JEAN BRETAGNOLLE
CATHERINE HUBER

Estimation des densités : risque minimax

Séminaire de probabilités (Strasbourg), tome 12 (1978), p. 342-363
<http://www.numdam.org/item?id=SPS_1978__12__342_0>

Université de Strasbourg
Séminaire de Probabilités 1976/77

Estimation des densités : Risque minimax

J. Bretagnolle , C. Huber

1. Introduction

Etant donnée une variable aléatoire X_1 à valeurs dans l'espace mesuré $(A,\mathcal{A},\mu)$,μ positive et σ-finie,et de loi de probabilité $f.\mu$,on veut estimer f grâce à l'observation d'un n-échantillon $X = (X_1,..,X_n)$ de la variable X_1.Un estimateur $\hat{f}_n$ est une application mesurable de $(A^n,\mathcal{A}^{\otimes n})\times(A,\mathcal{A})$ dans $\mathbb{R}$ (ou $\mathbb{R}^+$) muni de la tribu $\mathcal{B}$ (ou $\mathcal{B}^+$) de ses boréliens : $\hat{f}_n(X_1,..,X_n,\cdot)$ estime $f(\cdot)$ et peut aussi être considéré comme une application de $(A^n,\mathcal{A}^{\otimes n})$ dans l'ensemble $\mathcal{L}^o$ des applications mesurables de $(A,\mathcal{A})$ dans $(\mathbb{R},\mathcal{B})$ ou dans $(\mathbb{R}^+,\mathcal{B}^+)$.

Le risque d'un estimateur $\hat{f}_n$ au point f,soit $R(\hat{f}_n,f)$ est défini comme $R(\hat{f}_n,f) = E_{f^{\otimes n}}\, D(\hat{f}_n,f)$,où D est une application de $\mathcal{L}^o\times\mathcal{L}^o$ dans $\mathbb{R}^+$ telle que,pour une fonction positive φ , $\varphi\circ D$ possède à peu près les propriétés d'une distance (Voir les définitions (2.1) et (2.2)).

L'objet de cette étude est d'évaluer le risque minimax lorsqu'on sait,a priori,que f appartient à un sous ensemble F de $\mathcal{L}^o$,c'est à dire que l'on détermine des constantes réelles positives c,d,et u,qui ne dépendent que de la caractérisation de F,et telles que l'on ait simultanément

(1.1) $$\inf_{\hat{f}_n}\ \sup_{f\in F}\ R(\hat{f}_n,f) \geqslant c\, n^{-u}$$

(1.2) Pour un $\hat{f}_n$ au moins

$$\sup_{f\in F}\ R(\hat{f}_n,f) \leqslant d\, n^{-u}$$

Au paragraphe 2,on établit une inégalité de base,fondée sur l'information de Kullback,qui permet d'obtenir les résultats de type (1.1).Le paragraphe 3 donne ces résultats pour le risque correspondant à la p-norme des densités sur $\mathbb{R}$,et les résultats de type (1.2),dans ce même cas,font l'objet du paragraphe 4;on compare les évaluations de c et d au paragraphe 5.Au paragraphe 6 figurent les résultats analogues pour la distance de Hellinger.Ces résultats ne sont pas asymptotiques,mais valables pour tout n.

Les mêmes méthodes permettent d'obtenir les résultats analogues en dimension ^{k}.Cela fait l'objet d'une remarque à la fin du paragraphe 3.

Par souci de clarté dans la présentation des énoncés et démonstrations,les calculs de constantes n'ont pas toujours été reproduits.

Cette étude nous a été suggérée par les travaux de Farrell et Wahba (2,6) qui ont obtenu les résultats correspondants pour l'étude locale.Par ailleurs,des résultats du type (1.2) se trouvent dans Nadaraya (5) pour la norme 2,dans Abou-Jaoudé (1) pour la norme 1.Ce dernier obtient un encadrement du risque pour l'estimateur à noyau uniforme,et il est à notre connaissance le premier qui ait employé une jauge du type $\|f^{(\nu)}\|_p \|f\|_\alpha^\beta$ (Voir §3).

En ce qui concerne la distance de Hellinger,Le Cam a démontré des résultats analogues dans le cas où F est indexé par un paramètre de dimension finie (3).

Au moment où nous achevons la rédaction de ce travail, nous prenons connaissance de résultats de même nature dans Terry G. Meyer (4).

2. Inégalité fondamentale

Les notations étant celles de l'introduction, définissons le risque d'un estimateur $\hat{f}_n$ à partir d'une fonctionnelle D qui satisfera dans la suite les propriétés résumées dans la

Définition 2.1 : D est une application de $\mathcal{L}^0 \times \mathcal{L}^0$ (ou de $A \times A$, A partie de $\mathcal{L}^0$) dans $\mathbb{R}^+$, qui est

symétrique: Pour tout couple g,h, $D(g,h) = D(h,g)$.

sur-additive: Pour toute partition A_j de A, tout couple g,h ,

$$D(g,h) \geqslant \sum_j D(g.1_{A_j} , h.1_{A_j}) \quad .$$

φ-distance: Il existe une fonction de $\mathbb{R}^+$ dans $\mathbb{R}^+$, croissante, continue et nulle en 0, prolongée par parité, telle que pour tout triplet f,g,h, on ait la φ-inégalité triangulaire:

$$\varphi(D(f,h)) + \varphi(D(g,h)) \geqslant \varphi(D(f,g)) .$$

Exemples: $D(g,h) = \|g-h\|_p^p = \int |g-h|^p \, d\mu$ est de ce type pour tout $p > 0$, en posant $\varphi(u) = u^{(1/p)\wedge 1}$. Egalement, quand g et h sont positives, leur distance de Hellinger $d_H(g,h) = \frac{1}{2}\int (g^{\frac{1}{2}} - h^{\frac{1}{2}})^2 d\mu$ pour $\varphi(u) = u^{\frac{1}{2}}$. Enfin, pour g et h γfois dérivables, $\| g^{(\gamma)} - h^{(\gamma)} \|_p^p$.

Définition 2.2 : Le risque de l'estimateur $\hat{f}_n$ au point f, densité de Probabilité sur (A , $\mathcal{A}$, μ) est

$$R(\hat{f}_n, f) = \int_{A^n} D(\hat{f}_n(x_1,\ldots,x_n,.), f(.)) \prod_{i=1}^n f(x_i) d\mu(x_i) . \tag{2.2}$$

Soient maintenant P et Q deux Probabilités sur un même espace, Q absolument continue par rapport à P; on note E_P, E_Q les espérances relatives à P et Q, on pose $X = \frac{dQ}{dP}$; leur information relative (de Kullback) est

(2.3) $\quad I(Q,P) = E_P(\ -\mathrm{Log} X\ ) \ = \int -\mathrm{Log}(\ \frac{dQ}{dP}\ )\ dP$

<u>Lemme 2.1</u> Soient U et V deux v.a.r. définies sur le même espace, de lois respectives P et Q; si, pour une fonction φ nulle en 0, croissante continue de $\mathbb{R}^+$ dans $\mathbb{R}^+$, pour un réel positif d $\varphi(U) + \varphi(V) \geqslant \varphi(d)$, on a, φ^{-1} étant la réciproque de φ.

(2.4) $\quad \mathrm{Max}(\ E_P(U)\ ,\ E_Q(V)\ ) \geqslant \frac{1}{4}\ \varphi^{-1}(\ \frac{1}{2}\varphi(d)\ ).\exp(-I(Q,P)\ ).$

<u>démonstration</u>: Pour tout évènement B, $Q(B)=P(B)+\ E_P(1_B(X-1)\ )$, soit $|Q(B) - P(B)| \leqslant \frac{1}{2}\ \|X-1\|_1$ (pour la P-mesure).

Soit alors u défini par $\varphi(u) = \frac{1}{2}\varphi(d)$; d'après la φ-inégalité triangulaire, $\quad V \geqslant u.(\ 1 - 1_{U>u})$, et donc

$E_Q(V) \geqslant u.(\ 1 - P(\ U>u\ ) - \frac{1}{2}\|X-1\|_1\ )$. Mais $E_P(U) \geqslant u.P(\ U>u\ )$.

Le minimax de ces deux expressions, linéaires en $P(\ U>u\ )$, vaut $\frac{1}{2}u.(\ 1 - \frac{1}{2}\|X-1\|_1)$. Si maintenant on pose $Y = (X-1)^+$, $X=1+Y-Z$, on a $E_P(Y)= E_P(Z) = \frac{1}{2}\|X-1\|_1$. Comme $\mathrm{Log}X = \mathrm{Log}(1+U)+\mathrm{Log}(1-V)$, l'inégalité de Jensen donne $\ -I(Q,P) \leq \mathrm{Log}(1-(E_PU)^2)$, soit

$\frac{1}{2}\|X-1\|_1 \leqslant (\ 1 - \exp(-I(Q,P))\ )^{\frac{1}{2}}$. La formule (2.4) résulte alors de ce que $\ 1 - (1-x)^{\frac{1}{2}} \geqslant \frac{1}{2}x$ pour x compris entre 0 et 1 .

Dans le cas où P et Q sont les n-puissances tensorielles de Probabilités p et q, avec $\frac{dq}{dp} = 1 + Z$, on a $I(Q,P) = nI(q,p)$. Si on suppose que $|Z| \leqslant \frac{1}{2}$, comme alors $|\mathrm{Log}(1+Z)-Z| \leqslant Z^2$, on a

(2.5) $\quad \exp(\ -\ I(Q,P)\ ) \geqslant \exp(\ -\ n\int Z^2 dp\ ) \quad .$

<u>Proposition 1</u> : Soit $(A, \mathcal{O}, \mu)$ un espace mesuré, avec μ positive et σ-finie, D comme dans (2.1), R le risque associé. Choisissons $\underline{f}$ densité de référence, A_j partition de A, Z_j des v.a.r. de support A_j, bornées en valeur absolue par $\frac{1}{2}$, et telles que $\int Z_j.\underline{f}.d\mu = 0$. Soit Θ l'ensemble de densités de Probabilité $\Theta = \left\{ g \mid g = \underline{f}.(1 + \sum b_j.Z_j) \; ; \; b_j = 0 \text{ ou } 1 \right\}$. Alors, si

(2.6) $\quad D_j = \varphi^{-1}\left(\frac{1}{2} \varphi\left(D(\underline{f}.1_{A_j} , \underline{f}.(1+Z_j).1_{A_j}) \right) \right)$, on a

(2.7) $\quad \bigwedge_{\hat{f}_n} \bigvee_{g \in \Theta} R_g(\hat{f}_n) \geqslant \frac{1}{4} \sum D_j \, . \exp(-n \int Z_j^2.\underline{f}.d\mu)$.

<u>Commentaire</u>: Dans le cas où φ^{-1} est à croissance modérée (il existe K tel que $\varphi^{-1}(2x) \leqslant K\varphi^{-1}(x)$), on remplacera D_j par $K^{-1}.D(\underline{f}1_{A_j}, \underline{f}(1+Z_j)1_{A_j})$

<u>Démonstration</u>: Soit $b = (b_j)$ variant dans un espace produit $B = \prod B_j$.

Notons $B^j = \prod_{i \neq j} B_i$, d'élément générique b^j. Alors, si $R(b) \geqslant \sum R_j(b)$, on a $\bigvee_{b \in B} R(b) \geqslant \sum \bigwedge_{b^j \in B^j} \bigvee_{b_j \in B_j} R_j(b)$.

Fixons maintenant $\hat{f}_n$, notons g_j la restriction de g à A_j, $g^j = g - g_j$. Ces deux fonctions ne dépendent de b que par b_j, respectivement b^j. Appliquons la sur-additivité de D, en posant $R_j(b) = E_g(D(\hat{f}_n 1_{A_j}, g_j))$ En vertu de la remarque précédente, on obtient comme minorant de (2.7) : $\sum \bigwedge_{g^j} \bigvee_{g_j} E_g(D(\hat{f}_n 1_{A_j}, g_j))$. Pour j et g^j fixés, appliquons la φ-inégalité triangulaire entre $\hat{f}_n 1_{A_j}$, $\underline{f}1_{A_j}$ et $\underline{f}(1+Z_j)1_{A_j}$ puis l'inégalité (2.4) du lemme précisée en (2.5). Il vient alors $\bigvee_{g_j} R_j(b) \geqslant 2^{-2}.D_j.\exp(-n \int Z_j^2. \underline{f}.d\mu)$ qui ne dépend plus de g^j ni de $\hat{f}_n$, d'où le résultat.

3. Risque minimax pour la norme p sur $\mathbb{R}$

Dans ce paragraphe, $(A,\mathcal{A},\mu)$ est $(\mathbb{R},\mathcal{B},dx)$, $D(f,g) = \int |f-g|^p dx$, qui répond à la définition (2.1) pour tout p positif en prenant $\varphi(x) = x^{\frac{1}{p}\wedge 1}$ et F est l'ensemble des densités de probabilité de classe C^{ν}, ν entier supérieur ou égal à 1, et telles que, pour un réel positif r, $\rho(f) \leqslant r$, où ρ est la jauge

$$(3.1) \qquad \rho(f) = \left(\|f\|_{\frac{p}{2}}^{\nu} \cdot \|f^{(\nu)}\|_p \right)^{\frac{p}{2\nu+1}}$$

en notant $D(f,g) = \|f-g\|_p^p$ même si p n'est pas supérieur à 1. Le risque au point f d'un estimateur $\hat{f}_n$ est donc dans toute la suite $R(\hat{f}_n,f) = E_{f^{\otimes n}}(\|\hat{f}_n - f\|_p^p)$, p positif.

On construit dans ce cadre un ensemble Θ de densités tel que celui de la poposition 1 :
Ayant choisi une densité de probabilité auxiliaire θ, de classe $C^{\nu-1}$ et de support contenu dans $(0,\frac{1}{2})$, on définit les fonctions

$$(3.2) \qquad \begin{aligned} \underline{f} &= 1_{(0,1)} * \theta \\ Z(x) &= \underline{f}(3x) - \underline{f}(3(x-\tfrac{1}{2})) \qquad & x\in\mathbb{R} \\ Z_j &= u.Z\circ\psi_j \qquad & j=1,2,..,J \end{aligned}$$

où u est un réel positif fixé inférieur à $\frac{1}{2}$ et ψ_j la bijection naturelle de (0,1), support de Z, sur A_j, $j^{\text{ème}}$ intervalle de la partition de $(\frac{1}{2},1)$ en J+1 intervalles disjoints, les J premiers de longueur d, le dernier de longueur moindre, éventuellement nulle, de sorte que 4Jd est compris entre 1 et 2. L'ensemble qui en résulte

$$(3.3) \qquad \Theta = \left\{ g = \underline{f}(1 + \sum_j b_j Z_j) \mid b_j \in \{0,1\} \right\}$$

est un ensemble de densités de probabilité de classe C^{ν}, de norme infinie inférieure à 1 puisque $\|\underline{f}\|_\infty = \|Z\|_\infty = 1$,et de support contenu dans $(0,\frac{3}{2})$.Appliquant à Θ la proposition 1 et tenant compte de ce que $\|f\|_2^2 \leqslant 1$ et $\|f\|_q^q$ reste compris entre $\frac{1}{2}$ et 3/2 pour tout q positif (car $\underline{f}$ vaut 1 sur $(\frac{1}{2},1)$),on trouve

$$(3.4) \qquad \bigwedge_{\hat{f}_n} \bigvee_{g\in\Theta} E_g\|\hat{f}_n - g\|_p^p \geqslant 2^{-(p+7)}u^p \exp(-\tfrac{2}{3}\, ndu^2)$$

Majorons la jauge de l'élément générique de Θ :Puisque $g \leqslant (1+u)\underline{f}$ et que $f^{(\nu)}$ et les $(f.Z_j)^{(\nu)}$ sont à supports disjoints,on a

$$\rho(g) \leqslant \rho(\underline{f})(1+u)^{\frac{\nu p}{2\nu+1}} \left(1 + \tfrac{2}{3}Jd(\tfrac{3}{d})^{\nu p}u^p\right)^{\frac{1}{2\nu+1}}$$

où $Jd \leqslant \frac{1}{2}$.

Pour stabiliser la jauge de Θ quand n croît,lions u et d par

$$(3.5) \qquad u = (\tfrac{d}{3})^{\nu}(3M)^{\frac{1}{p}} \qquad M \text{ fixé } \geqslant 1$$

obtenant ainsi

$$(3.6) \qquad \rho(g) \leqslant \rho(\underline{f})\,(1+u)^{\frac{\nu p}{2\nu+1}}\,(2M)^{\frac{1}{2\nu+1}}$$

Le minorant de la formule (3.4) prend alors la forme $Ax^X\exp(-Bx^Y)$ où $A = (3M).2^{-p-7}$, $X = \nu p$, $Y = 2\nu+1$, $B = (3M)^{2/p}.2n$ et $x = d/3$.

<u>Lemme 3.1</u>

Le maximum de la fonction $x \longmapsto x^X\exp(-Bx^Y)$,où X,Y et B sont des constantes positives,s'obtient en $x_0 = (X/BY)^{1/Y}$, et vaut,en posant $y=X/Y$, $B^{-y}y^y\exp(-y)$.

Pour chaque n,on choisit,en fonction de M, la famille Θ qui maximise le minorant de (3.4),et qui,d'après le lemme,correspond à $d=3x_0$ et $u=(3M)^{1/p}x_0$,valeurs qui,toutes les deux,tendent vers 0 quand n croît.Par suite,les contraintes $u \leqslant \frac{1}{2}$ et $d \leqslant \frac{1}{2}$ sont asymptotiquement vérifiées.

Notons

$$(3.7) \qquad r_o = \rho(\underline{f})$$

Pour tout r supérieur à $r_o.2^{1/(2\nu+1)}$,on peut,en choisissant un M tel que $(2M)^{1/(2\nu+1)}$ soit inférieur à r/r_o,avoir une suite de tels $\Theta = \Theta(n)$ qui ,asymptotiquement,ont une jauge inférieure ou égale à r,($(1+u)^{\nu p/(2\nu+1)} \xrightarrow[n\to\infty]{} 0$) ; d'où,en remarquant que $3.2^{1/(2\nu+1)}$ est inférieur à 4,le résultat asymptotique suivant

<u>Proposition 2</u>

Soit F l'ensemble des densités de probabilité par rapport à la mesure de Lebesgue sur $\mathbb{R}$,qui sont de classe C^ν,$\nu \geqslant 1$,et qui vérifient $\rho(f) \leqslant r$,où ρ est la jauge définie en (3.1) et r un nombre supérieur à $2^{1/(2\nu+1)}.r_o$, r_o défini en (3.7).Alors,pour tout p positif,notant

$$y = \frac{\nu p}{2\nu+1}$$

$$(3.8) \qquad \underline{\lim}_n \inf_{\hat{f}_n} \sup_{f\in F} E_f(\|\hat{f}_n - f\|_p^p) \geqslant \frac{r}{r_o}.y^y.e^{-y}.2^{-9-p}$$

<u>Remarque</u>

On peut transformer ce résultat en minoration non asymptotique,c'est à dire valable pour tout entier n supérieur ou égal à 1,si l'on multiplie le membre de gauche par $\underline{3^{-y}.2^{y-p}.6^{-\nu p}}$:

En effet,les contraintes $2u \leqslant 1, 6x \leqslant 1$ se traduisent par $x \leqslant x_1$,où $x_1 = \frac{1}{6} \wedge 2^{-1/\nu}(3M)^{-\nu/p}$.Si x_1 est supérieur ou égal à x_o,la minoration précédente est correcte,une fois multipliée par $(1+u)^{-y}$,soit,a fortiori,par $3^{-y}2^y$.Si x_1 est inférieur strictement à x_o,en prenant x_1 comme valeur,comme $\exp(-\frac{2}{3}ndu^2)$ est,en x_1, supérieur ou égal à e^{-y},on obtient un minorant en multipliant encore par $2^{-p}6^{-\nu p}$,puisque l'inf de trois réels positifs est certainement supérieur à leur produit dès que deux d'entre eux sont moindres que 1.

Deux cas se séparent :

Si $p \neq 1$ en utilisant l'homogénéité $(p-1)$ de la norme et de la jauge pour la transformation $h(.) \longrightarrow \frac{1}{a} h(\frac{\cdot}{a})$,$a > 0$ fixé, toute densité a une a-homothétique de jauge r_o et le résultat prend la forme du théorème 1 ci-dessous.

Si $p = 1$ on ne peut plus utiliser le raisonnement précédent car la norme et la jauge sont maintenant invariantes par changement d'échelle.Mais,dans ce cas,on se convainct aisément que la jauge est minorée par un réel strictement positif,sur l'ensemble des densités de probabilité.Le théorème 1 bis ne donne de résultats que pour les densités suffisamment éloignées de la densité minimale (au point de vue de la jauge).

Théorème 1

Soit p positif différent de 1, ν entier positif non nul,F l'ensemble des densités de probabilité par rapport à la mesure de Lebesgue sur $\mathbb{R}$,qui sont de classe C^{ν} et telles que $\rho(f) \leqslant r$,où r est positif et ρ est la jauge $\rho(f)=(\int f^{p/2})^{2\nu/2\nu+1} . (\int (f^{(\nu)})^p)^{1/(2\nu+1)}$.
Alors il existe une constante $C=C(\nu,p)$ telle que

$$\underline{\lim}_{n}\; n^{\nu p/(2\nu+1)} \inf_{\hat{f}_n} \sup_{f\in F} E_f \|\hat{f}_n - f\|_p^p \geqslant C.r \tag{3.9}$$

La constante C vaut

$$C(\nu,p) = 2^{-9-2p} . \left(\frac{\nu p}{2\nu+1}\right)^{\frac{\nu p}{2\nu+1}} . e^{\frac{-\nu p}{2\nu+1}} . \|\theta^{(\nu-1)}\|_p^{\frac{-p}{2\nu+1}} \tag{3.10}$$

où on peut prendre pour θ n'importe quelle densité de probabilité auxiliaire de support contenu dans $(0,\frac{1}{2})$ et de classe $C^{\nu-1}$.

Théorème 1 bis

Dans le cas où $p = 1$,on a les mêmes conclusions pourvu que l'on se restreigne aux valeurs de r supérieures à r_o,défini en (3.7) et (3.2).

Remarque

On obtient le résultat en dimension k en se fondant sur le même principe : $\|f^{(\nu)}\|_p^p$ est maintenant remplacé par $\int_{\mathbb{R}^k} \sum_{i=1}^{k} \left(\frac{\partial^\nu f}{\partial x_i^\nu}\right)^p \prod_{i=1}^{k} dx_i$.

On prend pour densité de base $\underline{f}$ une densité de classe C^ν, uniforme sur le cube $(\frac{1}{2},1)^k$, et on obtient Θ en remplaçant $\underline{f}$, sur les J^k cubes $Q_j = \{ x_i \in A_{j_i} \mid j_i \in \{1,2,..,J\}, i = 1,..,k\}$, par $\underline{f}\,(1 + b_j Z_j)$, où les Z_j sont les perturbations produits de celles qui sont utilisées sur $\mathbb{R}$.

On a alors pour une constante C qui dépend de k, ν, p mais non du majorant r de la jauge,

$$\underline{\lim}_n \; n^{\frac{\nu p}{2\nu+k}} \inf_{\hat{f}_n} \sup_{f\in F} E_f \|f_n - f\|_p^p \geqslant C.r \tag{3.11}$$

Ce résultat admet comme la proposition 2 (Voir la remarque qui la suit), une version non asymptotique.

4. Vitesse atteinte par les estimateurs à noyau

Ce paragraphe contient des résultats bien connus, mais, comme nous voulons les comparer à ceux du paragraphe précédent, nous avons préféré en donner une démonstration unitaire.

Soit η un noyau, alors η_Δ sera défini par $\eta_\Delta(x)=\Delta^{-1}\eta(\Delta^{-1}x)$.

L'estimateur à noyau correspondant à η et de pas Δ est ainsi défini à partir de l'échantillon (X) :

(4.1) $$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \eta_\Delta(x - X_i)$$. On pose

(4.2) $$\bar{f}(x) = E_f(\hat{f}_n)(x) = f * \eta_\Delta .$$

On sait bien que dans le cas où on veut utiliser au mieux la γ-dérivabilité de f, quand γ est strictement supérieur à 2, on ne peut assurer la positivité du noyau (noyaux de Parzen).

Si $\gamma = 1, 2$, on demande à η d'être pair, borné, d'intégrale 1, et enfin $\int x^2|\eta(x)|\,dx < \infty$. On peut ajouter $\eta \geqslant 0$; par exemple $\eta(x)=1_{|x|\leqslant\frac{1}{2}}$

Sinon, pour tout autre γ, on utilise un noyau, pair, borné, avec $\int \eta(x)dx = 1$, $\int x^j\eta(x)dx = 0$ ($1 \leqslant j < \gamma$), enfin $\|x^\gamma\eta\|_1 = \int |x|^\gamma|\eta(x)|\,dx$ fini .

(par exemple, $\eta(x) = (2\pi)^{-\frac{1}{2}}\,\tilde{\underline{f}}(x)\exp(-\frac{3}{4}ix)$, où $\tilde{\underline{f}}$ est la transformée de Fourier de $\underline{f}$ définie dans la formule (3.2))

Démontrons d'abord deux lemmes:

Lemme 4.1 Avec les hypothèses et notations précédentes, on a

(4.3) $$\|\bar{f}_\Delta - f\|_p \leqslant \frac{\Delta^\gamma}{\gamma!}\|x^\gamma\eta\|_1\,\|f^{(\gamma)}\|_p$$

démonstration: Par homogénéité, il suffit de prouver le résultat pour $\Delta = 1$. Ecrivant alors la formule de Taylor:

$$f(x+t)-f(x) = \sum_{j=1}^{\gamma-1} \frac{t^j}{j!} f^{(j)}(x) + \int_x^{x+t} \frac{(x+t-u)^{\gamma-1}}{(\gamma-1)!} f^{(\gamma)}(u)du,$$

il vient en intégrant : $(\bar{f}_1-f)(x) = f^{(\gamma)} * \phi\,(x)$, où le noyau pair ϕ défini pour $y>0$ par. $\phi(y) = \int_y^{\infty} \frac{(s-y)^{\gamma-1}}{(\gamma-1)!} \eta(s)ds$, a une 1-norme majorée par $(\gamma!)^{-1} \|x^{\gamma}\eta\|_1$. Le résultat découle alors de l'inégalité $\|f^{(\gamma)} * \phi\|_p \leq \|f^{(\gamma)}\|_p . \|\phi\|_1$.

<u>Lemme 4.2</u> (inégalité de Khintchine) : Soient Y_i des v.a.r. de même distribution, indépendantes, centrées et bornées par 1 en valeur absolue. On a alors

$$(4.4) \quad E\left|\sum_{i=1}^{n} Y_i\right|^p \leq \left\{nE(|Y|)\right\}^{\frac{1}{2}p} \quad \text{pour } 0 < p \leq 2.$$

$$(4.5) \quad E\left|\sum_{i=1}^{n} Y_i\right|^p \leq C_p\left[\left\{nE(|Y|)\right\} \vee \left\{nE(|Y|)\right\}^{\frac{1}{2}p}\right] \quad \text{pour } 2 < p.$$

On peut prendre pour C_p: $2^{k+1}(\,k!\,)^2$ si $2k-2 < p \leq 2k$.

<u>Démonstration</u>:Soit S_n la somme des Y_i. Pour $0 < p \leq 2$, on utilise l'inégalité $\|S_n\|_p < \|S_n\|_2 \leq [nE(|Y|)]^{\frac{1}{2}}$, car $|Y| \leq 1$.

sinon: une construction élémentaire dans le carré $[0, 1] \times [0, 1]$ muni de la mesure de Lebesgue montre que toute Y satisfaisant les conditions du lemme peut s'interpréter comme $E^{\mathcal{B}}(\varepsilon.1_B)$ où: ε est un signe ($P(\varepsilon = 1) = P(\varepsilon = -1) = \frac{1}{2}$) , <u>indépendant</u> de l'indicatrice 1_B, avec $E\,Y = P(B)$, et $\mathcal{B}$ σ-algèbre. L'inégalité de Jensen montrequ'alors, pour toute ϕ convexe, $E(\,\phi(S_n)\,) \leq E(\,\phi(\,\sum \varepsilon_i 1_{B_i}))$.

On est donc ramené au cas où $Y = \varepsilon 1_B$. Dans ce cas, si $p = 2k$,

$$E(S_n)^{2k} \leq 2^{k+1}.\,k!\,E\left(\sum 1_{B_i}\right)^{k} \quad (\text{ Formule de Khintchine })$$

Une majoration des moments factoriels $E(\ Z(Z-1)..(Z-j+1)\ )$ de la binomiale $X = B(\ n,\ E|Y|\ )$ donne alors le résultat.

Enfin, si $2k-2 < p \leq 2k$, on utilise alternativement les inégalités $\|S_n\|_p \leq \|S_n\|_{2k}$, quand $nE(|Y|) > 1$, et $\|S_n\|_p^p \leq \|S_n\|_{2k}^{2k}$, quand $nE(|Y|) \leq 1$, puisqu'on s'est placé dans le cas où S_n est une entier relatif.

<u>Corollaire du Lemme 4.2</u> Sous les mêmes hypothèses sur les Y_i, plus $E|Y_1| \leq \frac{1}{2}$, soit $\phi_p(u) = \sup_{nE(|Y|) \leq u} E\left|\sum_{i=1}^{n} Y_i\right|^p$. Alors

Pour $p > 2$, à tout $\varepsilon > 0$ on peut associer M tel que

$$(4.6) \qquad \phi_p(u) \leq M.|u| + |u|^{\frac{1}{2}p} \left(\frac{\Gamma(\frac{1}{2}(p+1))}{\Gamma(\frac{1}{2})} + \varepsilon \right).$$

<u>Démonstration</u>: Il suffit de remarquer que, le sup étant obtenu sur les Y du type $\varepsilon 1_B$, quand $nP(B)$ tend vers l'infini, on peut remplacer le C_p du lemme 4.2 , asymptotiquement, par $\Gamma(\ \frac{1}{2}(p+1)\ ) / \Gamma(\frac{1}{2})$, en vertu du théorème de limite centrale. La majoration du Lemme 4.2 est convexe, on peut donc la majorer au début par une fonction du type $M.|u|$.

Pour simplifier l'énoncé qui suit, faisons sur la densité f l'

$$(4.7)\ \underline{\text{Hypothèse } H_p} \ : \quad \lim_{\Delta \to 0} \int (f * |\eta_\Delta|)^{\frac{1}{2}p}(x)dx \leq \|\eta\|_1^{\frac{1}{2}p} \int f^{\frac{1}{2}p}(x)dx \quad (p<2)$$

Cette propriété sera en particulier vérifiée si f est à support compact, et par extension si elle est monotone à l'infini.

D'après l'inégalité de Jensen, elle l'est bien sûr automatiquement pour $p \geq 2$, puisque $|\eta_\Delta| / \|\eta\|_1$ est un noyau positif et d'intégrale 1 .

<u>Théorème 2</u>: Soit p un réel positif, γ un entier strictement positif.

Dans le cas $p<2$, on suppose que f satisfait (H_p). Il existe alors une suite d'estimateurs à noyaux $\hat{f}_n$, ne dépendant que des nombres $\|f^{(\gamma)}\|_p^p$ et $\int f^{\frac{1}{2}p}dx$, et une constante $D(\gamma,p)$ ne dépendant que de p et γ tels que

(4.8) $\overline{\lim}_{n}\, n^{\gamma p/(2\gamma+1)} E_f \|\hat{f}_n - f\|_p^p \leq D(\gamma,p).\,\rho(f)$ où

(3.1) $\rho(f) = \left\{\int f^{\frac{1}{2}p}dx\right\}^{2\gamma/(2\gamma+1)} .\, \| f^{(\gamma)}\|_p^{p/(2\gamma+1)}$

Pour le choix des $\hat{f}_n$ et une majoration de $D(\gamma,p)$, voir la remarque qui suit la démonstration.

<u>Démonstration</u>: Traitons d'abord le cas $p>2$. η et Δ étant choisis,

$$E_f\|\hat{f}_n - f\|_p^p \leq 2^{p-1}\left\{\|f-\bar{f}_\Delta\|_p^p + E_f\|\hat{f}_n - \bar{f}_\Delta\|_p^p\right\} = 2^{p-1}(B+A).$$

- B est majoré, d'après le lemme 2, par $\left\{\frac{\Delta^\gamma}{\gamma!}\|x^\gamma \eta\|_1\, \|f^{(\gamma)}\|_p\right\}^p$.

- Si l'on pose $Y_i(x) = (2\|\eta_\Delta\|_\infty)^{-1}.(\eta_\Delta(x-X_i) - f*\eta_\Delta(x))$, on peut appliquer le lemme 3 avec $E\,|Y|\,(x) \leq (2\|\eta\|_\infty)^{-1}.\,\Delta.f*|\eta_\Delta|(x)$.

A est donc majoré par

$$2^p\,\|\eta\|_\infty^p\,(n\Delta)^{-p}\int \phi_p\left(\frac{n\Delta}{2\|\eta\|_\infty} f*|\eta_\Delta|\right)(x)\,dx.$$

D'après le corollaire de ce lemme,

$$(2\|\eta\|_\infty)^{-p}(n\Delta)^{\frac{1}{2}p}B \leq \frac{M}{2\|\eta\|_\infty}(n\Delta)^{1-\frac{1}{2}p}\int f*|\eta_\Delta|\,dx$$
$$+\left(\frac{\Gamma(\frac{1}{2}(p+1))}{\Gamma(\frac{1}{2})} + \varepsilon\right).(2\|\eta\|_\infty)^{-\frac{1}{2}p}\int (f*|\eta_\Delta|)^{\frac{1}{2}p}dx.$$

Il en résulte que

$$\overline{\lim}_{n}\,(n\Delta)^{\frac{1}{2}p}\,B \leq \frac{\Gamma(\frac{1}{2}(p+1))}{\Gamma(\frac{1}{2})}\,(2\,\|\eta\|_1\|\eta\|_\infty)^{\frac{1}{2}p}.\int f^{\frac{1}{2}p}(x)\,dx\,.$$

En égalant asymptotiquement les deux termes, il vient

$$\overline{\lim_n} \; n^{\gamma p/(2\gamma+1)} \; E_f \|\hat{f}_n - f\|_p^p \leq$$

$$\leq \; 2^p \cdot \rho(f) \left\{ \frac{\Gamma(\frac{1}{2}(p+1))}{\Gamma(\frac{1}{2})} \right\}^{2\gamma/(2\gamma+1)} \left\{ 2 \|\eta\|_1 \|\eta\|_\infty \right\}^{\gamma p/(2\gamma+1)} \left\| \frac{x^\gamma \eta}{\gamma !} \right\|_1^{p/(2\gamma+1)}$$

Dans le cas $p \leq 2$, on obtient le même résultat en utilisant l'hypothèse (H_p) à celà près que le quotient des deux fonctions Γ est remplacé par 1.

<u>Remarques</u>: On a une expression explicite de

$$\mathcal{D}(\gamma, p) = 2^p \cdot \left\{ (2\|\eta\|_1 \|\eta\|_\infty)^{\gamma p} \left(\left\| \frac{x^\gamma \eta}{\gamma !} \right\|_1 \right)^p \left(\frac{\Gamma(\frac{1}{2}(p+1))}{\Gamma(\frac{1}{2})} \right)^{2\gamma} \right\}^{1/(2\gamma+1)}$$

et le pas optimal Δ_n est déterminé par

$$n.\Delta_n^{2\gamma+1} = \left\{ \left\| \frac{x^\gamma \eta}{\gamma !} \right\|_1 \|f^{(\gamma)}\|_p \right\}^{-2} \left\{ \frac{\Gamma(\frac{1}{2}(p+1))}{\Gamma(\frac{1}{2})} \; (2\|\eta\|_1 \|\eta\|_\infty)^{\frac{1}{2}p} \int f^{\frac{1}{2}p} \right\}^{2/p},$$

du moins pour $p > 2$. Pour $p \leq 2$, remplacer le quotient des fonctions Γ par 1

On peut obtenir également une égalité non asymptotique, c'est-à-dire valable pour tout n, en remplaçant le quotient des fonctions Γ par la constante C_p, dans le cas $p > 2$. Dans l'autre cas, il faudra faire entrer dans le majorant la quantité $\int (f * |\eta_\Delta|)^{\frac{1}{2}p} dx$.

5. Comparaison des résultats des paragraphes 3 et 4:

Les théorèmes 1 et 2 signifient que les estimateurs à noyaux sont optimaux du point de vue de l'ordre en n et ρ sur les espaces $\mathbb{L}_p^{(\gamma)}$ (à la restriction près de la remarque qui suit). Dans ce paragraphe, on étudie le comportement relatif des constantes $C(\gamma, p)$ et $D(\gamma, p)$ comme fonctions de γ et p.

Remarque: L'estimateur optimal à noyau fait intervenir séparément les deux facteurs composant la jauge. D'une part, on peut aisément adapter la démonstration du théorème 1 pour séparer ces deux facteurs (du moins pour $p \neq 1$), d'autre part, par adaptativité (comme dans Nadaraya (3) par exemple), on peut certainement construire une suite d'estimateurs ayant la même vitesse asymptotique sous le seul renseignement que la jauge est finie.

Proposition 3: Il existe une constante C (ne dépendant ni de γ ni de p) telle que

$$(5.1) \quad \frac{D(\gamma, p)}{C(\gamma, p)} \leqslant (C \cdot \gamma)^{\frac{1}{2}p} \qquad (p \geqslant 1, \ \gamma \geqslant 1).$$

Commentaire: 1.En réalité, on devrait prendre pour risque effectif la quantité $(E_f \|\hat{f}_n - f\|_p^p)^{1/p}$, auquel cas la Proposition 3 signifie que le rapport des vitesses est en $\gamma^{\frac{1}{2}}$.

2.On verra que la démonstration permet de déduire une majoration similaire non asymptotique du rapport des vitesses.

Démonstration: On utilise ici un noyau θ dont on suppose qu'il appartient à $\mathcal{C}^{(\gamma-1)}$, et η est choisi comme dans le paragraphe 4, en fonction de $\underline{f} = 1_{[0,1]} * \theta$. On va effectuer sur $\underline{f}$ une translation de -3/4, ce qui ne change évidemment rien aux résultats des théorèmes 1 et 2.

Posons $\gamma = j+k+1$ (j, k entiers positifs, avec en fait j = k si γ est impair, j = k+1 sinon). On peut remarquer que les fonctions auxiliaires $\underline{f}$ et η n'interviennent dans le rapport étudié que par l'intermédiaire de

$$(\|\eta\|_1 \|\eta\|_\infty)^{p/(2\gamma+1)} \left(\left\|\frac{x^\gamma \eta}{\gamma !}\right\|_1 \|f^{(\gamma)}\|_p\right)^{p/(2\gamma+1)} .$$

Soit $\Phi_j(x) = \frac{(2j+1)!}{2^{2j+1}(j!)^2}(1-x^2)^j.1_{|x|\leq 1}$ $(j\geq 0)$. Ce sont des densités de Probabilité, $\int \Phi_j^2\, dx \leq \frac{1}{2}$, enfin

$$\|\Phi_j^{(j)}\|_2 = \frac{(2j+1)!}{j!} 2^{-j-1} \sqrt{\frac{2}{2j+1}}$$ (en effet, $\Phi_j^{(j)}$ est proportionnelle au $j^{\text{ième}}$ polynôme de Legendre, ce qui permet de donner une expression explicite de sa norme quadratique). On pose $\varphi_j = 8\,\Phi_j(8x)$, et $\theta = \varphi_j * \varphi_k$ satisfait alors les conditions imposées, d'où, pour $p \geq 1$

$\|\theta^{(\gamma-1)}\|_p \leq 2^{-2/p} \|\varphi_j^{(j)}\|_2 \|\varphi_k^{(k)}\|_2$, en vertu de l'inégalité $\|f * g\|_p \leq \|f\|_t \|g\|_t$, si $1/p = 2/t - 1$, et du fait que $\varphi_j^{(j)}$ et $\varphi_k^{(k)}$ étant à support compact, on peut comparer leur 2-norme à leur t-norme .(Pour $p < 1$, on obtient la majoration

$\|\theta^{(\gamma-1)}\|_p \leq 2^{-1-1/p} \|\varphi_j^{(j)}\|_2 \|\varphi_k^{(k)}\|_2$, en passant par la 1- norme.) De même, $\|x^\gamma \eta\|_1 = (2\pi)^{-\frac{1}{2}} \| 2\sin(\frac{1}{2}x)\, x^j \varphi_j x^k \varphi_k\|_1$ $\leq 2(2\pi)^{\frac{1}{2}} \|\varphi_j^{(j)}\|_2 \|\varphi_k^{(k)}\|_2$, et $\|\eta\|_\infty = (2\pi)^{-\frac{1}{2}}$. Il suffit de reporter ces évaluations dans le rapport étudié pour obtenir le résultat.

<u>Commentaire</u>: Il est bien sûr possible de donner une majoration explicite de la constante C, ainsi que des constantes $C(\gamma,p)$ et $D(\gamma,p)$.

6. Application à la distance de Hellinger

Dans ce paragraphe, nous prenons $(A, \mathcal{A}, \mu) = (\mathbb{R}, \mathcal{B}, dx)$.
La fonctionnelle étudiée est la distance de Hellinger

$D(f,g) = \frac{1}{2}\int(\sqrt{f} - \sqrt{g})^2\, dx$, qui possède les propriétés (2.1), avec $\varphi(x) = \sqrt{x}$, et F est l'ensemble des densités de Probabilité de classe $\mathcal{C}^{\nu}$, avec $\nu = 1$ ou 2 , enfin telles que leur jauge $\rho_1(f)$ soit finie, ou ici

$$(6.1) \qquad \rho_1(f) = \| f \|_{\alpha}^{\frac{\alpha(p(\nu+1)-1)}{p(\nu+1)-\alpha}} \, \| f^{(\nu)} \|_p^{\frac{p(1-\alpha)}{p(\nu+1)-\alpha}}$$

Remarques:

1. $\nu = 1$ ou 2 : La distance de Hellinger n'étant définie que pour les fonctions positives, on est limité aux estimateurs positifs. Celà nous impose donc, pour les estimateurs de Parzen, $\nu = 1$ ou 2. Cependant, le résultat de minoration est valable pour ν quelconque.

2. Jauge : Dans le cas précédent de la p-norme, au lieu de prendre une jauge générale (la seule limitation étant le degré d'homogénéité), nous avons choisi directement la jauge qui donne l'ordre maximum en n. Ici, il correspond à $\alpha = 0$ et vaut $\frac{\nu}{\nu+1}$. Il n'y a rien d'étonnant à ce que cet ordre(mais non les constantes bien sûr) ne dépende pas de p, puisque dans le cas $\alpha = 0$, une densité de jauge finie est à support compact.

Nous choisissons comme famille de densités la famille

$$(6.2)\quad \Theta_1 = (1 - K.a)\underline{f} + a.\sum_{k=1}^{K} \Theta\,(\,.\, - 3k)$$

(somme disjointe de translatées des familles Θ introduites au paragraphe 3) où on choisit $u = \frac{1}{2}$, les autres paramètres satisfaisant : J et K entiers positifs. A, d réels positifs.

$$\frac{1}{4} \leqslant J.d \leqslant \frac{1}{2}\ . \qquad K.a \leqslant 1.$$ Nous allons les <u>lier</u>, de façon à contrôler la jauge, par, M étant choisi supérieur à 1,

$$K.a^{\alpha} = \left(\frac{2}{3}\right)^{\alpha} \quad \text{et} \quad K.a^{p} = 2^{p} \left(\frac{d}{3}\right)^{\gamma p} M^{(\,p(\gamma+1)-\alpha)/(1-\alpha)}$$

Dans ces conditions, on a asymptotiquement pour toute g de Θ_1 (on vérifiera à la fin du calcul que la valeur choisie de a tend vers 0 avec $1/n$)

$$\rho_1(g) \quad \leqslant \quad 2\, M\, \rho_1(\underline{f})$$

Sur cette famille, la Proposition 1 donne comme minorant du risque, compte-tenu de l'inégalité $(\sqrt{1+u} - 1)^2 \geqslant 0{,}2.\ u^2$, pour $u \leqslant \frac{1}{2}$,

$$2^{-12}.\ K\ .\ a\ .\ \exp\left(\ -\ \frac{n}{12}\ .\ a\ .\ d\ \right).$$

On pose maintenant , comme précédemment, $d = 3x$, on applique le Lemme 3.1 en tenant compte des liaisons. Il vient, de la même manière que dans les démonstrations de la Proposition 2

<u>Proposition 4</u>: Soit p un réel supérieur ou égal à 1, α réel compris entre 0 et 1. Il existe deux constantes C et r_o, telles que, si $r \geqslant r_o$

$$(6.3)\quad \liminf_{n}\ n^{\frac{\gamma\, p(1-\alpha)}{p(\gamma+1)-\alpha}} \bigwedge_{\hat{f}_n}\ \bigvee_{\rho_1(h) \leqslant r} E_f\, d_H^2(\hat{f}_n, f) \geqslant C.r.$$

<u>Remarques</u> 1. Comme précédemment, on pourra prendre pour valeur de r_o, $\rho_1(\underline{f})$ multipliée par une puissance convenable de 2.

2. La distance de Hellinger, comme la 1-norme, étant invariante par changement d'échelle, les normes ρ_1 sont minorées sur l'ensemble des densités de probabilité.

3. De même que dans les résultats précédents, on peut expliciter sans peine une formule non asymptotique.

Passons maintenant à la <u>majoration</u>: nous limitant à $\gamma = 1$ ou 2, nous prenons pour η le <u>noyau uniforme</u> $\eta(x) = 1_{|x|\leq \frac{1}{2}}$, et la formule (4.3) devient

$$\|\bar{f}_\Delta - f\|_p \leqslant (\tfrac{1}{2}\Delta)^\gamma \cdot \left\|\frac{f^{(\gamma)}}{(\gamma+1)!}\right\|_p \quad .$$

Nous majorons $E_f(\ d_H^2(\hat{f}_n, f)\ )$ par $A + B$ (aléa et biais) où

(6.4) $A : \mathbb{E}(\ \|\sqrt{\hat{f}_n} - \sqrt{\bar{f}_\Delta}\|_2^2\ ) \leqslant (n\Delta)^{\alpha-1}\|2\,\bar{f}_\Delta\|_\alpha^\alpha \quad (\ 0 \leqslant \alpha \leqslant 1)$

en effet, $(\sqrt{a} - \sqrt{b}\ )^2 \leqslant b^{-1}(a-b)^2 \wedge (a+b)$, si $0 \leqslant a,b$.

Or $n.\Delta\,\hat{f}_n = \sum \Delta\,\eta_\Delta(x-X_i) = S_n$ est une variable de loi binomiale pour tout x, puisque somme d'indicatrices. $E(\ \sqrt{\hat{f}_n} - \sqrt{\bar{f}}\ )^2$ est donc majoré par $(n\Delta)^{-1} \wedge (2n\Delta\bar{f}_\Delta) \leqslant (n\Delta)^{\alpha-1}\ (2\,\bar{f}_\Delta)^\alpha$.

(6.5) $B = \|\sqrt{\bar{f}_\Delta} - \sqrt{f}\|_2^2 \leqslant \|\bar{f}_\Delta - f\|_p^{(1-\alpha)/(p-\alpha)}\ \|\bar{f} + f\|_\alpha^{\alpha(p-1)/(p-\alpha)}$

en effet, $(\sqrt{a} - \sqrt{b}\ )^2 \leqslant (a-b)^q(a+b)^{1-q}$ ($0 \leqslant q \leqslant 2$), on appliquera l'inégalité de Hölder d'exposants conjugués q/p et $(p-q)/p$, après avoir choisi $q/p = (1-\alpha)/(p-\alpha)$.

On reporte maintenant dans (6.5) l'inégalité (4.3), on égale enfin les deux majorations obtenues, ce qui détermine Δ.

Proposition 5: Soit p supérieur à 1, α compris entre 0 et 1, $\gamma = 1$ ou 2. On suppose de plus que la densité f satisfait l'hypothèse ($H_{2\alpha}$) ((4.7)). Il existe alors une suite d'estimateurs à noyaux, ne dépendant que des nombres $\|f^{(\gamma)}\|_p$ et $\|f\|_\alpha$, tels que

$$(6.6) \quad \limsup_n n^{\frac{\gamma p(1-\alpha)}{p(\gamma+1)-\alpha}} E_f(d_H^2(\hat{f}_n, f)) \leq 4 \rho_1(f) \quad .$$

Remarque: Comme d'habitude, on peut modifier la constante pour avoir un résultat non asymptotique.

J. Bretagnolle Université Paris 13
L.A. n°224 du C.N.R.S.

C. Huber Université Paris 13
E.R.A. n°532 du C.N.R.S.

Références

(1) Abou-Jaoudé (1977) Thèse, Université de Paris.

(2) Farrell R.H. (1972) "On best obtainable asymptotic rates of convergence in estimation of a density function at a point." A.M.S. 43, 170-180.

(3) Le Cam L. (1975) "On local and global properties in the theory of asymptotic normality of experiments". Stochastic Processes and related topics, Vol.1 A.P. New york.

(4) Meyer Terry G. (1977) "Bounds for estimation of density functions and their derivatives" A.S. 5, n°1, 136-142.

(5) Nadaraya E.A. (1974) "On the integral mean square error of some non parametric estimates for the density function" Theor. of prob. and its applications 19, 133-141.

(6) Wahba Grace (1975) "Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation" A.S. 3, n°1, 15-29.