f^2$ par l'une ou l'autre des quantités suivantes :

- a) $\displaystyle\int \hat{f}^2 = \frac{1}{n^2}\sum_{j,k=1}^{n} \int K_h(x - X_j)K_h(x - X_k)dx;$

- b) $\displaystyle\frac{1}{n}\sum_{j=1}^{n} \hat{f}(X_j) = \frac{1}{n^2}\sum_{j,k=1}^{n} K_h(X_j - X_k).$

Dans les deux cas, Hall et Marron constatent que les termes diagonaux, i.e. correspondant à $j = k$ introduisent un biais supplémentaire, ils les suppriment et proposent donc les estimateurs suivants pour $\int f^2$:

$$\begin{aligned}
\widehat{\theta}_1 &= \frac{1}{n(n-1)}\sum_{j\neq k=1}^{n} \int K_h(x - X_j)K_h(x - X_k)dx \\
\widehat{\theta}_2 &= \frac{1}{n(n-1)}\sum_{j\neq k=1}^{n} K_h(X_j - X_k).
\end{aligned}$$

Dans le cas où $f \in \Lambda_\alpha$, ces estimateurs convergent à la vitesse $1/\sqrt{n}$ dès que $\alpha \geq \frac{1}{2}$; lorsque $\alpha \leq \frac{1}{2}$ $E(\widehat{\theta}_i - \theta)^2_{i=1,2} = O(n^{\frac{-4\alpha}{1+2\alpha}})$ pour un choix optimal de la fenêtre h. Ces résultats ne sont pas optimaux ; l'indice de régularité critique pour la densité f permettant d'atteindre la vitesse semi-paramétrique $1/\sqrt{n}$ dans ce type de problème n'est pas $\frac{1}{2}$ mais $\frac{1}{4}$, comme l'ont démontré Bickel et Ritov en 1988.

Afin de mieux comprendre les résultats contenus dans l'article de Bickel et Ritov, nous allons tout d'abord évaluer le biais des estimateurs proposés par Hall et Marron.

$$E(\widehat{\theta}_1) = \int (f * K_h)^2, \quad E(\widehat{\theta}_2) = \int (f * K_h)f.$$

Ainsi, pour $f \in \Lambda_\alpha$, le biais de chacun des deux estimateurs ci-dessus est de l'ordre de h^α. Il est aisé de constater que le biais de l'estimateur $2\widehat{\theta}_2 - \widehat{\theta}_1$ est égal à $-\int (f * K_h - f)^2$, il est donc de l'ordre de $h^{2\alpha}$ pour $f \in \Lambda_\alpha$. Ainsi, cette combinaison linéaire de $\widehat{\theta}_1$ et $\widehat{\theta}_2$ va réaliser de meilleures performances que chacun des deux estimateurs pris séparément pour estimer $\int f^2$. C'est l'idée de la construction de Bickel et Ritov, mais leur estimateur a une expression plus compliquée car ils ont séparé l'échantillon en deux parties, ce qui n'était pas nécessaire. Ils proposent en fait une combinaison linéaire de trois estimateurs pour estimer $T_k = \displaystyle\int (f^{(k)})^2$. Ils font une hypothèse à priori du type

$$f \in F_{p,\alpha,C} = \{f, |f^{(p)}(x) - f^{(p)}(x+h)| \leq g(x)|h|^\alpha\}$$

avec $g \in \mathbb{L}^2 \cap \mathbb{L}^\infty$, $p \geq k$, $\alpha \in]0,1]$. Leur estimateur $\widehat{T_k}$ a les propriétés suivantes :

$$\text{i) Si } p+\alpha > 2k + \frac{1}{4} \qquad \sqrt{n}(\widehat{T_k} - T_k) \xrightarrow{\mathcal{L}} \mathcal{N}(0, C_k(f))$$
$$nE(\widehat{T_k} - T_k)^2 \to C_k(f)$$
$$\text{ii) Si } p+\alpha \leq 2k + \frac{1}{4} \qquad E(\widehat{T_k} - T_k)^2 = O(n^{\frac{-8(p+\alpha-k)}{1+4p+4\alpha}}).$$

La variance asymptotique $C_k(f) = 4\left[\int (f^{(2k)})^2 f - \left(\int (f^{(k)})^2\right)^2\right]$ est l'analogue en semi-paramétrique de la borne de Cramér-Rao d'après les résultats de Koshevnik et Levit (1976). L'estimateur est donc efficace. De plus, les deux auteurs montrent que la vitesse obtenue en ii) est optimale.

Il est également intéressant de citer un article de Donoho et Nussbaum (1990) qui traite de l'estimation de $\int (f^{(k)})^2$ dans le cadre du modèle du bruit blanc. Des vitesses de convergence du même ordre de grandeur que celles des estimateurs de Bickel et Ritov sont obtenues.

En 1991, Birgé et Massart ont démontré des résultats de minoration de vitesse de convergence pour le problème de l'estimation de $\int \phi(f,.)$ si f est une densité sur $\mathbb{R}^d$ et de l'estimation de $\int \phi(f,\ldots,f^{(k)},.)$ si f est une densité sur $\mathbb{R}$. Supposons que f soit à support dans $[0,1]^d$ et admette une régularité d'ordre s. Lorsque $s \leq 2k + \frac{d}{4}$, pour tout estimateur $\widehat{T_n}$ de $\int \phi(f, f', \ldots f^{(k)}, .)$ (avec $k = 0$ si $d > 1$), l'ordre de grandeur de $E(\widehat{T_n} - \int \phi(f, f', \ldots f^{(k)}, .)^2$ est minoré par $n^{\frac{-8(s-k)}{d+4s}}$. Ces résultats sont tout à fait intéressants dans le cadre de cette thèse car ils nous permettront dans un certain nombres de cas de voir que nos résultats sont optimaux.

1.2 Quelques applications

Nous allons donner un certain nombre d'exemples d'applications de l'estimation des fonctionnelles intégrales non linéaires. Nous présenterons successivement des applications de l'estimation de $\int f^2$, $\int (f^{(k)})^2$, $\int f \log f$ et $\int \frac{(f')^2}{f}$.

1) Estimation de $\int f^2$:

Supposons que l'on dispose de deux échantillons $X_1, \ldots, X_m$ et $Y_1, \ldots, Y_n$ issus d'un modèle de translation. Les variables X_i sont i.i.d. de densité f. Les variables Y_i sont i.i.d. et admettent la densité $f(.-\Delta)$. Nous nous intéressons au test de l'hypothèse $\Delta = 0$ contre $\Delta > 0$. Lehmann (1975) p. 371-374 montre que si l'on utilise la statistique de la somme des rangs de Wilcoxon notée U_{mn} pour bâtir le test, l'efficacité du test est proportionnelle à $\int f^2$. La puissance Π du test peut également être estimée à partir de l'estimation de $\int f^2$ d'après l'équivalence suivante démontrée par Lehmann :

$$\Pi \approx \Phi\left(\Delta\sqrt{\frac{12mn}{m+n}}\int f^2 - u_\alpha\right)$$

où Φ est la fonction de répartition de la loi normale $\mathcal{N}(0,1)$ et u_α est défini par $P(U_{mn} \leq u_\alpha) = \alpha$.

2) <u>Estimation de $\int(f^{(k)})^2$:</u>

L'estimation de $\int(f^{(k)})^2$ peut être utilisée pour déterminer la fenêtre optimale (en un sens à préciser) lors de l'estimation d'une densité par la méthode des noyaux ou des projections orthogonales.
Dans le cas de l'estimation par la méthode des noyaux, Deheuvels et Hominal (1980) s'intéressent à la détermination de la fenêtre optimale au sens le la minimisation du M.I.S.E. (mean integrated square error), défini par $E\left(\int(\hat{f}_n - f)^2\right)$ avec

$$\hat{f}_n = \frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

Il s'agit, pour un noyau K donné de déterminer la fenêtre optimale h. Deheuvels et Hominal utilisent le développement suivant :

$$E\left(\int(\hat{f}_n - f)^2\right) \simeq \frac{1}{nh}\int K^2 + \frac{1}{4}h^4\left(\int x^2K(x)dx\right)^2\int(f'')^2.$$

Il est alors aisé de voir que la valeur optimale de h sera obtenue en égalisant les deux termes de droite dans l'équivalence ci-dessus, ce qui conduit à choisir

$$h = \left(\frac{A}{n\int(f'')^2}\right)^{1/5} \quad \text{avec } A = \frac{\int K^2}{\left(\int x^2K(x)dx\right)^2}.$$

L'estimation de la densité par la méthode des noyaux peut alors s'effectuer en deux étapes:
a) on estime $\int(f'')^2$ ce qui nous permet d'avoir une estimation de la fenêtre qui minimise le M.I.S.E. ,

b) on estime la densité .
De même, si l'on estime la densité par la méthode des projections :

$$\tilde{f}_n = \frac{1}{n} \sum_{i=0}^{m} \sum_{j=1}^{n} p_i(X_j) p_i(x),$$

où $(p_i)_{i \in \mathbb{N}}$ est une base orthonormée, le choix optimal de m fait intervenir des fonctionnelles du type $\int (f^{(k)})^2$.

3) Estimation de l'entropie :

Vasicek (1976) propose un estimateur de l'entropie qu'il utilise pour construire un test de normalité. La légitimité d'un tel test est liée au fait que l'entropie de la loi normale $\mathcal{N}(0, \sigma^2)$ est supérieure à l'entropie de toute loi de même variance. Soit $X_1, \ldots, X_n$ un n-échantillon et $\widehat{T_n}$ un estimateur qui converge en probabilité vers l'entropie $-\displaystyle\int f \log(f)$. La quantité

$$\Lambda_n = \frac{\exp(\widehat{T_n})}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2}}$$

converge en probabilité vers $\sqrt{2\pi e}$ sous l'hypothèse de normalité des X_i ; en effet, l'entropie de la loi $\mathcal{N}(0, \sigma^2)$ est égale à $\log(\sqrt{2\pi e}\sigma)$. Sous toute hypothèse alternative, Λ_n va converger vers une valeur inférieure strictement à $\sqrt{2\pi e}$. L'hypothèse de normalité est rejetée si $\Lambda_n \leq u_\alpha$, u_α est déterminé à l'aide de simulations dans le cas où la loi de Λ_n sous l'hypothèse de normalité est inconnue. Sous certaines contre-hypothèses, Vasicek donne des résultats de simulations permettant d'évaluer la puissance du test. Il est clair que les performances de $\widehat{T_n}$ en tant qu'estimateur de l'entropie vont influer sur la puissance du test.

Dans le même ordre d'idée, Dudewicz et Van der Meulen (1981) proposent d'utiliser des estimateurs de l'entropie pour tester l'uniformité d'un $n-$ échantillon dont la densité est à support dans l'intervalle $[0,1]$. Si f est à support dans $[0,1]$, l'entropie $-\int f \log f$ est négative ; la valeur maximale de l'entropie, à savoir zéro, etant atteinte seulement dans le cas de la loi uniforme sur $[0,1]$.

4) Estimation de l'information :

L'application la plus simple de l'estimation de l'information de Fisher est dûe à l'inégalité de Cramér-Rao. Supposons que l'on dispose d'un $n-$échantillon issu d'une loi de densité $f(. - \theta)$ où θ est un paramètre réel inconnu. Il s'agit là d'un modèle de translation. Soit

Y un estimateur sans biais de θ. Si $\theta \to f_\theta$ est dérivable, si $\displaystyle\int \frac{(f'_\theta)^2}{f_\theta}$ existe et est non nulle, alors

$$nE_\theta(Y-\theta)^2 \geq \frac{1}{\int \frac{(f'_\theta)^2}{f_\theta}}.$$

Il est donc intéressant de savoir estimer la borne de Cramér-Rao : $\dfrac{1}{\frac{\int (f'_\theta)^2}{f_\theta}}$ afin d'évaluer les performances d'un estimateur Y de θ. Y sera dit efficace si la quantité $E_\theta(Y-\theta)^2$ est égale à la borne de Cramér-Rao.

Après ces préliminaires, qui nous ont permis de faire le point sur les résultats antérieurs sur le sujet, et de donner quelques motivations au problème de l'estimation des fonctionnelles intégrales non linéaires, nous allons décrire les résultats contenus dans la thèse.

1.3 Description de notre travail

Nous traitons le problème de l'estimation d'une fonctionnelle du type $\displaystyle\int \phi(f, f', \ldots, f^{(k)}, .)$ dans le cas d'un modèle i.i.d..

Le Chapitre 2 concerne l'estimation de fonctionnelles du type $\displaystyle\int \phi(f, .)$ dans un cadre très général. La première étape de la construction des estimateurs consiste à effectuer un développement de Taylor à l'ordre 2 de $\phi(f(x), x)$ au voisinage de $(\hat{f}(x), x)$ où $\hat{f}$ est un estimateur préliminaire de la densité f, qui sera construit sur une petite partie de l'échantillon. Il faut prendre un certain nombre de précautions lors du choix de l'estimateur préliminaire afin de garantir que les quantités $\phi(\hat{f}, .)$, $\phi'(\hat{f}, .)$ et $\phi''(\hat{f}, .)$ ont un sens. Ceci nous conduit à supposer que ϕ, ϕ' et ϕ'' sont définies sur un voisinage de l'image de f. En particulier, si nous estimons l'entropie, nous supposons que f est minorée par une constante strictement positive, et donc que f est à support compact. Cette hypothèse est assez restrictive ; elle exclut notamment l'application de ces techniques en vue d'un test de normalité. Par contre, nos estimateurs pourront être utilisés pour tester l'uniformité d'un n échantillon dont la densité est à support dans l'intervalle $[0, 1]$. En regroupant les termes du même type, le développement de Taylor s'écrit sous la forme $\displaystyle\int \phi(f, .) = \int G(\hat{f}, .) + \int H(\hat{f}, .)f + \int K(\hat{f}, .)f^2 + \Gamma_n$ où Γ_n est le terme de reste. La seconde étape de la construction consiste à estimer avec les données indépendantes de $\hat{f}$ les fonctionnelles linéaires et quadratiques qui apparaissent dans le développement. Les fonctionnelles linéaires sont estimées par un estimateur empirique. Les fonctionnelles quadratiques sont du type $\int K(\hat{f}, .)f^2$. Dans la mesure où elles sont estimées avec des données indépendantes de $\hat{f}$, nous pouvons considérer qu'elles sont du type $\int f^2\psi$ où ψ

est une fonction fixe.

Pour permettre de mieux comprendre la construction des estimateurs de $\int f^2\psi$, attachons-nous dans un premier temps à l'estimation de $\int f^2$. Supposons que $f \in \mathbb{L}^2(d\mu)$, μ étant une mesure queconque. Soit $(p_i)_{i\in D}$ une base orthonormée de $\mathbb{L}^2(d\mu)$, D est dénombrable. Soit $a_i = \int f p_i$, alors $\displaystyle\int f^2 = \sum_{i\in D} a_i^2$. Le coefficient a_i est estimé par l'estimateur empirique $\displaystyle\hat{a_i} = \frac{1}{n}\sum_{j=1}^{n} p_i(X_j)$, il est alors naturel d'estimer $\int f^2$ par $\displaystyle\sum_{i\in M_n} \hat{a_i}^2$; M_n étant un sous-ensemble fini de D, dépendant de n. Ceci nous conduit dans un premier temps à l'estimateur suivant pour $\int f^2$:

$$\frac{1}{n^2}\sum_{i\in M_n}\sum_{j,k=1}^{n} p_i(X_j)p_i(X_k).$$

Les calculs de biais de cet estimateurs montrent que les termes diagonaux (i.e. correspondant à $j = k$) introduisent un biais supplémentaire ; nous les supprimons donc, ce qui nous donne :

$$\hat{\theta} = \frac{1}{n(n-1)}\sum_{i\in M_n}\sum_{j\neq k=1}^{n} p_i(X_j)p_i(X_k).$$

L'ensemble M_n est déterminé de manière à minimiser la quantité $E(\hat{\theta} - \int f^2)^2$, c'est-à-dire la somme de la variance et du carré du biais de l'estimateur. Le choix optimal de M_n dépend bien entendu des hypothèses à priori que l'on fait sur f.

Supposons que f soit à support dans un compact de $\mathbb{R}$, disons $[-\pi, \pi]$, soit $(p_i)_{i\in\mathbb{N}}$ la base Fourier de $\mathbb{L}^2([-\pi,\pi])$. En prenant $M_n = \{i \in \mathbb{N}, i \leq m_n\}$, $\hat{\theta}$ s'écrit sous la forme suivante

$$\hat{\theta} = \frac{1}{n(n-1)}\sum_{j\neq k=1}^{n} D_{m_n}(X_j - X_k)$$

où D_{m_n} est le noyau de Dirichlet : $\displaystyle D_{m_n}(t) = \frac{\sin(m_n + \frac{1}{2})t}{2\pi \sin\frac{t}{2}}$.

On ne peut alors s'empêcher de noter l'analogie entre cet estimateur et l'estimateur $\hat{\theta}_2$ proposé par Hall et Marron :

$$\hat{\theta}_2 = \frac{1}{n(n-1)}\sum_{j\neq k=1}^{n} K_{h_n}(X_j - X_k).$$

Le noyau K_{h_n} a simplement été remplacé par le noyau de Dirichlet. Cette opération permet d'effectuer une réduction de biais ; le biais de $\hat{\theta}$ est de l'ordre de grandeur du carré du biais de $\hat{\theta}_2$.

Venons-en au problème initial de l'estimation de $\int f^2\psi$. En notant $S_{M_n}f = \sum_{i\in M_n} a_i p_i$, nous constatons que le biais de l'estimateur de $\int f^2$ proposé ci-dessus est égal à $-\int (S_{M_n} f - f)^2$. Pour l'estimation de $\int f^2 \psi$, nous cherchons à obtenir un biais du même ordre de grandeur, plus précisément un biais égal à $-\int (S_{M_n} f - f)^2 \psi$. Ceci nous conduit à prendre comme estimateur :

$$\hat{\theta} = \frac{2}{n(n-1)} \sum_{i \in M_n} \sum_{j \neq k = 1}^{n} p_i(X_j)(p_i \psi)(X_k)$$

$$-\frac{1}{n(n-1)} \sum_{i,i' \in M_n} \sum_{j \neq k = 1}^{n} p_i(X_j) p_{i'}(X_k) \int p_i p_{i'} \psi.$$

Nous sommes désormais en mesure de décrire l'estimateur de $\int \phi(f,.)$. Supposons que l'estimateur préliminaire soit construit avec les n_1 dernières données ; les parties linéaires et quadratiques du développement de Taylor sont estimées avec les n_2 premières données. Nous estimons $\int \phi(f,.)$ par

$$\widehat{T_n} = \int G(\hat{f},.) + \frac{1}{n_2} \sum_{j=1}^{n_2} H(\hat{f},.)(X_j) + \frac{2}{n(n-1)} \sum_{i \in M_n} \sum_{j \neq k = 1}^{n} p_i(X_j)(p_i K(\hat{f},.))(X_k)$$

$$-\frac{1}{n(n-1)} \sum_{i,i' \in M_n} \sum_{j \neq k = 1}^{n} p_i(X_j) p_{i'}(X_k) \int p_i p_{i'} K(\hat{f},.)$$

Le Théorème 2.2 du Chapitre 2 donne les propriétés asymptotiques de cet estimateur dans le cadre assez général où f est supposée à priori appartenir à l'ellipsoïde

$$\mathcal{E} = \{f = \sum_{i \in D} a_i p_i, \sum_{i \in D} |\frac{a_i}{c_i}|^2 \leq 1\}.$$

Cette étude couvre en particulier le cadre multidimensionnel. Dans le cas où f est à support dans $[0,1]^d$ et admet une régularité d'ordre s nous obtenons les résultats suivants:

$$\text{Si } s > \frac{d}{4} \qquad \sqrt{n}(\widehat{T_n} - \int \phi(f,.)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, C(f,\phi))$$

$$nE(\widehat{T_n} - \int \phi(f,.))^2 \to C(f,\phi)$$

avec

$$C(f,\phi) = \int (\phi'(f))^2 f - (\int \phi'(f) f)^2.$$

En nous appuyant sur des résultats de Koshevnik et Levit (1976) et de Ibragimov et Hasminskii (1991), nous démontrons que $C(f,\phi)$ est la variance asymptotique minimale pour ce problème. Nos estimateurs sont donc efficaces. Par ailleurs, Birgé et Massart (1991) ont démontré que lorsque $s < \frac{d}{4}$ on ne peut pas atteindre la vitesse semi-paramétrique $1/\sqrt{n}$. Dans le cas particulier de l'estimation de $\int f^2\psi$, nous obtenons la vitesse de convergence optimale si $s \leq \frac{d}{4}$, par contre lors de l'estimation de $\int \phi(f,.)$, le terme de reste du développement de Taylor Γ_n devient prépondérant devant les autres termes lorsque $s < \frac{d}{4}$ et nous n'atteignons pas la vitesse optimale dans ce cas.

Le Chapitre 3 traite de l'estimation de $T(f) = \int \phi(f, f', \ldots f^{(k)}, .)$. Nous restreignons le cadre de l'étude : f est supposée être une fonction à support dans $[-\pi, \pi]$, périodique ainsi que ses dérivées. De plus, nous faisons une hypothèse de régularité sur f :

$$f \in F_{p,\alpha,C} = \{f, |f^{(p)}(x) - f^{(p)}(y)| \leq C|x-y|^\alpha\}.$$

Dans la suite nous noterons s la quantité $p + \alpha$.
L'idée de la construction des estimateurs repose comme au chapitre précédent sur un développement de Taylor au second ordre de $\phi(f(x), \ldots f^{(k)}(x), x)$ au voisinage de $\phi(\hat{f}(x), \ldots \hat{f}^{(k)}(x), x)$, où $\hat{f}$ est un estimateur préliminaire de f construit avec les n_1 dernières données.
Après réarrangement des termes du développement de Taylor, la fonctionnelle $T(f)$ s'écrit sous la forme

$$T(f) = \int G(\hat{f}) + \sum_{j=0}^{k} \int_{-\pi}^{\pi} H_j(\hat{f}) f^{(j)} + \sum_{j,j'=0}^{k} \int_{-\pi}^{\pi} K_{jj'}(\hat{f}) f^{(j)} f^{(j')} + \Gamma_n.$$

En prenant soin de choisir $\hat{f}$ périodique ainsi que ses dérivées, des intégrations par parties successives nous permettent d'écrire :

$$\int_{-\pi}^{\pi} H_j(\hat{f}) f^{(j)} = (-1)^j \int_{-\pi}^{\pi} (H_j(\hat{f}))^{(j)} f.$$

Cette quantité est estimée par l'estimateur empirique correspondant bâti sur les n_2 premières données (avec $n_1 + n_2 = n$).
Une étape intermédiaire consiste à estimer les fonctionnelles du type $\int_{-\pi}^{\pi} f^{(j)} f^{(j')} \psi$ et à décrire les propriétés asymptotiques des estimateurs. C'est l'objet du Théorème 3.1 du Chapitre 3. Pour estimer cette quantité, nous généralisons les résultats du chapitre précédent concernant l'estimation de $\int f^2\psi$. Soit $(p_i)_{i\in\mathbb{N}}$ la base Fourier de $\mathbb{L}^2([-\pi, \pi])$; dans la mesure où les densités sont périodiques, il est assez naturel de choisir cette base orthonormée pour construire nos estimateurs. Soit $S_m f = \sum_{i=0}^{m} a_i p_i$; nous allons chercher à construire un estimateur de $\int_{-\pi}^{\pi} f^{(j)} f^{(j')} \psi$ dont le biais sera :

$$-\int (S_m f^{(j)} - f^{(j)})(S_m f^{(j')} - f^{(j')})\psi.$$

Ceci nous conduit à l'estimateur

$$\widehat{T_\psi^{jj'}} = \frac{1}{n(n-1)} \sum_{i=0}^{m} \sum_{l_1 \neq l_2 = 1}^{n} p_i(X_{l_1})(-1)^j (p_i^{(j')}\psi)^{(j)}(X_{l_2})$$

$$+\frac{1}{n(n-1)} \sum_{i=0}^{m} \sum_{l_1 \neq l_2 = 1}^{n} p_i(X_{l_1})(-1)^{j'} (p_i^{(j)}\psi)^{(j')}(X_{l_2})$$

$$-\frac{1}{n(n-1)} \sum_{i,i'=0}^{m} \sum_{l_1 \neq l_2 = 1}^{n} p_i(X_{l_1}) p_{i'}(X_{l_2}) \int_{-\pi}^{\pi} p_i^{(j)} p_{i'}^{(j')} \psi.$$

Remarquons que lorsque $j = j' = k$, $\psi = 1$, nous estimons $\int (f^{(k)})^2$ par

$$\widehat{T^k} = \frac{1}{n(n-1)} \sum_{i=0}^{m} \sum_{l_1 \neq l_2 = 1}^{n} (-1)^k p_i(X_{l_1}) p_i^{(2k)}(X_{l_2}).$$

Dans ce cas en effet les trois termes qui apparaissent dans l'expression de $\widehat{T_\psi^{jj'}}$ coïncident. L'estimateur $\widehat{T^k}$ a une expression très simple ; il réalise les mêmes performances que l'estimateur proposé par Bickel et Ritov (1988). Cet estimateur est asymptotiquement gaussien et efficace si $s > 2k+\frac{1}{4}$; il atteint la vitesse optimale de convergence si $s \leq 2k+\frac{1}{4}$.

Après cette étape intermédiaire, nous revenons au problème initial de l'estimation de $T(f) = \int \phi(f, \dots, f^{(k)}, .)$. Nous proposons l'estimateur suivant :

$$\begin{aligned}
\widehat{T_n} &= \int G(\hat{f}) + \frac{1}{n_2} \sum_{j=0}^{k} \sum_{l=1}^{n_2} (-1)^j (H_j(\hat{f}))^{(j)}(X_l) \\
&+ \sum_{j,j'=0}^{k} \frac{2}{n_2(n_2-1)} \sum_{i=0}^{m} \sum_{l_1 \neq l_2 = 1}^{n_2} p_i(X_{l_1})(-1)^j (p_i^{(j')} K_{jj'}(\hat{f}))^{(j)}(X_{l_2}) \\
&- \sum_{j,j'=0}^{k} \frac{1}{n_2(n_2-1)} \sum_{i,i'=0}^{m} \sum_{l_1 \neq l_2 = 1}^{n_2} p_i(X_{l_1}) p_{i'}(X_{l_2}) \int_{-\pi}^{\pi} p_i^{(j)} p_{i'}^{(j')} K_{jj'}(\hat{f})
\end{aligned}$$

Le résultat essentiel du Chapitre 3 est le Théorème 3.2 qui annonce les propriétés suivantes pour $\widehat{T_n}$:

i) Si $s > 2k + \frac{1}{4}$ $\quad \sqrt{n}(\widehat{T_n} - T(f)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, C(f, \dots f^{(k)}, \phi))$

$\quad nE(\widehat{T_n} - T(f))^2 \to C(f, \dots f^{(k)}, \phi)$

ii) Si $1 < s \leq 2k + \frac{1}{4}$ $\quad E(\widehat{T_n} - T(f))^2 = O(n^{\frac{-8(s-k)}{1+4s}})$.

Nous montrons à l'aide des résultats de Koshevnik et Levit (1976) ou de Ibragimov et Hasminskii (1991) que la constante $C(f, \ldots f^{(k)}, \phi)$ est la variance asymptotique minimale pour ce problème. De plus, d'après Birgé et Massart (1991), la vitesse obtenue en ii) est optimale.
Notons que l'hypothèse de périodicité n'est pas nécessaire si l'on estime une fonctionnelle qui ne dépend que de la densité, ceci est démontré au Chapitre 2. Par contre, nous ne pouvons éviter cette hypothèse dans le cas de l'estimation d'une fonctionnelle du type $\int \phi(f, \ldots, f^{(k)}, .)$. En effet, si nous cherchons à estimer la fonctionnelle $\int_{-\pi}^{\pi} f' = f(\pi) - f(-\pi)$, nous sommes ramenés au problème de l'estimation de la densité en un point. Farell (1972) a démontré que cette fonctionnelle s'estime à la même vitesse que la densité elle-même, c'est-à-dire à une vitesse inférieure à celle que nous avons énoncé ci-dessus.
Dans le cadre de ce théorème, de même que lors de l'estimation de $\int \phi(f, .)$, nous exigeons que ϕ et ses dérivées soient définies sur un ouvert contenant $f([-\pi, \pi]) \times \ldots f^{(k)}([-\pi, \pi]) \times [-\pi, \pi]$. Si nous estimons l'information de Fisher $\int \frac{(f')^2}{f}$, cette hypothèse se traduit, comme dans le cas de l'entropie, par une hypothèse de minoration de la densité par une constante strictement positive. Pour ces deux exemples particuliers (entropie et information de Fisher), il conviendrait d'éviter cette hypothèse mais actuellement nous n'avons obtenu aucun résultat dans cette direction.

Le Chapitre 4 donne des résulats de simulations concernant l'estimation de $\int f^2$ et celle de l'entropie. Pour l'estimation de $\int f^2$ nous comparons l'estimateur par projection proposé dans la thèse :

$$\widehat{\theta} = \frac{1}{n(n-1)} \sum_{j \neq k = 1}^{n} D_m(X_j - X_k)$$

où D_m est le noyau de Dirichlet, avec l'un des estimateurs proposés par Hall et Marron :

$$\widehat{\theta}_2 = \frac{1}{n(n-1)} \sum_{j \neq k = 1}^{n} K_h(X_j - X_k).$$

En ce qui concerne l'estimation de l'entropie $- \int f \log f$, l'estimateur proposé au Chapitre 2 de cette thèse est comparé avec un estimateur étudié par Györfi et Van der Meulen (1990):

$$-\frac{1}{n_2} \sum_{j=1}^{n_2} \log \hat{f}(X_j).$$

Cet estimateur correspond à un développement de Taylor au premier ordre, alors que notre estimateur est construit à partir d'un développement de Taylor au second ordre.
Les résultats des simulations font apparaître très nettement la réduction de biais réalisée par nos estimateurs. Lorsque les paramètres de lissage varient, la moyenne des estimateurs

proposés dans la thèse et beaucoup plus stable et proche de la vraie valeur du paramètre à estimer. Nous estimons également $E(\widehat{\theta}-\theta)^2$ par la moyenne des carrés des écarts entre le paramètre estimé et la vraie valeur. Dans la majorité des cas, nos estimateurs se comportent mieux que ceux auxquels ils sont comparés.

Pour conclure, notons qu'il serait intéressant à l'issue de ce travail de répondre aux questions suivantes non résolues dans la thèse :

a) Lors de l'estimation de $\int \phi(f,.)$, en supposant que f admette une régularité d'ordre s sur $\mathbb{R}^d$, nous n'obtenons pas les résultats optimaux annoncés par Birgé et Massart (1991) si $s < \frac{d}{4}$. Pour cela, il faudrait faire un développement de Taylor à l'ordre 3 et atteindre des résultats optimaux pour l'estimation de fonctionnelles du type $\int f^3\psi$. En utilisant des techniques du même genre que celles qui nous ont permis d'estimer les fonctionnelles quadratiques, nous n'obtenons pas les résultats escomptés. En fait les résultats récents de Kerkyacharian et Picard (1992), achevés en même temps que cette thèse et concernant l'estimation de $\displaystyle\int f^3$, permettrons sans doute de résoudre ce problème.

b) La seconde perspective de recherche consiste à essayer de supprimer l'hypothèse de minoration de la densité par une constante positive dans le cas de l'estimation de l'entropie et de l'information de Fisher. Il serait également intéressant de savoir estimer ces fonctionnelles dans le cas d'une densité à support non compact, ce qui nous permettrait de mettre en oeuvre des tests de normalité, tels qu'ils ont été proposés par Vasicek (1976).

c) Enfin, il serait utile d'obtenir des résultats concernant l'estimation des fonctionnelles intégrales non linéaires dans le cadre d'un modèle de données censurées, ce qui pourrait avoir des applications dans le domaine des statistiques médicales.

Chapitre 2

Efficient estimation of integral functionals of a density

Abstract

We consider the problem of estimating a functional of a density of the type $\int \phi(f,.)$. Starting from efficient estimators of linear and quadratic functionals of f and using a Taylor expansion of ϕ, we build estimators which achieve the $n^{-1/2}$ rate whenever f is smooth enough. Moreover, we show that these estimators are efficient. Concerning the estimation of quadratic functionals, more precisely of integrated squared density, Bickel and Ritov have already built efficient estimators, we propose here an alternative construction based on projections, which seems more natural.

2.1 Introduction

Let $X_1, \ldots, X_n$ be i.i.d. with common density f with respect to some measure μ. When μ is the Lebesgue measure on the real line, Bickel and Ritov (1988) have studied the problem of estimating $\int (f^{(k)})^2$ where f is supposed to belong to a nonparametric set of densities Θ_s included in some compact set of smooth functions of order s. They built an efficient estimator if $s > 2k + \frac{1}{4}$. If $s \leq 2k + \frac{1}{4}$, they showed that the best order of convergence is $n^{\frac{-4(s-k)}{1+4s}}$. It is a quite remarkable result in the sense that the critical regularity $2k + \frac{1}{4}$ is completely unusual ; actually, one could think at first glance that this critical regularity should be $2k + \frac{1}{2}$, (see Hall and Marron (1987) where some statistical motivations for studying these functionals are also provided).
This problem has also been treated by Donoho and Nussbaum (1990) for the white noise model ; it is also worth to mention the paper by Ibragimov, Nemirovskii and Khas'minskii (1987) which deals with differentiable functionals in the same framework.

Bickel and Ritov's estimator is a quite intricate expression based on kernel estimators of the density. In this paper, we propose an alternative and somehow simpler method of estimation based on orthogonal projections. This method will allow us to treat the more general problem of estimating $\int f^2 \psi d\mu$ when f belongs to some ellipsoid $\mathcal{E} = \{\sum_{i \in D} a_i p_i ; \sum_{i \in D} |\frac{a_i^2}{c_i^2}| \leq 1\}$ where $(p_i)_{i \in D}$ is an orthonormal basis of $\mathbb{L}^2(d\mu)$.
This generalisation is crucial to achieve the aim of this paper which is to construct efficient estimators of functionals of the type $T(f) = \int \phi(f(x), x) d\mu(x)$, $f \in \mathcal{E}$, when it is possible. This problem was first studied by Levit (1978), who built efficient estimators of this kind of functionals, under regularity properties for the density which are more restrictive than our conditions.
Some typical motivating example of such functionals is the Shannon entropy $\int f \log(f)$. Dudewicz and Van der Meulen (1981) show how estimators of the entropy may be used to test uniformity of a n-sample with density f concentrated on the interval $[0, 1]$. Moreover, Vasicek (1976) proposes a test of normality which is also based on estimators of the entropy.

Before stating the results that we get, let us explain how the two problems mentioned above are connected. ϕ is assumed to be a smooth function. So, expanding ϕ up to the second order with Taylor's formula provides an expansion of $T(f) - T(\widehat{f})$, where $\widehat{f}$ is a nonparametric preliminary estimator of the density f, constructed with a small part of the $n-$ sample. With the remainder of the sample, we build estimators of the terms, up to the second order, which appear in the Taylor expansion. Some of these terms are linear functionals of f while others are quadratic functionals of the type $\int f^2 K(\widehat{f})$ which has been studied in the first section. Our main result is proved in Section 3 and may be summarized as follows : when f belongs to some Hölder's space with index s over $\mathbb{R}^d$, we can built an estimator $\widehat{T_n}$ of $T(f)$ such that

i) If $s > \frac{d}{4}$ $\sqrt{n}(\widehat{T_n} - T(f)) \to \mathcal{N}(0, C(f,\phi))$ where $\mathcal{N}(0,\sigma^2)$ denotes the normal distribution and $nE(\widehat{T_n} - T(f))^2 \to C(f,\phi)$ where

$$C(f,\phi) = \int (\phi_1'(f,.))^2 f - (\int \phi_1'(f,.)f)^2$$

and $\phi_1'(u,v) = \frac{\partial \phi}{\partial u}(u,v)$. Moreover, $C(f,\phi)$ is the semiparametric information bound for the problem of estimating $T(f)$ as will be shown in the Appendix, hence our estimator is asymptotically efficient.

When $s < \frac{d}{4}$ we do not know what the optimal rate is, except that it is smaller than $n^{\frac{-3s}{d+2s}}$. Actually, in this very case the remainder term in the Taylor expansion is precisely of order $n^{\frac{-3s}{d+2s}}$. Birgé and Massart (1991) have proved that the rate cannot be smaller than $n^{\frac{-4s}{d+4s}}$. So, it would be necessary to do the Taylor expansion up to the third order and to estimate $\int f^3 \psi$ but we do not know how to estimate $\int f^3$ at a rate faster than the one we just mentioned. After the completion of this paper, Kerkyacharian and Picard (1992) have found that the rate of convergence for the estimation of $\int f^3$ is actually $n^{\frac{-4s}{d+4s}}$.

2.2 Estimation of $\int f^2 \psi$

Suppose $X_1, \ldots, X_n$ are i.i.d. random variables with common density $f \in \mathbb{L}^2(d\mu)$. Let $(p_i)_{i \in D}$ be an orthonormal basis of $\mathbb{L}^2(d\mu)$, where D is a countable set ; $a_i = \int f p_i$ and consider the ellipsoid $\mathcal{E} = \{\sum_{i \in D} a_i p_i ; \sum_{i \in D} |\frac{a_i^2}{c_i^2}| \leq 1\}$.

Our purpose is to estimate $\theta = \int f^2 \psi$ when $f \in \mathcal{E}$. Let us first look at the case $\psi = 1$. Since $\int f^2 = \sum_{i \in D} a_i^2$, a natural idea would be to estimate this integral by $\tilde{\theta} = \sum_{i \in M} \widehat{a_i}^2$, where $\widehat{a_i}$ is the empirical estimator of a_i : $\widehat{a_i} = \frac{1}{n} \sum_{j=1}^{n} p_i(X_j)$ and M is a subset of D. Therefore $\tilde{\theta} = \frac{1}{n^2} \sum_{i \in M} \sum_{j,l=1}^{n} p_i(X_j) p_i(X_l)$.

The computation of the bias of this estimator shows that it can be reduced by removing the diagonal terms, i.e. the terms of the type $\frac{1}{n^2} p_i^2(X_l)$. This leads to an estimator of $\int f^2$ with bias $-\int (S_M f - f)^2$ where $S_M f$ denotes $\sum_{i \in M} a_i p_i$.

We intend to built an estimator with a similar bias for the estimation of $\int f^2\psi$. More precisely, we wish the bias to be equal to

$$-\int (S_M f - f)^2 \psi = 2\int (S_M f) f\psi - \int (S_M f)^2 \psi - \int f^2 \psi.$$

Hence, the problem is to find an estimator with expectation $2\int (S_M f) f\psi - \int (S_M f)^2 \psi$. For the part $\int (S_M f) f\psi$, we propose the following estimator :

$$\widehat{\theta}_1 = \frac{1}{n(n-1)} \sum_{i \in M} \sum_{j \neq k = 1}^{n} p_i(X_j)(p_i\psi)(X_k). \tag{2.1}$$

To get the term $\int (S_M f)^2 \psi$, we propose the estimator

$$\widehat{\theta}_2 = \frac{1}{n(n-1)} \sum_{i,i' \in M} \sum_{j \neq k = 1}^{n} p_i(X_j) p_{i'}(X_k) \int p_i p_{i'} \psi(x) dx \tag{2.2}$$

with $p_i p_{i'} \psi(x) = p_i(x) p_{i'}(x) \psi(x)$. This explains the expression of the estimator $\widehat{\theta}$ proposed below :

Theorem 2.1 *Let $X_1, \ldots, X_n$ i.i.d. random variables with common density f belonging to some Hilbert space $\mathbb{L}^2(d\mu)$. Let $(p_i)_{i \in D}$ an orthonormal basis of $\mathbb{L}^2(d\mu)$. We assume that f is uniformly bounded and belongs to the ellipsoid $\mathcal{E} = \{\sum_{i \in D} a_i p_i; \sum_{i \in D} |\frac{a_i^2}{c_i^2}| \leq 1\}$. Suppose that the following condition holds : we can find a subset M_n of D such that*

$$\left(\sup_{i \notin M_n} |c_i|^2\right)^2 \approx \frac{|M_n|}{n^2}$$

where $|M_n|$ denotes the cardinality of M_n and

$$\forall\, g \in \mathbb{L}^2(d\mu) \quad \int (S_{M_n} g - g)^2 d\mu \to 0 \quad as \quad n \to \infty.$$

Let $\theta = \int f^2 \psi$ to be estimated where ψ is a bounded function and let

$$\widehat{\theta} = \frac{2}{n(n-1)} \sum_{i \in M_n} \sum_{j \neq k = 1}^{n} p_i(X_j)(p_i\psi)(X_k)$$

$$-\frac{1}{n(n-1)} \sum_{i,i' \in M_n} \sum_{j \neq k = 1}^{n} p_i(X_j) p_{i'}(X_k) \int p_i p_{i'} \psi(x) dx,$$

$\widehat{\theta}$ has the following asymptotic properties as an estimator of $\int f^2 \psi$:
i) If $\frac{|M_n|}{n} \to 0$ as $n \to \infty$ then

$$\sqrt{n}(\widehat{\theta}-\theta) \quad \rightarrow \quad \mathcal{N}(0,\Lambda(f,\psi)) \tag{2.3}$$

$$|nE(\widehat{\theta}-\theta)^2-\Lambda(f,\psi)| \quad \leq \quad \gamma_1\left[\frac{|M_n|}{n}+\|S_{M_n}f-f\|_2+\|S_{M_n}(f\psi)-f\psi\|_2\right] \tag{2.4}$$

where

$$\Lambda(f,\psi)=4[\int f^3\psi^2-(\int f^2\psi)^2].$$

ii) Otherwise

$$E(\widehat{\theta}-\theta)^2 \leq \gamma_2\frac{|M_n|}{n^2}$$

where γ_1 and γ_2 depend only on $\|f\|_\infty$ and $\|\psi\|_\infty$. Moreover, they are both increasing functions of $\|f\|_\infty$ and $\|\psi\|_\infty$.

The notation $A_n \approx B_n$ used in Theorem 2.1 means that $\lambda_1 \leq \frac{A_n}{B_n} \leq \lambda_2$ where λ_1 and λ_2 are positive constants.

<u>Comments :</u> Of course, it follows from (2.4) that $\lim_{n\to\infty} nE(\widehat{\theta}-\theta)^2 = 4[\int f^3\psi^2-(\int f^2\psi)^2]$.
In the next section, ψ will be a random function depending on n, which explains the fact that we need a bound depending explicitly on ψ.
We shall prove in the Appendix that the asymptotic variance is optimal when f is assumed to belong to the ellipsoid $\mathcal{E}$ and to be bounded from below by some positive constant, or when f is assumed to belong to a class of regular densities as it will be defined in the next section. Our estimator is therefore efficient in these cases.

<u>Example 1 :</u>
Let $X_1,\ldots,X_n$ be n i.i.d. d dimensional random variables with density f belonging to $\mathbb{L}^2(d\mu)$ where μ is the Lebesgue measure over $\mathbb{R}^d$. f is supposed to belong to the ellipsoid $\mathcal{E}=\{\sum_{i\in\mathbb{Z}^d} a_ip_i;\ \sum_{(i_1\ldots i_d)\in\mathbb{Z}^d}(|i_1|^{2s_1}+\ldots+|i_d|^{2s_d})|a_{i_1\ldots i_d}|^2\leq 1\}$ where $s_j>0\ \forall\, j\in\{1\ldots d\}$.
Let $s\in\mathbb{R}$ be defined by

$$\frac{d}{s}=\sum_{j=1}^{d}\frac{1}{s_j}$$

and $M_n=\{(i_1,\ldots,i_d)\in\mathbb{Z}^d, |i_j|\leq m^{\frac{s}{ds_j}}\}$. The cardinality of M_n has the same order as m. Moreover,

$$\sup_{i\notin M_n}|c_i|^2\approx\left(m^{\frac{-2s}{d}}\right).$$

Let $m \approx n^{\frac{2}{1+4s/d}}$. Then,

$$\frac{(\sup_{i \notin M_n} |c_i|^2)^2}{|M_n|} \approx \frac{1}{n^2}$$

Hence, as soon as $s > \frac{d}{4}$ the estimator $\widehat{\theta}$ defined in Theorem 2.1 has the properties defined by (2.3) and (2.4). When $s \leq \frac{d}{4}$ we get $E(\widehat{\theta} - \theta)^2 \leq \lambda n^{\frac{-8s}{d+4s}}$ for some $\lambda \in \mathbb{R}$.
In the framework of this example, when f belongs to $\mathbb{L}^2([0,1]^d)$ and when $(p_i)_{i \in \mathbb{Z}^d}$ is the Fourier orthonormal basis of $\mathbb{L}^2([0,1]^d)$ we shall prove (see proof of Corollary 2.1) that the condition $f \in \mathcal{E}$ generalizes a condition of the type : f belongs to some Hölder space of any index bigger than s. In this particular framework, and when ψ is either always positive or always negative, Birgé and Massart (1991) have proved lower bounds for the rates of convergence which agree with the rates of Theorem 2.1. Our result is therefore optimal.

Example 2 :
Suppose that $f \in \mathbb{L}^2([0,1])$. We give another example where $(p_i)_{i \in \mathbb{N}}$ will be the Haar orthonormal basis of $\mathbb{L}^2([0,1])$. Let

$$\begin{aligned} p_{-1,0}(x) &= \mathbb{1}_{[0,1]}(x) \\ p_{0,0}(x) &= \mathbb{1}_{[0,1/2]}(x) - \mathbb{1}_{]1/2,1]}(x) \\ p_{j,k}(x) &= 2^{j/2} p_{0,0}(2^j x - k) \quad \forall\, 0 \leq k \leq 2^j - 1, \quad j \in \mathbb{N} \end{aligned}$$

i.e.

$$p_{j,k}(x) = 2^{j/2} \left\{ \mathbb{1}_{[\frac{k}{2^j}, \frac{k}{2^j} + \frac{1}{2^{j+1}}]} - \mathbb{1}_{]\frac{k}{2^j} + \frac{1}{2^{j+1}}, \frac{k+1}{2^j}]} \right\} (x).$$

Let $\alpha_{j,k} = \int_0^1 f(x) p_{j,k}(x) dx$ and let the ellipsoid $\mathcal{E}$ be defined by :

$$\mathcal{E} = \{ \sum_{j \geq -1} \sum_{k=0}^{2^j - 1} \alpha_{j,k} p_{j,k}, \sum_{j \geq -1} \sum_{k=0}^{2^j - 1} 2^{2js} \alpha_{j,k}^2 \leq 1 \}.$$

In order to apply Theorem 2.1, we set

$$M_n = \{(j,k), 0 \leq k \leq 2^j - 1, -1 \leq j \leq j_0 = (\frac{2}{1+4s}) \log_2(n)\}$$

and

$$c_{j,k} = 2^{-js}, \quad \forall\, j, \forall\, k \in \{0, \ldots 2^j - 1\}.$$

Since

$$|M_n| \approx n^{\frac{2}{1+4s}}$$

and

$$\sup_{(j,k)\notin M_n} |c_{j,k}|^2 \approx 2^{-2j_0 s} = n^{\frac{-4s}{1+4s}}$$

we get :

$$\begin{array}{lll} \text{if } s > \frac{1}{4} & \text{then} & \sqrt{n}(\widehat{\theta} - \theta) \to \mathcal{N}(0, \Lambda(f, \psi)) \\ & & nE(\widehat{\theta} - \theta)^2 \to \Lambda(f, \psi)) \\ \text{if } s \leq \frac{1}{4} & \text{then} & E(\widehat{\theta} - \theta)^2 = O(n^{\frac{-8s}{1+4s}}). \end{array}$$

2.3 Estimation of $\int \phi(f)$.

2.3.1 Main results

The purpose of this section is to estimate $T(f) = \int \phi(f(x), x)dx$ efficiently when it is possible. As in the previous section, we assume that f belongs to the ellipsoid $\mathcal{E} = \{\sum_{i\in D} a_i p_i ; \sum_{i\in D} |\frac{a_i^2}{c_i^2}| \leq 1\}$.

We would like to start with some preliminary estimator $\widehat{f}$ of the density f built on a small part of the initial sample and do a Taylor expansion of ϕ in a neighbourhood of $(\widehat{f}(x), x)$. In order to give a sense to this expansion we shall assume the following :

• A_1 : The function $u \to \phi(u, x)$ belongs to $C^3(\Omega)$ $\forall\ x$, where $C^p(\Omega)$ denotes the class of p times continuously differentiable functions over Ω. $\forall\ x \quad a \leq f(x) \leq b$, where $a, b \in \mathbb{R}$, with $[a, b] \subset \Omega$.

• A_2 : We can find a preliminary estimator $\widehat{f}$ of f constructed with $n_1 \approx \frac{n}{log(n)}$ data, such that $\forall\ x \quad a - \epsilon \leq \widehat{f}(x) \leq b + \epsilon$ with $[a - \epsilon, b + \epsilon] \subset \Omega$ for some $\epsilon > 0$. Moreover, $\forall\ 2 \leq q < +\infty, \quad \forall\ l \in \mathbb{N}^*\ E_f \|\widehat{f} - f\|_q^l \leq C(q, l) n_1^{-l\alpha}$ for some $\alpha > \frac{1}{6}$ and for some constant $C(q, l)$ independent of f belonging to the ellipsoid $\mathcal{E}$.

We shall give an example of such an estimator in the case where f is a density defined over a compact set S of $\mathbb{R}^d$ satisfying some regularity assumptions.
Assuming that A1 and A2 are verified, it is now legitimate to make a Taylor expansion of ϕ in a neighbourhood of $(\widehat{f}(x), x)$. We shall use the following notation for partial derivatives

$$\phi_1' = \frac{\partial \phi}{\partial u}(u,v), \quad \phi_1'' = \frac{\partial^2 \phi}{\partial u^2}(u,v), \quad \|\phi_1^{(3)}\|_\infty = \sup_{x,u \in K_\epsilon} |\frac{\partial^3 \phi}{\partial u^3}(u,x)|$$

where $K_\epsilon = [a - \epsilon, b + \epsilon]$.

$$T(f) = \int \phi(\widehat{f}(x), x)dx + \int \phi_1'(\widehat{f}(x), x)(f - \widehat{f})(x)dx$$

$$+\frac{1}{2}\int \phi_1''(\widehat{f}(x), x)(f - \widehat{f})^2(x)dx + \Gamma_n,$$

where Γ_n is a remainder term which will be proved to be negligible compared to the linear and quadratic terms. It is convenient to write $T(f)$ as follows :

$$T(f) = \int G(\widehat{f},.) + \int H(\widehat{f},.)f + \int K(\widehat{f},.)f^2 + \Gamma_n$$

where

$$G(\widehat{f},.) = \phi(\widehat{f},.) - \phi_1'(\widehat{f},.)\widehat{f} + \frac{1}{2}\phi_1''(\widehat{f},.)\widehat{f}^2 \quad (2.5)$$

$$H(\widehat{f},.) = \phi_1'(\widehat{f},.) - \widehat{f}\phi_1''(\widehat{f},.) \quad (2.6)$$

$$K(\widehat{f},.) = \frac{1}{2}\phi_1''(\widehat{f},.) \quad (2.7)$$

We have to estimate two types of functionals :

- $L = \int H(\widehat{f},.)f$ which is a linear functional of f.
- $Q = \int K(\widehat{f},.)f^2$ which is a quadratic functional of f of the type studied in Section 2 .

If $\widehat{f}$ is based on the n_1 last observations, then L and Q are estimated with the n_2 first data where $n_2 = n - n_1$. The following theorem gives the expression of the estimator $\widehat{T_n}$ of $T(f)$ and its properties.

Theorem 2.2 *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with common density f belonging to some Hilbert space $\mathbb{L}^2(d\mu)$. Let $(p_i)_{i \in D}$ be an orthonormal basis of $\mathbb{L}^2(d\mu)$, $a_i = \int f p_i$ and suppose that f belongs to the ellipsoid*

$$\mathcal{E} = \{\sum_{i \in D} a_i p_i; \sum_{i \in D} |\frac{a_i^2}{c_i^2}| \leq 1\}$$

Let $T(f) = \int \phi(f,.)$ to be estimated.
We assume that the hypotheses A1 and A2 hold and that $\|\phi_1'\|_\infty$, $\|\phi_1''\|_\infty$ and $\|\phi_1^{(3)}\|_\infty$ are finite.

Suppose that we can find a subset M_n of D such that $\left(\sup_{i\notin M_n} |c_i^2|\right)^2 \approx \frac{|M_n|}{n^2}$ *and such that* $\forall\ g \in \mathbb{L}^2(d\mu)\quad \|S_{M_n}g - g\|_2 \to 0 \quad$ *as* $n \to \infty$. *Let*

$$\widehat{T_n} = \int G(\widehat{f},.) + \frac{1}{n_2}\sum_{j=1}^{n_2} H(\widehat{f},.)(X_j) + \frac{2}{n_2(n_2-1)}\sum_{i\in M_n}\sum_{j\neq k=1}^{n} p_i(X_j)(K(\widehat{f},.)p_i)(X_k)$$

$$-\frac{1}{n_2(n_2-1)}\sum_{i,i'\in M_n}\sum_{j\neq k=1}^{n} p_i(X_j)p_{i'}(X_k)\int p_i p_{i'} K(\widehat{f},.)$$

where G, H, K are defined by (2.5), (2.6), and (2.7). The following properties hold :

$$\text{If } \frac{|M_n|}{n} \to 0 \text{ then } \sqrt{n}(\widehat{T_n} - T(f)) \;\to\; \mathcal{N}(0, C(f,\phi)) \tag{2.8}$$

$$\lim_{n\to\infty} nE(\widehat{T_n} - T(f))^2 \;=\; C(f,\phi) \tag{2.9}$$

where $C(f,\phi)$ is defined by

$$C(f,\phi) = \int (\phi_1'(f,.))^2 f - \left(\int \phi_1'(f,.) f\right)^2.$$

Comments :

1) The asymptotic constant $C(f,\phi)$ appearing in Theorem 2.2 is optimal when f is supposed to belong to the class $F_{r,\alpha,C}$ which is defined below, or belongs to the ellipsoid $\mathcal{E}$ and is bounded from below by some positive constant. This follows from the general theory of efficient estimation as explained in the Appendix.

2) We shall give some corollary of Theorem 2.2 with an explicit construction of the preliminary estimator $\widehat{f}$ in the particular case where f is a density defined over a compact set of $\mathbb{R}^d$, for example $S = [0,1]^d$ and satisfying regularity conditions. When it will be useful, the function f will be extended by periodicity. The regularity conditions are defined as follows : let $r = (r_1, \ldots, r_d) \in \mathbb{N}^d$ and $\alpha = (\alpha_1, \ldots, \alpha_d) \in]0,1]^d$. Let D_j the derivation operator with respect to the j^{th} variable. We shall denote by $F_{r,\alpha,C}$ the set of densities f defined over $[0,1]^d$ such that

$$D_j^{r_j} f \quad \text{exists} \quad \forall\ j \in \{0, \ldots d\} \quad \text{and} \quad D_j^l f \quad \text{is periodic for} \quad l = 0 \ldots r_j.$$

$$\sup_{x,y\in S, x_j\neq y_j} \frac{|D_j^{r_j} f(x) - D_j^{r_j} f(y)|}{|x_j - y_j|^{\alpha_j}} \le C$$

We define $s_j = r_j + \alpha_j$ and s by $\frac{d}{s} = \sum_{j=1}^{d} \frac{1}{s_j}$. We will show in Section 4 that if f belongs to $F_{r,\alpha,C}$ then f belongs to some ellipsoid

$$\mathcal{E} = \{ \sum_{i \in \mathbb{Z}^d} a_i p_i; \quad \sum_{(i_1 \ldots i_d) \in \mathbb{Z}^d} (|i_1|^{2s'_1} + \ldots + |i_d|^{2s'_d}) |a_{i_1 \ldots i_d}|^2 \leq \gamma \}$$

for all $(s'_1, \ldots, s'_d)$ such that $\forall\ j \quad 0 < s'_j < s_j = r_j + \alpha_j$; for some $\gamma > 0$.

In this case, the preliminary estimator $\hat{f}$ is defined as follows : let $\tilde{f}$ be a kernel estimator of f based on the n_1 last observations where $n_1 \approx \frac{n}{\log n}$. Let $f_0 \in F_{r,\alpha,C}$ such that $f_0(S) \in [a - \epsilon, b + \epsilon]$ let $A_n = \{\tilde{f}(S) \subset [a - \epsilon, b + \epsilon]\}$ and

$$\hat{f}(x) = \tilde{f}(x)\ \mathbb{I}_{A_n} + f_0(x)\ \mathbb{I}_{A_n^c}. \tag{2.10}$$

Theorem 2.2, together with rates of convergence results by Ibragimov and Has'minskii ensuring A2 when $s > \frac{d}{4}$ (see Ibragimov and Has'minskii (1980) and (1981)), imply the following corollary:

Corollary 2.1 *Let $X_1, X_2, \ldots, X_n$ be i.i.d. d dimensional random variables with density f belonging to the set $F_{r,\alpha,C}$, $\quad r \in \mathbb{N}^d$, $\alpha \in]0,1]^d$. Let $s_j = r_j + \alpha_j$ and $\frac{d}{s} = \sum_{j=1}^{d} \frac{1}{s_j}$. Suppose that $s > \frac{d}{4}$. Let $(s'_1, \ldots, s'_d)$ be any element of $\mathbb{R}^d$ such that $\forall\ j \quad 0 < s'_j < s_j$ and such that s' defined by $\frac{d}{s'} = \sum_{j=1}^{d} \frac{1}{s'_j}$ satisfies $s' > \frac{d}{4}$. Let $(p_i)_{i \in \mathbb{Z}^d}$ be the orthonornal Fourier basis of $\mathbb{L}^2([0,1]^d)$, and $\hat{f}$ the estimator of f defined by (2.10). Let $M'_n = \{(i_1, \ldots, i_d) \in \mathbb{Z}^d; \forall\ j, |i_j| \leq m^{\frac{s'}{ds'_j}}\}$ and $m \approx n^{\frac{2d}{d+4s'}}$. $\widehat{T_n}$ is defined as in Theorem 2.2 with M'_n instead of M_n. Then*

$$\sqrt{n}(\widehat{T_n} - T(f)) \rightarrow \mathcal{N}(0, C(f, \phi))$$

and

$$\lim_{n \to \infty} nE(\widehat{T_n} - T(f))^2 = C(f, \phi)$$

Comment : We don't need the periodicity condition for the partial derivatives of f when $d = 1$ as will be shown in the next section.

2.3.2 Entropy estimation

As an example, let us give the precise shape of our estimator of $\int f \log(f)$. Condition A1 means in this case that $\forall\ x$, $0 < \epsilon \leq f(x) \leq b$ i.e. that f is bounded from below by some positive constant. Since f is a density, this condition implies that f is defined over a compact set. Hence, this estimator will not be suitable to test the normality but we may use it to test the uniformity of a density defined over $[0,1]$, (see Dudewicz and Van der Meulen (1981)).

Using Taylor expansion as above, we get :

$$\int f\log(f) = -\frac{1}{2}\int \widehat{f} + \int \log(\widehat{f})f + \frac{1}{2}\int \frac{f^2}{\widehat{f}} + \Gamma_n.$$

The general estimator proposed in Theorem 2.2 has the following expression :

$$\widehat{T_n} = -\frac{1}{2}\int \widehat{f} + \frac{1}{n_2}\sum_{l=1}^{n_2}\log \widehat{f}(X_l) + \frac{1}{n_2(n_2-1)}\sum_{l_1\neq l_2=1}^{n_2}\sum_{i\in M_n} p_i(X_{l_1})(p_i/\widehat{f})(X_{l_2})$$

$$-\frac{1}{2n_2(n_2-1)}\sum_{l_1\neq l_2=1}^{n_2}\sum_{i,i'\in M_n} p_i(X_{l_1})p_{i'}(X_{l_2})\int (p_ip_{i'}/\widehat{f})(x)dx.$$

If $f \in F_{r,\alpha,C}$ and $s > \frac{d}{4}$ then $\widehat{T_n}$ has the following properties :

$$\sqrt{n}(\widehat{T_n} - T(f)) \to \mathcal{N}\left(0, \int \log^2(f)f - \left(\int f\log(f)\right)^2\right)$$

$$\lim_{n\to\infty} nE(\widehat{T_n} - \int f\log(f))^2 = \int \log^2(f)f - \left(\int f\log(f)\right)^2.$$

2.4 Proofs

2.4.1 Proof of Theorem 2.1

We start from the usual decomposition :

$$E(\widehat{\theta} - \int f^2\psi)^2 = \text{Bias}^2(\widehat{\theta}) + \text{Var}(\widehat{\theta})$$

where $\text{Bias}(\widehat{\theta}) = E(\widehat{\theta}) - \theta$.
We shall write M instead of M_n for short.
$\widehat{\theta}$ has been constructed in such a way that $\text{Bias}(\widehat{\theta}) = -\int (S_Mf - f)^2\psi$. Hence,

$$|\text{Bias}(\widehat{\theta})| \leq \|\psi\|_\infty \int (S_Mf-f)^2 = \|\psi\|_\infty \sum_{i\notin M}|a_i|^2 \leq \|\psi\|_\infty \sup_{i\notin M}|c_i|^2 \quad \text{since } f\in\mathcal{E}.$$

Let us now evaluate the variance of our estimator. We recall that $\widehat{\theta} = 2\widehat{\theta}_1 - \widehat{\theta}_2$ where $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are respectively defined by (2.1) and (2.2). We notice that $\widehat{\theta}$ is a U-statistic, more precisely, we can write $\widehat{\theta}$ under the following form :

$$\widehat{\theta} = \frac{1}{n(n-1)} \sum_{j \neq k=1}^{n} h(X_j, X_k) \quad \text{where } h \text{ is symetric.}$$

Actually,

$$2\widehat{\theta}_1 = \frac{1}{n(n-1)} \sum_{j \neq k=1}^{n} H_1(X_j, X_k)$$

where $H_1(x,t) = (\psi(x) + \psi(t)) \sum_{i \in M} p_i(x) p_i(t)$.

Moreover,

$$\widehat{\theta}_2 = \frac{1}{n(n-1)} \sum_{j \neq k=1}^{n} H_2(X_j, X_k)$$

where $H_2(x,t) = \sum_{i,i' \in M} p_i(x) p_{i'}(t) \int p_i p_{i'} \psi$, and h is defined by $h(x,t) = H_1(x,t) - H_2(x,t)$.

It follows from Hoeffding's results concerning the computation of the variance of U-statistics (see Hoeffding (1948) or Serfling (1980) p. 183) that

$$\mathrm{Var}(\widehat{\theta}) = \frac{4(n-2)}{n(n-1)} \xi_1 + \frac{2}{n(n-1)} \xi_2$$

where

$$\xi_1 = \mathrm{Var}(h_1(X_1)), \quad \xi_2 = \mathrm{Var}(h(X_1, X_2))$$

and

$$h_1(x_1) = E(h(x_1, X_2)).$$

It is easy to see that

$$h_1(x_1) = (S_M f)\psi(x_1) + S_M(f\psi)(x_1) - S_M[(S_M f)\psi](x_1).$$

Let us first bound $\mathrm{Var}(\widehat{\theta})$.

$$\begin{aligned} \xi_1 &\leq E(h_1^2(X_1)) \\ &\leq 3E\left[((S_M f)\psi(X_1))^2 + (S_M(f\psi)(X_1))^2 + (S_M(\psi S_M f)(X_1))^2\right]. \end{aligned}$$

Our purpose is to show that ξ_1 is bounded by some constant depending only on $\|f\|_\infty$ and $\|\psi\|_\infty$.

$$E\left[((S_M f)\psi(X_1))^2\right] \leq \|\psi\|_\infty^2 \int (S_M f)^2 f$$

$$
\begin{aligned}
&\leq \|\psi\|_\infty^2 \|f\|_\infty \|f\|_2^2 \quad \text{since } S_M \text{ is a projection} \\
&\leq \|\psi\|_\infty^2 \|f\|_\infty^2 \quad \text{since } \int f^2 \leq \|f\|_\infty \int f.
\end{aligned}
$$

$$
E\left[(S_M(f\psi)(X_1))^2\right] = \int S_M^2(f\psi)f \leq \|f\|_\infty \|f\psi\|_2^2 \leq \|\psi\|_\infty^2 \|f\|_\infty^2.
$$

$$
\begin{aligned}
E\left[(S_M(\psi S_M f)(X_1))^2\right] &= \int S_M^2[(S_M f)\psi]f \\
&\leq \|f\|_\infty \|(S_M f)\psi\|_2^2 \leq \|f\|_\infty^2 \|\psi\|_\infty^2.
\end{aligned}
$$

Collecting the above evaluations, we obtain $\xi_1 \leq 9\|f\|_\infty^2 \|\psi\|_\infty^2$. We now bound ξ_2.

$$
\xi_2 = \mathrm{Var}[h(X_1, X_2)] \leq E[h^2(X_1, X_2)] \leq 2E\left[H_1^2(X_1, X_2) + H_2^2(X_1, X_2)\right].
$$

$$
\begin{aligned}
E(H_1^2(X_1, X_2)) &\leq 4E\left[\psi^2(X_1)(\sum_{i\in M} p_i(X_1)p_i(X_2))^2\right] \\
&\leq 4\int \psi^2(x)(\sum_{i\in M} p_i(x)p_i(t))^2 f(x)f(t)d\mu(x)d\mu(t) \\
&\leq 4\|f\|_\infty^2 \|\psi\|_\infty^2 \int\int (\sum_{i\in M} p_i(x)p_i(t))^2 d\mu(x)d\mu(t) \\
&\leq 4\|f\|_\infty^2 \|\psi\|_\infty^2 \int\int \sum_{i,i'\in M} p_i p_{i'}(x) p_i p_{i'}(t) d\mu(x)d\mu(t) \\
&\leq 4\|f\|_\infty^2 \|\psi\|_\infty^2 |M| \quad \text{by orthogonality.}
\end{aligned}
$$

We make the same type of computations for the term $E[H_2^2(X_1, X_2)]$.

$$
\begin{aligned}
E(H_2^2(X_1, X_2)) &= \int \left(\sum_{i,i'\in M} p_i(x)p_{i'}(t)\int p_i p_{i'}\psi\right)^2 f(x)f(t)d\mu(x)d\mu(t) \\
&\leq \|f\|_\infty^2 \int (\sum_{i,i'\in M} p_i(x)p_{i'}(t)\int p_i p_{i'}\psi)(\sum_{i_1,i'_1\in M} p_{i_1}(x)p_{i'_1}(t)\int p_{i_1}p_{i'_1}\psi)d\mu(x)d\mu(t) \\
&\leq \|f\|_\infty^2 \int \sum_{i,i'\in M} p_i^2(x)p_{i'}^2(t)(\int p_i p_{i'}\psi)^2 d\mu(x)d\mu(t) \quad \text{by orthogonality.} \\
&\leq \|f\|_\infty^2 \sum_{i,i'\in M} (\int p_i p_{i'}\psi)^2 \leq \|f\|_\infty^2 \int\int (\sum_{i\in M} p_i(x)p_i(t))^2 \psi(x)\psi(t)d\mu(x)d\mu(t) \\
&\leq \|f\|_\infty^2 \|\psi\|_\infty^2 \int\int (\sum_{i\in M} p_i(x)p_i(t))^2 d\mu(x)d\mu(t) \\
&\leq \|f\|_\infty^2 \|\psi\|_\infty^2 |M| \quad \text{by orthogonality.}
\end{aligned}
$$

Finally,

$$\xi_2 \leq 10\|f\|_\infty^2\|\psi\|_\infty^2|M|. \tag{2.11}$$

It follows from the above computations that

$$\mathrm{Var}(\widehat{\theta}) \leq \gamma\|f\|_\infty^2\|\psi\|_\infty^2[\frac{1}{n}+\frac{|M|}{n^2}] \quad \text{for some } \gamma \in \mathbb{R}.$$

We recall that $\mathrm{Bias}^2(\widehat{\theta}) \leq \|\psi\|_\infty^2 \left(\sup_{i\notin M}|c_i|^2\right)^2$.

By assumption, $\dfrac{(\sup_{i\notin M}|c_i|^2)^2}{|M|} \approx \dfrac{1}{n^2}$. Hence we get

- If $\frac{|M|}{n} \to 0$ then $E(\widehat{\theta}-\theta)^2 = O(\frac{1}{n})$
- else $E(\widehat{\theta}-\theta)^2 \leq \gamma_2\dfrac{|M|}{n^2}$ where γ_2 depends only on $\|f\|_\infty$ and $\|\psi\|_\infty$.

Let us now look more precisely at the semi-parametric case, that is the case where $E(\widehat{\theta}-\theta)^2 = O(\frac{1}{n})$ and show that the estimator is asymptotically normal. To do this, we will show that the expectation of the square of

$$R = \sqrt{n}\left(\widehat{\theta}-\theta-2(\frac{1}{n}\sum_{j=1}^{n}(f\psi)(X_j)-\int f^2\psi)\right)$$

converges towards zero. This will clearly imply (2.3).
We will prove (2.4) simultaneously.

$$E(R^2) = nE(\widehat{\theta}-\theta)^2 + \frac{4}{n}E\left[\left(\sum_{j=1}^{n} f\psi(X_j) - \int f^2\psi\right)^2\right]$$
$$-4E\left[(\widehat{\theta}-\theta)(\sum_{j=1}^{n} f\psi(X_j) - \int f^2\psi)\right]$$

We first compute $nE(\widehat{\theta}-\theta)^2$.

$$n|\mathrm{Bias}^2(\widehat{\theta})| \leq \lambda\|\psi\|_\infty^2\frac{|M|}{n} \quad \text{for some } \lambda \in \mathbb{R}.$$

We will now evaluate $n\mathrm{Var}(\widehat{\theta}) - 4[\int f^3\psi^2 - (\int f^2\psi)^2]$.

$$n\mathrm{Var}(\widehat{\theta}) = 4\xi_1 - \frac{4}{n-1}\xi_1 + \frac{2}{n-2}\xi_2.$$

ξ_1 is bounded by $9\|f\|_\infty^2\|\psi\|_\infty^2$ and from (2.11) we get $\frac{\xi_2}{(n-2)} \leq 20\|f\|_\infty^2\|\psi\|_\infty^2\frac{|M|}{n}$, to show (2.4) we just have to prove that

$$\left|\xi_1 - [\int f^3\psi^2 - (\int f^2\psi)^2]\right| \leq \gamma_1\,[\|S_M f - f\|_2 + \|S_M(f\psi) - f\psi\|_2]\,.$$

We recall that

$$\xi_1 = \mathrm{Var}\left((S_M f)\psi(X_1) + S_M(f\psi)(X_1) - S_M[(S_M f)\psi(X_1)]\right).$$

Denote

$$Y_1 = (S_M f)\psi(X_1), \quad Y_2 = S_M(f\psi)(X_1), \quad Y_3 = -S_M[(S_M f)\psi](X_1).$$

Then $\xi_1 = \sum_{i,j=1}^{3} \mathrm{Cov}(Y_i, Y_j)$. In fact, $\forall\ i,j \in \{1,2,3\}^2$, we claim that :

$$\left|\mathrm{Cov}(Y_i,Y_j) - \epsilon_{ij}[\int f^3\psi^2 - (\int f^2\psi)^2]\right| \leq \gamma_1\,[\|S_M f - f\|_2 + \|S_M(f\psi) - f\psi\|_2] \quad (2.12)$$

where $\epsilon_{ij} = -1$ if $(i,j) = (1,3),(2,3),(3,1)$ or $(3,2)$ and $\epsilon_{ij} = 1$ otherwise. (2.4) follows immediately from the above property. We give a complete proof for $i = j = 3$ since the computations are similar in the other cases.

$$\mathrm{Var}(Y_3) = \int S_M^2[(S_M f)\psi]f - \left(\int S_M[(S_M f)\psi]f\right)^2.$$

The computation will be done in two steps : we first bound the quantity

$$\left|\int S_M^2[(S_M f)\psi]f - \int f^3\psi^2\right|.$$

This expression is bounded by

$$\int |S_M^2[(S_M f)\psi]f - S_M^2(f\psi)f| + \int |S_M^2(f\psi)f - f^3\psi^2|$$

$$\begin{aligned}&\leq \|f\|_\infty\|S_M[(S_M f)\psi] + S_M(f\psi)\|_2\|S_M[(S_M f)\psi] - S_M(f\psi)\|_2\\&\quad+\|f\|_\infty\|S_M(f\psi) + f\psi\|_2\|S_M(f\psi) - f\psi\|_2\end{aligned}$$

by Cauchy-Schwarz inequality.
Using repeatedly the fact that since S_M is a projection, $\|S_M g\|_2 \leq \|g\|_2$ the sum is bounded by

$$\begin{aligned}&\|f\|_\infty\|(S_M f)\psi + f\psi\|_2\|(S_M f)\psi - f\psi\|_2 + 2\|f\|_\infty\|f\psi\|_2\|S_M(f\psi) - f\psi\|_2\\ \leq\ & 2\|f\|_\infty\|\psi\|_\infty^2\|f\|_2\|S_M f - f\|_2 + 2\|f\|_\infty\|\psi\|_\infty\|f\|_2\|S_M(f\psi) - f\psi\|_2.\end{aligned}$$

The second step consists in bounding the quantity

$$\left|\left(\int S_M[(S_M f)\psi]f\right)^2 - \left(\int f^2\psi\right)^2\right|.$$

This term is equal to

$$\left|\left(\int f(S_M[(S_M f)\psi] + f\psi)\right)\left(\int S_M[(S_M f)\psi]f - \int f^2\psi\right)\right|$$

which by Cauchy-Schwarz inequality is bounded by

$$(\|f\|_2\|S_M[(S_M f)\psi] + f\psi\|_2)\left(\int |S_M[(S_M f)\psi]f - S_M(f\psi)f| + \int |S_M(f\psi)f - f^2\psi|\right)$$

$$\begin{aligned}
&\leq \|f\|_2(\|(S_M f)\psi\|_2 + \|f\psi\|_2) \times (\|f\|_2\|S_M[(S_M f)\psi - f\psi]\|_2 + \|f\|_2\|S_M(f\psi) - f\psi\|_2) \\
&\leq 2\|f\|_2^2\|\psi\|_\infty(\|f\|_2\|(S_M f)\psi - f\psi\|_2 + \|f\|_2\|S_M(f\psi) - f\psi\|_2) \\
&\leq 2\|f\|_2^3\|\psi\|_\infty(\|\psi\|_\infty\|S_M f - f\|_2 + \|S_M(f\psi) - f\psi\|_2).
\end{aligned}$$

Collecting the above inequalities and noting that $\|f\|_2 \leq \|f\|_\infty^{1/2}$ we finally get (2.12) for $i = j = 3$, hence (2.4).

To show that $E(R^2) \to 0$, we have to compute now $\frac{4}{n}E\left(\sum_{j=1}^{n}(f\psi(X_j) - \int f^2\psi)\right)^2$. This quantity equals $4\left[\int f^3\psi^2 - (\int f^2\psi)^2\right]$.
At last, it remains to compute

$$-4E\left[(\hat{\theta} - \theta)(\sum_{j=1}^{n} f\psi(X_j) - \int f^2\psi)\right]$$

which is equal to

$$-4E\left(\hat{\theta}\sum_{j=1}^{n} f\psi(X_j)\right) + 4n\int f^2\psi E(\hat{\theta}).$$

We recall that $\hat{\theta}$ has been written under the following form

$$\hat{\theta} = \frac{1}{n(n-1)}\sum_{j\neq k=1}^{n} h(X_j, X_k) \quad \text{where h is symetric.}$$

Hence

$$\begin{aligned}\widehat{\theta}(\sum_{l=1}^{n} f\psi(X_l)) &= \frac{1}{n(n-1)} \sum_{j\neq k=1,l}^{n} h(X_j, X_k) f\psi(X_l) \\ &= \frac{2}{n(n-1)} \sum_{j\neq k=1}^{n} h(X_j, X_k) f\psi(X_j) + \frac{1}{n(n-1)} \sum_{j,k,l\neq}^{n} h(X_j, X_k) f\psi(X_l)\end{aligned}$$

This implies that

$$-4E\left(\widehat{\theta}\sum_{j=1}^{n} f\psi(X_j)\right) = -8E\left(h(X_j, X_k) f\psi(X_j)\right) - 4(n-2)E(\widehat{\theta})\int f^2\psi.$$

So,

$$-4E\left(\widehat{\theta}\sum_{j=1}^{n} f\psi(X_j)\right) + 4n\int f^2\psi E(\widehat{\theta}) = -8E\left(h(X_j, X_k) f\psi(X_j)\right) + 8E(\widehat{\theta})\int f^2\psi.$$

We have to determine the limit of this expression as $n \to \infty$.
As it was already proved $E(\widehat{\theta}) \to \theta$, so $8E(\widehat{\theta})\int f^2\psi \to 8(\int f^2\psi)^2$.
Let us now compute the other term.

$$-8E\left(h(X_j, X_k) f\psi(X_j)\right) = -8\int\int h(x,t) f^2(x)\psi(x) f(t) d\mu(x) d\mu(t).$$

We want to show that as $n \to \infty$ this term converges towards $-8\int f^3\psi^2$.
We recall that

$$h(x,t) = [\psi(x) + \psi(t)] \sum_{i\in M} p_i(x) p_i(t) - \sum_{i,i'\in M} p_i(x) p_{i'}(t) \int p_i p_{i'}\psi.$$

So, we have to compute

$$\begin{aligned} &- 8\int\int \sum_{i\in M} \psi(x) p_i(x) p_i(t) f^2(x)\psi(x) f(t) d\mu(x) d\mu(t) \\ &- 8\int\int \sum_{i\in M} \psi(t) p_i(x) p_i(t) f^2(x)\psi(x) f(t) d\mu(x) d\mu(t) \\ &+ 8\int\int \sum_{i,i'\in M} p_i(x) p_{i'}(t) f^2(x)\psi(x) f(t) d\mu(x) d\mu(t) \int p_i p_{i'}\psi \end{aligned}$$

which is equal to

$$-8\left[\int (S_M f)\psi^2 f^2 + \int S_M(f\psi) f^2\psi - \int (S_M f) S_M(f^2\psi)\psi\right].$$

Using the fact that $\psi \in \mathbb{L}^\infty$, that $f \in \mathbb{L}^\infty \cap \mathbb{L}^2$, and that $\forall\, g \in \mathbb{L}^2 \quad \|S_M g - g\|_2 \to 0$ as $n \to \infty$, we see that the above expression converges towards $-8\int f^3\psi^2$.
Finally, we get

$$\lim_{n\to\infty} -4E\left[(\widehat{\theta}-\theta)(\sum_{j=1}^{n} f\psi(X_j) - \int f^2\psi)\right] = -8\left[\int f^3\psi^2 - (\int f^2\psi)^2\right].$$

Hence, $E(R^2) \to 0$, this achieves the proof of Theorem 2.1.□

2.4.2 Proof of Theorem 2.2

We will first control the remainder term Γ_n.

$$|\Gamma_n| \leq \frac{1}{6}\|\phi_1^{(3)}\|_\infty \int |f-\widehat{f}|^3.$$

$\|\phi_1^{(3)}\|_\infty$ is finite so, $E(\Gamma_n^2) = O\left(E\left[(\int |f-\widehat{f}|^3)^2\right]\right) = O\left(E\left[\|\widehat{f}-f\|_3^6\right]\right)$.
Since $\widehat{f}$ is assumed to satisfy condition A_2 this quantity have order $O({n_1}^{-6\alpha})$ where $n_1 \approx \frac{n}{\log(n)}$ and $\alpha > \frac{1}{6}$. So,

$$E(\Gamma_n^2) = o(\frac{1}{n}).$$

This proves that the remainder term is negligible.
We are now going to prove the asymptotic efficiency. Let

$$R = \sqrt{n}\left[\widehat{T_n} - T(f) - \left(\frac{1}{n_2}\sum_{l=1}^{n_2}\phi_1'(f,.)(X_l) - \int \phi_1'(f,.)f\right)\right].$$

Of course, to ensure that both (2.8) and (2.9) hold, it is enough to show that $E(R^2) \to 0$.
We notice that $R = R_1 + R_2$ where :

$$R_1 = \sqrt{n}\left[\widehat{T_n} - T(f) - \left(\frac{1}{n_2}\sum_{l=1}^{n_2}\phi_1'(\widehat{f},.)(X_l) - \int \phi_1'(\widehat{f},.)f\right)\right]$$

$$R_2 = \sqrt{n}\left[\frac{1}{n_2}\sum_{l=1}^{n_2}\left(\phi_1'(\widehat{f},.)(X_l) - \int \phi_1'(\widehat{f},.)f\right)\right] - \sqrt{n}\left[\frac{1}{n_2}\sum_{l=1}^{n_2}\left(\phi_1'(f,.)(X_l) - \int \phi_1'(f,.)f\right)\right].$$

We shall prove that both $E(R_1^2)$ and $E(R_2^2) \to 0$. Pluging the expression of $\widehat{T_n}$ and $T(f)$ in R_1 we get

$$R_1 = \sqrt{n}[\widehat{L'} - L' + \widehat{Q} - Q + \Gamma_n]$$

where

$$L' = -\int \widehat{f}\phi_1''(\widehat{f},.)f; \quad \widehat{L'} = -\frac{1}{n_2}\sum_{j=1}^{n_2} \widehat{f}\phi_1''(\widehat{f},.)(X_j); \quad Q = \frac{1}{2}\int \phi_1''(\widehat{f},.)f^2$$

and $\widehat{Q}$ is the corresponding estimator. Since $E(\Gamma_n^2) = o(\frac{1}{n})$, we just have to control the expectation of the square of $\sqrt{n}[\widehat{L'} - L' + \widehat{Q} - Q]$.

• Computation of $\lim_{n\to\infty} nE(\widehat{L'} - L')^2$:

$$nE\left[(\widehat{L'} - L')^2|\widehat{f}\right] = \frac{n}{n_2}\left[\int (\widehat{f}\phi_1''(\widehat{f},.))^2 f - \left(\int \widehat{f}\phi_1''(\widehat{f},.)f\right)^2\right].$$

$\frac{n}{n_2} \to 1$ as $n \to \infty$. Moreover, we will show that the expectation of the above expression converges towards the same expression with f instead of $\widehat{f}$. We shall only give the proof for the term $\int (\widehat{f}\phi_1''(\widehat{f},.))^2 f$.

$$|E\left(\int (\widehat{f}\phi_1''(\widehat{f},.))^2 f - \int f^3(\phi_1'')^2(f,.)\right)|$$

$$\leq \|f\|_\infty E\left(\int |\widehat{f}\phi_1''(\widehat{f},.) + f\phi_1''(f,.)||\widehat{f}\phi_1''(\widehat{f},.) - f\phi_1''(f,.)|\right).$$

We recall that $a \leq f(x) \leq b$ and $a - \epsilon \leq \widehat{f}(x) \leq b + \epsilon$, for some $\epsilon > 0$, so the above difference is bounded by

$$\|f\|_\infty \left[2(\sup(|a|,|b|) + \epsilon)\|\phi_1''\|_\infty\right] E\left(|\widehat{f}\phi_1''(\widehat{f},.) - f\phi_1''(f,.)|\right).$$

Finally,

$$E\left(|\widehat{f}\phi_1''(\widehat{f},.) - f\phi_1''(f,.)|\right)$$

$$\leq E\left(|\widehat{f}\phi_1''(\widehat{f},.) - \widehat{f}\phi_1''(f,.)|\right) + E\left(|\widehat{f}\phi_1''(f,.) - f\phi_1''(f,.)|\right)$$

Since $E(\|\widehat{f} - f\|_1) \to 0$ and since $\|\phi_1^{(3)}\|_\infty$ and $\|\phi_1''\|_\infty$ are finite, each of these terms converges towards zero and we get

$$\lim_{n\to\infty} E(\widehat{L'} - L')^2 = \int (\phi_1''(f,.))^2 f^3 - (\int \phi_1''(f,.)f^2)^2.$$

• Computation of $\lim_{n\to\infty} nE(\widehat{Q} - Q)^2$:

Denote $\psi = \frac{1}{2}\phi_1''(\widehat{f},.)$. It follows from (2.4) that for $\frac{|M_n|}{n} \to 0$,

$$\left| nE[(\hat{Q}-Q)^2|\hat{f}] - \left[\int (\phi_1''(\hat{f},.))^2 f^3 - (\int \phi_1''(\hat{f},.)f^2)^2\right]\right|$$

$$\leq \gamma_1(\|f\|_\infty, \|\psi\|_\infty)\left[\frac{|M|}{n} + \|S_M f - f\|_2 + \|S_M(f\phi_1''(\hat{f},.)) - f\phi_1''(\hat{f},.)\|_2\right].$$

Hence,

$$\left| nE[(\hat{Q}-Q)^2] - E\left[\int (\phi_1''(\hat{f},.))^2 f^3 - (\int \phi_1''(\hat{f},.)f^2)^2\right]\right|$$

$$\leq \gamma_1(\|f\|_\infty, \|\psi\|_\infty)\left[\frac{|M|}{n} + \|S_M f - f\|_2 + E\left(\|S_M(f\phi_1''(\hat{f},.)) - f\phi_1''(\hat{f},.)\|_2\right)\right].$$

We have to prove that

$$E\left(\|S_M(f\phi_1''(\hat{f},.)) - f\phi_1''(\hat{f},.)\|_2\right) \to 0 \quad \text{as} \quad n \to \infty.$$

Since, by similar arguments as above, the expectation of the term $\int (\phi_1''(\hat{f},.))^2 f^3 - (\int \phi_1''(\hat{f},.)f^2)^2$ converges as $n \to \infty$ towards $\int (\phi_1''(f,.))^2 f^3 - (\int \phi_1''(f,.)f^2)^2$, this will imply that $nE(\hat{Q}-Q)^2$ converges towards the same limit.

$$E\left(\|S_M(f\phi_1''(\hat{f},.)) - f\phi_1''(\hat{f},.)\|_2\right) \leq E\left(\|S_M(f\phi_1''(\hat{f},.)) - S_M(f\phi_1''(f,.)\|_2\right)$$

$$+\|S_M(f\phi_1''(f,.) - f\phi_1''(f,.)\|_2 + E\left(\|f\phi_1''(f,.) - f\phi_1''(\hat{f},.)\|_2\right)$$

$$\leq 2E\left(\|f\phi_1''(f,.) - f\phi_1''(\hat{f},.)\|_2\right) + \|S_M(f\phi_1''(f,.) - f\phi_1''(f,.)\|_2$$

since S_M is a projection.
The second term converges towards zero since $f\phi_1''(f,.) \in \mathbb{L}^2(d\mu)$ and since $\forall\, g \in \mathbb{L}^2(d\mu)$ $\|S_M g - g\|_2 \to 0$ moreover the first term converges also towards zero by the arguments already used above.

To complete the proof we shall show that the sum of the two terms we have just computed is the opposite of the covariance term.

- Computation of $\lim_{n\to\infty} 2nE(\widehat{L'} - L')(\hat{Q}-Q)$:

Since conditionally to $\hat{f}$, $\widehat{L'}$ is an unbiased estimator of L' we get

$$E[(\widehat{L'} - L')(\hat{Q}-Q)|\hat{f}] = E(\widehat{L'}\hat{Q}|\hat{f}) - L'E(\hat{Q}|\hat{f}).$$

We recall that $\hat{Q} = 2\widehat{Q_1} - \widehat{Q_2}$ where

$$\widehat{Q_1} = \frac{1}{n_2(n_2-1)} \sum_{i\in M} \sum_{j\neq k=1}^{n_2} p_i(X_j) \frac{p_i\phi_1''(\widehat{f},.)}{2}(X_k)$$

$$\widehat{Q_2} = \frac{1}{n_2(n_2-1)} \sum_{i,i'\in M} \sum_{j\neq k=1}^{n_2} p_i(X_j)p_{i'}(X_k) \int p_ip_{i'}\frac{\phi_1''(\widehat{f},.)}{2}(x)dx.$$

We shall first develop the calculations for $\widehat{Q_1}$.

$$E(\widehat{L'}\widehat{Q_1}|\widehat{f}) = -\frac{1}{n_2}\sum_{i\in M}\int \widehat{f}\phi_1''(\widehat{f},.)fp_i \int \phi_1''(\widehat{f},.)\frac{fp_i}{2}$$

$$-\frac{1}{n_2}\sum_{i\in M} a_i \int \widehat{f}(\phi_1''(\widehat{f},.))^2\frac{fp_i}{2} + \left(1-\frac{2}{n_2}\right) L'E(\widehat{Q_1}|\widehat{f}).$$

So

$$E(\widehat{L'}\widehat{Q_1}|\widehat{f}) - L'E(\widehat{Q_1}|\widehat{f}) = -\frac{1}{2n_2}\int S_M[f\phi_1''(\widehat{f},.)]\widehat{f}\phi_1''(\widehat{f},.)f$$

$$-\frac{1}{2n_2}\int (S_Mf)\widehat{f}(\phi_1''(\widehat{f},.))^2 f - \frac{2}{n_2}L'E(\widehat{Q_1}|\widehat{f}).$$

Let us first show that

$$E\left(S_M[f\phi_1''(\widehat{f},.)]\widehat{f}\phi_1''(\widehat{f},.)f\right) \to \int f^3(\phi_1''(f,.))^2,$$

of course, an analogous result is valid for the second term. We shall first show that the expectation of

$$\int S_M[f\phi_1''(\widehat{f},.)]\left[\widehat{f}\phi_1''(\widehat{f},.)f - f^2\phi_1''(f,.)\right] \to 0.$$

This integral is bounded by

$$\|S_M(f\phi_1''(\widehat{f},.))\|_2\|\widehat{f}\phi_1''(\widehat{f},.) - f\phi_1''(f,.)\|_2\|f\|_\infty.$$

Moreover,

$$\|S_M(f\phi_1''(\widehat{f},.))\|_2 \leq \|f(\phi_1'')(\widehat{f},.)\|_2 \leq \|f\|_2\|\phi_1''\|_\infty \leq \|f\|_\infty^{1/2}\|\phi_1''\|_\infty$$

and as it was already shown, the expectation of the term $\|\widehat{f}\phi_1''(\widehat{f},.) - f\phi_1''(f,.)\|_2$ converges towards zero. It remains to show that the expectation of

$$\int \left(S_M[f\phi_1''(\widehat{f},.)] - f\phi_1''(f,.)\right) f^2\phi_1''(f,.)$$

converges towards zero. This expression is bounded by

$$\|\phi_1''f\|_\infty\|f\|_2\|S_M[f\phi_1''(\hat{f},.)] - f\phi_1''(f,.)\|_2.$$

So, we just have to prove that

$$E\left(\|S_M[f\phi_1''(\hat{f},.)] - f\phi_1''(f,.)\|_2\right) \to 0.$$

This quantity is bounded by

$$E\left(\|S_M[f\phi_1''(\hat{f},.)] - S_M[f\phi_1''(f,.)]\|_2\right) + \|S_M[f\phi_1''(f,.)] - f\phi_1''(f,.)\|_2.$$

$$\leq E\left(\|f\phi_1''(\hat{f},.) - f\phi_1''(f,.)\|_2\right) + \|S_M(f\phi_1''(f,.)) - f\phi_1''(f,.)\|_2.$$

Since each of these terms converges towards zero, we get

$$E\left(S_M[f\phi_1''(\hat{f},.)]\hat{f}\phi_1''(\hat{f},.)f\right) \to \int f^3(\phi_1''(f,.))^2.$$

We will now prove that

$$\lim_{n\to\infty} E\left(-2L'E(\widehat{Q}_1|\hat{f})\right) = (\int \phi_1''(f,.)f^2)^2.$$

We have to show that

$$E\left[\int \hat{f}\phi_1''(\hat{f},.)f\int(S_Mf)\phi_1''(\hat{f},.)f\right] - (\int \phi_1''(f,.)f^2)^2 \to 0.$$

The above difference is equal to :

$$E\left[\int \hat{f}\phi_1''(\hat{f},.)f\left(\int(S_Mf)\phi_1''(\hat{f},.)f - \int \phi_1''(f,.)f^2\right)\right]$$
$$+E\left[\int \phi_1''(f,.)f^2\left(\hat{f}\phi_1''(\hat{f},.)f - \int \phi_1''(f,.)f^2\right)\right].$$

The first part of the sum is bounded by

$$\|\hat{f}\|_\infty\|\phi_1''\|_\infty E\left[\int f\left|(S_Mf)\phi_1''(\hat{f},.) - f\phi_1''(f,.)\right|\right]$$

$$\leq \|\hat{f}\|_\infty\|\phi_1''\|_\infty\|f\|_2 E\left[\|(S_Mf)\phi_1''(\hat{f},.) - f\phi_1''(f,.)\|_2\right]$$

$$= O\left[E\left(\|(S_Mf)\phi_1''(\hat{f},.) - f\phi_1''(\hat{f},.)\|_2 + \|f\phi_1''(\hat{f},.) - f\phi_1''(f,.)\|_2\right)\right]$$

$$= O\left[\|\phi_1''\|_\infty\|S_Mf - f\|_2 + \|f\|_\infty E\left(\|\phi_1''(\hat{f},.) - \phi_1''(f,.)\|_2\right)\right]$$

and each of this term converges towards zero. It follows that

$$\lim_{n\to\infty} nE(\widehat{L'} - L')(\widehat{Q_1} - Q_1) = -\int (\phi_1''(f,.))^2 f^3 + (\int \phi_1''(f,.)f^2)^2.$$

Since $\widehat{Q} = 2\widehat{Q_1} - \widehat{Q_2}$, to show that the above result holds with $\widehat{Q}$ instead of $\widehat{Q_1}$, we have to prove it for $\widehat{Q_2}$.

$$nE(\widehat{L'}-L')(\widehat{Q_2}-Q_2) = \frac{n}{n_2}E\left[-\int S_M[(S_M f)\phi_1''(\hat{f},.)]\hat{f}\phi_1''(\hat{f},.)f + \int (S_M f)^2\phi_1''(\hat{f},.)\int f\phi_1''(\hat{f},.)\hat{f}\right].$$

We first want to prove that

$$E\left[\int S_M[(S_M f)\phi_1''(\hat{f},.)]\hat{f}\phi_1''(\hat{f},.)f\right] \to \int f^3(\phi_1''(f,.))^2.$$

The difference between these terms is

$$\begin{aligned}
& E\left\{\int S_M[(S_M f)\phi_1''(\hat{f},.)]\left(\hat{f}\phi_1''(\hat{f},.)f - f^2\phi_1''(f,.)\right)\right\} \\
+ \; & E\left\{\int f^2\phi_1''(f,.)\left(S_M[(S_M f)\phi_1''(\hat{f},.)] - S_M(f\phi_1''(f,.))\right)\right\} \\
+ \; & E\left\{\int f^2\phi_1''(f,.)\left(S_M(f\phi_1''(f,.)) - f\phi_1''(f,.)\right)\right\}.
\end{aligned}$$

Using the fact that S_M is a projection, this sum is bounded by

$$\begin{aligned}
\leq \; & \|S_M f\|_2 \|\phi_1''\|_\infty E\left(\|\hat{f}\phi_1''(\hat{f},.) - f\phi_1''(f,.)\|_2\right)\|f\|_\infty \\
+ \; & \|f\|_\infty \|\phi_1''\|_\infty \|f\|_2 E\left(\|(S_M f)\phi_1''(\hat{f},.) - f\phi_1''(f,.)\|_2\right) \\
+ \; & \|f\|_\infty \|\phi_1''\|_\infty \|f\|_2 \left(\|S_M(f\phi_1''(f,.)) - f\phi_1''(f,.)\|_2\right).
\end{aligned}$$

We already proved that the first two terms converge towards zero, as to the third term, it also converge towards zero since $f\phi_1''(f,.) \in \mathbb{L}^2(d\mu)$.
To conclude the proof, it remains to show that

$$E\left[\int (S_M f)^2\phi_1''(\hat{f},.)\int f\phi_1''(\hat{f},.)\hat{f}\right] \to \left(\int f^2\phi_1''(f,.)\right)^2.$$

The difference between these two terms is

$$\begin{aligned}
& E\left\{\int (S_M f)^2\phi_1''(\hat{f},.)\left[\int f\phi_1''(\hat{f},.)\hat{f} - \int f^2\phi_1''(f,.)\right]\right\} \\
+ \; & E\left\{\int f^2\phi_1''(f,.)\left[\int (S_M f)^2\phi_1''(\hat{f},.) - \int f^2\phi_1''(f,.)\right]\right\}.
\end{aligned}$$

The first term is bounded by

$$\|\phi_1''\|_\infty \|f\|_3^2 E\left[\|\phi_1''(\hat{f},.)\hat{f} - \int f\phi_1''(f,.)\|_2\right]$$

which converges towards zero as proved above in the computations for the term $\widehat{Q_1}$. The expectation of the second term is majorized by

$$\|\phi_1''\|_\infty\|f\|_2^2 E\left[\int f^2\left|\phi_1''(\hat{f},.)-\phi_1''(f,.)\right|+\int\left|\phi_1''(f,.)(f^2-(S_Mf)^2)\right|\right]$$

$$\begin{aligned}&\leq \|\phi_1''\|_\infty\|f\|_2^2\left[\|f\|_\infty\|f\|_2 E\left(\|\phi_1''(\hat{f},.)-\phi_1''(f,.)\|_2\right)\right]\\&+ \|\phi_1''\|_\infty^2\|f\|_2^2\|f+S_Mf\|_2\|f-S_Mf\|_2.\end{aligned}$$

Of course both terms converge towards zero. We now conclude that

$$\lim_{n\to\infty} 2nE(\widehat{L'}-L')(\hat{Q}-Q)=-2\int(\phi_1''(f,.))^2f^3+2(\int\phi_1''(f,.)f^2)^2$$

which implies that

$$\lim_{n\to\infty} E(R_1^2)=0.$$

It remains to prove that $E(R_2^2)\to 0$.

$$\begin{aligned}E(R_2^2)&=\frac{n}{n_2}E\left[\int\left(\phi_1'(\hat{f},.)-\phi_1'(f,.)\right)^2 f\right]\\&-\frac{n}{n_2}E\left[\int\phi_1'(\hat{f},.)f-\int\phi_1'(f,.)f\right]^2.\end{aligned}$$

Using as before the fact that $E(\|\hat{f}-f\|_1)\to 0$ and that $\|\phi_1'\|_\infty$ is bounded, it is easy to see that $E(R_2^2)\to 0$. This achieves the proof of Theorem 2.2. □

2.4.3 Proof of Corollary 2.1

Let us first show that condition A2 is verified by the preliminary estimator $\hat{f}$ defined by (2.10) as soon as $s>\frac{d}{4}$. This estimator is based on the kernel estimator $\tilde{f}$ studied by Ibragimov and Has'minskii, who showed the following properties for $\tilde{f}$:

$$\sup_{f\in F_{r,\alpha,C}} E_f(\|\tilde{f}-f\|_q^l)\leq A_1(q,l)n_1^{\frac{-ls}{d+2s}}\quad\forall\ 2\leq q<+\infty,\forall\ l\in\mathbb{N}^*$$

$$\sup_{f\in F_{r,\alpha,C}} E_f(\|\tilde{f}-f\|_\infty^l)\leq A_2(l)\left(\frac{n_1}{\log(n_1)}\right)^{\frac{-ls}{d+2s}}\qquad\forall\ l\in\mathbb{N}^*.$$

We recall that $\hat{f}=\tilde{f}\ \mathbb{1}_{A_n}+f_0\ \mathbb{1}_{A_n^c}$ where $A_n=\{\tilde{f}\subset[a-\epsilon,b+\epsilon]\}$.

$$E(\|\hat{f}-f\|_q^l)\leq E(\|\tilde{f}-f\|_q^l)+\|f_0-f\|_q^l P(A_n^c).$$

Hence, to prove condition A2, it is enough to show that $P(A_n^c)$ is small enough. We notice that $A_n^c \subset \{\|\tilde{f} - f\|_\infty \geq \epsilon\}$. Since

$$E_f\|\tilde{f} - f\|_\infty^{l'} \leq A_2(l')\left(\frac{n_1}{\log(n_1)}\right)^{\frac{-l's}{d+2s}} \quad \forall\, l' \in \mathbb{N}^*$$

we get

$$P(A_n^c) \leq P(\|\tilde{f} - f\|_\infty \geq \epsilon) \leq \frac{1}{\epsilon^{l'}} E_f(\|\tilde{f} - f\|_\infty^{l'})$$

So,

$$P(A_n^c) \leq A_3(l')\left(\frac{n_1}{\log(n_1)}\right)^{\frac{-l's}{d+2s}} \quad \forall\, l' \in \mathbb{N}^*.$$

For n_1 large enough,

$$P(A_n^c) \leq A_4(l) n_1^{\frac{-ls}{d+2s}}.$$

It follows that

$$E_f(\|\hat{f} - f\|_q^l) \leq C(q,l) n_1^{\frac{-ls}{d+2s}}.$$

This achieves the proof of condition A_2 since when $s > \frac{d}{4}$, $\frac{s}{d+2s} > \frac{1}{6}$.
Let us now prove the following lemma

Lemma 2.1 *Let $f \in F_{r,\alpha,C}$ with $r \in \mathbb{N}^d$ and $\alpha \in]0,1]^d$. Then f belongs to the ellipsoid*

$$\mathcal{E} = \{\sum_{i\in\mathbb{Z}^d} a_i p_i;\ \sum_{(i_1\ldots i_d)\in\mathbb{Z}^d} (|i_1|^{2s'_1} + \ldots + |i_d|^{2s'_d})|a_{i_1\ldots i_d}|^2 \leq \gamma\}$$

for all $(s'_1,\ldots,s'_d)$ such that $\forall\, j \quad 0 < s'_j < s_j = r_j + \alpha_j$, for some γ independent of $f \in F_{r,\alpha,C}$ and where $(p_i)_{i\in\mathbb{Z}^d}$ denotes the Fourier orthonornal basis of $\mathbb{L}^2([0,1]^d)$.

We have to show that $\forall\, j \in \{1,\ldots d\}$ the quantity $\sum_{(i_1\ldots i_d)\in\mathbb{Z}^d} (|i_j|^{2s'_j})|a_{i_1\ldots i_d}|^2$ is bounded by some constant γ_j independent of $f \in F_{r,\alpha,C}$. Since $D_j^l f$ is periodic for $l = 0\ldots r_j$ it is easy to see by integration by parts that

$$|\int (D_j^{r_j} f) p_i| = (2\pi)^{r_j}|i_j|^{r_j}|\int f p_i| = (2\pi)^{r_j}|i_j|^{r_j}|a_i|.$$

Using this remark, it is enough to show that if f satisfies the condition

$$|f(x_1,\ldots,x_j+h,..x_d) - f(x_1,\ldots,..x_d)| \leq C|h|^{\alpha_j}$$

then

$$\forall\, j \in \{1,\dots,d\} \quad \sum_{(i_1\dots i_d)\in\mathbb{Z}^d} (|i_j|^{2\beta_j})|a_{i_1\dots i_d}|^2 \le \gamma_j \quad \forall\, 0<\beta_j<\alpha_j.$$

The Fourier expansion of $f(x_1,\dots,x_j+h,..x_d)-f(x_1,\dots,x_j-h,..x_d)$ is

$$\sum_{(i_1,\dots i_d)\in\mathbb{Z}^d} a_{i_1\dots i_d} e^{2i\pi(i_1x_1+\dots i_dx_d)} 2i\sin(2\pi i_j h).$$

Hence

$$\begin{aligned}\int (f(x_1,\dots,x_j+h,..x_d)-f(x_1,\dots,x_j-h,..x_d))^2 &= 4\sum_{(i_1,\dots i_d)\in\mathbb{Z}^d} |a_{i_1\dots i_d}|^2\sin^2(2\pi i_j h)\\ &\le C^2|2h|^{2\alpha_j}.\end{aligned}$$

We shall denote by $(i_1,\dots\widehat{i_j},\dots i_d)$ the element $(i_1,\dots i_{j-1},i_{j+1},\dots i_d)$ of $\mathbb{Z}^{d-1}$. Let $q\in\mathbb{N}^*$ and $h=\frac{1}{8q}$. Then, $\forall\, i_j\in\{q,\dots,2q-1\}$ $\sin^2(2\pi i_j h)\ge\frac{1}{2}$ and

$$\sum_{i_j=q}^{2q-1}\ \sum_{(i_1,\dots\widehat{i_j},\dots id)\in\mathbb{Z}^{d-1}} |a_{i_1\dots i_d}|^2 \le \frac{C^2}{2}\frac{1}{(8q)^{2\alpha_j}}$$

By similar arguments,

$$\sum_{i_j=-2q+1}^{-q}\ \sum_{(i_1,\dots\widehat{i_j},\dots id)\in\mathbb{Z}^{d-1}} |a_{i_1\dots i_d}|^2 \le \frac{C^2}{2}\frac{1}{(8q)^{2\alpha_j}}$$

Let $0<\beta_j<\alpha_j$

$$\sum_{i_j=q}^{2q-1}\ \sum_{(i_1,\dots\widehat{i_j},\dots id)\in\mathbb{Z}^{d-1}} |i_j|^{2\beta_j}|a_{i_1\dots i_d}|^2 \le \frac{C^2}{2}\frac{(2q)^{2\beta_j}}{(8q)^{2\alpha_j}}$$

Now, let $q=2^l$,

$$\sum_{(i_1,\dots,i_d)\in\mathbb{Z}^d} |a_{i_1\dots i_d}|^2 \le C^2\frac{4^{\beta_j}}{8^{2\alpha_j}}\sum_{l=0}^{+\infty}\left(\frac{1}{4^{\alpha_j-\beta_j}}\right)^l.$$

This achieves the proof of the lemma, and we can therefore apply the results of Theorem 2.2 for some well chosen $(s'_1,\dots,s'_j)$. □

Comment : of course, if $d=1$ we don't need to assume that f is periodic. Actually, we use the same proof as above (except that we do not integrate by parts) to show that, for some $\alpha>\frac{1}{4}$ f belongs to the ellipsoid

$$\mathcal{E}=\{\sum a_ip_i, \sum_{i\in\mathbb{Z}}|i|^{2s}|a_i|^2\le\gamma\} \quad \text{for some constant } \gamma.$$

2.5 Appendix

2.5.1 Semi-parametric information bound

A) We first suppose that f is a regular function belonging to the set $F_{r,\alpha,C}$ defined in Section 3, we want to show that the asymptotic variance appearing in Corollary 2.1 is optimal. To do this, we will apply the results of Koshevnik and Levit (1976), see also Ibragimov and Hasminskii (1991). We also refer to Levit (1978) where an application of Koshevnik and Levit (1976) is given, for a problem which is almost similar to our problem.

We suppose that f belongs to the class $F_{r,\alpha,C}$ and satisfies

$$\sup_{x_j \neq y_j} \frac{|D_j^{r_j} f(x) - D_j^{r_j} f(y)|}{|x_j - y_j|^{\alpha_j}} < C.$$

Let ξ a bounded function, infinitely differentiable, such that $\int f\xi = 0$, and let

$$f_t = f(1 + t\xi) \quad \text{where } t \in \mathbb{R}.$$

For small t, $f_t \in F_{r,\alpha,C}$ and is a density.

$$\begin{aligned} T(f_t) - T(f) &= \int \phi_1'(f,.)(f_t - f) + o(t) \\ &= \int \left(\phi_1'(f,.) - \int \phi_1'(f,.)f\right)(f_t - f) + o(t) \end{aligned}$$

since $\int (f_t - f) = 0$. Hence,

$$T(f_t) - T(f) = \int \left(\phi_1'(f,.) - \int \phi_1'(f,.)f\right)(\sqrt{f_t} - \sqrt{f})(\sqrt{f_t} + \sqrt{f}) + o(t).$$

Using the notations of Koshevnik and Levit (1976), p.741, we get

$$\|\sqrt{f_t} - \sqrt{f} - At\|_2 = o(t)$$

where

$$At = \frac{t\xi}{2}\sqrt{f}.$$

Moreover,

$$T(f_t) - T(f) = Bt + o(t)$$

where

$$Bt = E_f(G\frac{\xi}{2}t)$$

and

$$G = 2\left(\phi_1'(f,.) - \int \phi_1'(f,.)f\right).$$

G is called the canonical gradient of $T(f)$, it satisfies $E_f(G) = 0$.
Noting that functions ξ are dense in the set of functions g of $\mathbb{L}^2(d\mu)$ satisfying $E_f(G) = 0$, we can find a sequence ξ_n such that

$$\lim_{n\to\infty} \int (\frac{\xi_n(x)}{2} - G)^2 f(x)dx = 0.$$

This ensures the condition 8 in Koshevnik and Levit (1976) since all the functions $f_{t_n} = f(1 + t\xi_n)$ belong to $F_{r,\alpha,C}$ for small t.
According to Theorem 1 and Theorem 2 of Koshevnik and Levit (1976), we see that

$$\inf_\epsilon \liminf_{n\to\infty} \sup_{f_1\in F_{r,\alpha,C}, \|f_1 - f\|_2 \le \epsilon} nE(\widehat{T_n} - T(f))^2 \ge \frac{1}{4} E_f(G^2).$$

The right-hand term of the above inequality is equal to $\int (\phi_1'(f,.))^2 f - (\int \phi_1'(f,.)f)^2$ which achieves the proof.
B) We now assume that f belongs to the ellipsoid $\mathcal{E}$, which is more general as above, but f is supposed to be bounded from below by some positive constant ϵ. In fact we assume that f satisfies

$$\sum_{i\in D} \frac{a_i^2}{c_i^2} < 1.$$

We suppose that the orthonormal basis $(p_i)_{i\in D}$ of $\mathbb{L}^2(d\mu)$ is such that $\|p_i\|_\infty < +\infty$, and, (except the first function which is usually constant and will not be used in the following) that $\int p_i = 0$. Let

$$f_t = f + t(\sum_{i\in I} c_i p_i), \quad t, c_i \in \mathbb{R}, |I| \in \mathbb{N}.$$

For small t, f_t is a density since $f > \epsilon > 0$, and belongs to the ellipsoid $\mathcal{E}$ since $|I|$ is finite.

$$\|\sqrt{f_t} - \sqrt{f} - A't\|_2 = o(t)$$

where

$$A't = \frac{t(\sum_{i\in I} c_i p_i)}{2f}\sqrt{f}.$$

Moreover,

$$T(f_t) - T(f) = Bt + o(t)$$

where

$$Bt = E_f\left(G'\frac{(\sum_{i\in I} c_i p_i)}{2f}t\right)$$

and $G' = 2\left(\phi_1'(f,.) - \int \phi_1'(f,.)f\right)$; $E_f(G') = 0$. To conclude the proof, we remark that the functions of the type $\sum_{i\in I} c_i p_i$ are dense in the set of functions $g \in \mathbb{L}^2(d\mu)$ satisfying $\int g = 0$. Hence, there exists a sequence ξ_n of functions of the type $\sum_{i\in I} c_i p_i$ such that

$$\lim_{n\to\infty} \int \left(\frac{\xi_n}{2f} - G'\right)^2 f = 0$$

since f is bounded from below by $\epsilon > 0$. The end of the proof is similar as above.□

Chapitre 3

Estimation of integral functionals of a density and its derivatives

Abstract

We consider the problem of estimating a functional of a density of the type $\int \phi(f, f', \ldots, f^{(k)}, .)$. The estimation of $\int \phi(f, .)$ has already been studied by the author : starting from efficient estimators of linear and quadratic functionals of f and its derivatives and using a Taylor expansion of ϕ, we build estimators which achieve the $n^{-1/2}$ rate whenever f is smooth enough. Moreover, we show that these estimators are efficient. We also get the optimal rate of convergence when the $n^{-1/2}$ rate is not achievable and when $k > 0$. Concerning the estimation of quadratic functionals, more precisely of integrated squared density derivatives, Bickel and Ritov have already built efficient estimators, we propose here an alternative construction based on projections, which seems more natural.

3.1 Introduction

Let $X_1, \ldots, X_n$ be i.i.d. random variables with common density f defined over a compact set S of $\mathbb{R}$. Our purpose is to estimate quantities of the type $T(f) = \int \phi(f, f', \ldots, f^{(k)}, .)$ when f is assumed to belong to some Sobolev space of index $s > k$. This problem was first studied by Levit (1978) who built efficient estimators of this kind of functionals under regularity properties for the density f which are not optimal. It is also worth mentioning the paper by Ibragimov, Nemirovskii and Hasminskii (1987) which deals with differentiable functionals in gaussian white noise.

The problem of estimating $\int (f^{(k)})^2$ has already been studied by Bickel and Ritov (1988) and by Donoho and Nussbaum (1990). The same results are obtained in these papers : if $s > 2k + \frac{1}{4}$, $\int (f^{(k)})^2$ may be estimated at the semiparametric rate $\frac{1}{\sqrt{n}}$; if $s \leq 2k + \frac{1}{4}$, the rate of convergence have order $n^{\frac{-4(s-k)}{1+4s}}$. Moreover, Bickel and Ritov show that their estimator is efficient and that the nonparametric rate of convergence is optimal.

In this paper we generalize the results obtained in the previous chapter and concerning the problem of estimating $\int \phi(f, .)$. Before stating the results, let us explain how our methods work : ϕ is assumed to be a smooth function. So, expanding ϕ up to the second order with Taylor's formula provides an expansion of $T(f) - T(\widehat{f})$, where $\widehat{f}$ is a nonparametric preliminary estimator of the density f, constructed with a small part of the $n-$ sample. With the remainder of the sample, we build estimators of the terms, up to the second order, which appear in the Taylor expansion. Some of these terms are linear functionals of f and its derivatives while others are quadratic functionals of the type $\int f^{(j)} f^{(j')} K_{j,j'}(\widehat{f})$. So, a crucial point to make our methods work will be the construction of estimators of quantities of the type $\int f^{(j)} f^{(j')} \psi$ where ψ is a fixed smooth function since we shall work conditionnally to $\widehat{f}$. This will be the purpose of Section 2 . For the particular problem of estimating $\int (f^{(k)})^2$ we get the same results as Bickel and Ritov, and as Donoho and Nussbaum, moreover the formulation of the estimator is very simple. Our main result is stated in Section 3 and may be summarized as follows : we can build an estimator $\widehat{T_n}$ of $T(f)$ such that

i) If $s > 2k + \dfrac{1}{4}$ $\sqrt{n}(\widehat{T_n} - T(f)) \to \mathcal{N}(0, C(f,\phi))$ where $\mathcal{N}(0, \sigma^2)$ denotes the normal distribution and $nE(\widehat{T_n} - T(f))^2 \to C(f, \phi)$ where

$$C(f,\phi)) = \int \left[\sum_{j=0}^{k}(-1)^j \left(\phi'_j(f, f', \ldots, f^{(k)}, .)\right)^{(j)}\right]^2 f - \left[\sum_{j=0}^{k} \int \phi'_j(f, f', \ldots, f^{(k)}, .) f^{(j)}\right]^2$$

where

$$\phi'_j(x_0,\ldots,x_k,t) = \frac{\partial \phi}{\partial x_j}(x_0,\ldots x_k,t).$$

ii) If $k > 0$ and $k < s \leq 2k + \frac{1}{4}$ then $E(\widehat{T_n} - T(f))^2 = O(n^{\frac{-8(s-k)}{1+4s}})$.

In case i) $C(f,\phi)$ is the semiparametric information bound for the problem of estimating $T(f)$ as will be shown in the Appendix, hence our estimator is asymptotically efficient. On the other hand, the rates which appear in ii) cannot be improved for most functionals $T(f)$ as shown in Birgé and Massart (1991).
When $k = 0$ and $s < \frac{1}{4}$ we do not know what the optimal rate is, except that it is smaller than $n^{\frac{-3s}{1+2s}}$. Actually, in this very case the remainder term in the Taylor expansion is precisely of order $n^{\frac{-3s}{1+2s}}$. So, it would be necessary to do the Taylor expansion up to the third order and to estimate $\int f^3\psi$ but do not know how to estimate $\int f^3\psi$ at a rate faster than the one we just mentioned. Actually, the recent results of Kerkyacharian and Picard (1992) concerning the estimation of f^3 show that the rate of convergence for the estimation of this functional is $n^{\frac{-4s}{1+4s}}$.
The paper is organized as follows : in Section 2, we built estimators for the quantity $\int f^{(j)}f^{(j')}\psi$. In Section 3, we propose an estimator for $T(f)$. The proofs of the theorems are postponed to the fourth section.

3.2 Estimation of $\int f^{(j)}f^{(j')}\psi$

We will first set some notations.
The notation $A_n \approx B_n$ will mean that $\nu_1 \leq \frac{A_n}{B_n} \leq \nu_2$ where ν_1 and ν_2 are positive constants.
We shall denote (j,j') by J and $\int f^{(j)}f^{(j')}\psi$ by T_ψ^J. For any function g defined over $S = [-\pi,\pi]$, we define :

$$\|g\|_\infty = \sup_{x\in S}|g(x)|, \qquad \|g\|_\alpha = \sup_{x\neq y\in S}\frac{|g(x)-g(y)|}{|x-y|^\alpha} \quad \text{for} \quad 0<\alpha\leq 1.$$

Let $s = p+\alpha$ where $p \in \mathbb{N}$ and $\alpha \in]0,1]$. We shall denote by $F_{s,C}$ the set of densities f with support on S satisfying the following conditions :
1) $f^{(l)}(\pi) = f^{(l)}(-\pi) \quad \forall\, l \in \{0,\ldots p-1\}$
2) $\|f^{(p)}\|_\alpha \leq C$.
When it will be necessary, the function f will be extended by periodicity. $C^k(S)$ will be the space of k times differentiable functions on S.

Let $(p_i, i \in \mathbb{N})$ be the Fourier basis :

$$p_0(x) = \frac{1}{\sqrt{2\pi}}, \quad p_{2i}(x) = \frac{\cos ix}{\sqrt{\pi}}, \quad p_{2i-1}(x) = \frac{\sin ix}{\sqrt{\pi}} \quad \text{for} \quad i > 0,$$

we set $S_m f(x) = \sum_{i=0}^{m} a_i(f) p_i(x)$ with $a_i(f) = \int f p_i$. When no confusion is to be feared, we shall write a_i instead of $a_i(f)$.
Let $T_\psi^J = \int_{-\pi}^{\pi} f^{(j)} f^{(j')} \psi(x) dx$ to be estimated. We assume that $j \le j' < s$, $\psi \in C^{j'}(S)$ and that $\psi^{(l)}(\pi) = \psi^{(l)}(-\pi) \quad \forall\, l \in \{0, \dots, j'-1\}$.

The problem of estimating $\int f^2 \psi$ has already been studied in a more general framework, in the previous chapter. In particular, we do not have to suppose that f is a periodic function. The construction of the estimator of $\int_{-\pi}^{\pi} f^{(j)} f^{(j')} \psi(x) dx$ will be of the same type. We recall that the estimator of $\int f^2 \psi$ proposed in Chapter 2 is

$$\widehat{T_\psi^{00}} = \frac{2}{n(n-1)} \sum_{i=0}^{m(n)} \sum_{l_1 \ne l_2 = 1}^{n} p_i(X_{l_1}) p_{i'} \psi(X_{l_2})$$

$$-\frac{1}{n(n-1)} \sum_{i,i'=0}^{m(n)} \sum_{l_1 \ne l_2 = 1}^{n} p_i(X_{l_1}) p_{i'}(X_{l_2}) \int p_i p_{i'} \psi.$$

Let us write m instead of $m(n)$ for short. The bias of this estimator is equal to $-\int (S_m f - f)^2 \psi$. We shall look for a similar bias for the estimation of T_ψ^J. More precisely, we wish the bias to be equal to

$$-\int_{-\pi}^{\pi} ((S_m f)^{(j)} - f^{(j)})((S_m f)^{(j')} - f^{(j')}) \psi$$

$$= \int_{-\pi}^{\pi} (f^{(j)} (S_m f)^{(j')} + f^{(j')} (S_m f)^{(j)}) \psi - \int_{-\pi}^{\pi} (S_m f)^{(j)} (S_m f)^{(j')} \psi - \int_{-\pi}^{\pi} f^{(j)} f^{(j')} \psi.$$

Hence, the problem is to find an estimator with expectation

$$\int_{-\pi}^{\pi} (f^{(j)} (S_m f)^{(j')} + f^{(j')} (S_m f)^{(j)}) \psi - \int_{-\pi}^{\pi} (S_m f)^{(j)} (S_m f)^{(j')} \psi.$$

Since $f^{(l)}(\pi) = f^{(l)}(-\pi)$ and $\psi^{(l)}(\pi) = \psi^{(l)}(-\pi) \quad \forall\, l \in \{0, \dots, j'-1\}$, successive integrations by parts lead to

$$\int_{-\pi}^{\pi} f^{(j)} (S_m f)^{(j')} \psi = (-1)^j \int_{-\pi}^{\pi} ((S_m f)^{(j')} \psi)^{(j)} f.$$

It is therefore easy to see that the following estimator

$$\widehat{T^J_{\psi,1}} = \frac{1}{n(n-1)}\sum_{i=0}^{m(n)}\sum_{l_1\neq l_2=1}^{n} p_i(X_{l_1})(-1)^j(p_i^{(j')}\psi)^{(j)}(X_{l_2}) + \frac{1}{n(n-1)}\sum_{i=0}^{m(n)}\sum_{l_1\neq l_2=1}^{n} p_i(X_{l_1})(-1)^{j'}(p_i^{(j)}\psi)^{(j')}(X_{l_2}) \qquad (3.1)$$

will have expectation $\int_{-\pi}^{\pi}(f^{(j)}(S_m f)^{(j')} + f^{(j')}(S_m f)^{(j)})\psi$.

To get the term $-\int_{-\pi}^{\pi}(S_m f)^{(j)}(S_m f)^{(j')}\psi$, we propose the estimator

$$\widehat{T^J_{\psi,2}} = -\frac{1}{n(n-1)}\sum_{i,i'=0}^{m(n)}\sum_{l_1\neq l_2=1}^{n} p_i(X_{l_1})p_{i'}(X_{l_2})\int_{-\pi}^{\pi} p_i^{(j)}p_{i'}^{(j')}\psi(x)dx. \qquad (3.2)$$

Theorem 3.1 *Let $X_1, X_2, \dots, X_n$ be i.i.d. variables with density f belonging to the set $F_{s,C}$ and let ψ belong to $C^{j'}(S)$ with $j' < s$. and satisfy $\psi^{(l)}(\pi) = \psi^{(l)}(-\pi)\ \forall\ l \in \{0, \dots, j'-1\}$.*
Let $T^J_\psi = \int_{-\pi}^{\pi} f^{(j)}f^{(j')}\psi$ to be estimated with $j \leq j'$.
Define $\widehat{T^J_\psi} = \widehat{T^J_{\psi,1}} + \widehat{T^J_{\psi,2}}$ where $\widehat{T^J_{\psi,1}}$ and $\widehat{T^J_{\psi,2}}$ are given by (3.1) and (3.2) respectively with $m(n) \approx n^{\frac{2}{1+4s}}$. Then

- *i) If $s > j + j' + \frac{1}{4}$, and $\lambda_1 = \sup_{l=0..j'} \|\psi^{(l)}\|_\infty$ then*

$$E(\widehat{T^J_\psi} - T^J_\psi)^2 \leq \frac{C_1\lambda_1^2}{n}. \qquad (3.3)$$

Assuming that ξ satisfies the same assumptions as ψ, let $\widehat{T^{J_1}_\xi}$ be defined as $\widehat{T^J_\psi}$ replacing (j, j') by (j_1, j_1') and ψ by ξ and $T^{J_1}_\xi = \int_{-\pi}^{\pi} f^{(j_1)}f^{(j_1')}\xi$.
Let $s > \sup(j + j' + \frac{1}{4}, j_1 + j_1' + \frac{1}{4})$, $j_1' \geq j_1$ and

$$\mu_1 = \sup\left[\sup_{l=0..j'}\left(\|\psi^{(l)}\|_\infty, \|\psi^{(l)}\|_\alpha\right); \sup_{l=0,..j_1'}\left(\|\xi^{(l)}\|_\infty, \|\xi^{(l)}\|_\alpha\right)\right]$$

then

$$\left|nE(\widehat{T^J_\psi} - T^J_\psi)(\widehat{T^{J_1}_\xi} - T^{J_1}_\xi) - \Lambda_{JJ_1}(f,\psi,\xi)\right| \leq C_1'\mu_1^2(m(n)^{-\alpha} + m(n)^{\frac{1}{2}+j+j'+j_1+j_1'-2s}) \qquad (3.4)$$

where

$$\Lambda_{JJ_1}(f,\psi,\xi) = \int\left((-1)^j(f^{(j')}\psi)^{(j)} + (-1)^{j'}(f^{(j)}\psi)^{(j')}\right)\left((-1)^{j_1}(f^{(j_1')}\xi)^{(j_1)} + (-1)^{j_1'}(f^{(j_1)}\xi)^{(j_1')}\right)f$$

$$-4\int_{-\pi}^{\pi} f^{(j)}f^{(j')}\psi \int_{-\pi}^{\pi} f^{(j_1)}f^{(j_1')}\xi.$$

- *ii) If* $s \leq j+j'+\frac{1}{4}$ *and*

$$\lambda_2 = \sup\left(\sup_{l=0\ldots p-j}\frac{\|\psi^{(l)}\|_\infty + \|\psi^{(l)}\|_\alpha}{m(n)^{1/6}}, \sup_{l=0\ldots j'}\frac{\|\psi^{(l)}\|_\infty}{m(n)^{(j+j'-s+\frac{1}{4})\wedge l}}\right),$$

then

$$E(\widehat{T_\psi^J} - T_\psi^J)^2 \leq C_2\lambda_2^2 n^{\frac{-8s+4j+4j'}{1+4s}}. \tag{3.5}$$

C_1, C_1' *and* C_2 *are absolute constants .*

Comments :

1) If $j=j'$, $s<2j+\frac{1}{4}$ and ψ is either positive on S or negative on S, the rate of convergence is optimal (see Birgé and Massart (1991)). Otherwise, we do not know whether the optimal rate is $n^{\frac{-8s+4j+4j'}{1+4s}}$ or not.

2) In the next section, ψ will be a random function depending on n, this explains why we need bounds which depend explicitly on ψ.

3) The estimation of $\theta_k = \int (f^{(k)})^2$ has already been treated by Bickel and Ritov (1988). In this case $j=j'=k$, $\psi=1$ and the expression of our estimator is very simple :

$$\widehat{\theta}_k = \frac{1}{n(n-1)}\sum_{i=0}^{m(n)}\sum_{l_1\neq l_2=1}^{n} q_i(k)p_i(X_{l_1})p_i(X_{l_2})$$

where $q_i(k) = \int_{-\pi}^{\pi}(p_i^{(k)})^2 = (-1)^k\int_{-\pi}^{\pi} p_i^{(2k)}p_i$. Hence $q_{2i}(k) = q_{2i-1}(k) = i^{2k} \quad \forall\, i>0$, $q_0(k)=0$ if $k\geq 1$, $q_0(0)=1$.

It has the same properties as Bickel and Ritov's estimator.

4) The assumption of periodicity for f and its derivatives is necessary : suppose that we want to estimate $2\int_{-\pi}^{\pi} ff'$. It is equal to $f^2(\pi) - f^2(-\pi)$; hence, this problem is the same as estimating the density at one point. Farell (1972) showed that the rate of convergence for the problem of estimating the density at one point is not better than the rate obtained when we estimate the whole density, which never achieves the $n^{-1/2}$ rate.

5) Since the assumption of periodicity for f and its derivatives is necessary, it seems natural to use the Fourier basis to built our estimators. Moreover, the orthogonality of the derivatives of the p_i' s will be used in the proof of Theorem 3.1. Another orthonormal basis of $\mathbb{L}^2([-\pi,\pi])$ having this property could also be used.

3.3 Estimation of integrated functionals of f and its derivatives

3.3.1 Main results

The purpose of this section is to estimate $T(f) = \int \phi(f, f', \ldots, f^{(k)}, x)dx$ efficiently when it is possible. Assuming that $f \in F_{s,C}$ and $k < s$, we also give the rates of convergence of the estimators when the $n^{-1/2}$ rate is not achievable.
We would like to start with some preliminary estimator $\widehat{f}$ of the density f built on a small part of the initial sample and do a Taylor expansion of ϕ in a neighbourhood of $(\widehat{f}(x), \ldots, \widehat{f}^{(k)}(x), x)$. In order to give a sense to this expansion, we first need some assumptions.

Let $K_\epsilon = \prod_{i=0}^{k}[a_i - \epsilon, b_i + \epsilon] \times S$ be some given hyperrectangle , $\epsilon > 0$. We shall assume the following :

• A1 : ϕ belongs to $C^{k+3}(\Omega)$ for some open set Ω and $\forall\ 0 \le i \le k\ f^{(i)}(S) \subset [a_i, b_i]$ where $\prod_{i=0}^{k}[a_i, b_i] \times S \subset \Omega$.
For instance, if the functional to be estimated is the entropy ($\phi(x) = x\log(x)$) we assume that f has is bounded from below by some positive constant.

Since $\prod_{i=0}^{k} \tilde{f}^{(i)}(S) \times S$ is not almost surely included in Ω when $\tilde{f}$ is a standard preliminary estimator of f, such as kernel or projection based estimator, we shall have to modify it. We shall assume the following :

• A2 : We can find a preliminary estimator $\widehat{f}$ of f based on the n_1 last observations such that

$$a)\quad \widehat{f} \in C^{2k}(S),\quad \widehat{f}^{(l)}(\pi) = \widehat{f}^{(l)}(-\pi)\quad \forall\ l \in \{0, \ldots 2k-1\}$$

b) $\widehat{f}^{(i)}(S) \subset [a_i - \epsilon, b_i + \epsilon] \quad \forall\, i \in \{0, \dots k\}$

c) $E(\|\widehat{f}^{(l)} - f^{(l)}\|_q^q) \leq \gamma_1(q) n_1^{\frac{q(l-s)}{1+2s}} \quad \forall\, l \leq p, \quad \forall\, 2 \leq q < +\infty$

d) $E(\|\widehat{f}^{(l)}\|_\infty^q) \leq \gamma_2(q)(1 + n_1^{\frac{q(l-s')}{1+2s}}) \quad \forall\, s' < s \quad \forall\, 0 \leq l \leq 2k, q \geq 1$

e) $E(\|\widehat{f}^{(l)}\|_\alpha^q) \leq \gamma_3(q)(1 + n_1^{\frac{q(l+\alpha-s')}{1+2s}}) \quad \forall\, s' < s \quad \forall\, 0 \leq l \leq 2k, q \geq 1,$

where $\gamma_1(q), \gamma_2(q), \gamma_3(q)$ are absolute constants independent of $f \in F_{s,C}$; and where $K_\epsilon = \prod_{i=0}^{k} [a_i - \epsilon, b_i + \epsilon] \times S \subset \Omega$.
The existence of such estimators will be proved in the Appendix.

For $\widehat{f}$ satisfying condition A2, $\phi(\widehat{f}, \dots, \widehat{f}^{(k)}, .)$ is well defined and it is legitimate to make a Taylor expansion of ϕ in a neighbourhood of $(\widehat{f}(x), \dots \widehat{f}^{(k)}(x), x)$. We shall use the following notations for partial derivatives

$$\phi'_j = \frac{\partial \phi}{\partial y_j}(y_0, \dots y_{k+1}), \quad \phi''_{jj'} = \frac{\partial^2 \phi}{\partial y_j \partial y_{j'}}(y_0, \dots y_{k+1})$$

$$\|\phi^{(l)}\|_\infty = \sup_{j_1, \dots j_l \in \{0, \dots k\}} \sup_{(y_0, \dots y_{k+1}) \in K_\epsilon} |\phi^{(l)}_{j_1, \dots j_l}(y_0, \dots y_{k+1})|.$$

$$T(f) = \int \phi(\widehat{f}(x), \dots, \widehat{f}^{(k)}(x), x)dx + \sum_{j=0}^{k} \int \phi'_j(\widehat{f}(x), \dots, \widehat{f}^{(k)}(x), x)(f^{(j)} - \widehat{f}^{(j)})(x)$$

$$+\frac{1}{2} \sum_{j=0}^{k} \sum_{j'=0}^{k} \int \phi''_{jj'}(\widehat{f}(x), \dots \widehat{f}^{(k)}(x), x)(f^{(j)} - \widehat{f}^{(j)})(f^{(j')} - \widehat{f}^{(j')})(x) + \Gamma_n,$$

where Γ_n is a remainder term which will be proved to be negligible compared to the linear and quadratic terms. It is convenient to write $T(f)$ as follows :

$$T(f) = \int G(\widehat{f}) + \sum_{j=0}^{k} \int H_j(\widehat{f}) f^{(j)} + \sum_{j,j'=0}^{k} \int K_{jj'}(\widehat{f}) f^{(j)} f^{(j')} + \Gamma_n$$

where :

$$G(\widehat{f}) \;=\; \phi(\widehat{f}, \dots \widehat{f}^{(k)}, .) - \sum_{j=0}^{k} \phi'_j(\widehat{f}, \dots \widehat{f}^{(k)}, .)\widehat{f}^j$$

$$+ \quad \frac{1}{2}\sum_{j,j'=0}^{k} \phi''_{jj'}(\hat{f},\ldots\hat{f}^{(k)},.)\hat{f}^{(j)}\hat{f}^{(j')} \tag{3.6}$$

$$H_j(\hat{f}) \quad = \quad \phi'_j(\hat{f},\ldots,\hat{f}^{(k)},.) - \sum_{j'=0}^{k} \phi''_{jj'}(\hat{f},\ldots\hat{f}^{(k)},.)\hat{f}^{(j')} \tag{3.7}$$

$$K_{jj'}(\hat{f}) \quad = \quad K_J(\hat{f}) = \frac{1}{2}\phi''_{jj'}(\hat{f},\ldots,\hat{f}^{(k)},.). \tag{3.8}$$

We have to estimate two types of terms :

$\bullet \mathcal{H}_j = \int_{-\pi}^{\pi} H_j(\hat{f})f^{(j)}$ which is a linear functional of f.

$\bullet T^J(K_J(\hat{f})) = \int_{-\pi}^{\pi} K_J(\hat{f})f^{(j)}f^{(j')}$ which is a quadratic functional of the type we have studied in Section 2.

Before stating the results, let us give some ideas about the estimation of these terms. Since $f^{(l)}(\pi) = f^{(l)}(-\pi) \quad \forall\, l \in \{0,\ldots,k-1\}$ and $\hat{f}^{(l)}(\pi) = \hat{f}^{(l)}(-\pi) \quad \forall\, l \in \{0,\ldots,2k-1\}$, assuming that ϕ, ϕ'_j, and $\phi''_{jj'}$ are periodic with respect to the $k+2^{\text{th}}$ variable, we get by successive integrations by parts

$$\int_{-\pi}^{\pi} H_j(\hat{f})f^{(j)} = \int_{-\pi}^{\pi} (-1)^j[H_j(\hat{f})]^{(j)}f.$$

Setting $n_2 = n - n_1$, we can estimate $\mathcal{H}_j$ by

$$\widehat{\mathcal{H}_j} = \frac{1}{n_2}\sum_{l=1}^{n_2}(-1)^j[H_j(\hat{f})]^{(j)}(X_l).$$

As to $T^J(K_J(\hat{f}))$ its estimation has been studied in the previous section. $K_J(\hat{f})$ is a random function based on the n_1 last observations and $T^J(K_J(\hat{f}))$ has to be estimated with the remainder of the n sample which leads to

$$\begin{aligned}
\widehat{T^J}(K_J(\hat{f})) \quad &= \quad \frac{1}{n_2(n_2-1)}\sum_{i=0}^{m}\sum_{l_1\neq l_2=1}^{n_2} p_i(X_{l_1})(-1)^j(p_i^{(j')}K_J(\hat{f}))^{(j)}(X_{l_2}) \\
&+ \quad \frac{1}{n_2(n_2-1)}\sum_{i=0}^{m}\sum_{l_1\neq l_2=1}^{n_2} p_i(X_{l_1})(-1)^{j'}(p_i^{(j)}K_J(\hat{f}))^{(j')}(X_{l_2}) \\
&- \quad \frac{1}{n_2(n_2-1)}\sum_{i,i'=0}^{m}\sum_{l_1\neq l_2=1}^{n_2} p_i(X_{l_1})p_{i'}(X_{l_2})\int_{-\pi}^{\pi} p_i^{(j)}p_{i'}^{(j')}K_J(\hat{f})(x)dx.
\end{aligned}$$

Theorem 3.2 *Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with common density f belonging to $F_{s,C}$, let $T(f) = \int \phi(f(x),\ldots,f^{(k)}(x),x)dx$ to be estimated with $k < s$ and*

assume that A1 holds and that $\phi, \phi'_j, \phi''_{jj'}$ are 2π periodic with respect to the $k+2^{th}$ variable $\forall\, j, j' \in \{0, \ldots k\}$.
Consider a preliminary estimator $\widehat{f}$ of f satisfying A2 and based on the n_1 last observations where $n_1 \approx \frac{n}{\log n}$. Let

$$\begin{aligned}
\widehat{T_n} &= \int_{-\pi}^{\pi} G(\widehat{f}) + \sum_{j=0}^{k} \frac{1}{n_2} \sum_{l=1}^{n_2} (-1)^j [H_j(\widehat{f})]^{(j)}(X_l) \\
&+ \sum_{j,j'=0}^{k} \sum_{i=0}^{m(n)} \frac{2}{n_2(n_2-1)} \sum_{l_1 \neq l_2 = 1}^{n_2} (-1)^{j'} p_i(X_{l_1}) [K_J(\widehat{f}) p_i^{(j)}]^{(j')}(X_{l_2}) \\
&- \sum_{j,j'=0}^{k} \sum_{i,i'=0}^{m(n)} \frac{1}{n_2(n_2-1)} \sum_{l_1 \neq l_2 = 1}^{n_2} p_i(X_{l_1}) p_{i'}(X_{l_2}) \int_{-\pi}^{\pi} p_i^{(j)} p_{i'}^{(j')} K_J(\widehat{f}) dx
\end{aligned}$$

where G, H_j, K_J are defined by (3.6), (3.7) and (3.8), and $m(n) \approx n^{\frac{2}{1+4s}}$.
The following properties hold :

- *i) If $s > 2k + \frac{1}{4}$ then*

$$\sqrt{n}(\widehat{T_n} - T(f)) \to \mathcal{N}(0, C(f, \phi)) \tag{3.9}$$

and

$$nE(\widehat{T_n} - T(f))^2 \to C(f, \phi) \tag{3.10}$$

where $C(f, \phi)$ is equal to

$$C(f,\phi) = \int_{-\pi}^{\pi} \left[\sum_{j=0}^{k} (-1)^j \left(\phi'_j(f, ..f^{(k)}, .) \right)^{(j)} \right]^2 f - \left[\sum_{j=0}^{k} \int_{-\pi}^{\pi} \phi'_j(f, ..f^{(k)}, .) f^{(j)} \right]^2 .$$

- *ii) If $k > 0$ and $k < s \leq 2k + \frac{1}{4}$ then*

$$E(\widehat{T_n} - T(f))^2 = O(n^{\frac{-8(s-k)}{1+4s}}).$$

Comments :

1) When $k > 0$ and $k < s \leq 2k + \frac{1}{4}$, assuming that T is not degenerated, Birgé and Massart (1991) have proved that the rate of convergence that we get is optimal, moreover, in the semi-parametric case, the asymptotic variance is optimal as will be shown in the Appendix.

2) When $k = 0$ the problem has already been studied in a more general framework (see Chapter 2). In this case, we do not have to suppose that f is a periodic function to get the semi-parametric rate of convergence $\frac{1}{\sqrt{n}}$.

3.3.2 Fisher Information estimation

As an example, let us give the precise expression of our estimator of $\int_{-\pi}^{\pi} \frac{f'^2}{f}$. Using a the Taylor expansion of $\phi(f)$ in a neighbourhood of $(\widehat{f}, \widehat{f}')$, we get

$$\int_{-\pi}^{\pi} \frac{f'^2}{f} = -\int_{-\pi}^{\pi} \frac{\widehat{f'}^2}{\widehat{f}^2} f + 2\int_{-\pi}^{\pi} \frac{\widehat{f'}}{\widehat{f}} f' + \int_{-\pi}^{\pi} \frac{\widehat{f'}^2}{\widehat{f}^3} f^2 + \int_{-\pi}^{\pi} \frac{f'^2}{\widehat{f}} - 2\int_{-\pi}^{\pi} \frac{\widehat{f'}}{\widehat{f}^2} ff' + \Gamma_n$$

This expression is equal to

$$\int_{-\pi}^{\pi} \left(\frac{\widehat{f'}^2}{\widehat{f}^2} - 2\frac{\widehat{f''}}{\widehat{f}} \right) f + \int_{-\pi}^{\pi} \frac{\widehat{f'}^2}{\widehat{f}^3} f^2 + \int_{-\pi}^{\pi} \frac{f'^2}{\widehat{f}} - 2\int_{-\pi}^{\pi} \frac{\widehat{f'}}{\widehat{f}^2} ff' + \Gamma_n$$

We have to suppose that f is bounded from below by a positive constant. The estimator has the following expression :

$$\widehat{T_n} = \frac{1}{n_2} \sum_{l=1}^{n_2} \left(\frac{\widehat{f'}^2}{\widehat{f}^2} - 2\frac{\widehat{f''}}{\widehat{f}} \right)(X_l) + \sum_{j,j'=0}^{1} \sum_{i=0}^{m(n)} \frac{2}{n_2(n_2-1)} \sum_{l_1 \neq l_2 = 1}^{n_2} (-1)^{j'} p_i(X_{l_1}) [K_{j,j'}(\widehat{f}) p_i^{(j)}]^{(j')}(X_{l_2})$$

$$- \sum_{j,j'=0}^{1} \sum_{i,i'=0}^{m(n)} \frac{1}{n_2(n_2-1)} \sum_{l_1 \neq l_2 = 1}^{n_2} p_i(X_{l_1}) p_{i'}(X_{l_2}) \int_{-\pi}^{\pi} p_i^{(j)} p_{i'}^{(j')} K_{j,j'}(\widehat{f})$$

Where :

$$K_{0,0} = \frac{\widehat{f'}^2}{\widehat{f}^3}, \quad K_{0,1} = K_{1,0} = -\frac{\widehat{f'}}{\widehat{f}^2}, \quad K_{1,1} = \frac{1}{\widehat{f}}.$$

If $s > \frac{9}{4}$ we get

$$\lim_{n \to \infty} nE(\widehat{T_n} - \int_{-\pi}^{\pi} \frac{f'^2}{f})^2 = \int_{-\pi}^{\pi} \left(\frac{2f''}{f} - \frac{f'^2}{f^2} \right)^2 f - \left(\int_{-\pi}^{\pi} \frac{f'^2}{f} \right)^2.$$

If $1 < s \leq \frac{9}{4}$

$$E(\widehat{T_n} - \int_{-\pi}^{\pi} \frac{f'^2}{f})^2 = O(n^{\frac{-8(s-1)}{1+4s}}).$$

3.4 Proofs

3.4.1 Proof of Theorem 3.1

We recall that $s = p + \alpha$ where $p \in \mathbb{N}$ and $\alpha \in]0,1]$. The following lemmas will prove usefull :

Lemma 3.1

$$\sup \ \{\sum_{i=1}^{\infty} i^{2p+2\beta}(a_{2i}^2(f) + a_{2i-1}^2(f)), f \in F_{s,C}\} < \frac{\pi^{2\alpha+1}C^2}{4^{\alpha-\beta}-1} \quad \forall \quad 0 < \beta < \alpha$$

$$\sup \ \{\sum_{i=m+1}^{\infty} i^{2p}(a_{2i}^2(f) + a_{2i-1}^2(f)), f \in F_{s,C}\} \leq \frac{C^2}{m^{2\alpha}}\left(\frac{\pi^{2\alpha+1}}{4^{\alpha}-1}\right)$$

The proof of Lemma 3.1 is postponed to the Appendix.

Lemma 3.2

$$\sup\{\|f^{(l)}\|_\infty, l = 0,\dots,p, f \in F_{s,C}\} < +\infty.$$

We refer to Bickel and Ritov (1988) for the proof of Lemma 3.2.

Proof of Theorem 3.1: We use the classical decomposition :

$$E(\widehat{T_\psi^J} - T_\psi^J)^2 = (\mathrm{Bias}(\widehat{T_\psi^J}))^2 + \mathrm{Var}(\widehat{T_\psi^J})$$

In the proofs, we shall write m instead of $m(n)$ for short and always assume that m is even.

- **Control of the bias :**

We noticed above that

$$\begin{aligned} |\mathrm{Bias}(\widehat{T_\psi^J})| &= \left|\int_{-\pi}^{\pi}((S_m f)^{(j)} - f^{(j)})((S_m f)^{(j')} - f^{(j')})\psi\right| \\ &\leq \|\psi\|_\infty \left(\int_{-\pi}^{\pi}((S_m f)^{(j)} - f^{(j)})^2\right)^{1/2}\left(\int_{-\pi}^{\pi}((S_m f)^{(j')} - f^{(j')})^2\right)^{1/2} \end{aligned}$$

We claim that, since m is even

$$\sum_{i=0}^{m} p_i(x)p_i^{(j')}(y) = (-1)^{j'}\sum_{i=0}^{m} p_i^{(j')}(x)p_i(y). \qquad (3.11)$$

This equality implies that

$$S_m(g^{(l)}) = (S_m g)^{(l)} \quad \forall\, g \in C^l(S)).$$

Hence

$$|\mathrm{Bias}(\widehat{T_\psi^J})|^2 \leq \|\psi\|_\infty^2 \sum_{i=m+1}^{\infty} a_i^2(f^{(j)}) \sum_{i=m+1}^{\infty} a_i^2(f^{(j')}) = \|\psi\|_\infty^2 \sum_{i=m+1}^{\infty} q_i(j)a_i^2 \sum_{i=m+1}^{\infty} q_i(j')a_i^2.$$

Since $0 \leq q_i(j) \leq i^{2j}$, we get

$$|\mathrm{Bias}(\widehat{T_\psi^J})|^2 \leq \|\psi\|_\infty^2 \sum_{i=m+1}^{\infty} i^{2p}a_i^2 i^{2j-2p} \sum_{i=m+1}^{\infty} i^{2p}a_i^2 i^{2j'-2p}.$$

From Lemma 3.1 and the fact that $j - p \leq 0, \quad j' - p \leq 0$, we derive

$$|\mathrm{Bias}(\widehat{T_\psi^J})|^2 = O(\|\psi\|_\infty^2 (m^{2j+2j'-4s})).$$

Here, and in the remainder of the proof, the O's are independent of ψ and its derivatives, and of $f \in F_{s,C}$.

Let us now evaluate the order of magnitude of $\mathrm{Var}(\widehat{T_\psi^J})$.

- **Control of the variance :**

$$\widehat{T_\psi^J} = \widehat{T_{\psi,1}^J} + \widehat{T_{\psi,2}^J}$$

as defined by (3.1) and (3.2) .

We shall first prove (3.3) and (3.5), so we just have to determine the order of magnitude of $\mathrm{Var}(\widehat{T_\psi^J})$. We notice that $\widehat{T_\psi^J}$ can be written under the following form:

$$\widehat{T_\psi^J} = \frac{1}{n(n-1)} \sum_{l_1 \neq l_2 = 1}^{n} h_{jj'}(X_{l_1}, X_{l_2})$$

where $h_{jj'}$ is symetric. More precisely,

$$\widehat{T_{\psi,1}^J} = \frac{1}{2n(n-1)} \sum_{l_1 \neq l_2 = 1}^{n} (H_1^{jj'} + H_1^{j'j})(X_{l_1}, X_{l_2}) + (H_1^{jj'} + H_1^{j'j})(X_{l_2}, X_{l_1})$$

where $H_1^{jj'}(x,y) = \sum_{i=0}^{m} p_i(x)(-1)^j (p_i^{(j')}\psi)^{(j)}(y)$. Moreover,

$$\widehat{T_{\psi,2}^J} = -\frac{1}{2n(n-1)} \sum_{l_1 \neq l_2 = 1}^{n} (H_2^{jj'} + H_2^{j'j})(X_{l_1}, X_{l_2})$$

with $H_2^{jj'}(x,y) = \sum_{i,i'=0}^{m} p_i(x)p_{i'}(y)\int_{-\pi}^{\pi} p_i^{(j)}p_{i'}^{(j')}\psi$.
We notice that $H_2^{jj'}(x,y) = H_2^{j'j}(y,x)$.
$h_{jj'}$ is defined by

$$h_{jj'}(x,y) = \frac{1}{2}\left[(H_1^{jj'} + H_1^{j'j})(x,y) + (H_1^{jj'} + H_1^{j'j})(y,x) - (H_2^{jj'} + H_2^{j'j})(x,y)\right].$$

It follows from Hoeffding's results concerning the computation of the variance for U-statistics (see Hoeffding (1948) or Serfling (1980) p. 183) that

$$\mathrm{Var}(\widehat{T_\psi^J}) = \frac{4(n-2)}{n(n-1)}\xi_1 + \frac{2}{n(n-1)}\xi_2$$

where

$$\xi_1 = \mathrm{Var}(\bar{h}_{jj'}(X_1)), \quad \xi_2 = \mathrm{Var}(h_{jj'}(X_1,X_2))$$

and $\bar{h}_{jj'}(x) = E(h_{jj'}(x,X_2))$. We first compute $\bar{h}_{jj'}(X_1)$.

- Computation of $E(H_1^{jj'}(x,X_2))$.

$$\begin{aligned}
E(H_1^{jj'}(x,X_2)) &= E\left[\sum_{i=0}^{m} p_i(x)(-1)^j(p_i^{(j')}\psi)^{(j)}(X_2)\right] \\
&= \sum_{i=0}^{m} p_i(x)\int p_i^{(j')}\psi f^{(j)} \quad \text{integrating by parts.} \\
&= \sum_{i=0}^{m}(-1)^{j'}p_i^{(j')}(x)\int \psi f^{(j)}p_i \quad \text{by (3.11).}
\end{aligned}$$

Hence,

$$E(H_1^{jj'}(x,X_2)) = (-1)^{j'}S_m^{(j')}[\psi f^{(j)}](x).$$

- Computation of $E(H_1^{jj'}(X_2,x))$.

$$\begin{aligned}
E(H_1^{jj'}(X_2,x)) &= E\left[\sum_{i=0}^{m} p_i(X_2)(-1)^j(p_i^{(j')}\psi)^{(j)}(x_1)\right] \\
&= \sum_{i=0}^{m}(\int fp_i)(-1)^j(p_i^{(j')}\psi)^{(j)}(x) \\
&= (-1)^j(S_m f^{(j')}\psi)^{(j)}(x)
\end{aligned}$$

• Computation of $E(H_2^{jj'}(x, X_2))$.

$$\begin{aligned} E(H_2^{jj'}(x, X_2)) &= E\left[\sum_{i,i'=0}^{m} p_i(x)p_{i'}(X_2) \int p_i^{(j)} p_{i'}^{(j')} \psi\right] \\ &= \sum_{i=0}^{m} p_i(x) \int p_i^{(j)} (S_m f)^{(j')} \psi \\ &= (-1)^j \sum_{i=0}^{m} p_i(x) \int ((S_m f)^{(j')} \psi)^{(j)} p_i \quad \text{integrating by parts} \\ &= (-1)^j S_m \left[((S_m f)^{(j')} \psi)^{(j)}\right](x). \end{aligned}$$

Denote

$$Y_1^{jj'} = (-1)^{j'} S_m^{(j')}[f^{(j)}\psi](X_1) \tag{3.12}$$

$$Y_2^{jj'} = (-1)^j (S_m f^{(j')}\psi)^{(j)}(X_1) \tag{3.13}$$

$$Y_3^{jj'} = -(-1)^j S_m \left[((S_m f)^{(j')}\psi)^{(j)}\right](X_1) \tag{3.14}$$

$$\bar{h}_{jj'}(X_1) = \frac{1}{2} \sum_{l=1}^{3} (Y_l^{jj'} + Y_l^{j'j}).$$

It follows that

$$\begin{aligned} \xi_1 = \mathrm{Var}[\bar{h}_{jj'}(X_1)] &\leq E[(\bar{h}_{jj'}(X_1))^2] \\ &\leq 3 \sum_{l=1}^{3} E(Y_l^{jj'} + Y_l^{j'j})^2 \\ &\leq 6 \sum_{l=1}^{3} E[(Y_l^{jj'})^2] + E[(Y_l^{j'j})^2] \end{aligned}$$

We will now compute $E[(Y_l^{jj'})^2]$ for $l \in \{1, 2, 3\}$. We first majorize these quantities.

• Computation of $E[(Y_1^{jj'})^2]$

$$E[(Y_1^{jj'})^2] = \int \left[S_m^{(j')}[\psi f^{(j)}]\right]^2 f.$$

i) If $s > j + j' + \frac{1}{4}$ then $p \geq j + j'$ and f is $j + j'$ times continuously differentiable so, we can write using the fact that $S_m(g^{(l)}) = (S_m g)^{(l)}$,

$$\begin{aligned} E[(Y_1^{jj'})^2] &= \int S_m^2[(\psi f^{(j)})^{(j')}] f \\ &\leq \|f\|_\infty \|(\psi f^{(j)})^{(j')}\|_2^2 \quad \text{since } S_m \text{ is a projection} \\ &\leq \|f\|_\infty \| \sum_{l=0}^{j'} \binom{j'}{l} \psi^{(l)} f^{(j+j'-l)} \|_2^2 \end{aligned}$$

Since by Lemma 3.2, $\sup\{\|f^{(l)}\|_\infty, l = 0, \ldots p, f \in F_{s,C}\} < +\infty$ we obtain

$$E[(Y_1^{jj'})^2] \leq C_1 \lambda_1^2$$

where C_1 is independent of ψ and $f \in F_{s,C}$.

ii) If $s \leq j + j' + \frac{1}{4}$ then $p \leq j + j'$. We first consider the case where $p = j + j'$. As above, we can write

$$\begin{aligned} E[(Y_1^{jj'})^2] &= \int S_m^2[(\psi f^{(j)})^{(j')}]f \\ &\leq \|f\|_\infty \|(\psi f^{(j)})^{(j')}\|_2^2 \\ &= O(\sup_{l=0,\ldots,j'} \|\psi^{(l)}\|_\infty^2) \end{aligned}$$

Moreover, by definition of λ_2

$$\sup_{l=0,\ldots j'} \|\psi^{(l)}\|_\infty^2 \leq \lambda_2^2 m^{2(j+j'-s+\frac{1}{4})}.$$

Let us now consider the case where $j + j' > p$. In this case we also have $j + j' \geq s$. Using the fact that $S_m(g^{(l)}) = (S_m g)^{(l)}$, for m even, we get :

$$\begin{aligned} E[(Y_1^{jj'})^2] &= \int \left(S_m^{(j+j'-p)}[(\psi f^{(j)})^{(p-j)}]\right)^2 f \\ &\leq \|f\|_\infty \sum_{i=0}^{m} i^{2j+2j'-2p} b_i^2 \quad \text{by orthogonality of the } p_i'\text{s and their derivatives,} \end{aligned}$$

where b_i denotes the i^{th} Fourier coefficient of the function $(\psi f^{(j)})^{(p-j)}$. We will evaluate the $\alpha-$ norm of this function in order to apply Lemma 3.1.

$$(\psi f^{(j)})^{(p-j)} = \sum_{l=0}^{p-j} \binom{p-j}{l} \psi^{(l)} f^{(p-l)}.$$

Using the fact that for any functions g, h

$$\|gh\|_\alpha \leq \|g\|_\infty \|h\|_\alpha + \|h\|_\infty \|g\|_\alpha$$

we get

$$\|(\psi f^{(j)})^{(p-j)}\|_\alpha \leq \sum_{l=0}^{p-j} \binom{p-j}{l} \left(\|\psi^{(l)}\|_\infty \|f^{(p-l)}\|_\alpha + \|\psi^{(l)}\|_\alpha \|f^{(p-l)}\|_\infty\right).$$

By Lemma 3.2 this quantity is $O\left(\sup_{l=0,\dots p-j}(\|\psi^{(l)}\|_\infty+\|\psi^{(l)}\|_\alpha)\right)$.
By definition of λ_2,

$$\sup_{l=0,\dots,p-j}\left(\|\psi^{(l)}\|_\infty+\|\psi^{(l)}\|_\alpha\right)\le\lambda_2 m^{1/6}.$$

From Lemma 3.1 and since $j+j'\ge s>p+\beta\quad\forall\, 0<\beta<\alpha$, we derive

$$\begin{aligned}\sum_{i=0}^m i^{2j+2j'-2p}b_i^2 &= \sum_{i=0}^m i^{2j+2j'-2p-2\beta}i^{2\beta}b_i^2\\ &= O\left(m^{2j+2j'-2p-2\beta}\|(\psi f^{(j)})^{(j')}\|_\alpha^2\right)\\ &= O\left(\lambda_2^2 m^{1/3}m^{2j+2j'-2s+1/2}m^{-\frac{1}{2}+2(\alpha-\beta)}\right).\end{aligned}$$

Since β is arbitrary close to α, $m^{1/3}m^{-\frac{1}{2}+2(\alpha-\beta)}\to 0$. Collecting the above evaluations, we get for $s\le j+j'+\frac{1}{4}$

$$E[(Y_1^{jj'})^2]\le C_2\lambda_2^2(m^{2j+2j'-2s+1/2}).$$

• Computation of $E[(Y_2^{jj'})^2]$

$$\begin{aligned}E[(Y_2^{jj'})^2] &= \int\left[\left(S_m f^{(j')}\psi\right)^{(j)}\right]^2 f\le\|f\|_\infty\|(S_m f^{(j')}\psi)^{(j)}\|_2^2\\ &\le \|f\|_\infty\int\left(\sum_{l=0}^{j}\binom{j}{l}(S_m f)^{(j+j'-l)}\psi^{(l)}\right)^2\\ &= O\left[\sup_{l=0,\dots j'}\|\psi^{(l)}\|_\infty^2\|(S_m f)^{(j+j'-l)}\|_2^2\right]\end{aligned}$$

As above, we shall consider the following cases :
i) If $s>j+j'+\frac{1}{4}$ then $p\ge j+j'$ and

$$(S_m f)^{(j+j'-l)}=S_m[f^{(j+j'-l)}]\quad\forall\, l\in\{0,\dots j\}.$$

Since S_m is a projection, it follows that

$$\|(S_m f)^{(j+j'-l)}\|_2^2\le\|f^{(j+j'-l)}\|_2^2\le 2\pi\sup_{l=0,\dots p}\|f^{(l)}\|_\infty^2$$

Hence, possibly enlarging C_1

$$E[(Y_2^{jj'})^2]\le C_1\lambda_1^2.$$

ii) If $s\le j+j'+\frac{1}{4}$, we shall consider two cases :

if $j + j' - l < s$ then $j + j' - l \leq p$ and $\|S_m(f^{(j+j'-l)})\|_2^2 \leq 2\pi \sup_{l=0,\ldots p} \|f^{(l)}\|_\infty^2$.

It follows that

$$E[(Y_2^{jj'})^2] = O\left(\sup_{l=0\ldots j'} \|\psi^{(l)}\|_\infty^2\right) = O\left(\lambda_2^2 m^{2j+2j'-2s+\frac{1}{2}}\right).$$

If $j + j' - l \geq s$ then

$$\begin{aligned}\|\psi^{(l)}\|_\infty^2 \|(S_m f)^{(j+j'-l)}\|_2^2 &\leq \|\psi^{(l)}\|_\infty^2 \sum_{i=0}^{m} i^{2j+2j'-2l} a_i^2 \\ &\leq \|\psi^{(l)}\|_\infty^2 \sum_{i=0}^{m} i^{2j+2j'-2l-2p-2\beta} i^{2p+2\beta} a_i^2 \quad \forall\, 0 < \beta < \alpha.\end{aligned}$$

Since $j + j' - l \geq s > p + \beta, \quad \forall\, 0 < \beta < \alpha$, we get

$$\begin{aligned}\|\psi^{(l)}\|_\infty^2 \|(S_m f)^{(j+j'-l)}\|_2^2 &\leq \|\psi^{(l)}\|_\infty^2 m^{2j+2j'-2l-2p-2\beta} \sum_{i=0}^{\infty} i^{2p+2\beta} a_i^2 \\ &= O\left[(\|\psi^{(l)}\|_\infty m^{-l})^2 m^{2j+2j'-2p-2\beta}\right] \\ &= O\left[\lambda_2^2 m^{2j+2j'-2p-2\beta}\right].\end{aligned}$$

Possibly enlarging C_2 the above equality leads to

$$E[(Y_2^{jj'})^2] \leq C_2 \lambda_2^2 (m^{2j+2j'-2s+1/2}).$$

- Computation of $E[(Y_3^{jj'})^2]$.

$$\begin{aligned}E[(Y_3^{jj'})^2] &= \int \left[S_m \left(S_m f^{(j')} \psi\right)^{(j)}\right]^2 f \\ &\leq \|f\|_\infty \| \left(S_m f^{(j')} \psi\right)^{(j)} \|_2^2\end{aligned}$$

since S_m is a projection and the computation of this term has just been done above.So, we get

i) if $s > j + j' + \frac{1}{4}$ then $E[(Y_3^{jj'})^2] \leq C_1 \lambda_1^2$

ii) if $s \leq j + j' + \frac{1}{4}$, then $E[(Y_3^{jj'})^2] \leq C_2 \lambda_2^2 (m^{2j+2j'-2s+1/2})$.

Of course, the above results hold for $E[(Y_l^{j'j})^2], \quad l = 1, 2, 3$. It follows, enlarging C_1 and C_2 that

$$\text{i) if } s > j + j' + \frac{1}{4} \quad \xi_1 \leq C_1 \lambda_1^2 \tag{3.15}$$

$$\text{ii) if } s \leq j + j' + \frac{1}{4} \quad \xi_1 \leq C_2 \lambda_2^2 (m^{2j+2j'-2s+1/2}) \tag{3.16}$$

We now bound ξ_2.

$$\begin{aligned}\xi_2 &= \mathrm{Var}\left(h_{jj'}(X_1,X_2)\right) \le E\left(h^2_{jj'}(X_1,X_2)\right)\\ &\le 3E\left[\left(H_1^{jj'}(X_1,X_2)\right)^2 + \left(H_1^{j'j}(X_1,X_2)\right)^2 + \left(H_2^{jj'}(X_1,X_2)\right)^2\right]\end{aligned}$$

The three inner bracketed terms will be evaluated separately.

• Computation of $E\left[\left(H_1^{jj'}(X_1,X_2)\right)^2\right]$.

$$\begin{aligned}E\left[\left(H_1^{jj'}(X_1,X_2)\right)^2\right] &= \int\int\left[\sum_{i=0}^{m} p_i(x)(p_i^{(j')}\psi)^{(j)}(y)\right]^2 f(x)f(y)dxdy\\
&\le \|f\|_\infty^2\int\int\left[\sum_{i=0}^{m} p_i(x)(p_i^{(j')}\psi)^{(j)}(y)\right]^2 dxdy\\
&\le \|f\|_\infty^2\int\int\sum_{i,i'=0}^{m} p_ip_{i'}(x)(p_i^{(j')}\psi)^{(j)}(y)(p_{i'}^{(j')}\psi)^{(j)}(y)dxdy\\
&\le \|f\|_\infty^2\sum_{i=0}^{m}\int\left[(p_i^{(j')}\psi)^{(j)}(y)\right]^2 dy \quad \text{by orthonormality}\\
&\le \|f\|_\infty^2\sum_{i=0}^{m}\int\left[\sum_{l=0}^{j}\binom{j}{l}p_i^{(j+j'-l)}\psi^{(l)}(y)\right]^2 dy\\
&\le \|f\|_\infty^2\sum_{i=0}^{m}\int\sum_{l,l'=0}^{j}\binom{j}{l}\binom{j}{l'}p_i^{(j+j'-l)}p_i^{(j+j'-l')}\psi^{(l)}\psi^{(l')}(y)dy\end{aligned}$$

Since $\|p_i^{(l)}\|_\infty \le i^l$ our expression has order

$$O\left[\sum_{l,l'=0}^{j}\sum_{i=0}^{m} i^{2j+2j'-l-l'}\|\psi^{(l)}\|_\infty\|\psi^{(l')}\|_\infty\right] = O\left[m^{1+2j+2j'}\left(\sum_{l=0}^{j} m^{-l}\|\psi^{(l)}\|_\infty\right)^2\right].$$

Moreover

i) if $s > j + j' + \frac{1}{4}$ then $\sum_{l=0}^{j} m^{-l}\|\psi^{(l)}\|_\infty = O(\lambda_1)$ which implies, possibly enlarging C_1

$$E\left[\left(H_1^{jj'}(X_1,X_2)\right)^2\right] \le C_1\lambda_1^2 m^{1+2j+2j'}.$$

ii) If $s \le j + j' + \frac{1}{4}$ then $\sum_{l=0}^{j} m^{-l}\|\psi^{(l)}\|_\infty = O(\lambda_2)$ which implies, possibly enlarging C_2

$$E\left[\left(H_1^{jj'}(X_1,X_2)\right)^2\right] \le C_2\lambda_2^2 m^{1+2j+2j'}.$$

It is clear that the same result holds for $E\left[\left(H_1^{j'j}(X_1,X_2)\right)^2\right]$.

- Computation of $E\left[\left(H_2^{jj'}(X_1,X_2)\right)^2\right]$.

$$E\left[\left(H_2^{jj'}(X_1,X_2)\right)^2\right] = E\left[\sum_{i,i',i_1,i_1'=0}^{m} p_i p_{i_1}(X_1) p_{i'} p_{i_1'}(X_2) \int p_i^{(j)} p_{i'}^{(j')} \psi \int p_{i_1}^{(j)} p_{i_1'}^{(j')} \psi\right]$$

$$\begin{aligned}
&= \sum_{i,i',i_1,i_1'=0}^{m} \int p_i p_{i_1} f \int p_{i'} p_{i_1'} f \int p_i^{(j)} p_{i'}^{(j')} \psi \int p_{i_1}^{(j)} p_{i_1'}^{(j')} \psi \\
&= \int\int \left(\sum_{i,i'=0}^{m} (\int p_i^{(j)} p_{i'}^{(j')} \psi) p_i(x) p_{i'}(y)\right)^2 f(x) f(y) dx dy \\
&\leq \|f\|_\infty^2 \int\int \left(\sum_{i,i'=0}^{m} (\int p_i^{(j)} p_{i'}^{(j')} \psi) p_i(x) p_{i'}(y)\right)^2 dx dy \\
&\leq \|f\|_\infty^2 \sum_{i,i',i_1,i_1'=0}^{m} \int p_i^{(j)} p_{i'}^{(j')} \psi \int p_{i_1}^{(j)} p_{i_1'}^{(j')} \psi \int p_i p_{i_1} \int p_{i'} p_{i_1'} \\
&\leq \|f\|_\infty^2 \sum_{i,i'=0}^{m} \int p_i^{(j)} p_{i'}^{(j')} \psi \int p_i^{(j)} p_{i'}^{(j')} \psi \quad \text{by orthogonality} \\
&\leq \|f\|_\infty^2 \int\int \left(\sum_{i=0}^{m} p_i^{(j)}(z) p_i^{(j)}(t)\right) \left(\sum_{i'=0}^{m} p_{i'}^{(j')}(z) p_{i'}^{(j')}(t)\right) \psi(z)\psi(t) dt dz
\end{aligned}$$

By Cauchy-Schwarz inequality, this expression is bounded by

$$\|f\|_\infty^2 \|\psi\|_\infty^2 \left[\int\int \left(\sum_{i=0}^{m} p_i^{(j)}(z) p_i^{(j)}(t)\right)^2 dt dz \int\int \left(\sum_{i'=0}^{m} p_{i'}^{(j')}(z) p_{i'}^{(j')}(t)\right)^2 dt dz\right]^{1/2}$$

Moreover,

$$\begin{aligned}
\int\int \left(\sum_{i=0}^{m} p_i^{(j)}(z) p_i^{(j)}(t)\right)^2 dt dz &= \int\int \sum_{i,i_1=0}^{m} p_i^{(j)}(z) p_{i_1}^{(j)}(z) p_i^{(j)}(t) p_{i_1}^{(j)}(t) dt dz \\
&\leq \sum_{i=0}^{m} i^{4j} \quad \text{by orthogonality of } p_i^{(j)}. \\
&\leq m^{1+4j}
\end{aligned}$$

It follows that

$$E\left[(H_2^{jj'})^2(X_1,X_2)\right] \leq \|f\|_\infty^2\|\psi\|_\infty^2(m^{1+4j})^{1/2}(m^{1+4j'})^{1/2}$$
$$\leq \|f\|_\infty^2\|\psi\|_\infty^2 m^{1+2j+2j'}$$

Hence, we obtain

$$\text{i) If } s > j+j'+\frac{1}{4} \quad \xi_2 \leq C_1\lambda_1^2 m^{1+2j+2j'} \tag{3.17}$$
$$\text{ii) If } s \leq j+j'+\frac{1}{4} \quad \xi_2 \leq C_2\lambda_2^2 m^{1+2j+2j'} \tag{3.18}$$

We recall that $\mathrm{Var}(\widehat{T_\psi^J}) = \frac{4(n-2)}{n(n-1)}\xi_1 + \frac{2}{n(n-1)}\xi_2$.
From (3.15) and (3.17) we obtain, enlarging C_1

$$\text{if } s > j+j'+\frac{1}{4} \text{ then } \mathrm{Var}(\widehat{T_\psi^J}) \leq C_1\lambda_1^2\left(\frac{1}{n}+\frac{m^{1+2j+2j'}}{n^2}\right).$$

From (3.16) and (3.18) we obtain, enlarging C_2

$$\text{if } s \leq j+j'+\frac{1}{4} \text{ then } \mathrm{Var}(\widehat{T_\psi^J}) \leq C_2\lambda_2^2\left(\frac{m^{2j+2j'-2s+1/2}}{n}+\frac{m^{1+2j+2j'}}{n^2}\right).$$

We recall that $\mathrm{Bias}^2(\widehat{T_\psi^J}) = O\left(\|\psi\|_\infty^2 m^{2j+2j'-4s}\right)$ and that $m \approx n^{\frac{2}{1+4s}}$, hence

$$\text{i) If } s > j+j'+\frac{1}{4} \quad E(\widehat{T_\psi^J}-T_\psi^J)^2 \leq C_1\frac{\lambda_1^2}{n}$$
$$\text{ii) If } s \leq j+j'+\frac{1}{4} \quad E(\widehat{T_\psi^J}-T_\psi^J)^2 \leq C_2\lambda_2^2 n^{\frac{-8s+4j+4j'}{1+4s}}.$$

This achieves the proof of (3.3) and (3.5). We shall now prove (3.4).

• **Control of the covariances :**

We first notice that when $s > \sup(j+j'+\frac{1}{4}, j_1+j_1'+\frac{1}{4})$ then $j+j' \leq p$ and $j_1+j_1' \leq p$. We recall that

$$\widehat{T_\psi^J} = \frac{1}{n(n-1)}\sum_{l_1\neq l_2=1}^{n} h_{jj'\psi}(X_{l_1},X_{l_2})$$

where $h_{jj'\psi}$ is symetric. We now denote $h_{jj'\psi}$ instead of $h_{jj'}$ since $\widehat{T_\psi^J}$ is an estimator of $\int f^{(j)}f^{(j')}\psi$ while $\widehat{T_\xi^{J_1}}$ estimates $\int f^{(j_1)}f^{(j_1')}\xi$. Since

$$E(\widehat{T_\psi^J}-T_\psi^J)(\widehat{T_\xi^{J_1}}-T_\xi^{J_1}) = \mathrm{Bias}(\widehat{T_\psi^J})\mathrm{Bias}(\widehat{T_\xi^{J_1}}) + \mathrm{Cov}(\widehat{T_\psi^J},\widehat{T_\xi^{J_1}})$$

we shall use the following lemma which will be proved in the Appendix :

Lemma 3.3 *Let T and U be U-statistics respectively defined by*

$$T = \frac{1}{n(n-1)} \sum_{j \neq k = 1}^{n} a(X_j, X_k), \quad U = \frac{1}{n(n-1)} \sum_{j \neq k = 1}^{n} b(X_j, X_k)$$

where a and b are symetric. Then

$$Cov(T, U) = \frac{4(n-2)}{n(n-1)}\xi_3 + \frac{2}{n(n-1)}\xi_4$$

where

$$\xi_3 = Cov\left[\bar{a}(X_1), \bar{b}(X_1)\right], \quad \xi_4 = Cov[a(X_1, X_2), b(X_1, X_2)]$$

and

$$\bar{a}(x) = E\left[a(x, X_2)\right]; \bar{b}(x) = E\left[b(x, X_2)\right].$$

This equality is a generalization for the covariance of U-statistics of Hoeffding's formula for the variance. It follows from Lemma 3.3 that

$$\mathrm{Cov}(\widehat{T_\psi^J}, \widehat{T_\xi^{J_1}}) = \frac{4(n-2)}{n(n-1)}\xi_3 + \frac{2}{n(n-1)}\xi_4 \tag{3.19}$$

where

$$\xi_3 = \mathrm{Cov}\left[\bar{h}_{jj'\psi}(X_1), \bar{h}_{j_1,j_1'\xi}(X_1)\right], \quad \xi_4 = \mathrm{Cov}\left[h_{jj'\psi}(X_1, X_2), h_{j_1,j_1'\xi}(X_1, X_2)\right]$$

and

$$\bar{h}_{jj'\psi}(x) = E\left[h_{jj'\psi}(x, X_2)\right].$$

(3.19) implies that

$$nE(\widehat{T_\psi^J} - T_\psi^J)(\widehat{T_\xi^{J_1}} - T_\xi^{J_1}) = n\mathrm{Bias}(\widehat{T_\psi^J})\mathrm{Bias}(\widehat{T_\xi^{J_r}}) + \frac{4(n-2)}{n-1}\xi_3 + \frac{2}{n-1}\xi_4.$$

Using the results obtained in the first part of the proof, we get

$$\begin{aligned} n|\mathrm{Bias}(\widehat{T_\psi^J})\mathrm{Bias}(\widehat{T_\xi^{J_1}})| &\leq C_1'\|\psi\|_\infty\|\xi\|_\infty n m^{j+j'+j_1+j_1'-4s} \\ &\leq C_1'\mu_1^2 m^{1/2+j+j'+j_1+j_1'-2s} \end{aligned}$$

since $n \approx m^{1/2+2s}$. We want to prove that $\frac{2}{n-1}\xi_4$ is bounded by a similar quantity.

$$\xi_4 \leq \left(E(h_{jj'\psi}^2(X_1, X_2))\right)^{1/2}\left(E(h_{j_1j_1'\xi}^2(X_1, X_2))\right)^{1/2}$$

From (3.17) which is actually an upper bound for $E(h_{jj'\psi}^2(X_1, X_2))$ we derive, possibly enlarging C_1'

$$\frac{2}{n-1}\xi_4 \leq C_1' \frac{\mu_1^2}{n} m^{1+j+j'+j_1+j_1'} = C_1' \mu_1^2 m^{1/2+j+j'+j_1+j_1'-2s}.$$

It follows that the asymptotic covariance $\Lambda_{JJ_1}(f,\psi)$ can only come from the term $\frac{4(n-2)}{n-1}\xi_3$. Since, $|\xi_3| \leq C_1'\mu_1^2$, (this result follows from the computation of ξ_1), to show (3.4), we just have to prove that

$$|4\xi_3 - \Lambda_{JJ_1}(f,\psi,\xi)| \leq C_1'\mu_1^2 m^{-\alpha}.$$

We recall that

$$\xi_3 = \mathrm{Cov}\left(\bar{h}_{jj'\psi}(X_1), \bar{h}_{j_1j_1'\xi}(X_1)\right).$$

Moreover,

$$\bar{h}_{jj'\psi}(X_1) = \frac{1}{2}\sum_{l=1}^{3}(Y_l^{jj'} + Y_l^{j'j})$$

where $Y_l^{jj'}$, $l = 1,2,3$, are respectively defined by (3.12), (3.13) and (3.14). We have to keep in mind that $Y_l^{jj'}$ depends on ψ while $Y_l^{j_1j_1'}$ depends on ξ.

$$4\mathrm{Cov}\left(\bar{h}_{jj'\psi}(X_1), \bar{h}_{j_1j_1'\xi}(X_1)\right) = \sum_{l,l'=1}^{3} \mathrm{Cov}\left((Y_l^{jj'} + Y_l^{j'j}), (Y_{l'}^{j_1j_1'} + Y_{l'}^{j_1'j_1})\right).$$

We claim that $\forall\ l,l' \in \{1,2,3\}^2$ that

$$\left|\mathrm{Cov}\left((Y_l^{jj'} + Y_l^{j'j}), (Y_{l'}^{j_1j_1'} + Y_{l'}^{j_1'j_1})\right) - \epsilon_{ll'}\Lambda_{JJ_1}(f,\psi,\xi)\right| \leq C_1'\mu_1^2 m^{-\alpha} \tag{3.20}$$

where

$$\begin{aligned}
\epsilon_{ll'} &= -1 \quad \text{if } (l,l') = (1,3),(2,3),(3,1),(3,2) \\
\epsilon_{ll'} &= 1 \quad \text{if } (l,l') = (1,1),(2,2),(3,3),(1,2),(2,1)
\end{aligned}$$

Of course, this will achieve the proof of (3.4). We give a complete proof for $l = 1$, and $l' = 3$, since the computations are similar for the other cases. We recall that

$$\begin{aligned}
Y_1^{jj'} &= (-1)^j S_m\left[(f^{(j')}\psi)^{(j)}\right](X_1) \\
Y_3^{j_1j_1'} &= -(-1)^{j_1} S_m\left[\left(S_m f^{(j_1')}\xi\right)^{(j_1)}\right](X_1)
\end{aligned}$$

Then,

$$\mathrm{Cov}\left((Y_1^{jj'} + Y_1^{j'j}), (Y_3^{j_1j_1'} + Y_3^{j_1'j_1})\right)$$

$$
\begin{aligned}
= & -\int \left[(-1)^j S_m\left((f^{(j')}\psi)^{(j)}\right) + (-1)^{j'} S_m\left((f^{(j)}\psi)^{(j')}\right)\right] \\
\times & \left[(-1)^{j_1} S_m\left((S_m f^{(j_1')}\xi)^{(j_1)}\right) + (-1)^{j_1'} S_m\left((S_m f^{(j_1)}\xi)^{(j_1')}\right)\right] f \\
+ & \left[\int S_m(f^{(j')}\psi) f^{(j)} + S_m(f^{(j)}\psi) f^{(j')}\right] \times \left[\int S_m(S_m f^{(j_1')}\xi) f^{(j_1)} + S_m(S_m f^{(j_1)}\xi) f^{(j_1')}\right],
\end{aligned}
$$

using integrations by parts for the last line.
As $n \to \infty$, this quantity converges towards $-\Lambda_{JJ_1}(f,\psi,\xi)$. More precisely, the difference between this and $-\Lambda_{JJ_1}(f,\psi,\xi)$ is bounded by $C_1'\mu_1^2 m^{-\alpha}$, possibly enlarging C_1'. The proof will be done in two steps :
a) We will first show that

$$
\left|\int S_m\left((f^{(j')}\psi)^{(j)}\right) S_m\left((S_m f^{(j_1')}\xi)^{(j_1)}\right) f - \int \left(f^{(j')}\psi\right)^{(j)} \left(f^{(j_1')}\xi\right)^{(j_1)} f\right| \leq C_1'\mu_1^2 m^{-\alpha}.
$$

b) After that, we will show that

$$
\left|\int S_m\left(f^{(j')}\psi\right) f^{(j)} \int S_m\left(S_m f^{(j_1')}\xi\right) f^{(j_1)} - \int f^{(j)} f^{(j')}\psi \int f^{(j_1)} f^{(j_1')}\xi\right| \leq C_1'\mu_1^2 m^{-\alpha}.
$$

Of course, the computations are similar for the other terms of the covariance since we just have to exchange j and j', j_1 and j_1'.

- Proof of a)

Introducing the auxiliary term $S_m\left(f^{(j')}\psi\right)^{(j)}\left(f^{(j_1')}\xi\right)^{(j_1)} f$, the absolute value of the difference is bounded by

$$
\begin{aligned}
& \|f\|_\infty \int \left|S_m\left(f^{(j')}\psi\right)^{(j)} S_m\left((S_m f^{(j_1')}\xi)^{(j_1)}\right) - S_m\left(f^{(j')}\psi\right)^{(j)}\left(f^{(j_1')}\xi\right)^{(j_1)}\right| \\
& + \|f\|_\infty \int \left|S_m\left(f^{(j')}\psi\right)^{(j)}\left(f^{(j_1')}\xi\right)^{(j_1)} - \left(f^{(j')}\psi\right)^{(j)}\left(f^{(j_1')}\xi\right)^{(j_1)}\right|
\end{aligned}
$$

which, by Cauchy-Schwarz inequality is bounded by

$$
\begin{aligned}
& \|f\|_\infty \|S_m\left(f^{(j')}\psi\right)^{(j)}\|_2 \|S_m\left((S_m f^{(j_1')}\xi)^{(j_1)}\right) - \left(f^{(j_1')}\xi\right)^{(j_1)}\|_2 & (3.21)\\
+ & \|f\|_\infty \|\left(f^{(j_1')}\xi\right)^{(j_1)}\|_2 \|S_m\left(f^{(j')}\psi\right)^{(j)} - \left(f^{(j')}\psi\right)^{(j)}\|_2. & (3.22)
\end{aligned}
$$

$$
\begin{aligned}
\bullet \|S_m\left(\psi f^{(j')}\right)^{(j)}\|_2 & \leq \|\left(\psi f^{(j')}\right)^{(j)}\|_2 \quad \text{since } S_m \text{ is a projection} \\
& = O\left(\sup_{l=0\ldots j'} \|\psi^{(l)}\|_\infty\right) = O(\mu_1).
\end{aligned}
$$

$$\bullet \| S_m\left((S_m f^{(j_1')}\xi)^{(j_1)}\right) - \left(f^{(j_1')}\xi\right)^{(j_1)} \|_2 \;\leq\; \| S_m\left((S_m f^{(j_1')}\xi)^{(j_1)}\right) - S_m\left((f^{(j_1')}\xi)^{(j_1)}\right) \|_2 + \| S_m\left((f^{(j_1')}\xi)^{(j_1)}\right) - \left(f^{(j_1')}\xi\right)^{(j_1)} \|_2.$$

Since S_m is a projection, the first $\mathbb{L}^2$ norm is bounded by

$$\begin{aligned} &\| \left(S_m f^{(j_1')}\xi\right)^{(j_1)} - \left(f^{(j_1')}\xi\right)^{(j_1)} \|_2 \\ = &\; \|\sum_{l=0}^{j_1} \binom{j_1}{l} \left(S_m f^{(j_1+j_1'-l)} - f^{(j_1+j_1'-l)}\right) \xi^{(l)} \|_2 \\ = &\; O\left[\sup_{l=0,..j_1'} \|\xi^{(l)}\|_\infty \|S_m f^{(j_1+j_1'-l)} - f^{(j_1+j_1'-l)}\|_2\right] \end{aligned}$$

Since $j_1 + j_1' - l \leq p, \forall\, l \in \{0, \ldots, j_1'\}$ by Lemma 3.1,

$$\|S_m f^{(j_1+j_1'-l)} - f^{(j_1+j_1'-l)}\|_2 = O(m^{-\alpha})$$

and the above $\mathbb{L}^2$ norm is $O(\mu_1 m^{-\alpha})$.

We now bound the second term

$$\|S_m\left((f^{(j_1')}\xi)^{(j_1)}\right) - \left(f^{(j_1')}\xi\right)^{(j_1)} \|_2 = \left(\sum_{m+1}^{\infty} b_i^2\right)^{1/2}$$

where b_i denotes the i^{th} Fourier coefficient of the function $(f^{(j_1')}\psi)^{(j_1)}$.
By Lemma 3.1,

$$\left(\sum_{m+1}^{\infty} b_i^2\right)^{1/2} = O\left(m^{-\alpha}\|(f^{(j_1')}\xi)^{(j_1)}\|_\alpha\right)$$

moreover $\|(f^{(j_1')}\xi)^{(j_1)}\|_\alpha = O(\mu_1)$.

It follows from the above computations that (3.21) is bounded by $C_1' \mu_1^2 m^{-\alpha}$. Of course, the same result holds for (3.22) by analoguous computations. This achieves the proof of a).

• Proof of b)
As above, we introduce an auxiliary term, so that the expression defined in b) is equal to

$$| \int S_m\left(f^{(j')}\psi\right) f^{(j)} \int S_m\left(S_m f^{(j_1')}\xi\right) f^{(j_1)} - \int S_m\left(f^{(j')}\psi\right) f^{(j)} \int f^{(j_1)} f^{(j_1')}\xi$$

$$+ \int S_m\left(f^{(j')}\psi\right) f^{(j)} \int f^{(j_1)} f^{(j_1')}\xi - \int f^{(j)} f^{(j')}\psi \int f^{(j_1)} f^{(j_1')}\xi|$$

$$\leq \int \left|S_m\left(f^{(j')}\psi\right) f^{(j)}\right| \left[\int \left|S_m\left(S_m f^{(j_1')}\xi\right) f^{(j_1)} - f^{(j_1)} f^{(j_1')}\xi\right|\right]$$

$$+ \left[\int \left|S_m\left(f^{(j')}\psi\right) f^{(j)} - f^{(j)} f^{(j')}\psi\right|\right] \int \left|f^{(j_1)} f^{(j_1')}\xi\right|$$

By Cauchy-Schwarz inequality, this expression is bounded by

$$\|f^{(j)}\|_2 \|S_m\left(f^{(j')}\psi\right)\|_2 \|f^{(j_1)}\|_2 \|S_m\left(S_m f^{(j_1')}\xi\right) - f^{(j_1')}\xi\|_2 \tag{3.23}$$

$$+\|\xi\|_2 \|f^{(j_1)} f^{(j_1')}\|_2 \|f^{(j)}\|_2 \|S_m\left(f^{(j')}\psi\right) - f^{(j')}\psi\|_2 \tag{3.24}$$

We first majorize (3.23) by

$$2\pi \|f^{(j)}\|_\infty \|f^{(j_1)}\|_\infty \|f^{(j')}\psi\|_2 \|S_m\left(S_m f^{(j_1')}\xi\right) - f^{(j_1')}\xi\|_2$$

Moreover

$$\|f^{(j')}\psi\|_2 \leq \sqrt{2\pi}\|\psi\|_\infty \|f^{(j')}\|_\infty = O(\mu_1).$$

We now compute

$$\|S_m\left(S_m f^{(j_1')}\xi\right) - f^{(j_1')}\xi\|_2.$$

It is bounded by

$$\|S_m\left(S_m f^{(j_1')}\xi\right) - S_m(f^{(j_1')}\xi)\|_2 + \|S_m(f^{(j_1')}\xi) - f^{(j_1')}\xi\|_2.$$

Since S_m is a projection

$$\|S_m\left(S_m f^{(j_1')}\xi\right) - S_m(f^{(j_1')}\xi)\|_2 \leq \|S_m f^{(j_1')}\xi - f^{(j_1')}\xi\|_2$$

$$\leq \|\xi\|_\infty \|S_m f^{(j_1')} - f^{(j_1')}\|_2 = O(\mu_1 m^{-\alpha}).$$

As to the term $\|S_m(f^{(j_1')}\xi) - f^{(j_1')}\xi\|_2$, it is bounded by Lemma 3.1 by

$$\|f^{(j_1')}\xi\|_\alpha m^{-\alpha} = O(\mu_1 m^{-\alpha}).$$

Finally, the expression defined by (3.23) has order $O(\mu_1^2 m^{-\alpha})$. The computations are similar for (3.24); this ahieves the proof of b). We finally get (3.20) for $l = 1$ and $l' = 3$, hence (3.4). This achieves the proof of Theorem 3.1. □

3.4.2 Proof of Theorem 3.2

Before specializing to case i) or ii), we first control the remainder term Γ_n.

$$|\Gamma_n| \leq \frac{1}{6}\|\phi^{(3)}\|_\infty \sum_{j,j',j''=0}^{k} \int |f^{(j)} - \widehat{f}^{(j)}||f^{(j')} - \widehat{f}^{(j')}||f^{(j'')} - \widehat{f}^{(j'')}|(x)dx.$$

Hence,

$$E(\Gamma_n^2) = O\left(\sup_{jj'j''} E\left[\left(\int_{-\pi}^{\pi} |\widehat{f}^{(j)} - f^{(j)}||\widehat{f}^{(j')} - f^{(j')}||\widehat{f}^{(j'')} - f^{(j'')}|\right)^2\right]\right).$$

Using first Cauchy-Schwarz, and then the generalized Hölder's inequality we get

$$E(\Gamma_n^2) = O\left(\sup_{j=0,\dots,k} E\left(\|\widehat{f}^{(j)} - f^{(j)}\|_6^6\right)\right).$$

Using Assumption A2, we see that $E(\Gamma_n^2) = O(n_1^{\frac{-6(s-k)}{1+2s}})$.

$$\text{i)}\quad \text{If}\quad s > 2k + \frac{1}{4} \text{ then } E(\Gamma_n^2) = o(\frac{1}{n}). \tag{3.25}$$

$$\text{ii)}\quad \text{If}\quad \frac{1}{4} < s \leq 2k + \frac{1}{4} \text{ then } E(\Gamma_n^2) = o(n^{\frac{-8(s-k)}{1+4s}}). \tag{3.26}$$

This proves that in both cases the remainder term is negligible.

• Proof in case ii) :

We will now look at the order of magnitude of $E(\widehat{T_n} - T(f))^2$ in order to obtain the rate of convergence in the nonparametric case.

$$\widehat{T_n} - T(f) = \sum_{j=0}^{k} \frac{1}{n_2} \sum_{l=1}^{n_2} \left((-1)^j (H_j(\widehat{f}))^{(j)}(X_l) - \int_{-\pi}^{\pi} H_j(\widehat{f}) f^{(j)}\right)$$

$$+ \sum_{j,j'=0}^{k} \left(\widehat{T^J}(K_J(\widehat{f})) - T^J(K_J(\widehat{f}))\right) + \Gamma_n$$

where $K_J(\widehat{f}) = \frac{1}{2}\phi''_{jj'}(\widehat{f},\dots \widehat{f}^{(k)},.,.)$, $\quad T^J(K_J(\widehat{f})) = \int_{-\pi}^{\pi} f^{(j)} f^{(j')} K_J(\widehat{f})$.

The expectation of the square of the linear term

$\sum_{j=0}^{k} \frac{1}{n_2} \sum_{l=1}^{n_2} \left((-1)^j (H_j(\widehat{f}))^{(j)}(X_l) - \int_{-\pi}^{\pi} H_j(\widehat{f}) f^{(j)}\right)$ is $O\left[\frac{1}{n_2} \sup_{j\in\{0,\dots k\}} E\left(\|(H_j(\widehat{f}))^{(j)}\|_\infty^2\right)\right]$.

As to the estimation of the quadratic term $\sum_{j,j'=0}^{k} \int_{-\pi}^{\pi} f^{(j)} f^{(j')} K_J(\widehat{f})$ we shall apply the results of Section 2. In Section 2, we studied the properties of estimators of $\int_{-\pi}^{\pi} f^{(j)} f^{(j')} \psi$. Here, ψ is the random function $K_J(\widehat{f})$ and $T^J(K_J(\widehat{f}))$ is estimated independently of $\widehat{f}$. Therefore, working conditionnally to $\widehat{f}$ we can apply the results of Section 2 . $\widehat{T^J}(K_J(\widehat{f}))$ will now be denoted by $\widehat{T^J}$.

We have to determine the order of magnitude of $E(\|(K_J(\widehat{f}))^{(l)}\|_{\infty}^2)$ and $E(\|K_J(\widehat{f}))^{(l)}\|_{\alpha}^2)$ for $l \in \{1, \ldots k\}$.
We recall that $K_J(\widehat{f}) = \frac{1}{2}\phi''_{jj'}(\widehat{f}, \ldots \widehat{f}^{(k)}, .)$ hence we can see that $(K_J(\widehat{f}))^{(l)}$ involve sum of terms of the type

$$\sum_{i=1}^{l} \phi^{(2+i)}_{jj'j_1\ldots j_i}(\widehat{f}, \ldots \widehat{f}^{(k)}, .)\widehat{f}^{(j_1+n_1)} \times \ldots \times \widehat{f}^{(j_i+n_i)} \, \mathbb{I}_{n_1+\ldots n_i = l}.$$

Since $\phi \in C^{k+3}(K_\epsilon)$, $\|\phi^{(2+i)}\|_{\infty} < \infty \quad \forall\, i \in \{1, \ldots k\}$ and we have to dertermine the order of magnitude of $E(\|\widehat{f}^{(j_1+n_1)} \times \ldots \widehat{f}^{(j_i+n_i)}\|_{\infty})$ and $E(\|\widehat{f}^{(j_1+n_1)} \times \ldots \widehat{f}^{(j_i+n_i)}\|_{\alpha})$ where $n_1 + \ldots + n_i = l$.
Put $h = n_1^{\frac{-1}{1+2s}}$. By condition d) of assumption A2 and Hölder's inequality, it follows that

$$E(\|\widehat{f}^{(j_1+n_1)} \ldots \widehat{f}^{(j_i+n_i)}\|_{\infty}^2) = O\left(h^{2(s'-j_1-n_1)} + 1\right) \times \ldots \left(h^{2(s'-j_i-n_i)} + 1\right) \quad \forall\, s' < s.$$

Let $k < s' < s$, since $\sum_{\nu=1}^{i} n_\nu = l$ and $j_\nu \leq k$, we get

$$E(\|\widehat{f}^{(j_1+n_1)} \ldots \widehat{f}^{(j_i+n_i)}\|_{\infty}^2) = O(h^{2(s'-k-l)} + 1). \tag{3.27}$$

Moreover, since

$$\widehat{f}^{(j_1+n_1)} \ldots \widehat{f}^{(j_i+n_i)}(x) - \widehat{f}^{(j_1+n_1)} \ldots \widehat{f}^{(j_i+n_i)}(y) =$$

$$\sum_{i_0=0}^{i-1} \left(\left[\prod_{i_1=1}^{i-i_0} \widehat{f}^{(j_{i_1}+n_{i_1})}(x) \prod_{i_1=i-i_0+1}^{i} \widehat{f}^{(j_{i_1}+n_{i_1})}(y) \right] - \left[\prod_{i_1=1}^{i-i_0-1} \widehat{f}^{(j_{i_1}+n_{i_1})}(x) \prod_{i_1=i-i_0}^{i} \widehat{f}^{(j_{i_1}+n_{i_1})}(y) \right] \right)$$

we get

$$\|\widehat{f}^{(j_1+n_1)} \ldots \widehat{f}^{(j_i+n_i)}\|_{\alpha} \leq \sum_{i_0=1}^{i} (\prod_{i_1 \neq i_0} \|\widehat{f}^{(j_{i_1}+n_{i_1})}\|_{\infty}) \|\widehat{f}^{(j_{i_0}+n_{i_0})}\|_{\alpha}.$$

It follows from (3.27) that

$$E\left(\prod_{i_1\neq i_0}\|\widehat{f}^{(j_{i_1}+n_{i_1})}\|_\infty^2\ \mathbb{1}_{\sum_{i_1\neq i_0} n_{i_1}=l-i_0}\right)=O(h^{2(s'-k-l+n_{i_0})}+1).$$

Moreover, using condition e) of assumption A2 we get

$$\begin{aligned} E(\|\widehat{f}^{(j_1+n_1)}\dots\widehat{f}^{(j_i+n_i)}\|_\alpha^2) &= O\left[\sum_{i_0=1}^{i}\left(h^{2(s'-k-l+n_{i_0})}+1\right)\left(h^{2(s'-k-\alpha-n_{i_0})}+1\right)\right] \\ &= O(h^{2(s'-k-\alpha-l)}+1) \quad \text{since} \quad s'-k>0. \end{aligned}$$

So,

$$E(\|(K_J(\widehat{f}))^{(l)}\|_\infty^2)=O\left(1+h^{2(s'-k-l)}\right)$$
$$E(\|(K_J(\widehat{f}))^{(l)}\|_\alpha^2)=O\left(1+h^{2(s'-k-l-\alpha)}\right)$$

By similar arguments, we can see that

$$\forall\ j\in\{0,\dots k\}\quad E(\|(H_j(\widehat{f}))^{(j)}\|_\infty^2)=O(h^{2s'-4k}+1).$$

Hence, the expectation of the square of the linear term has order $O\left(\frac{h^{2s'-4k}+1}{n_2}\right)=O\left(n^{\frac{-8(s-k)}{1+4s}}+\frac{1}{n}\right)$ since s' is arbitrary close to s and $\frac{n}{n_2}\to 1$.

We can now apply the results of Theorem 3.1 to get the order of magnitude of $E(\widehat{T^J}-T^J)^2$. Since $s\le 2k+\frac{1}{4}$ we have to consider the following cases

• if $j+j'+\frac{1}{4}<s$ then

$$\sup_{l=0\dots j'} E\left(\|(K_J(\widehat{f}))^{(l)}\|_\infty^2\right)=O(h^{2(s'-k-j')})=O(h^{2s'-4k})$$

since $j'\le k$. Hence, by Theorem 3.1,

$$E(\widehat{T^J}-T^J)^2\le C_1\frac{h^{2s'-4k}}{n_2}\le C_1 n^{\frac{-8(s-k)}{1+4s}}.$$

• If $j+j'+\frac{1}{4}\ge s$ we have to evaluate the order of magnitude of $E(\lambda_2^2)\ \forall\ j,j'$. We want to prove that $E(\lambda_2^2)=O(h^{2j-2k})$. To do this, we shall prove that

$$\text{a)}\quad \sup_{l=0,\dots,p-j} E\left[\frac{(K_J(\widehat{f}))^{(l)}\|_\infty^2+\|(K_J(\widehat{f}))^{(l)}\|_\alpha^2)}{m^{1/3}}\right]=O(h^{2j-2k})$$

$$\text{b)}\quad \sup_{l=0\dots j'} E\left[\frac{\|(K_J(\widehat{f}))^{(l)}\|_\infty^2}{m^{(2j+2j'-2s+\frac{1}{2})\wedge 2l}}\right]=O(h^{2j-2k}).$$

Proof of a) :

$$\forall\, l=0,\dots p-j,\quad E\left[\frac{\|(K_J(\widehat{f}))^{(l)}\|_\infty^2+\|(K_J(\widehat{f}))^{(l)}\|_\alpha^2}{m^{1/3}}\right] = O\left(\frac{h^{2(s'-k-p+j-\alpha)}}{m^{1/3}}\right) = O\left(h^{2j-2k}\right)\left(h^{2s'-2p-2\alpha}m^{-1/3}\right).$$

Since s' is arbitrary close to $s=p+\alpha$, we get $\lim_{n\to\infty} h^{2s'-2p-2\alpha}m^{-1/3}=0$, this conclude the proof of a).

Proof of b) : We first assume that $2j+2j'-2s+\frac{1}{2}\le l$, then we have to show that

$$E\left(\frac{\|K_J(\widehat{f}))^{(l)}\|_\infty^2}{m^{(2j+2j'-2s+\frac{1}{2})}}\right)=O(h^{2j-2k})\quad \forall\, l\le j'.$$

From the above evaluations of $E(\|K_J(\widehat{f}))^{(l)}\|_\infty^2)$ we derive

$$E\left(\frac{\|K_J(\widehat{f}))^{(l)}\|_\infty^2}{m^{(2j+2j'-2s+\frac{1}{2})}}\right) = O\left(\frac{1+h^{2s'-2k-2l}}{m^{(2j+2j'-2s+\frac{1}{2})}}\right) = O\left(h^{2j-2k}\right)\left(h^{2k-2j}m^{-2j-2j'+2s-1/2}+h^{2s'-2j-2l}m^{-2j-2j'+2s-1/2}\right).$$

Since $j\le k$ and since $s\le j+j'+\frac{1}{4}$, $h^{2k-2j}m^{-2j-2j'+2s-1/2}\to 0$. Moreover, $\forall\, l\le j'$

$$h^{2s'-2j-2l}m^{-2j-2j'+2s-1/2} \le h^{2s'-2j-2j'}m^{-2j-2j'+2s-1/2} \le (mh)^{2s-2j-2j'}h^{2s'-2s}m^{-1/2}.$$

Using the fact that mh is a positive power of n and that $s\le j+j'+\frac{1}{4}$, the above expression is bounded by

$$(mh)^{1/2}h^{2s'-2s}m^{-1/2}=h^{1/2+2s'-2s}$$

which converges towards zero since s' is arbitrary close to s.

To conclude the proof of b), we have to consider the case of $l>2j+2j'-2s+\frac{1}{4}$, and show that

$$E\left(\frac{(\|K_J(\widehat{f}))^{(l)}\|_\infty^2}{m^{2l}}\right)=O(h^{2j-2k}).$$

$$E\left(\frac{(\|K_J(\widehat{f}))^{(l)}\|_\infty^2}{m^{2l}}\right) = O\left((h^{2s'-2k-2l}+1)m^{-2l}\right) = O\left(h^{2j-2k}\right)\left(h^{2s'-2j}(mh)^{-2l}+h^{2k-2j}m^{-2l}\right) = O(h^{2j-2k}).$$

This achieves the proof of b).

Finally, if $s \leq 2k + \frac{1}{4}$

$$
\begin{aligned}
E\left(\left(\sum_{j,j'=0}^{k} \widehat{T^J} - T^J\right)^2\right) &= O\left(\sum_{j,j'=0}^{k} h^{2j-2k} n^{\frac{-8s+4j+4j'}{1+4s}} + n^{\frac{-8(s-k)}{1+4s}}\right) \\
&= O\left(\sum_{j,j'=0}^{k} h^{2j-2k} m^{-4s+2j+2j'} + n^{\frac{-8(s-k)}{1+4s}}\right) \\
&= O\left(\sum_{j,j'=0}^{k} (mh)^{2j-2k} m^{-4s+2j'+2k} + n^{\frac{-8(s-k)}{1+4s}}\right) \\
&= O\left(m^{4k-4s} + n^{\frac{-8(s-k)}{1+4s}}\right) \quad \text{since} \quad mh \to \infty, j \leq k, j' \leq k \\
&= O\left(n^{\frac{-8(s-k)}{1+4s}}\right)
\end{aligned}
$$

Collecting the above evaluations and the computation of the remainder and linear terms, we get in case ii)

$$E(\widehat{T_n} - T(f))^2 = O(n^{\frac{-8(s-k)}{1+4s}}).$$

• Proof in case i) :

We are now going to prove the asymptotic efficienty. Let

$$R = \sqrt{n}\left[\widehat{T_n} - T(f) - \frac{1}{n_2}\sum_{l=1}^{n_2}\sum_{j=0}^{k}(-1)^j \left(\phi_j'(f, ..f^{(k)}, .)\right)^{(j)}(X_l) + \int \phi_j'(f, ..f^{(k)}, .)f^{(j)}\right].$$

Of course, to ensure that both (3.9) and (3.10) hold, it is enough to show that $E(R^2) \to 0$. We notice that $R = R_1 + R_2$ where

$$R_1 = \sqrt{n}\left[\widehat{T_n} - T(f) - \frac{1}{n_2}\sum_{l=1}^{n_2}\left(\sum_{j=0}^{k}(-1)^j(\phi_j'(\hat{f}, ..\hat{f}^{(k)}, .))^{(j)}(X_l) - \int_{-\pi}^{\pi}\phi_j'(\hat{f}, ..\hat{f}^{(k)}, .)f^{(j)}\right)\right]$$

and

$$R_2 = -\sqrt{n}\sum_{j=0}^{k}\left[\frac{1}{n_2}\sum_{l=1}^{n_2}(-1)^j \left(\phi_j'(f, \ldots f^{(k)}, .) - \phi_j'(\hat{f}, \ldots \hat{f}^{(k)}, .)\right)^{(j)}(X_l)\right]$$

$$+\sqrt{n}\sum_{j=0}^{k}\left[\int_{-\pi}^{\pi}\phi_j'(f,\dots f^{(k)},.)f^{(j)}-\int_{-\pi}^{\pi}\phi_j'(\widehat{f},\dots\widehat{f}^{(k)},.)f^{(j)}\right].$$

We want to prove that both $E(R_1^2)$ and $E(R_2^2)\to 0$. Pluging the expression of $\widehat{T_n}$ and $T(f)$ in R_1 we get

$$R_1=\sqrt{n}\sum_{j,j'=0}^{k}\left[\frac{-2}{n_2}\sum_{l=1}^{n_2}(-1)^j\left(K_J(\widehat{f})\widehat{f}^{(j')}\right)^{(j)}(X_l)+2\int_{-\pi}^{\pi}K_J(\widehat{f})\widehat{f}^{(j')}f^{(j)}\right]$$

$$+\sqrt{n}\sum_{j,j'=0}^{k}\left[\widehat{T^J}-T^J+\Gamma_n\right].$$

Because of (3.25), we just have to prove that the expectation of the square of $\sqrt{n}\sum_{jj'=0}^{k}(\widehat{L^J}-L^J+\widehat{T^J}-T^J)$ converges towards zero, where $L^J=2\int K_J(\widehat{f})\widehat{f}^{(j')}f^{(j)}$ and $\widehat{L^J}$ is the corresponding estimator. We will first evaluate $nE(\sum_{j,j'=0}^{k}\widehat{L^J}-L^J)^2$ and $nE(\sum_{j,j'=0}^{k}\widehat{T^J}-T^J)^2$, and then we will show that the sum of these terms is the opposite of the covariance term $2nE(\sum_{j,j'=0}^{k}\widehat{L^J}-L^J)(\sum_{j,j'=0}^{k}\widehat{T^J}-T^J)$.

- Calculation of $\lim_{n\to\infty}nE(\sum_{j,j'=0}^{k}\widehat{L^J}-L^J)^2$:

$$\begin{aligned}nE\left[(\widehat{L^J}-L^J)(\widehat{L^{J_1}}-L^{J_1})|\widehat{f}\right] &= 4\frac{n}{n_2}\int_{-\pi}^{\pi}(-1)^{j+j_1}\left(K_J(\widehat{f})\widehat{f}^{(j')}\right)^{(j)}\left(K_{J_1}(\widehat{f})\widehat{f}^{(j_1')}\right)^{(j_1)}f\\ &- 4\frac{n}{n_2}\int_{-\pi}^{\pi}K_J(\widehat{f})\widehat{f}^{(j')}f^{(j)}\int_{-\pi}^{\pi}K_{J_1}(\widehat{f})\widehat{f}^{(j_1')}f^{(j_1)}.\end{aligned}$$

The expectation of this expression converges towards the same expression with f instead of $\widehat{f}$. So,

$$\begin{aligned}\lim_{n\to\infty}nE\left[(\sum_{j,j'=0}^{k}\widehat{L^J}-L^J)^2\right] &= 4\sum_{j,j',j_1,j_1'=0}^{k}\int_{-\pi}^{\pi}(-1)^{j+j_1}(K_J(f)f^{(j')})^{(j)}(K_{J_1}(f)f^{(j_1')})^{(j_1)}f\\ &- 4\sum_{j,j'=0}^{k}\sum_{j_1,j_1'=0}^{k}\int_{-\pi}^{\pi}K_J(f)f^{(j')}f^{(j)}\int_{-\pi}^{\pi}K_{J_1}(f)f^{(j_1')}f^{(j_1)}\\ &= \sum_{j,j',j_1,j_1'=0}^{k}\Lambda_{JJ_1}(f,K_J(f),,K_{J_1}(f)).\end{aligned}$$

• Calculation of $\lim_{n\to\infty} nE\left[(\sum_{j,j'=0}^{k} \widehat{T^J} - T^J)^2\right]$.

Using the results of Section 2 we see that

$$\left|nE\left((\widehat{T^J} - T^J)(\widehat{T^{J_1}} - T^{J_1})|\hat{f}\right) - \Lambda_{JJ_1}(\hat{f}, K_J(\hat{f}), K_{J_1}(\hat{f}))\right| \leq C_1'\mu_1^2(m^{-\alpha}+m^{1/2+j+j'+j_1+j_1'-2s})$$

where

$$\mu_1 = \sup\left[\sup_{l=0,\ldots j'}\left(\|(K_J(\hat{f}))^{(l)}\|_\infty + \|(K_J(\hat{f}))^{(l)}\|_\alpha\right); \sup_{l=0\ldots j_1'}\left(\|(K_{J_1}(\hat{f}))^{(l)}\|_\infty + \|(K_{J_1}(\hat{f}))^{(l)}\|_\alpha\right)\right].$$

As already proved, $E(\mu_1^2) = O(h^{2s'-4k-2\alpha}+1)$. Since s' is arbitrary close to s and $s > 2k+\frac{1}{4}$ we get $\lim_{n\to\infty} E\left(C_1'\mu_1^2(m^{-\alpha}+m^{1/2+j+j'+j_1+j_1'-2s})\right) = 0$. So,

$$\begin{aligned}\lim_{n\to\infty} nE(\widehat{T^J} - T^J)(\widehat{T^{J_1}} - T^{J_1}) &= \lim_{n\to\infty} E(\Lambda_{JJ_1}(\hat{f}, K_J(\hat{f}), K_{J_1}(\hat{f})) \\ &= \Lambda_{JJ_1}(f, K_J(f), K_{J_1}(f)),\end{aligned}$$

and

$$\lim_{n\to\infty} nE\left[(\sum_{j,j'=0}^{k} \widehat{T^J} - T^J)^2\right] = \lim_{n\to\infty} nE\left[(\sum_{j,j'=0}^{k} \widehat{L^J} - L^J)^2\right] = \sum_{j,j',j_1,j_1'=0}^{k} \Lambda_{JJ_1}(f, K_J(f), K_{J_1}(f)).$$

• Calculation of $\lim_{n\to\infty} 2nE\left[(\sum_{j,j'=0}^{k} \widehat{L^J} - L^J)(\sum_{j,j'=0}^{k} \widehat{T^J} - T^J)\right]$.

We will show that

$$\lim_{n\to\infty} 2nE\left[(\sum_{j,j'=0}^{k} \widehat{L^J} - L^J)(\sum_{j,j'=0}^{k} \widehat{T^J} - T^J)\right] = -2\sum_{j,j',j_1,j_1'=0}^{k} \Lambda_{JJ_1}(f, K_J(f), K_{J_1}(f)).$$

We first notice that

$$E\left[(\widehat{T^J} - T^J)(\widehat{L^{J_1}} - L^{J_1})|\hat{f}\right] = E(\widehat{T^J}\widehat{L^{J_1}}|\hat{f}) - E(\widehat{T^J}|\hat{f})L^{J_1}.$$

We recall that

$$\widehat{L^J} = \frac{-2}{n_2}\sum_{l=1}^{n_2}(-1)^j(K_J(\hat{f})\hat{f}^{(j')})^{(j)}(X_{l_2}), \quad L^J = -2\int K_J(\hat{f})\hat{f}^{(j')}f^{(j)}$$

and that $\widehat{T^J} = \widehat{T_1^J} + \widehat{T_2^J}$ where $\widehat{T_1^J} = \widehat{T_0^{jj'}} + \widehat{T_0^{j'j}}$, and

$$\widehat{T_0^{jj'}} = \frac{1}{n_2(n_2-1)} \sum_{i=0}^{m} \sum_{l_1 \neq l_2 = 1}^{n_2} (-1)^j p_i(X_{l_1})(K_J(\hat{f})p_i^{(j')})^{(j)}(X_{l_2});$$

$$\widehat{T_2^J} = -\frac{1}{n_2(n_2-1)} \sum_{i,i'=0}^{m} \sum_{l_1 \neq l_2 = 1}^{n_2} p_i(X_{l_1})p_{i'}(X_{l_2}) \int_{-\pi}^{\pi} p_i^{(j)} p_{i'}^{(j')} K_J(\hat{f}).$$

We first do the computations for $\widehat{T_0^{jj'}}$ instead of $\widehat{T^J}$, of course, the computations are simular for $\widehat{T_0^{j'j}}$.

• Computation of $E(\widehat{T_0^{jj'}}\widehat{L^{J_1}}|\hat{f}) - E(\widehat{T_0^{jj'}}|\hat{f})L^{J_1}$:

$$\begin{aligned} E(\widehat{T_0^{jj'}}\widehat{L^{J_1}}|\hat{f}) = & - \frac{2}{n_2}\left(\sum_{i=0}^{m} \int_{-\pi}^{\pi} (-1)^{j_1}(K_{J_1}(\hat{f})\hat{f}^{(j_1')})^{(j_1)} p_i f \int_{-\pi}^{\pi} K_J(\hat{f}) p_i^{(j')} f^{(j)}\right) \\ & - \frac{2}{n_2}\left(\sum_{i=0}^{m} a_i (-1)^{j+j_1} \int_{-\pi}^{\pi} (K_J(\hat{f})p_i^{(j')})^{(j)}(K_{J_1}(\hat{f})\hat{f}^{(j_1')})^{(j_1)} f\right) \\ & - \frac{2(n_2-2)}{n_2}\left(\sum_{i=0}^{m} a_i \int_{-\pi}^{\pi} K_J(\hat{f}) p_i^{(j')} f^{(j)} \int_{-\pi}^{\pi} K_{J_1}(\hat{f})\hat{f}^{(j_1')} f^{(j_1)}\right) \end{aligned}$$

by integration by parts. It follows that

$$\begin{aligned} E(\widehat{T_0^{jj'}}\widehat{L^{J_1}}|\hat{f}) - E(\widehat{T_0^{jj'}}|\hat{f})L^{J_1} = & - \frac{2}{n_2}\left(\int_{-\pi}^{\pi} (-1)^{j'+j_1} S_m[(K_{J_1}(\hat{f})\hat{f}^{(j_1')})^{(j_1)} f](K_J(\hat{f})f^{(j)})^{(j')}\right) \\ & - \frac{2}{n_2}\left(\int_{-\pi}^{\pi} (-1)^{j+j_1}(K_J(\hat{f})S_m f^{(j')})^{(j)}(K_{J_1}(\hat{f})\hat{f}^{(j_1')})^{(j_1)} f\right) \\ & + \frac{4}{n_2}\int_{-\pi}^{\pi} K_J(\hat{f}) S_m f^{(j')} f^{(j)} \int_{-\pi}^{\pi} K_{J_1}(\hat{f})\hat{f}^{(j_1')} f^{(j_1)}. \end{aligned}$$

Finally, $2n \sum_{jj'j_1j_1'=0}^{k} \left[E(\widehat{T_0^{jj'}}\widehat{L^{J_1}}) - E(\widehat{T_0^{jj'}}L^{J_1})\right]$ converges as $n \to \infty$ towards $-2 \sum_{j,j',j_1,j_1'=0}^{k} \Lambda_{JJ_1}(f, K_J(f), K_{J_1}(f))$.

The same result holds with $\widehat{T_0^{j'j}}$ replacing $\widehat{T_0^{jj'}}$. By similar computations for $E(\widehat{T_2^{jj'}}\widehat{L^{J_1}}|\hat{f}) - E(\widehat{T_2^{jj'}}|\hat{f})L^{J_1}$ the above result holds for $\widehat{T^J}$ instead of $\widehat{T_0^{jj'}}$.

Finally,

$$\lim_{n\to\infty} nE\left[\left(\sum_{jj'=0}^{k} \widehat{T^J} - T^J + \widehat{L^J} - L^J\right)^2\right] = 0.$$

It remains to prove that $R_2 \overset{\mathbb{L}^2}{\to} 0$.

$$E(R_2^2) = \frac{n}{n_2} E\left[\int_{-\pi}^{\pi}\left(\sum_{j=0}^{k}(-1)^j(\phi_j'(\hat{f}\dots\hat{f}^{(k)},.))^{(j)} - \sum_{j=0}^{k}(-1)^j(\phi_j'(f\dots f^{(k)},.))^{(j)}\right)^2 f\right]$$
$$-\frac{n}{n_2} E\left[\left(\sum_{j=0}^{k}\int_{-\pi}^{\pi}\phi_j'(\hat{f}\dots\hat{f}^{(k)},.)f^{(j)} - \int_{-\pi}^{\pi}\phi_j'(f\dots f^{(k)},.)f^{(j)}\right)^2\right].$$

Using the fact that ϕ and its derivatives are bounded over K_ϵ and that $\hat{f}$ and its p first derivatives converge in $\mathbb{L}^1-$ norm towards f and its derivatives respectively, we show as above that

$$E\left[\int_{-\pi}^{\pi}\left((\phi_j'(\hat{f}\dots,\hat{f}^{(k)},.))^{(j)}\right)^2 f\right] \to \int_{-\pi}^{\pi}\left((\phi_j'(f\dots f^{(k)},.))^{(j)}\right)^2 f.$$

By similar arguments for all the terms which appear in the expression of $E(R_2^2)$ and since $\frac{n}{n_2} \to 1$ $\quad R_2 \overset{\mathbb{L}^2}{\to} 0$.
This achieves the proof of Theorem 3.2. □

3.5 Appendix

3.5.1 Proof of Lemma 3.1

Let us fix a function f in $F_{\alpha,C}$. We first want to bound $\sum_{m}^{2m-1}(a_{2i-1}^2 + a_{2i}^2)$.
The Fourier series expansion of $f(x+h) - f(x-h)$ is :

$$2\sum_{i=1}^{\infty}(a_{2i}\frac{\sin ix}{\sqrt{\pi}} - a_{2i-1}\frac{\cos ix}{\sqrt{\pi}})\sin(ih).$$

Therefore

$$\int_{-\pi}^{\pi}(f(x+h) - f(x-h))^2 = 4\sum_{i=1}^{\infty}(a_{2i}^2 + a_{2i-1}^2)\sin^2 ih$$
$$\leq C^2\int_{-\pi}^{\pi}(2h)^{2\alpha} = 2\pi C^2(2h)^{2\alpha}.$$

Hence

$$\forall m \in \mathbb{N}, \sum_{i=m}^{2m-1}(a_{2i}^2 + a_{2i-1}^2)\sin^2 ih \leq \frac{\pi}{2}4^\alpha C^2 h^{2\alpha}.$$

If we take $h = \frac{\pi}{4m}$, since $\sin ih \geq \sin\frac{\pi}{4} = \frac{\sqrt{2}}{2}$ for $i \in \{m, m+1, \ldots 2m-1\}$

$$\sum_{i=m}^{2m-1} (a_{2i}^2 + a_{2i-1}^2) \leq \frac{\pi^{2\alpha+1}C^2}{4^\alpha m^{2\alpha}}.$$

Now, for $f \in F_{s,C}$ the same arguments show, since $f^{(p)} \in F_{\alpha,C}$, that

$$\sum_{i=m}^{2m-1} i^{2p}(a_{2i}^2 + a_{2i-1}^2) \leq \frac{\pi^{2\alpha+1}C^2}{4^\alpha m^{2\alpha}}. \tag{3.28}$$

Therefore, for $\alpha > \beta > 0$,

$$\sum_{i=m}^{2m-1} i^{2p+2\beta}(a_{2i}^2 + a_{2i-1}^2) \leq \frac{\pi^{1+2\alpha}C^2}{4^{\alpha-\beta}} m^{2\beta-2\alpha}.$$

Let $m = 2^j$ then

$$\sum_{j=0}^{\infty} \sum_{i=2^j}^{2^{j+1}-1} i^{2p+2\beta}(a_{2i}^2 + a_{2i-1}^2) \leq \frac{\pi^{2\alpha+1}C^2}{4^{\alpha-\beta}-1} \quad \text{if} \quad \beta < \alpha$$

which gives the first part of the lemma.

From (3.28), we deduce that

$$\begin{aligned}\sum_{i=m}^{\infty} i^{2p}(a_{2i}^2 + a_{2i-1}^2) &= \sum_{j=0}^{\infty} \sum_{i=m2^j}^{m2^{j+1}-1} i^{2p}(a_{2i}^2 + a_{2i-1}^2) \\ &\leq \pi^{1+2\alpha}4^{-\alpha}C^2 \sum_{j=0}^{\infty} (\frac{1}{m2^j})^{2\alpha} = \left(\frac{\pi^{1+2\alpha}}{4^\alpha - 1}\right)\frac{C^2}{m^{2\alpha}} \square.\end{aligned}$$

3.5.2 Proof of Lemma 3.3

To prove Lemma 3.3, we first compute $E(TU)$.

$$\begin{aligned}TU &= \frac{1}{(n(n-1))^2} \sum_{j \neq k=1}^{n} \sum_{j' \neq k'=1}^{n} a(X_j, X_k)b(X_{j'}, X_{k'}) \\ &= \frac{2}{(n(n-1))^2} \sum_{j \neq k=1}^{n} a(X_j, X_k)b(X_j, X_k) \\ &+ \frac{4}{(n(n-1))^2} \sum_{j,k,k' \neq} a(X_j, X_k)b(X_j, X_{k'}) \\ &+ \frac{1}{(n(n-1))^2} \sum_{j,j',k,k' \neq} a(X_j, X_k)b(X_{j'}, X_{k'})\end{aligned}$$

It follows that

$$\begin{aligned}\mathrm{Cov}(T,U) &= \frac{2}{n(n-1)}E\left[a(X_1,X_2)b(X_1,X_2)\right] \\ &+ \frac{4(n-2)}{n(n-1)}E\left[\bar{a}(X_1)\bar{b}(X_1)\right] \\ &+ \left(\frac{(n-2)(n-3)}{n(n-1)}-1\right)E[a(X_1,X_2)]E[b(X_1,X_2)]\end{aligned}$$

Since

$$\frac{(n-2)(n-3)}{n(n-1)}-1 = -\frac{2}{n(n-1)}-\frac{4(n-2)}{n(n-1)},$$

and since

$$E[a(X_1,X_2)] = E[\bar{a}(X_1)],$$

we obtain

$$\begin{aligned}\mathrm{Cov}(T,U) &= \frac{2}{n(n-1)}\left[E\left(a(X_1,X_2)b(X_1,X_2)\right)-E(a(X_1,X_2))E(b(X_1,X_2))\right] \\ &+ \frac{4(n-2)}{n(n-1)}\left[E\left(\bar{a}(X_1)\bar{b}(X_1)\right)-E(\bar{a}(X_1))E(\bar{b}(X_1))\right].\square\end{aligned}$$

This achieves the proof of Lemma 3.3.

3.5.3 Proof of Condition A2

Let $\tilde{f}$ an estimator of the density f based on projection methods :

$$\tilde{f}(x) = \frac{1}{n_1}\sum_{j=n-n_1}^{n}\sum_{i=0}^{m_1}p_i(X_j)p_i(x)$$

where $m_1 \approx n_1^{\frac{1}{1+2s}}$. Let

$$A_n = \{\tilde{f}^{(i)}(S)\subset[a_i-\epsilon,b_i+\epsilon] \quad \forall\, i\in\{0,\dots k\}\}$$

and put

$$\hat{f} = \tilde{f}\ \mathbb{I}_{A_n} + f_0\ \mathbb{I}A_n^c$$

where f_0 satisfies properties a) and b) of Condition A2.
It is clear that $\hat{f}$ also satisfies a) and b). We will first prove condition c) with $\tilde{f}$ instead of $\hat{f}$.
Let us first evaluate $||\bar{f}^{(l)}-f^{(l)}||_q^q$ where $\bar{f}^{(l)} = E(\tilde{f}^{(l)})$.

$$\bar{f}^{(l)} = (S_{m_1}f)^{(l)} = S_{m_1}(f^{(l)}) \quad \text{for } m_1 \quad \text{even} .$$

Hence,

$$\|\bar{f}^{(l)} - f^{(l)}\|_q^q = \int_{-\pi}^{\pi} |S_{m_1}(f^{(l)}) - f^{(l)}|_q^q.$$

Let $\mathcal{P}_{m_1}$ be the set of trigonometric polynomials with degree not bigger than m_1. It follows from M. Riesz's theorem (see Zygmund) that for $1 < q < +\infty$ and for $P \in \mathcal{P}_{m_1}$

$$\begin{aligned} \|S_{m_1}(f^{(l)}) - f^{(l)}\|_q &= \|S_{m_1}(f^{(l)} - P) - (f^{(l)} - P)\|_q \leq C_q\|f^{(l)} - P\|_q \\ &\leq (2\pi)^{1/q}C_q\|f^{(l)} - P\|_\infty \quad \forall\, P \end{aligned}$$

hence,

$$\|S_{m_1}(f^{(l)}) - f^{(l)}\|_q \leq (2\pi)^{1/q}C_q \inf_{P\in\mathcal{P}_{m_1}} \|f^{(l)} - P\|_\infty \leq \frac{C'_q}{m_1^{s-l}}.$$

$$\|\bar{f}^{(l)} - f^{(l)}\|_q^q = O(n_1^{\frac{q(l-s)}{1+2s}}).$$

Here, and in the remainder of the proof, the O's do not depend on $f \in F_{s,C}$ but depend on q.

We will show that $E(\|\tilde{f}^{(l)} - \bar{f}^{(l)}\|_q^q)$ is of the same order. Denote

$$Y_j = \frac{\sum_{i=0}^{m_1} p_i(X_j)p_i^{(l)}(x) - \sum_{i=0}^{m_1} a_i(f)p_i^{(l)}(x)}{m_1^{1+l}}.$$

$$\|Y_j\|_\infty \leq 2(1 + \|f\|_\infty), \qquad \sum_{j=n-n_1}^{n} E(Y_j^2) = O(\frac{n_1}{m_1}).$$

So, by Corollary 1 in Bretagnolle-Huber (1979), we get $E(\|\sum_{j=n-n_1}^{n} Y_j\|_q^q) = O((\frac{n_1}{m_1})^{q/2})$, which leads to

$$E(\|\tilde{f}^{(l)} - \bar{f}^{(l)}\|_q^q) = O(n_1^{\frac{q(l-s)}{1+2s}}).$$

It remains to show that the same result holds when $\tilde{f}$ is replaced by $\hat{f}$.

Since $f \in F_{s,C}$, $\|S_{m_1}(f^{(l)}) - f^{(l)}\|_\infty = O(\frac{\log(m_1)}{m_1^\alpha}) \quad \forall\, l \leq p$. (See Zygmund, vol. 1 p. 120).

For n large enough independently of $f \in F_{s,C}$, $\|S_{m_1}(f^{(l)}) - f^{(l)}\|_\infty = \|\bar{f}^{(l)} - f^{(l)}\|_\infty \leq \frac{e}{2}$ $\forall\, l \in \{0,\ldots p\}$. If $\|\tilde{f}^{(l)} - f^{(l)}\|_\infty \leq \frac{e}{2} \quad \forall\, l \in \{0,\ldots p\}$, then $\hat{f} = \tilde{f}$.

So, $P(A_n^c) \le \sum_{l=0}^{p} P(\|\tilde{f}^{(l)} - \bar{f}^{(l)}\|_\infty \ge \frac{\epsilon}{2})$. Denote by D_{m_1} Dirichlet's kernel :
$D_{m_1}(t) = \frac{\sin(m_1 + \frac{1}{2})t}{2\pi \sin\frac{t}{2}}$.

$$\tilde{f}^{(l)}(x) = \frac{1}{n_1} \sum_{j=n-n_1}^{n} D_{m_1}^{(l)}(x - X_j), \quad \bar{f}^{(l)}(x) = f * D_{m_1}^{(l)}(x).$$

It follows from Bernstein's inequality that

$$P(|\tilde{f}^{(l)} - \bar{f}^{(l)}|(x) \ge \frac{\epsilon}{2}) \le 2\exp\left\{-\frac{1}{2}\left[\frac{n_1\epsilon^2/4}{m_1^{1+2l} + \frac{1}{6}\epsilon m_1^{1+l}}\right]\right\}.$$

Let $x_0, \ldots x_N$ be n points from $[-\pi, \pi]$ such that $x_0 = -\pi$, $x_N = \pi$, $\quad x_i \le x_{i+1}$ $|x_{i+1} - x_i| = \delta$.

$$|\tilde{f}^{(l)}(x) - \bar{f}^{(l)}(x) - (\tilde{f}^{(l)}(x_i) - \bar{f}^{(l)}(x_i))| \le (\|f^{(\tilde{l}+1)}\|_\infty + \|f^{(\bar{l}+1)}\|_\infty)|x - x_i| \le 2m^{2+l}|x - x_i|.$$

Let δ such that $\delta m_1^{2+l} = \frac{\epsilon}{4}$, then

$$P\left(\sup_{|x-x_i|\le\delta} |\tilde{f}^{(l)} - \bar{f}^{(l)}|(x) > \frac{\epsilon}{2}\right) \le P(|\tilde{f}^{(l)} - \bar{f}^{(l)}|(x_i) > \frac{\epsilon}{4})$$
$$\le 2\exp\left\{-\frac{1}{2}\left[\frac{n_1\epsilon^2/16}{m_1^{1+2l} + \frac{1}{12}\epsilon m_1^{1+l}}\right]\right\}.$$

At last,

$$P(\|\tilde{f}^{(l)} - \bar{f}^{(l)}\|_\infty \ge \frac{\epsilon}{2}) \le \frac{2\pi}{\delta} P(|\tilde{f}^{(l)} - \bar{f}^{(l)}|(x_i) > \frac{\epsilon}{4}).$$

$$P(\|\tilde{f}^{(l)} - \bar{f}^{(l)}\|_\infty \ge \frac{\epsilon}{2}) \le K_0 m_1^{2+l} \exp\left\{-\left(\frac{K_1 n_1 \epsilon^2}{m_1^{1+2l} + \frac{1}{12}\epsilon m_1^{1+l}}\right)\right\} \tag{3.29}$$

where K_0, K_1 are positive absolute constants. Since $m_1 = n_1^{\frac{1}{1+2s}}$, for n_1 large enough,

$$\sum_{l=0}^{p} P(\|\tilde{f}^{(l)} - \bar{f}^{(l)}\|_\infty \ge \frac{\epsilon}{2}) \le K_0 n_1^{\frac{-qs}{1+2s}}$$

$$E(\|\hat{f}^{(l)} - f^{(l)}\|_q^q) \le E(\|\tilde{f}^{(l)} - f^{(l)}\|_q^q) + \|f^{(l)} - f_0^{(l)}\|_q^q K_0 n_1^{\frac{-qs}{1+2s}}$$
$$\le \gamma_1 n_1^{\frac{-q(s-l)}{1+2s}}.$$

This achieves the proof of c). Let us now control the order of magnitude of $E(||\bar{f}^{(l)}||_\infty^q) \quad \forall\, l \leq 2k$. From (3.29) we deduce

$$P\left(\frac{||\tilde{f}^{(l)} - \bar{f}^{(l)}||_\infty^q}{m_1^{q(l-s')}} \geq t\right) \leq K_0 m_1^{2+l} \exp\left[-K_1 n_1 t^{2/q} m_1^{-1-2s'}\right].$$

Since the above inequality holds for $t \geq t_0 > 0$,

$$E\left(\frac{||\tilde{f}^{(l)} - \bar{f}^{(l)}||_\infty^q}{m_1^{q(l-s')}}\right) \leq t_0 + \int_{t_0}^{+\infty} K_0 m_1^{2+l} \exp\left[-K_1 n_1^{\frac{2(s-s')}{1+2s}} t^{2/q}\right] dt$$

$$\leq 2t_0 \quad \text{for } n_1 \text{ large enough.}$$

To conclude the proof of d), it remains to prove that

$$||\bar{f}^{(l)}||_\infty = O(1 + m_1^{l-s'}) \quad \forall\, s' < s, \quad \forall\, 0 \leq l \leq 2k.$$

Using Lemma 3.2 and the fact that $\forall\, l \leq p\ ||\bar{f}^{(l)} - f^{(l)}||_\infty = O(\frac{\log m_1}{m_1^\alpha})$, we see that $\forall\, l \leq p$, $||\bar{f}^{(l)}||_\infty$ is bounded by some constant independent of $f \in F_{s,C}$. We shall now prove that $||\bar{f}^{(p+1)}||_\infty = O(m_1^{1-\alpha'}) \quad \forall\, 0 < \alpha' < \alpha$.

$$\bar{f}^{(p+1)}(x) = \int_{-\pi}^{\pi} D'_{m_1}(x-t) f^{(p)}(t)dt = \int_{x-\pi}^{x+\pi} D'_{m_1}(t)[f^{(p)}(x-t) - f^{(p)}(x)]dt.$$

$$= \int_{x-\pi}^{x+\pi} \left[\frac{(m_1+\frac{1}{2})\cos(m_1+\frac{1}{2})t}{2\pi \sin\frac{t}{2}} - \frac{\sin(m_1+\frac{1}{2})t \cos\frac{t}{2}}{4\pi(\sin\frac{t}{2})^2}\right] [f^{(p)}(x-t) - f^{(p)}(x)]dt.$$

So,

$$|\bar{f}^{(p+1)}(x)| \leq (m_1 + \frac{1}{2}) \left| \int_{x-\pi}^{x+\pi} \left[\frac{\cos(m_1+\frac{1}{2})t}{2\pi(\sin\frac{t}{2})}\right] [f^{(p)}(x-t) - f^{(p)}(x)]dt \right|$$

$$+ C m_1^{-\alpha} \int_{m_1(x-\pi)}^{m_1(x+\pi)} |\frac{\sin(1+\frac{1}{2m_1})y}{(sin\frac{y}{2m_1})^2}| |y|^\alpha \frac{dy}{m_1} \quad \text{since } f \in F_{s,C}.$$

The second term of the sum is bounded by

$$C' m_1^{1-\alpha} \int_{m_1(x-\pi)}^{m_1(x+\pi)} |y|^{\alpha-2} |\sin(1+\frac{1}{2m_1})y| = O(m_1^{1-\alpha}).$$

As to the term $\int_{x-\pi}^{x+\pi} \left[\frac{(\cos(m_1+\frac{1}{2})t)}{2\pi \sin\frac{t}{2}}\right] (f^{(p)}(x-t) - f^{(p)}(x))dt$ we notice that it is equal to $\int_{x-\pi}^{x+\pi} (S_{m_1}(\check{f^{(p)}})(t) - \check{f^{(p)}})(t))dt$ where $S_m(\check{f^{(p)}})$ denotes the conjugate Fourier serie of

$f^{(p)}$ and $\check{f^{(p)}}$ the conjugate function of $f^{(p)}$ (see Bary (1964), vol. 2, p.52). Since $\check{f^{(p)}}$ is also an $\alpha-$ hölderian function, as proved in Bary (1964),

$$\int_{x-\pi}^{x+\pi} (S_{m_1}(\check{f^{(p)}})(t) - \check{f^{(p)}})(t))dt = O(m_1^{-\alpha}\log(m_1))$$

and $\|\bar{f}^{(p+1)}\|_\infty = O(m_1^{1-\alpha'}) \quad \forall\ \alpha' < \alpha$.
From Bernstein's inequality concerning trigonometric polynomials it follows that for $l \geq 1$

$$\|\bar{f}^{(p+l)}\|_\infty \leq m_1^{l-1}\|\bar{f}^{(p+1)}\|_\infty = O(m_1^{l-s'}) \quad \forall\ s' < s.$$

This achieves the proof of d). So,

$$E(\|\hat{f}^{(l)}\|_\infty^q) = O(m_1^{q(l-s')}) \quad \forall\ s' < s. \tag{3.30}$$

At last, we have to show that

$$E(\|\hat{f}^{(l)}\|_\alpha^q) \leq \gamma_3(1 + n_1^{\frac{q(l+\alpha-s')}{1+2s}}).$$

Since $s-\alpha$ is an integer, we first notice that if $s-l-\alpha > 0$ then $s-l-1 > 0$ and for $s' < s$ close to s $s'-l-1 > 0$. In this case, $E(\|\hat{f}^{(l+1)}\|_\infty^q)$ is bounded and $E(\|\hat{f}^{(l)}\|_\alpha^q)$ is also bounded.
If $s-l-\alpha \leq 0$

$$|\hat{f}^{(l)}(x) - \hat{f}^{(l)}(y)| \leq \|\hat{f}^{(l+1)}\|_\infty|x-y|$$

$$\frac{|\hat{f}^{(l)}(x) - \hat{f}^{(l)}(y)|}{|x-y|^\alpha} \leq \|\hat{f}^{(l+1)}\|_\infty|x-y|^{1-\alpha}.$$

If $|x-y| \leq \frac{1}{m_1}$, then

$$\frac{|\hat{f}^{(l)}(x) - \hat{f}^{(l)}(y)|}{|x-y|^\alpha} \leq \|\hat{f}^{(l+1)}\|_\infty m_1^{\alpha-1}.$$

If $|x-y| \geq \frac{1}{m_1}$, then

$$\frac{|\hat{f}^{(l)}(x) - \hat{f}^{(l)}(y)|}{|x-y|^\alpha} \leq 2\|\hat{f}^{(l)}\|_\infty m_1^\alpha.$$

Finally,

$$\|\hat{f}^{(l)}\|_\alpha^q \leq \|\hat{f}^{(l+1)}\|_\infty^q m_1^{q(\alpha-1)} + 2^q\|\hat{f}^{(l)}\|_\infty^q m_1^{q\alpha}$$
$$E\left(\|\hat{f}^{(l)}\|_\alpha^q\right) \leq E\left[\|\hat{f}^{(l+1)}\|_\infty^q m_1^{q(\alpha-1)} + 2^q\|\hat{f}^{(l)}\|_\infty^q m_1^{q\alpha}\right].$$

From (3.30) we derive,

$$E(\|\hat{f}^{(l)}\|_\alpha^q) = O(1 + m_1^{q(l+\alpha-s')}) \quad \forall\ s' < s.$$

3.5.4 Semi-parametric information bound

This proof will be almost similar as the proof which was done in Chapter 2 for regular functions. We still refer to Koshevnik and Levit (1976) and Levit (1978). We suppose that f belongs to $F_{s,C}$ and satisfies

$$\sup_{x \neq y \in [-\pi,\pi]} \frac{|f^{(p)}(x) = f^{(p)}(y)|}{|x-y|^{\alpha}} < C.$$

Let ξ be a bounded function, infinitely differentiable, such that $\int f\xi = 0$. We define

$$f_t = f(1+t\xi).$$

$f_t \in F_{s,C}$ and is a density as soon as t is small enough.

$$T(f_t) - T(f) = \int \sum_{j=0}^{k} \frac{\partial \phi}{\partial x_j}(f, \dots, f^{(k)}, .)(f_t^{(j)} - f^{(j)}) + o(t)$$

$$= \int \sum_{j=0}^{k} (-1)^j \left(\frac{\partial \phi}{\partial x_j}(f, \dots, f^{(k)}, .) \right)^{(j)} (f_t - f) + o(t) \quad \text{by integration by parts.}$$

$$= \int \sum_{j=0}^{k} \left[(-1)^j \left(\frac{\partial \phi}{\partial x_j}(f, \dots, f^{(k)}, .) \right)^{(j)} - \int \left(\frac{\partial \phi}{\partial x_j}(f, \dots, f^{(k)}, .) f^{(j)} \right) \right] (f_t - f) + o(t)$$

since $\int f_t = \int f = 1$. Moreover,

$$\|\sqrt{f_t} - \sqrt{f} - At\|_2 \to 0 \quad \text{as } n \to \infty$$

where

$$At = t\frac{\xi}{2}\sqrt{f}$$

and

$$T(f_t) - T(f) = Bt + o(t)$$

where

$$Bt = E_f(G\frac{\xi}{2}t)$$

where

$$G = 2\left[\sum_{j=0}^{k} (-1)^j \left(\frac{\partial \phi}{\partial x_j}(f, \dots, f^{(k)}, .) \right)^{(j)} - \sum_{j=0}^{k} \int \frac{\partial \phi}{\partial x_j}(f, \dots, f^{(k)}, .) f^{(j)} \right].$$

So the function G satifies $E_f(G) = 0$. Since the function ξ are dense in the set of functions $g \in \mathbb{L}^2([-\pi,\pi])$ such that $E_f(g) = 0$, there exists a sequence ξ_n such that

$$\lim_{n\to\infty} \int_{-\pi}^{\pi} (\frac{\xi_n}{2} - G)^2 f(x)dx = 0.$$

According to Theorem 1 and Theorem 2 of Koshevnik and Levit (1976),

$$\inf_{\epsilon} \liminf_{n\to\infty} \sup_{f_1 \in F_{s,C}, \|f_1 - f\|_2 \leq \epsilon} nE(\widehat{T_n} - T(f))^2 \geq \frac{1}{4} E_f(G^2).$$

At last,

$$\frac{1}{4}E_f(G^2) = \int \left[\sum_{j=0}^{k}(-1)^j \left(\frac{\partial\phi}{\partial x_j}(f,\ldots,f^{(k)},.)\right)^{(j)}\right]^2 f d\mu - \left(\int \sum_{j=0}^{k} \frac{\partial\phi}{\partial x_j}(f,\ldots f^{(k)},.) f^{(j)} d\mu\right)^2$$

For the problem of estimating $\int (f^{(k)})^2$, we get for the analogue of the Cramer Rao bound:

$$4[\int (f^{(2k)})^2 f - (\int (f^{(k)})^2)^2].$$

Chapitre 4

Simulations

4.1 Estimation de $\int f^2$

4.1.1 Introduction

Toutes les simulations ont été réalisées à l'aide du logiciel MATLAB.
Nous présentons dans cette section les résultats des simulations concernant l'estimation de $\int f^2$. Nous comparons les performances de deux estimateurs:

- 1) L'estimateur proposé par Hall et Marron (voir Hall et Marron (1987))

$$\frac{1}{n(n-1)} \sum_{j \neq k=1}^{n} K_h(X_j - X_k)$$

où K sera ou bien un noyau gaussien, ou bien un noyau uniforme sur $[-1, 1]$.

- 2) L'estimateur par projection proposé dans la thèse

$$\frac{1}{n(n-1)} \sum_{i=0}^{m} \sum_{j \neq k=1}^{n} p_i(X_j) p_i(X_k)$$

où $(p_i)_{i \in \mathbb{N}}$ est la base Fourier. Dans les deux cas, il faut ajuster le paramètre m ou h. Nous allons donner les résultats des simulations pour différentes valeurs de ces paramètres. Les simulations sont réalisées aves $n = 300$ données. Pour chaque paramètre (m ou h), 500 simulations sont effectuées. Nous donnons la moyenne et la variance de ces 500 simulations et nous estimons le risque quadratique $nE(\hat{\theta} - \theta)^2$ par la quantité $n \times \frac{1}{500} \sum_{j=1}^{500} (\hat{\theta}_j - \theta)^2$, les $\hat{\theta}_j$ correspondant aux 500 simulations.

Nous avons choisi deux densités à support dans l'intervalle [0, 1].

- A) $f(x) = 1 + 1/2(\sin 4\pi x)$.
- B) $f(x) = (-4x + \frac{3}{2})\ \mathbb{1}_{[0,\frac{1}{4}]} + (4x - \frac{1}{2})\ \mathbb{1}_{[\frac{1}{4},\frac{1}{2}]} + (-4x + \frac{7}{2})\ \mathbb{1}_{[\frac{1}{2},\frac{3}{4}]} + (4x - \frac{5}{2})\ \mathbb{1}_{[\frac{3}{4},1]}$.

4.1.2 Résultats des simulations

- A) $f(x) = 1 + (1/2)\sin(4\pi x)$; $\int f^2 = 1,125$, $4[\int f^3 - (\int f^2)^2] = 0,4375$; $n = 300$.

Estimateur proposé par Hall et Marron

	NOYAU UNIFORME			NOYAU GAUSSIEN		
h	Moyenne	Variance	Risque	Moyenne	Variance	Risque
1/1000	1,1318	$1,54.10^{-2}$	4,615	1,1310	$2,9.10^{-3}$	2,8479
5/1000	1,1238	$4,3.10^{-3}$	1,277	1,1211	$2,9.10^{-3}$	0,8416
1/100	1,1233	$2,7.10^{-3}$	0,8202	1,1167	$2,1.10^{-3}$	0,6446
5/100	1,0935	$1,6.10^{-3}$	0,7876	1,0694	$1,3.10^{-3}$	1,4559
1/10	1,0481	$1,2.10^{-3}$	2,1416	0,9842	$6,7.10^{-4}$	6,1490

Estimateur par projection

m	MOYENNE	VARIANCE	BORNE
3	1,0063	$5,05.10^{-5}$	4,6808
5	1,1267	$1,5.10^{-3}$	0,4515
10	1,1248	$1,6.10^{-3}$	0,4665
60	1,1278	$3,1.10^{-3}$	0,9293
200	1,1271	$6,9.10^{-3}$	2,0729

• B) $f(x) = (-4x + \frac{3}{2})\, \mathbb{1}_{[0,\frac{1}{4}]} + (4x - \frac{1}{2})\, \mathbb{1}_{[\frac{1}{4},\frac{1}{2}]} + (-4x + \frac{7}{2})\, \mathbb{1}_{[\frac{1}{2},\frac{3}{4}]} + (4x - \frac{5}{2})\, \mathbb{1}_{[\frac{3}{4},1]}.$

$$\int f^2 = 13/12 = 1,0833; \quad 4[\int f^3 - (\int f^2)^2] = 0,3056.$$

Estimateur proposé par Hall et Marron

	NOYAU GAUSSIEN			NOYAU UNIFORME		
h	Moyenne	Variance	Risque	Moyenne	Variance	Risque
1/1000	1,0889	$8,1.10^{-3}$	2,4366	1,1310	$2,9.10^{-3}$	2,8479
5/1000	1,0798	$2,4.10^{-3}$	0,7207	1,1211	$2,9.10^{-3}$	0,8416
1/100	1,0692	$1,6.10^{-3}$	0,5350	1,1167	$2,1.10^{-3}$	0,6446
3/100	1,0315	$9,7.10^{-4}$	1,0954	1,0927	$1,5.10^{-3}$	0,7692
1/10	0,9153	$5,02.10^{-4}$	8,6235	0,9842	$6,7.10^{-4}$	6,1490

Estimateur par projection

m	MOYENNE	VARIANCE	BORNE
7	$1,0844$	$1,1.10^{-3}$	$0,3178$
20	$1,0855$	$1,3.10^{-3}$	$0,4044$
60	$1,0892$	$2,6.10^{-3}$	$0,7740$
200	$1,0867$	$6,0.10^{-3}$	$1,7876$
600	$1,0918$	$1,6.10^{-2}$	$4,8151$

4.2 Estimation de $\int f \log f$.

4.2.1 Introduction

Nous proposons ici les résultats des simulations concernant l'estimation de l'entropie. Comme précedemment, nous comparons les performances de deux estimateurs :
• 1) L'estimateur qui correspondrait à un développement de Taylor au premier ordre (voir le chapitre 2), cet estimateur a été proposé par Györfi et Van der Meulen (1990) :

$$\frac{1}{n_2}\sum_{j=1}^{n_2}\log(\widehat{f})(X_j)$$

• 2) L'estimateur proposé dans la thèse et qui correspond à un développement de Taylor au second ordre :

$$\frac{1}{n_2}\sum_{j=1}^{n_2}\log(\widehat{f})(X_j) - \frac{1}{2}\int \widehat{f} + \frac{1}{n_2(n_2-1)}\sum_{j\neq k=1}^{n_2} p_i(X_j)(\frac{p_i}{\widehat{f}})(X_k)$$

Dans ces deux exemples, l'estimateur préliminaire $\widehat{f}$ sera dans un premier temps basé sur les n_1 dernières données de l'echantillon avec $n_1 = n_2 = \frac{n}{2}$. Dans un deuxième temps, nous utiliserons l'echantillon complet pour construire l'estimateur préliminaire de la densité et les estimateurs de l'entropie. L'estimateur préliminaire sera tantôt un estimateur à noyau (le noyau choisi est l'uniforme sur [-1,1])

$$\widehat{f}(x) = \frac{1}{n_1}\sum_{j=n-n_1+1}^{n} K_h(x - X_j)$$

tantôt un estimateur par projection $\widehat{f}(x) = \frac{1}{n_1}\sum_{j=n-n_1+1}^{n}\sum_{i=0}^{m_1} p_i(X_j)p_i(x)$.
Nous avons choisi deux densités à support dans l'intervalle $[0,10]$.

$$\bullet \text{ A) } f(x) = \frac{1}{10}(1 + \frac{1}{2}\sin \pi x)$$

Cette densité comporte cinq périodes sur $[0,10]$.
La deuxième densité est affine par morceaux et comporte deux périodes sur $[0,10]$:

$$\begin{aligned}\bullet \text{ B) } f(x) &= \frac{1}{10}(-\frac{2}{5}x + \frac{3}{2})\ \mathbb{1}_{[0,\frac{5}{2}]} + \frac{1}{10}(\frac{2}{5}x - \frac{1}{2})\ \mathbb{1}_{[\frac{5}{2},5]} \\ &+ \frac{1}{10}(-\frac{2}{5}x + \frac{7}{2})\ \mathbb{1}_{[5,\frac{15}{2}]} + \frac{1}{10}(\frac{2}{5}x - \frac{5}{2})\ \mathbb{1}_{[\frac{15}{2},10]}.\end{aligned}$$

4.2.2 Résultats des simulations

A) $f(x) = \frac{1}{10}(1 + \frac{1}{2}\sin \pi x)$. $-\int f\log(f) = 2,238$, $\int f(\log f)^2 - (\int f \log f)^2 = 0,1185$.
$n = 400, n_1 = n_2 = 200$. L'estimateur préliminaire est un estimateur à noyau.

	ESTIMATEUR A L'ORDRE 1			ESTIMATEUR A L'ORDRE 2		
(m,h)	Moyenne	Variance	Risque	Moyenne	Variance	Risque
$(7,\frac{1}{10})$	2,3385	$2,4.10^{-3}$	5,001	2,2638	$4,9.10^{-4}$	0,4620
$(40,\frac{1}{10})$	"	"	"	2,2201	$1,9.10^{-3}$	0,9070
$(100,\frac{1}{10})$	"	"	"	2,2101	$3,7.10^{-3}$	1,7934
$(7,\frac{3}{10})$	2,3008	$1,3.10^{-3}$	2,0922	2,2684	$3,7.10^{-4}$	0,5166
$(40,\frac{3}{10})$	"	"	"	2,2385	$1,5.10^{-3}$	0,5998
$(100,\frac{3}{10})$	"	"	"	2,2380	$2,7.10^{-3}$	1,0692
$(7,1)$	2,3702	$3,4.10^{-4}$	7,1274	2,3229	$1,3.10^{-4}$	2,9314
$(40,1)$	"	"	"	2,2582	$1,4.10^{-3}$	0,7124
$(100,1)$	"	"	"	2,2591	$2,4.10^{-3}$	1,1559

A) $f(x) = \frac{1}{10}(1+\frac{1}{2}\sin \pi x)$. $-\int f\log(f) = 2,238$, $\int f(\log f)^2 - (\int f\log f)^2 = 0,1185$.
$n = 400 = n_1 = n_2$. L'estimateur préliminaire est un estimateur à noyau.
Les 400 données sont utilisées à la fois pour construire $\widehat{f}$ et pour construire les estimateurs de l'entropie.

	ESTIMATEUR A L'ORDRE 1			ESTIMATEUR A L'ORDRE 2		
(m,h)	Moyenne	Variance	Risque	Moyenne	Variance	Risque
$(7, \frac{1}{10})$	2,1799	$6,6.10^{-4}$	1,616	2,2170	$3,0.10^{-4}$	0,2989
$(40, \frac{1}{10})$	"	"	"	2,2267	$5,72.10^{-4}$	0,2796
$(100, \frac{1}{10})$	"	"	"	2,2495	$7,86.10^{-4}$	0,3669
$(7, \frac{3}{10})$	2,238	$4,4.10^{-4}$	0,1774	2,2581	$2,0.10^{-4}$	0,2418
$(40, \frac{3}{10})$	"	"	"	2,2475	$5,09.10^{-4}$	0,2393
$(100, \frac{3}{10})$	"	"	"	2,2463	$7,04.10^{-4}$	0,3085
(7, 1)	2,3535	$1,6.10^{-4}$	5,3978	2,3248	$3,3.10^{-5}$	3,0251
(40, 1)	"	"	"	2,2610	$5,3.10^{-4}$	0,4225
(100, 1)	"	"	"	2,2610	$8,0.10^{-4}$	0,5421

A) $f(x) = \frac{1}{10}(1 + \frac{1}{2}\sin \pi x)$. $-\int f\log(f) = 2,238$, $\int f(\log f)^2 - (\int f \log f)^2 = 0,1185$.
$n = 400 = n_1 = n_2$. L'estimateur préliminaire est un estimateur par projection.
Les 400 données sont utilisées à la fois pour construire $\widehat{f}$ et pour construire les estimateurs de l'entropie.

	ESTIMATEUR A L'ORDRE 1			ESTIMATEUR A L'ORDRE 2		
(m_1,m)	Moyenne	Variance	Risque	Moyenne	Variance	Risque
$(5,5)$	2,2978	$1,1.10^{-5}$	1,4360	2,3028	$1,1.10^{-5}$	1,6856
$(5,40)$	"	"	"	2,2403	$4,74.10^{-4}$	0,1915
$(5,100)$	"	"	"	2,2409	$7,03.10^{-4}$	0,2842
$(20,5)$	2,2143	$3,6.10^{-4}$	0,3664	2,2391	$2,4.10^{-4}$	0,096
$(20,20)$	"	"	"	2,2383	$3,8.10^{-4}$	0,1519
$(20,100)$	"	"	"	2,2371	$8,6.10^{-4}$	0,3432
$(60,5)$	2,1643	$5,93.10^{-4}$	2,4076	2,1985	$3,9.10^{-4}$	0,7798
$(60,40)$	"	"	"	2,2176	$5,8.10^{-4}$	0,3959
$(60,100)$	"	"	"	2,2351	$7,6.10^{-4}$	0,3078

B) f est la densité affine par morceaux comportant deux périodes sur $[0,10]$ décrite ci-dessus. L'estimateur préliminaire est un estimateur par projections.
$-\int f\log(f) = 2,2598$, $\int f(\log f)^2 - (\int f\log f)^2 = 0,0803$. $n = n_1 = n_2 = 200$.

	ESTIMATEUR A L'ORDRE 1			ESTIMATEUR A L'ORDRE 2		
(m_1, m)	Moyenne	Variance	Risque	Moyenne	Variance	Risque
$(5,5)$	$2,2498$	$4,1.10^{-4}$	$0,1005$	$2,2595$	$4,1.10^{-4}$	$0,0813$
$(5,40)$	"	"	"	$2,2593$	$8,3.10^{-4}$	$0,1655$
$(5,100)$	"	"	"	$2,2588$	$1,4.10^{-3}$	$0,2883$
$(20,5)$	$2,2105$	$5,9.10^{-4}$	$0,6042$	$2,2288$	$5,4.10^{-4}$	$0,3001$
$(20,20)$	"	"	"	$2,2596$	$6,1.10^{-4}$	$0,1214$
$(20,100)$	"	"	"	$2,2567$	$1,5.10^{-3}$	$0,3094$
$(60,5)$	$2,1059$	$1,2.10^{-3}$	$4,9728$	$2,1463$	$1,2.10^{-3}$	$2,8144$
$(60,40)$	"	"	"	$2,2132$	$1,2.10^{-3}$	$0,6651$
$(60,100)$	"	"	"	$2,2409$	$1,8.10^{-3}$	$0,4321$

4.3 Interprétation des résultats

La correction de biais réalisée par nos estimateurs est très nette dans tous les cas. La moyenne de nos estimateurs est assez stable, et en général très proche de la vraie valeur de la fonctionnelle à estimer. La variance de nos estimateurs et la variance des estimateurs auxquels ils sont comparés sont du même ordre de grandeur. Par conséquent, la borne, qui estime la quantité $nE(\widehat{\theta} - \theta)^2$, se comporte mieux pour nos estimateurs dans la plupart des cas : elle semble être moins sensible aux variations des paramètres de lissage m ou h. D'autre part, pour une valeur optimale des paramètres de lissages nos estimateurs atteignent pratiquement la variance asymptotique minimale, c'est-à-dire l'analogue de la borne de Cramér-Rao. Ceci apparait clairement dans le cas de l'estimation de $\int f^2$. Nous constatons également que les résultats optimaux sont obtenus pour de petites valeurs de m, en effet, conformément à la théorie, la variance est une fonction croissante de m. Dans la mesure où les calculs sont beaucoup plus rapides pour de petites valeurs du paramètre m, ce résultat est intéressant.
En ce qui concerne l'estimation de l'entropie, il apparait que lorsque l'estimateur au premier ordre est fortement biaisé, la correction au second ordre nécessite de prendre une plus grande valeur pour m.

References

BARY N.K. (1964) : *A treatise on trigonometric series.* Pergamon Press.

BASHARIN G.P. (1956) : *On statistical estimate for the entropy of a sequence of independent random variables.* Theory Prob. Appl. 4, 333-336.

BHATTACHARYA P.K. (1967) : *Estimation of a probability density and its derivatives.* Sankhya, ser. A, 29, 373-382.

BICKEL P.J and RITOV Y. (1988) : *Estimating integrated squared density derivatives : sharp best order of convergence estimates.* Sankhya, 50, 381-393.

BIRGÉ L. and MASSART P. (1991) : *Estimation of integral functionals of a density.* To appear in MSRI series.

BRETAGNOLLE J. and HUBER C. (1979) : *Estimation des densités : risque minimax.* Z. Warsscheinlichkeitstheorie verw Gebiete, 47, 119-137.

DEHEUVELS P. et HOMINAL P. (1980) : *Estimation automatique de la densité.* Revue de statistique appliquée, vol. 28, 25-55.

DEVROYE L. and GYORFI L. (1985) : *Nonparametric density estimation, the L^1 view.* Wiley series in probability and mathematical statistics.

DEVROYE L. (1987) : *A course in density estimation.* Progress in probability and statistics. Birkhäuser.

DMITRIEV YU. G. and TARASENKO F.P. (1973) *On the estimation of functionals of a probability density function and its derivatives.* Theory Prob. Appl., 18, 628-633.

DONOHO D. and NUSSBAUM M. (1990) : *Minimax quadratic estimation of a quadratic functional.* J. Complexity 6, 290-323.

DUDEWICZ E.J. and VAN DER MEULEN E.C. (1981) : *Entropy-based tests of uniformity.* Journal of American Statistical Association, 76, 967-974.

FARELL R.H. (1972) : *On the best obtainable asymptotic rates of convergence in estimation of a density function at a point.* Ann. Math. Statis., 43, 170-180

GYÖRFI L. and VAN DER MEULEN E.C. (1990) *On the nonparametric estimation of the entropy functional.* Manuscript.

HALL P. and MARRON J.S. (1987) : *Estimation of integrated squared density derivatives.* Statistic and probability letters 6, 109-115.

HOEFFDING W. (1948) : *A class of statistics with asymptotically normal distribution.* Ann. Math. Statist., 19, 293-325.

IBRAGIMOV I.A and HAS'MINSKII R.Z. (1980) : *On the estimation of distribution density.* Investigation in the mathematical statistics. Zap. Nauch. Sem.LOMI , vol. 98, 61-85.

IBRAGIMOV I.A. and HAS'MINSKII R.Z. (1981) : *On the nonparametric density estimates.* Investigation in the mathematical statistics. Zap. Nauch. Sem.LOMI, vol. 108, 73-89.

IBRAGIMOV I.A. and HAS'MINSKII R.Z. (1981) : *Statistical estimation : asymptotic theory.* Springer Verlag.

IBRAGIMOV I.A., NEMIROVSKII A.S. and HAS'MINSKII R.Z. (1987) : *Some problems on nonparametric estimation in gaussian white noise.* Theory Probab. Appl., 31, 391-406.

IBRAGIMOV I.A and HAS'MINSKII R.Z. (1991) : *Asymptotic normal families of distribution and effective estimation.* The Annals of Statistics, 19, 1681-1724.

KERKYACHARIAN G. and PICARD D. (1992) : *An estimator of the integral of the third power of the density.* Preprint.

KOSHEVNIK YU. A. and LEVIT B.YA. (1976) : *On a non parametric analogue of the information matrix.* Theor. Probab. Appl., 21, 738-753.

LEHMANN E.L. (1975) : *Nonparametrics : statistical methods based on ranks.* Holden-day series in probability and statistics.

LEVIT B.YA. (1978) : *Asymptotically efficient estimation of nonlinear functionals.* Translated from Problemy Peredachi Informatsii, 14, 65-72.

MEYER Y. (1989) : *Ondelettes et opérateurs* Hermann.

PARZEN E. (1962) : *On estimation of a probability density function and mode.* Ann. Math. Statistics, 33, 1065-1076.

PRAKASA RAO B.L. (1983) : *Nonparametric functional estimation.* Probability and Mathematical Statistics. Academic Press.

ROSENBLATT M. (1956) : *Remarks on some nonparametric estimates of a density function.* Ann. Math. Statistics, 27, 832-837.

SCHUSTER E.F. (1969) : *Estimation of a probability density function and its derivatives.* Ann. Math. Statistics, 40, 1187-1195.

SERFLING R.J. (1980) *Approximation theorems of mathematical statistics.* Wiley series in probability and mathematical statistics.

VASICEK O. (1976) : *A test of normality based on sample entropy.* Journal of the Royal Statistical Society, 38, 54-59.

ZYGMUND A. (1968) : *Trigonometric series.* Cambridge at the University Press.