Z. Wahrscheinlichkeitstheorie verw. Gebiete
47, 119–137 (1979)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Estimation des densités: risque minimax

J. Bretagnolle et C. Huber

Université Paris-Nord, C.S.P. Département de Mathématiques,
Av. Jean Baptiste Clément, F-93430 Villetaneuse

1. Introduction

La plupart des travaux sur l'estimation des densités établissent les propriétés asymptotiques de certaines classes d'estimateurs, qu'il s'agisse de l'étude des vitesses de convergence comme dans [1, 4, 8, 10, 14, 15] ou d'études de convergence en loi, comme dans [2, 3, 9, 11]. Ces propriétés sont en général uniformes sur certaines classes de densités. Ces travaux concernent soit l'estimation d'une densité en un point (estimation ponctuelle), soit l'estimation de la densité toute entière (estimation globale).

En ce qui concerne l'estimation ponctuelle, le premier qui ait abordé le problème inverse, c'est-à-dire celui de la minoration du risque, est Farrell: ayant remarqué que les estimateurs optimaux dans chaque classe, qu'il s'agisse des estimateurs à noyau de Parzen [9], de ceux du maximum de vraisemblance [17], ou de ceux construits à partir de séries orthogonales [12, 16], atteignaient tous le même ordre en n, taille de l'échantillon, il a démontré dans [5] qu'il était impossible de faire mieux sur certaines classes de densités proches d'un polynôme, établissant ainsi une borne de Cramer-Rao en dimension infinie. Sa méthode consiste à employer un couple de densités de l'ensemble choisi, globalement voisines mais perturbées au voisinage du point où se fait l'estimation. Le couple est adapté à n.

Wahba dans [14] a étendu ce résultat ponctuel à des espaces de Sobolev de densités, en calculant par ailleurs la vitesse effectivement atteinte sur ces mêmes espaces pour quatre classes distinctes d'estimateurs.

C'est ce travail qui nous a suggéré de faire l'étude analogue pour l'estimation globale. La méthode est similaire, à cela près que la perturbation n'est évidemment plus locale, et qu'au lieu de prendre un couple de densités dans l'ensemble considéré, on en prend un nombre croissant avec n. La minoration repose sur la Proposition 2.1, utilisant l'information de Kullback, et inspirée des idées de Le Cam. Dans le §3, on obtient un résultat analogue à celui de Wahba pour le risque correspondant à la p-norme. La jauge introduite est la jauge naturelle du problème, comme on le voit dans le §4, où on démontre que la vitesse maximale

0044-3719/79/0047/0119/$03.80

est effectivement atteinte par des estimateurs de Parzen bien choisis. A notre connaissance, Abou-Jaoudé [1] est le premier ayant introduit de telles jauges, dans le cas $p=1$. Le §4 se termine par une estimation adaptative du paramètre d'échelle du problème, en suivant les idées de Nadaraya [8] qui a traité le cas $p=2$ pour les densités ayant une dérivée de carré intégrable. Le §5 traite rapidement de la distance de Hellinger. Récemment, Meyer [7] a obtenu des minorants du risque minimax, dans le cas ponctuel et dans le cas global, pour certaines classes de densités, fondés sur le comportement à l'infini de la fonction caractéristique; sa construction semble plus adaptée au problème ponctuel qu'au problème global.

On notera que les résultats obtenus sont en fait non asymptotiques, bien que, pour la commodité, ils soient donnés pour n infiniment grand.

Dans la suite, $(X_1, X_2, \ldots, X_n)$ est un n-échantillon d'une variable aléatoire (v.a.) à valeurs dans l'espace mesuré $(A, \mathfrak{A}, \mu)$, de loi $f\cdot\mu$, μ positive et σ-finie. $\hat{f}_n = \hat{f}_n(x, X_1, X_2, \ldots, X_n)$, application mesurable de $(A, \mathfrak{A}) \times (A^n, \mathfrak{A}^{\otimes n})$ dans $\mathbb{R}$ ou $\mathbb{R}^+$, estime f au point x.

D étant une application de $\mathscr{L}^0 \times \mathscr{L}^0$ dans $\mathbb{R}^+$, ($\mathscr{L}^0$ désignant les fonctions mesurables), on lui associe son risque R par la formule $R(\hat{f}_n, f) = E_f D(\hat{f}_n, f)$, E_f signifiant ici Espérance pour la loi $[f\cdot\mu]^{\otimes n}$.

Le p-risque, noté R_p, correspond au cas de la norme dans $\mathbb{L}^p(A, \mathfrak{A}, \mu)$, soit $D(f, g) = \|f-g\|_p^p$.

L'objet de l'étude est le suivant: Sachant a priori que la densité f appartient à un sous-ensemble F de $\mathscr{L}^0$, déterminer des constantes réelles positives c, d, u ne dépendant que de F telles que l'on ait simultanément

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

et, pour un estimateur $\hat{f}_n$ au moins

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

Au moment où nous achevons la correction de la version définitive de ce travail nous prenons connaissance de l'article de Boyd et Steele [19] qui proposent une minoration asymptotique de $\operatorname{Sup}_f \operatorname{Inf}_{\{\hat{f}_n\}} \limsup_n n R(\hat{f}_n, f)$, où le risque est le risque quadratique.

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 $\mathscr{L}^0 \times \mathscr{L}^0$ (ou de $L\times L$, L partie de $\mathscr{L}^0$) 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,

$$D(g,h) \geqq \sum_j D(g \cdot 1_{A_j}, h \cdot 1_{A_j})$$

ϕ-distance: il existe une fonction de $\mathbb{R}^+$ dans $\mathbb{R}^+$, croissante, de fonction réciproque notée Φ^{-1}, telle que pour tout triplet f, g, h, on ait la ϕ-inégalité triangulaire:

$$\phi(D(f,h)) + \phi(D(g,h)) \geqq \phi(D(f,g)).$$

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

Soient Q, Q' deux probabilités sur le même espace. L'*information de Kullback* de Q' par rapport à Q est notée $K(Q', Q)$, définie par

$$(2.1) \quad K(Q',Q) = -E_Q\left(\mathrm{Log}\frac{dQ'}{dQ}\right)$$

si Q' est absolument continue par rapport à Q, $+\infty$ sinon. La parenté $\pi(Q',Q)$ vaut $\int dQ' \wedge dQ$. Elles sont reliées par:

$$\exp(-K(Q',Q)) \leqq 2\pi(Q',Q)$$

qu'on démontre en appliquant l'inégalité de Jensen (pour la Q-Probabilité) à la v.a. $\mathrm{Log}\left(\frac{dQ'}{dQ}\right) = \mathrm{Log}\left(\frac{dQ'}{dQ} \wedge 1\right) + \mathrm{Log}\left(\frac{dQ'}{dQ} \vee 1\right)$, soit

$$(2.2) \quad \exp(-K) \leqq [\textstyle\int dQ' \wedge dQ] \cdot [\int dQ' \vee dQ] = \pi \cdot (2-\pi) \leqq 2\pi.$$

Lemme 2.1. *Soit ϕ concave croissante de $\mathbb{R}^+$ dans $\mathbb{R}^+$, de fonction réciproque ϕ^{-1}, D un réel positif. Soient U et U' deux v.a. positives définies sur le même espace, de lois respectives Q et Q', telles que:*

$$\phi(U) + \phi(U') \geqq \phi(D).$$

Alors:

$$(2.3) \quad E_Q(U) + E_{Q'}(U') \geqq \phi^{-1}(\tfrac{1}{2}\phi(D)) \cdot \exp(-K(Q',Q)).$$

Démonstration. Un minorant du premier membre est $\int (U+U') \cdot (dQ \wedge dQ')$; un minorant de $(U+U')$ est $2 \cdot \phi^{-1}(\frac{1}{2}\phi(D))$, en utilisant la concavité.

Remarque. Ne gardant que ϕ croissante, (2.3) reste vraie, à condition de multiplier le membre de droite par $\frac{1}{2}$, mais, dans les applications suivantes, ϕ est concave.

Soit maintenant P une Probabilité, Z une v.a. telle que $|Z| \leqq 1$, $E_p(Z)=0$. Si $dQ=(1+Z)\cdot dP$, $dQ'=(1-Z)\cdot dP$, Q et Q' sont deux Probabilités, et comme

$$\left|(1+z)\operatorname{Log}\left(\frac{1-z}{1+z}\right)-2z\right| \leqq (1-|z|)^{-1}\cdot 2z^2,$$

pour $|z| \leqq 1$, on a dans ce cas:

(2.4) $K(Q',Q) \leqq 2\cdot(1-\|Z\|_\infty)^{-1}\cdot E_P(Z^2)$.

Proposition 2.1.[1] *Soit, sur* $(A, \mathfrak{A}, \mu)$, μ *positive* σ*-finie*, $dP=g_0\cdot d\mu$ *une Probabilité. On se donne des* Z_j *de supports* A_j *disjoints, telles que* $|Z_j| \leqq 1$, $E_P(Z_j)=0$, *pour* $1 \leqq j \leqq J$. *Soit alors* $\Theta=\{g \mid g=g_0\cdot(1+\sum \varepsilon_j\cdot Z_j)$, *où* (ε_j) *varie dans* $\{-1, +1\}^J\}$. *Alors*

(2.5) $\operatorname*{Inf}_{\hat{f}_n}\operatorname*{Sup}_{g\in\Theta} R(\hat{f}_n, g) \geqq \frac{1}{2}\cdot\sum \phi^{-1}(\frac{1}{2}\phi(D_j))\cdot\exp(-n\cdot I_j)$,

où

$$I_j=2\cdot(1-\|Z_j\|_\infty)^{-1}\cdot E_P(Z_j^2), \qquad D_j=D(g_0\cdot(1+Z_j)\cdot 1_{A_j}, \quad g_0\cdot(1-Z_j)\cdot 1_{A_j}).$$

Remarque. Là encore, on suppose ϕ concave, sinon multiplier le minorant par $\frac{1}{2}$.

Démonstration. La sur-additivité du risque donne:

$$R(\hat{f}_n, g) \geqq \sum E_g(D(\hat{f}_n\cdot 1_{A_j}, g_0\cdot(1+\varepsilon_j Z_j)\cdot 1_{A_j})).$$

Evaluons le risque Bayésien du second membre, pour la mesure uniforme 2^{-J} sur $\{-1, +1\}^J$. Pour le $j^{\text{ème}}$ terme, on prend d'abord l'espérance en ε_j: on applique (2.3), (avec un facteur $\frac{1}{2}$), à

$$dQ=\{g_0\cdot(1+\sum_{k\neq j}\varepsilon_k\cdot Z_k+Z_j)\cdot d\mu\}^{\otimes n}, \qquad dQ'$$

(la même, en remplaçant Z_j par $-Z_j$), $U=D(\hat{f}_n\cdot 1_{A_j}, g_0\cdot(1+Z_j)\cdot 1_{A_j})$, U' la même en remplaçant Z_j par $-Z_j$; on applique la ϕ-inégalité triangulaire, enfin on remarque que $K(R^{\otimes n}, S^{\otimes n})=n\cdot K(R,S)$, et on applique (2.4). La minoration obtenue ne dépend plus des ε_k, pour $k\neq j$, ni de $\hat{f}_n$. Enfin, le risque Bayésien minore le risque maximum.

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

Dans ce paragraphe, $D(f,g)=\|f-g\|_p^p$, $\phi^{-1}(x)=x^p$, $\phi(x)=x^{1/p}$, et p est un réel $\geqq 1$. $(A, \mathfrak{A}, \mu)$ est $(\mathbb{R}, \mathscr{B}, dx)$. Les densités f sont supposées avoir une dérivée d'ordre s (au sens de Lebesgue) dans $\mathbb{L}^p$, s entier $\geqq 1$. On les munit de la *jauge*

[1] La démonstration incorrecte de ce résultat, qui figurait dans la première version de ce travail [Lecture Notes in Mathematics, Vol. 649, Springer Verlag] a été rétablie sur les indications de P. Assouad et de l'un des rapporteurs

(3.1) $\rho_{s,p}(f) = \|f^{(s)}\|_p^{\frac{p}{2s+1}} \cdot \|f\|_{p/2}^{\frac{sp}{2s+1}}$,

et, si F est un ensemble de densités, sa jauge est $\rho_{s,p}(F) = \operatorname{Sup}_{f \in F} \rho_{s,p}(f)$. La h-homothétique d'une densité f est définie par $f_h(x) = \frac{1}{h} \cdot f\left(\frac{x}{h}\right)$, c'est une densité de Probabilité, et $\rho_{s,p}(f_h) = h^{p-1} \cdot \rho_{s,p}(f)$. Cette jauge a donc la même homogénéité en h que la (p-norme)^p, ce qui nous permettra de limiter l'étude aux densités de jauge fixée, du moins pour $p \neq 1$.

On pourrait faire le même type de calcul pour toute jauge ayant la même homogénéité que D; celle-ci est choisie car elle apparaît comme la jauge naturelle du problème, comme on le verra dans le chapitre suivant.

Pour minorer le p-risque, on va appliquer P.2.1. à une densité centrale g_0 autour de laquelle on va adapter un ensemble Θ en fonction de la taille n de l'échantillon.

Choix de g_0. On choisit une fonction v positive, d'intégrale 1, à support contenu dans $[-\frac{1}{4}, \frac{1}{4}]$, de classe $\mathscr{C}^{(s-1)}$. I étant l'intervalle $[-\frac{1}{2}, \frac{1}{2}]$, on pose:

(3.2) $g_0 = 1_I * v$.

g_0 est alors de classe $\mathscr{C}^{(s)}$, positive et bornée par 1, nulle en dehors de $[-\frac{3}{4}, \frac{3}{4}]$, égale à 1 sur $[-\frac{1}{4}, \frac{1}{4}]$, et donc $\frac{1}{2} \leqq \|g_0\|_q^q \leqq \frac{3}{2} (0 \leqq q)$; puisque bornée par 1, $\|g_0\|_q \leqq 1$ $(q \geqq 1)$; enfin, pour $s \geqq 1$,

$$\|g_0^{(s)}\|_q^q = 2 \cdot \|v^{(s-1)}\|_q^q \qquad (0 \leqq q < \infty).$$

Choix de Θ. On pose $Z_{u,h}(x) = u \cdot \left(g_0\left(\frac{x}{h} + \frac{3}{4}\right) - g_0\left(\frac{x}{h} - \frac{3}{4}\right)\right)$, cette fonction est de somme nulle, bornée par u, de support $[-\frac{3}{2}h, +\frac{3}{2}h]$, somme disjointe de ses deux composantes. On a donc:

$$\|Z_{u,h}^{(s)}\|_q^q = 2 \cdot u^q \cdot h^{-sq+1} \cdot \|g_0^{(s)}\|_q^q \qquad (\text{y compris pour } s=0).$$

On s'impose dans la suite:

(3.3) $6 \cdot \sum h_j < 1$ et naturellement $0 \leqq u_j \leqq 1$.

Il existe alors des x_j tels que les $A_j = [x_j - \frac{3}{2} \cdot h_j, x_j + \frac{3}{2} \cdot h_j]$ soient disjoints et tous contenus dans $\{g_0 = 1\}$, de sorte que les $Z_j(x) = Z_{u_j, h_j}(x - x_j)$ forment avec g_0 une famille Θ dont les paramètres sont:

(3.4)
$$\phi^{-1}(\tfrac{1}{2}\phi(D_j)) = \|Z_j\|_p^p = 2 \cdot u_j^p \cdot h_j \cdot \|g_0\|_p^p \geqq u_j^p \cdot h_j$$
$$I_j \leqq 2 \cdot (1-u_j)^{-1} \cdot \|Z_j\|_2^2 \leqq 4 \cdot u_j^2 \cdot h_j \cdot (1-u_j)^{-1}.$$

La proposition 2.1. nous donne comme minorant pour la p-norme, *sous* (3.3)

(3.5) $\operatorname*{Inf}_{\hat{f}_n} \operatorname*{Sup}_{g \in \Theta} R_p(\hat{f}_n, g) \geqq \frac{1}{2} \cdot \sum u_j^p \cdot h_j \cdot \exp(-4n \cdot u_j^2 \cdot h_j \cdot (1-u_j)^{-1})$.

Jauge de Θ. Toutes les g de Θ ont même jauge, et si $u=\bigvee u_j$, on a

$$(1-u)^{p/2}\cdot\|g_0\|_{p/2}^{p/2}\leqq\|g\|_{p/2}^{p/2}\leqq(1+u)^{p/2}\cdot\|g_0\|_{p/2}^{p/2},$$

et

$$\|g^{(s)}\|_p^p=(1+2\cdot\sum u_j^p\cdot h_j^{-sp+1})\cdot\|g_0^{(s)}\|_p^p$$

soit:

$$(3.6)\quad (1-u)^{\frac{sp}{2s+1}}\cdot(1+2\cdot\sum u_j^p\cdot h_j^{-sp+1})^{\frac{1}{2s+1}}\leqq\frac{\rho_{s,p}(\Theta)}{\rho_{s,p}(g_0)}$$

$$\leqq(1+u)^{\frac{sp}{2s+1}}\cdot(1+2\cdot\sum u_j^p\cdot h_j^{-sp+1})^{\frac{1}{2s+1}}.$$

Calcul asymptotique. Si on prend tous les h_j égaux à h, en nombre J tendant vers l'infini, vérifiant (3.3), tous les u_j égaux à u, lié à h par:

$$(3.7)\quad u^p\cdot h^{-sp}=a>0$$

h tend vers 0 ainsi que u, et $\rho(\Theta)$ tend vers $\rho(g_0)\cdot\left(1+\frac{a}{3}\right)^{\frac{1}{2s+1}}$. Combinant (3.5) et (3.7), on est conduit à choisir h^{2s+1} équivalent à $\frac{sp}{2s+1}\cdot(4\cdot n\cdot a^{2/p})^{-1}$ (ce qui assure que h et u tendent vers 0 quand n tend vers l'infini) et qui, compte-tenu de $6\cdot J\cdot h$ équivalent à 1, donne comme minorant:

$$\frac{1}{12}\cdot\left\{\frac{sp}{4\cdot n\cdot e\cdot(2s+1)}\right\}^{\frac{sp}{2s+1}}\cdot a^{\frac{1}{2s+1}}.$$

Proposition 3.1. *Soit* $s\geqq 1$, $p\geqq 1$, g_0 *une densité auxiliaire* (*définie en* (3.2)). *Soit:*

$$r_0=\rho_{s,p}(g_0),\quad \textit{et}\quad r>r_0.$$

Alors:

$$(3.8)\quad \liminf_n n^{\frac{sp}{2s+1}}\cdot\operatorname*{Inf}_{\hat{f}_n}\ \operatorname*{Sup}_{\rho_{s,p}(f)<r}\ R_p(\hat{f}_n,f)$$

$$\geqq\frac{1}{12}\cdot\left(\frac{sp}{4\cdot e\cdot(2s+1)}\right)^{\frac{sp}{2s+1}}\cdot\left(3\cdot\left(\left(\frac{r}{r_0}\right)^{2s+1}-1\right)\right)^{\frac{1}{2s+1}}.$$

Démonstration. On choisit a de façon qu'asymptotiquement, Θ soit dans

$$F_r=\{f\,|\,\rho_{s,p}(f)\leqq r\},$$

soit $a=3\cdot\left(\left(\frac{r}{r_0}\right)^{2s+1}-1\right)$, ce qui donne bien un $a>0$, si $r>r_0$.

Deux cas se séparent alors. Si $p\neq 1$, on a remarqué que par homothétie, on peut transformer F_r en $F_{r'}$, ce qui entraîne la linéarité du minorant en r, donc en

$\frac{r}{r_0}$. La meilleure constante se calcule en multipliant le second membre de (3.8) par $\frac{r_0}{r}$ et en faisant tendre r vers l'infini.

On ne peut le faire pour $p=1$, l'homothétie laissant F_r invariant. La jauge $\rho_{s,1}$ a un infimum strictement positif, mais la technique employée ne permet d'obtenir un minorant que pour $r>r_0$, jauge d'une g_0 auxiliaire. On minorera dans ce cas

$$\left(\left(\frac{r}{r_0}\right)^{2s+1}-1\right)^{1/(2s+1)} \quad \text{par } \frac{r}{r_0}-1.$$

Pour donner une forme lisible au résultat, on remarque que:

$$\left\{\frac{sp}{4\cdot e\cdot(2s+1)}\right\}^{\frac{sp}{2s+1}} \geqq e^{-4}, \quad \text{que } \frac{1}{12}\cdot 3^{\frac{1}{2s+1}} \geqq 2^{-4},$$

et on a donc démontré le:

Théorème 3.1. *Soit $s\geqq 1$, $p\geqq 1$ et F_r l'ensemble des densités de Probabilité sur $\mathbb{R}$ qui vérifient:*

$$\|f^{(s)}\|_p^{\frac{1}{2s+1}}\cdot\|f\|_{p/2}^{\frac{sp}{2s+1}}=\rho_{s,p}(f)\leqq r.$$

Soit g_0 une densité auxiliaire de type (3.2), *de jauge $\rho_{s,p}(g_0)=r_0$. Alors:*

$$\liminf_n n^{\frac{sp}{2s+1}}\cdot\operatorname*{Inf}_{\hat f_n}\operatorname*{Sup}_{f\in F_r} E_f\|\hat f_n-f\|_p^p\geqq(2\cdot e)^{-4}\cdot\frac{r}{r_0} \qquad \textit{pour } p>1 \tag{3.9}$$

$$\geqq(2\cdot e)^{-4}\cdot\left(\frac{r}{r_0}-1\right) \qquad \textit{pour } p=1.$$

Calcul non asymptotique. Les évaluations précédentes sont bien sûr non asymptotiques. On peut montrer par exemple que pour $1<p\leqq 2$, et uniformément en s, la minoration précédente est effective dès que $n\geqq 10$, à condition de multiplier le membre de droite par 1/3. (Le calcul se simplifie notablement pour $p\leqq 2$, car, par un argument de convexité, le facteur $(1+u)^{\frac{sp}{2s+1}}$ disparait dans le membre de droite de (3.6).)

4. Vitesse atteinte par les estimateurs à noyau

Dans ce paragraphe, f est une densité de Probabilité sur $\mathbb{R}$, s fois dérivable (au sens de Lebesgue) avec $f^{(s)}$ dans $\mathbb{L}^p$, s étant un entier, p un réel, $s\geqq 1$, $p\geqq 1$. Le problème de l'estimation de f en norme p est complètement traité dans la littérature dans le cas $p=2$, s quelconque (voir par exemple [8]), ainsi que le cas $p=1$, $s=1,2$ (voir [1]) mais, voulant comparer les résultats à ceux du paragraphe précédent, nous avons préféré donner une démonstration unitaire. Les cas $p \geqq 2$ et $1 \leqq p \leqq 2$ sont sensiblement différents, nous traitons d'abord le cas $p \geqq 2$; cependant, les définitions et résultats intermédiaires sont, sauf mention explicite, valables pour $p \geqq 1$, $s \geqq 1$.

$\mathscr{C}_c^{(k)}$ désignera l'espace des fonctions k fois continûment dérivables, à support compact, $*$ la convolution.

$\hat{\mu}_n$ est la mesure empirique associée à un n-échantillon, soit

$$\hat{\mu}_n = n^{-1} \cdot \sum_{i=1}^{n} \partial_{X_i}$$

K étant alors une fonction, le $n-K$ estimateur à noyau est défini par la formule

(4.1) $\hat{f}_{n,k} = \hat{\mu}_n * K$, noté aussi $\hat{f}$ et

(4.2) $\bar{f}_{n,K} = E_f(\hat{f}_{n,K}) = f * K$, noté aussi $\bar{f}_K$, ou $\bar{f}$.

L'Aléa est la quantité $\hat{f} - \bar{f}$, et le *Biais* est $f - \bar{f}$.

L'inégalité $E_f \|\hat{f} - f\|_p^p \leqq 2^{p-1}(\|f - \bar{f}\|_p^p + E_f \|\hat{f} - \bar{f}\|_p^p)$ nous conduit à majorer séparément les termes de Biais et d'Aléa.

Le choix du noyau K est dicté par l'ordre de dérivabilité s: on sait bien (voir par exemple [9]), qu'on ne peut assurer la positivité de K pour $s > 2$. On dira que K est un *noyau de Parzen d'ordre s* (ou *s-noyau*) si

(4.3) K est continu, borné, pair, $\int K(x)\,dx = 1$, $\int x^j K(x)\,dx = 0$ pour $1 \leqq j < s$, enfin $\int |x|^s |K(x)|\,dx$ est fini.

L'existence de tels noyaux est prouvée dans l'appendice, où on voit que l'on peut de plus imposer que K soit dans $\mathscr{C}_c^{(\infty)}$.

Au s-noyau K, on associe le noyau ${}_sK$ défini par

(4.4) $${}_sK(x) = (-1)^s \int_x^\infty \frac{(y-x)^{s-1}}{(s-1)!} K(y)\,dy \quad \text{pour } x > 0, \quad {}_sK(-x) = -(-1)^s {}_sK(x).$$

${}_sK$ est alors dans $\mathbb{L}^1$ et $\mathscr{C}^{(s)}$, et ${}_s(K_h) = h^s \cdot ({}_sK)_h$.

Le cas échéant, on notera $K^{(s)}$ le noyau s-dérivé de K. Les trois noyaux seront dans $\mathscr{C}_c^{(\infty)}$ *et donc dans tous les* $\mathbb{L}^q (0 \leqq q \leqq \infty)$ *si* $K \in \mathscr{C}_c^{(\infty)}$.

Evaluation du terme de Biais. La formule de Taylor:

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

convoluée par le *s-noyau K*, donne

(4.5) $f * K - f = f^{(s)} * {}_sK$

et donc

(4.6) $\|f * K_h - f\|_p \leqq h^s \cdot \|f^{(s)}\|_p \, \|{}_sK\|_1 \quad (p \geqq 1)$.

Choisissant de plus K dans $\mathscr{C}_c^{(\infty)}$, et notant p' le conjugué de p (soit $(p')^{-1}+p^{-1}=1$), une autre conséquence de (4.5) est

$$\|f\|_\infty \leqq \|f*K\|_\infty + \|f^{(s)} *_s K\|_\infty \leqq \|f\|_1 \, \|K\|_\infty + \|f^{(s)}\|_p \, \|_s K\|_{p'},$$

et donc: si $f^{(s)}$ est dans $\mathbb{L}^p$ pour un $s \geqq 1$, un $p \geqq 1$, alors la densité de Probabilité f est *bornée* et dans tous les $\mathbb{L}^q$, pour $q \geqq 1$.

Evaluation de l'Aléa. Si K est un noyau borné, en tout x l'aléa $(\hat{f}-\bar{f})(x)$ apparait comme une somme de v.a.r. centrées bornées dont on va évaluer le moment d'ordre p. Pour cela, établissons le lemme suivant:

Lemme 4.1. (d'après [6] et [18]).

Soient

$$\mathscr{Z}_u = \{Z \mid Z = \textstyle\sum z_i, \ z_i \ \textit{indépendantes}, \ |z_i| \leqq 1, \ \sum E(|z_i|) \leqq u\},$$
$$\mathscr{Y}_u = \{Y \mid Y = \textstyle\sum y_i, \ y_i \ \textit{indépendantes centrées}, \ |y_i| \leqq 1, \ \sum E(y_i^2) \leqq u^2\},$$
$$\Psi_q(u) = \operatorname{Sup}\{E(|Z|^q) \mid Z \in \mathscr{Z}_u\}, \qquad \Phi_p(u) = \operatorname{Sup}\{E(|Y|^p) \mid Y \in \mathscr{Y}_u\}.$$

(4.7) *Pour* $0 \leqq q \leqq 1, \ 0 \leqq p \leqq 2, \ \Psi_q(u) = u^q, \ \Phi_p(u) \leqq u^p$

(4.8) *Pour* $1 \leqq q, \ \Psi_q(u) = E(P_u^q)$, *où* P_u *est une v.a. de Poisson avec* $E(P_u) = u$.

(4.9) *Pour* $2 \leqq p, \ \Phi_p(u) \leqq 2^p \cdot C_p \cdot \Psi_{p/2}\left(\frac{u^2}{2}\right)$, *où* C_p *est la constante de Khintchine, ici* $C_p = 2^{p/2} \cdot \Gamma\left(\frac{p+1}{2}\right) \cdot \left(\Gamma\left(\frac{1}{2}\right)\right)^{-1}$.

Démonstration. Voir Appendice.

Corollaire 1. *Soient* $p > 2, \ q > 1, \ a > 0$. *Il existe deux constantes* $N_{a,q}$ *et* $M_{a,p}$ *telles que:*

(4.10) $\quad \Psi_q(u) \leqq N_{a,q} \cdot u + (1+a) \cdot u^q,$

(4.11) $\quad \Phi_p(u) \leqq M_{a,p} \cdot u^2 + (1+a) \cdot C_p \cdot u^p.$

Démonstration. Pour k entier, $E(P_u^k)$ est un polynôme en u, à coefficients positifs, équivalent à u ou u^q quand u est équivalent à 0 ou l'infini. Le résultat est donc vrai pour $q=k$. Si $q \leqq k$, on a $E(P_u^q) \leqq E(P_u^k)$ et $E(P_u^q) \leqq (E(P_u^k))^{q/k}$, d'où (4.10). Si on reporte ce résultat dans (4.9), on obtient seulement

$$\Phi_p(u) \leqq M_{a,p} \cdot u^2 + (1+a) \cdot 2^{p/2} \cdot C_p \cdot u^p.$$

Pour éliminer ce $2^{p/2}$ parasite, on peut remarquer que d'après le Théorème de limite centrale avec convergence des moments, $\Phi_p(u)$ est équivalente à l'infini à $C_p \cdot u^p$, d'où la majoration (4.11) en changeant de constante $M_{a,p}$.

Corollaire 2. *Avec les mêmes hypothèses et les mêmes notations que dans le Corollaire précédent*

(4.12) $\quad E\int |\hat{f}_{n,K}|^q \leqq N_{a,q} \cdot n^{1-q} \cdot \|K\|_\infty^{q-1} \cdot \|K\|_1 + (1+a) \cdot \|K\|_q^q$

(4.13) $\quad E\int |\hat{f}_{n,K} - \bar{f}_K|^p \leqq M_{a,p} \cdot n^{1-p} \cdot \|2K\|_\infty^{p-2} \cdot \|K\|_2^2$
$$+ (1+a) \cdot C_p \cdot n^{-p/2} \cdot \|K\|_2^p \cdot \|f\|_{p/2}^{p/2}.$$

On peut remplacer le dernier terme de (4.13) *par* $(1+a)\cdot C_p\cdot n^{-p/2}\cdot\|K\|_p^p$.

Démonstration. Pour tout x, $\|K\|_\infty^{-1}\cdot n\cdot\hat f_{n,K}$ est dans $\mathscr{X}_u$ avec

$$u=\|K\|_\infty^{-1}\cdot n\cdot(f*|K|)(x),\ \|2K\|_\infty^{-1}\cdot n\cdot(\hat f_{n,K}-\bar f_K)(x)$$

est dans $\mathscr{Y}_u$, avec $u^2=\|2K\|_\infty^{-2}\cdot n\cdot(f*K^2)(x)$. On applique le corollaire précédent, avec les majorations suivantes: (voir (A.2) dans l'Appendice) $\|f*|K|\|_r\leqq\|K\|_r$ (pour $r=1,q$), $\|f*K^2\|_r\leqq\|K\|_{2r}^2$ (pour $r=1$, $p/2$) $\|f*K^2\|_r\leqq\|K^2\|_1\cdot\|f\|_r$ (pour $r=p/2$).

Remarque. Pour $q=1$, $p=2$, les formules se réduisent à

$$(4.14)\quad E\int|\hat f_{n,K}|\leqq\|K\|_1;\quad E\int|\hat f_{n,K}-\bar f_K|^2\leqq n^{-\frac{1}{2}}\|K\|_2^2,$$

et pour $q<1$, $p<2$, elles prennent la forme:

$$(4.15)\quad E\int|\hat f_{n,K}|^q\leqq\|f*|K|\|_q^q;\quad E\int|\hat f_{n,K}-\bar f_K|^p\leqq n^{-p/2}\cdot\|f*K^2\|_{p/2}^{p/2}$$

mais on ne peut plus appliquer l'inégalité (A.2).

Première évaluation de la vitesse. Choisissons un s-noyau K, et remplaçons dans les formules (4.6) et (4.13) K par $K_h\left(K_h(x)=h^{-1}\cdot K\left(\frac{x}{h}\right)\right)$. Comme ${}_s(K_h)=h^s\cdot({}_sK)_h$, d'après (A.4)

$$(4.16)\quad E_f\|\hat f-f\|_p^p\leqq\alpha\cdot h^{sp}+\beta\cdot(nh)^{1-p}+\gamma\cdot(nh)^{-p/2},$$

où

$$\alpha=2^{p-1}\cdot\|{}_sK\|_1^p\cdot\|f^{(s)}\|_p^p,\qquad \beta=2^{2p-3}\cdot\|K\|_\infty^{p-2}\cdot\|K\|_2^2\cdot M_{a,p},$$
$$\gamma=2^{2p-1}\cdot(1+a)\cdot C_p\cdot\|K\|_2^p\cdot\|f\|_{p/2}^{p/2},\qquad \hat f=\hat f_{n,K_h}.$$

En liant n et h par

$$(4.17)\quad (nh)^{p/2}\cdot h^{sp}=R\qquad \text{il vient.}$$

Proposition 4.1. *Si n et h sont liés par* (4.17), *l'estimateur à noyau $\hat f=\hat f_{n,K_h}$, satisfait*

$$(4.18)\quad \limsup_n n^{\frac{sp}{2s+1}}E_f\|\hat f-f\|_p^p\leqq C\{\|f^{(s)}\|_p^p R^{\frac{2s}{2s+1}}+\|f\|_{p/2}^{p/2}R^{\frac{-1}{2s+1}}\}$$

où C ne dépend que de s, p et du s-noyau K ($s\geqq1$, $p>2$)
(le terme intermédiaire disparait car $(1-p)<-\frac{1}{2}p$ pour $p>2$). Si on connaissait les paramètres

$$A(f)=\|f^{(s)}\|_p^p\quad et\quad B(f)=\|f\|_{p/2}^{p/2},$$

le meilleur choix serait

$$(4.19)\quad R(f)=2^p\cdot C_p\cdot\|K\|_2^p\cdot\|{}_sK\|_1^{-p}\cdot B(f)\cdot(A(f))^{-1},$$

et donnerait comme majorant au second membre dans (4.18)

$$(4.20)\quad C' = 2^p \cdot \{4 \cdot C_p^{2/p} \cdot \|K\|_2^2 \cdot \|_s K\|_1^{1/s}\}^{\frac{sp}{2s+1}} \cdot \rho_{s,p}(f).$$

Estimation adaptative de l'échelle (d'après Nadaraya [8]).

Pour obtenir (4.20), il nous faudra une estimation du paramètre d'échelle de f, ici donc de $A(f)$ et $B(f)$. Si on reporte dans (4.16) la valeur $\hat{h}$ obtenue dans (4.17) en substituant dans (4.19) $\hat{A}$ et $\hat{B}$ à $A(f)$ et $B(f)$, on constate que le second membre de (4.16) ne dépend de $\hat{A}$ et $\hat{B}$ que par

$$(4.21)\quad \hat{U}_1 = (\hat{B} \cdot \hat{A}^{-1})^{\frac{2s}{2s+1}}, \qquad \hat{U}_2 = (\hat{B} \cdot \hat{A}^{-1})^{\frac{2-2p}{p(2s+1)}}, \qquad \hat{U}_3 = (\hat{B} \cdot \hat{A}^{-1})^{\frac{-1}{2s+1}}.$$

Découpons l'échantillon en trois parts, $(X) = (X)_1 + (X)_2 + (X)_3$, de cardinaux respectifs n_1, n_2, n_3 avec $n_1 + n_2 + n_3 = n$. Si $\hat{A}$ ne dépend que de $(X)_2$, $\hat{B}$ que de $(X)_3$, $\hat{h}$ est indépendant de $(X)_1$; on pose donc $\hat{f} = \hat{\mu}_{n_1} * K_{\hat{h}}$, (4.16) devient

$$E_f \|\hat{f} - f\|_p^p \leqq E_f(\alpha \cdot \hat{h}^{sp} + \beta \cdot (n_1 \hat{h})^{1-p} + \gamma \cdot (n_1 \hat{h})^{-p/2}).$$

Si $n/n_1 \to 1$, il nous suffira pour avoir le résultat asymptotique que $E_f(\hat{U}_i)$ tende vers $U_i(f)$, soit, comme $\hat{A}$ et $\hat{B}$ sont indépendants, que, n_2 et n_3 tendant vers l'infini, on ait démontré

Lemme 4.2. *Soit $p > 2$, $s \geqq 1$, $M > 0$. Il existe des estimateurs*

$$\hat{A}_n(X_1, X_2, \ldots, X_n) \quad et \quad \hat{B}_n(X_1, X_2, \ldots, X_n)$$

tels que, si $f^{(s)} \in \mathbb{L}^p$,

$$\lim_n E_f(\hat{A}_n^a) = A(f)^a, \qquad \lim_n E_f(\hat{B}_n^a) = B(f)^a$$

pour tout a tel que $-M \leqq a \leqq M$.

(*Il suffira d'appliquer ce lemme avec M majorant les valeurs absolues des trois exposants intervenant dans* (4.21).)

Pour démontrer ce lemme, on va d'abord établir la version suivante du lemme 2 de [8].

Lemme 4.3. *Soit Y_n une suite de fonctions positives mesurables sur $\mathbb{R}^n$, F une famille de Probabilités sur $\mathbb{R}$, d un réel > 0. Si*

(i) $\forall n, \forall f \in F, E_f(|Y_n(X_1, X_2, \ldots, X_n) - y_n(f)|) \leqq n^{-d} \cdot C(f)$ *avec*

(ii) $\forall f \in F$, $\lim_n y_n(f) = y(f)$, *où* $\forall f \in F$, $0 < y(f) < \infty$ *et* $C(f) < \infty$, *alors, pour tout $M > 0$, il existe Z_n, suite de fonctions mesurables positives sur $\mathbb{R}^n$ telles que, si $-M \leqq a \leqq M$, $f \in F$, $\lim_n E_f(Z_n^a) = y(f)^a$.*

Démonstration. Voir Appendice.

Démonstration du Lemme 4.2. Il nous suffit donc de trouver des estimateurs $\hat{A}_n, \hat{B}_n$ préliminaires vérifiant (i) et (ii), avec $F = \{f \mid f^{(s)} \in \mathbb{L}^p\}$, $y(f)$ étant soit $A(f)$ soit $B(f)$, puis on les modifiera suivant L.4.3.

On pose

$$\hat{g}=\hat{\mu}_n*(K_h)^{(s)}, \quad \bar{g}=E_f(\hat{g})=f*(K_h)^{(s)}=f^{(s)}*K_h, \quad \hat{f}=\hat{\mu}_n*K_h,$$
$$\bar{f}=f*K_h, \quad \hat{A}=\int|\hat{g}|^p, \ \bar{A}=\int|\bar{g}|^p, \ \hat{B}=\int|\hat{f}|^{p/2}, \ \bar{B}=\int|\bar{f}|^{p/2},$$

où cette fois $K\in\mathscr{C}_c^{(\infty)}$, $\int K=1$; n et h seront liés à la fin du calcul.

Tout d'abord, (ii) est vérifié avec $y_n(f)=\bar{A}$ ou $\bar{B}$, dès que h tend vers 0 quand n tend vers l'infini: c'est la continuité de la translation dans $\mathbb{L}^q(q\geqq 1)$, et $y(f)$ est fini, conformément à la remarque qui suit (4.6), puisque $p\geqq 2$. Appliquant le Corollaire 2, on constate ensuite que les quantités $E_f\int|\hat{g}|^b$, $E_f\int|\hat{f}|^b$, $n^{p/2}\cdot E_f\int|\hat{g}-\bar{g}|^p$, $n^{p/2}\cdot E_f\int|\hat{f}-\bar{f}|^p$ peuvent être simultanément majorées (pour b compris entre 1 et p) par $C'\cdot(1\vee h^{-D'})$, où les constantes positives finies C' et D' ne dépendent que de K et p: en effet, dans les majorants de (4.12) et (4.13) deuxième forme, h n'intervient que par des puissances négatives, les normes de K sont uniformément bornées, et l'exposant de n dans (4.13) est moindre que $-p/2$.

Utilisons l'inégalité, valable pour tout q, $0\leqq r\leqq 1$, $0\leqq x, y$:

$$(4.22) \quad |y^q-x^q|\leqq q\cdot|y-x|^r\cdot(y^{q-r}\vee x^{q-r}).$$

On y porte d'abord $q=p$, $r=1$, $y=|\hat{g}|$, $x=|\bar{g}|$, puis on applique Hölder au second membre d'exposants conjugués p et p'. Comme

$$E_f(|\hat{g}|^p)\geqq|\bar{g}|^p,$$
$$E_f|\hat{A}-\bar{A}|\leqq p\cdot(E_f\int|\hat{g}-\bar{g}|^p)^{1/p}\cdot(2\cdot E_f\int|\hat{g}|^p)^{(p-1)/p}\leqq n^{-d}\cdot C,$$

compte tenu des majorations simultanées, pour un choix $h=n^{-t}$, *t positif* (pour assurer (ii)) convenable. On procède de même pour $\hat{B}$, avec cette fois $q=p/2$, Hölder appliqué avec les exposant conjugués p/r et $p/(p-r)$ à condition qu'on puisse trouver un $r>0$ tel que $(p/2-r)\cdot p\cdot(p-r)^{-1}=s\geqq 1$ (pour pouvoir comparer $E_f|\hat{g}|^s$ à $|\bar{g}|^s$, et pour assurer qu'il reste une puissance strictement négative de n). Les deux conditions sont réalisables si $p>2$.

Théorème 4.1. *p étant un réel ($p\geqq 1$), s un entier ($s\geqq 1$), on peut leur associer une suite d'estimateurs $\hat{f}_n$ tels que*

$$(4.23) \quad \limsup_n n^{\frac{sp}{2s+1}}\cdot E_f\|\hat{f}_n-f\|_p^p\leqq D_{s,p}\cdot\rho_{s,p}(f)$$

où

$$\rho_{s,p}(f)=\|f^{(s)}\|_p^{\frac{p}{2s+1}}\cdot\|f\|_{p/2}^{\frac{sp}{2s+1}}.$$

On peut prendre pour $D_{s,p}$:

$$2^p\cdot\{2^2\cdot C_p^{2/p}\cdot\|K\|_2^2\cdot\|_sK\|_1^{1/s}\}^{\frac{sp}{2s+1}} \qquad \textit{pour } p>2,$$

$$2^p\cdot\{\|K\|_2^2\cdot\|_sK\|_1^{1/s}\}^{\frac{sp}{2s+1}} \qquad \textit{pour } 1\leqq p\leqq 2,$$

où K est un s-noyau auxiliaire.

Si $1<p<2$, il faut ajouter f et K à support compact pour que (4.23) *soit valable.*

Démonstration. Les pages précédentes démontrent le résultat pour $p>2$. Pour $p=2$, la démonstration est correcte en ce qui concerne l'estimation de $A(f)$, la restriction $p>2$ au lieu de $p \geqq 2$, n'intervenant que pour $B(f)$, à la fin de la démonstration de L.4.1. Mais si $p=2$, nul besoin d'estimer $B(f)$ qu'on sait être 1. Vu (4.15), pour $1 \leqq p \leqq 2$, (4.16) devient

$$E_f \|\hat{f}-f\|_p^p \leqq 2^{p-1} \cdot \{\|_s K\|_1^p \cdot \|f^{(s)}\|_p^p \cdot h^{sp} + (n\,h)^{-p/2} \cdot \|f*(K^2)_h\|_{p/2}^{p/2}\}.$$

Si f est à support compact, $\lim_h \|f*(K^2)_h\|_{p/2}^{p/2} = \|K\|_2^p \cdot \|f\|_{p/2}^{p/2}$ et l'analogue de la Proposition (4.1) est donc valable. Pour l'estimation adaptative, on utilisera $\|f*L\|_q \leqq \|f\|_\infty \cdot \|L\|_\infty \cdot (S(f)+2S(L))^{1/q}$ et $E|\hat{f}|^q \leqq |\bar{f}|^q$, pour $q<1$, où $S(f)$ désigne la longueur d'un intervalle contenant le support de f, et cette fois la constante $C(f)$ du L.4.3. dépendra effectivement de f.

Calcul non asymptotique. C'est pour pouvoir obtenir des majorations à distance finie *(cas non adaptatif)* que le lemme 4.1 a été établi. Si $p>2$, on utilise dans la formule (4.16) la majoration

(4.9) $\quad \Phi_p(u) \leqq 2^p \cdot C_p \cdot \Psi_{p/2}(\tfrac{1}{2}u^2),\qquad$ et non sa conséquence (4.11).

En effet, si $q=p/2$ est entier, on majore $\Psi_q(v)$ par $\Psi_q(1)\cdot(v \vee v^q)$, car $\Psi_q(v)$ est alors un polynôme en v à coefficients positifs. Sinon, comme dans la démonstration du Corollaire 1, on a $\Psi_q(v) \leqq \Psi_k(v) \wedge (\Psi_k(v))^{q/k}$ dès que $q \leqq k$.

Si $p<2$, on utilise (4.7).

Comparaison des résultats des §3 et 4. Les Théorèmes 3.1 et 4.1 peuvent s'énoncer, pour $p>1$:

$$\liminf_n n^{\frac{sp}{2s+1}} \operatorname*{Inf}_{\hat{f}_n} \operatorname*{Sup}_{f\in F_r} E_f \|\hat{f}_n - f\|_p^p \geqq C_{s,p} \cdot r,$$

$$\limsup_n n^{\frac{sp}{2s+1}} \operatorname*{Sup}_{f\in F_r} E_f \|\hat{f}_n - f\|_p^p \leqq D_{s,p} \cdot r,$$

pour $\hat{f}_n$ bien choisi.

On peut comparer $C_{s,p}$ à $D_{s,p}$ par le

Théorème 4.2. *Il existe une constante universelle C telle que*

(4.24) $$\left(\frac{D_{s,p}}{C_{s,p}}\right)^{1/p} \leqq C \cdot p^{1/2} \cdot s^{1/2}.$$

Démonstration. Voir Appendice.

Remarque 1. La restriction de compacité du support dans le cas $1 \leqq p < 2$ n'enlève rien au résultat de comparaison, le minorant du Théorème 3.1 ayant été calculé sur des densités à support compact. On peut ajouter: d'une part, on aurait pu faire le même calcul sur le groupe du Tore, sans problèmes de compacité. D'autre part, Abou-Jaoudé [1] a rencontré le même problème dans l'estimation de la vitesse de convergence des estimateurs à noyaux en 1-norme, et a pris l'hypothèse que f est intégrable Riemann sur $\mathbb{R}$, (impliquée par exemple par f monotone à l'infini), pour assurer que pour tout noyau d'intégrale 1, $\int (f * K_h)^q$ tend vers $\int f^q$ pour $q<1$, ce qui entraîne la validité du Théorème 4.1. Enfin, nous pensons qu'en serrant les évaluations (4.7) on doit pouvoir montrer le résultat sous des hypothèses très générales.

Remarque 2. Les calculs présentés ici, ainsi que ceux du § suivant, pourraient faire croire que les jauges sont intrinsèques au problème. Il n'en est rien, on peut faire les mêmes calculs avec toute jauge du même type, ayant bien sûr les mêmes propriétés d'invariance et d'homogénéité que la fonctionnelle étudiée. Les jauges utilisées sont simplement celles qui, dans la classe $f(s) \in \mathbb{L}^p$, d'une part donnent l'ordre maximum en n, et d'autre part donnent des calculs à peu près présentables pour le problème de majoration.

5. Application à la distance de Hellinger

Dans ce paragraphe $(A, \mathfrak{A}, \mu)$ est $(\mathbb{R}, \mathscr{B}, dx)$, et la fonctionnelle étudiée est le carré de la distance de Hellinger

$$D(f,g)=d_H^2(f,g)=\tfrac{1}{2}\int (\sqrt{f}-\sqrt{g})^2\, dx.$$

Elle répond à la définition 2.1, avec $\phi(x)=\sqrt{x}$, mais elle nous cantonne aux estimateurs positifs; voulant nous limiter aux estimateurs de Parzen, nous nous limiterons aux densités 1 ou 2 fois dérivables, et le noyau de base sera le noyau uniforme $K(x)=1_{[-\frac{1}{2},\frac{1}{2}]}(x)$. Pour pouvoir atteindre l'ordre maximum en n, il faut supposer de plus que f est à support compact, et nous posons

(5.1) $\quad S(f)=\operatorname{Inf}\{(b-a) \mid \text{le support de } f \text{ est contenu dans } [a,b]\}.$

Si $\hat{f}_n=\hat{\mu} * K_h$ et $\bar{f}=f * K_h$, on a $S(\hat{f}_n) \leqq S(f)+h$, $S(\bar{f}) \leqq S(f)+h$. Enfin, $(n h)\cdot(\hat{f}_n-\bar{f})$ suit une loi binômiale de paramètres $n, h\cdot\bar{f}$, et, comme pour toute variable binômiale X, $E((\sqrt{X}-\sqrt{E(X)})^2) \leqq 1$, le terme d'aléa est majoré par $(2 n h)^{-1}\cdot(S(f)+h)$.

On majore le terme de biais en utilisant l'inégalité (4.22), $q=r=\frac{1}{2}$, ce qui, avec Hölder au second membre pour les exposants p, p', donne au total, pour $s=1$ ou 2, p réel supérieur ou égal à 1

$$ED(\hat{f}_n,f) \leqq \{\tfrac{1}{4}\|{}_sK\|_1\cdot\|f^{(s)}\|_p\cdot h^s\cdot(S(f)+h)^{\frac{p-1}{p}}+\frac{1}{n h}\cdot(S(f)+h)\}.$$

En admettant qu'on puisse faire une estimation adaptative de l'échelle comme dans le précedent § (l'estimation du support est facile), on obtient le résultat asymptotique suivant:

Proposition 5.1. *Soit $s=1$ ou 2, $p \geqq 1$. Il existe alors un estimateur $\hat{f}_n$ tel que*

(5.2) $\quad \limsup_n n^{\frac{s}{s+1}}\cdot E_f\, d_H^2(\hat{f}_n,f) \leqq \tfrac{1}{2}\rho_{H,s,p}(f)$

où

$$(5.3)\quad \rho_{H,s,p}(f)=\|f^{(s)}\|_p^{\frac{1}{s+1}}\cdot(S(f))^{\frac{p(s+1)-1}{p(s+1)}}$$

Passons à la minoration. Nous prenons comme famille Θ_1 la famille

$$\Theta_1=\{g|g(x)=(1-t)\cdot g_0(x-3/2)+t\,g(x),\text{ où } g\in\Theta\}.$$

Θ est la famille décrite au §3, avec $h=h_j$ en nombre J, $u_j=u$, les contraintes sont $0<6Jh<1$, $0<u<1$. On évalue la jauge r de Θ_1 à partir de la jauge r_0 de g_0 en utilisant: $S(g)=2S(g_0)$ et

$$\|g^{(s)}\|_p^p\leqq\|g_0^{(s)}\|_p^p\cdot\{(1-t)^p+t^p\cdot(1+u^p\cdot h^{-sp})\}.$$

Le calcul de l'analogue de la formule (3.4) donne ici

$$\phi^{-1}(\tfrac{1}{2}\phi(D_j))\geqq\tfrac{1}{8}t\cdot h\cdot u^2,$$

$I_j\leqq 4t\cdot h\cdot u^2\cdot(1-u)^{-1}$, et le minorant est alors

$$\tfrac{1}{8}(J_h)\cdot t\cdot u^2\cdot\exp(-4n\cdot t\cdot h\cdot u^2\cdot(1-u)^{-1}).$$

Le calcul de minoration asymptotique donne alors.

Proposition 5.2. *Soit* $s=1$ *ou* 2, $p\geqq 1$. *Si on pose*

$$r_0=\rho_{H,s,p}(g_0),\qquad F_r=\{f|\rho_{H,s,p}(f)\leqq r\},$$

on a

$$(5.4)\quad \liminf_n n^{\frac{s}{s+1}}\operatorname*{Inf}_{\hat f_n}\operatorname*{Sup}_{f\in F_r}E_f\,d_h^2(\hat f_n,f)>2^{-10}\left(\frac{r}{r_0}-2\right)$$

(pour les mêmes raisons que dans le cas de la 1-norme, la technique ne permet pas de descendre en dessous d'une jauge liée au choix de la densité auxiliaire, mais en tout état de cause minorée par un nombre strictement positif).

Appendice

1. Convolution

L'espace de base étant $(\mathbb{R},\mathscr{B},dx)$, on note $*$ la convolution

$$f*g(x)=\int f(x-y)\,g(y)\,dy,\qquad f*\mu(x)=\int f(x-y)\,\mu(dy).$$

$\mathbb{L}^p$ désigne l'espace des fonctions de puissance $p^{\text{ème}}$ sommable, $\|f\|_p$ sa «norme» (pour $0<p<\infty$), abus de notation si $0<p<1$, sans conséquence si on n'utilise pas l'inégalité triangulaire. $\mathbb{L}^0$ désigne l'ensemble des fonctions à support intégrable.

L'exposant conjugé de $p\geqq 1$ est noté p': $1/p+1/p'=1$.

On utilise de manière répétitive les inégalités:

(A.1) $\|f*g\|_r \leqq \|f\|_p \cdot \|g\|_q$ si $p,q,r \geqq 1$ et $\frac{1}{p}+\frac{1}{q}=\frac{1}{r}+1$

et ses cas particuliers

(A.2) $\|f*g\|_p \leqq \|f\|_1 \cdot \|g\|_p$ si $p \geqq 1$,

(A.3) $\|f*g\|_\infty \leqq \|f\|_p \cdot \|g\|_{p'}$.

La h-homothétique de f, définie par $f_h(x)=\frac{1}{h}\cdot f\left(\frac{x}{h}\right)$ pour $h>0$ vérifie $(f_h)^{(s)} = h^{-s}\cdot (f^{(s)})_h$, et donc

(A.4) $\|f_h\|_q = h^{1/q-1}\cdot \|f\|_q$; $\|(f_h)^{(s)}\|_q = h^{1/q-s-1}\cdot \|f^{(s)}\|_q$ $(0<q\leqq\infty)$.

La dérivation est la dérivation au sens de Lebesgue: utiliser le symbole f' signifie que f est absolument continue, que f' est localement intégrable, et que $f(y) - f(x) = \int_x^y f'(t)\,dt$.

2. Fonctions auxiliaires

Si v est une fonction positive, d'intégrale 1 de classe $\mathscr{C}^{(s-1)}$, de support contenu dans un intervalle J, de longueur $|J|$ moindre que $\frac{1}{2}(b-a)$, si I désigne l'intervalle $[a,b]$, $(1_I * v)'(x) = v(x-a) - v(x-b)$, ces deux fonctions étant à supports disjoints. De là, la formule (3.2).

Un cas particulier. Soit Φ_j des fonctions positives, d'intégrale 1, paires et de support contenu dans $[-\frac{1}{8}, \frac{1}{8}]$, enfin $\Phi_j^{(j)} \in \mathbb{L}^2$. $g_0 = 1_I * \Phi_j * \Phi_k$ est alors une densité auxiliaire (3.2), pour $s=j+k+1$. Pour calculer $\|g_0^{(s)}\|_p = 2^{1/p}\|\Phi_j^{(j)} * \Phi_k^{(k)}\|_p$, on utilise (A.1) avec $r=p$, $p=q=t$, et $t=\frac{2p}{p+1}$ $(1<t\leqq 2$ si $1\leqq p)$; $\Phi_j^{(j)}$ étant à support compact, on peut comparer sa 2-norme à sa t-norme par

$$\|\Phi_j^{(j)}\|_t \leqq 4^{\frac{1}{2}-1/t}\cdot \|\Phi_j^{(j)}\|_2.$$

Dans ce cas, comme $\|g_0\|_{p/2} \leqq (\frac{3}{2})^{2/p}$, on a donc $(1\leqq p,\ s=j+k+1)$

(A.5) $\rho_{s,p}(g_0) \leqq \frac{1}{2}\cdot 3^{\frac{2s}{2s+1}} \cdot \{\|\Phi_j^{(j)}\|_2 \cdot \|\Phi_k^{(k)}\|_2\}^{\frac{p}{2s+1}}$.

3. Noyaux de Parzen

a) Soit v continue à support compact, d'intégrale non nulle. Soient des x_i distincts, en nombre s. Il existe alors un choix des a_i tels que $K(x) = \sum_{0\leqq i<s} a_i \cdot v(x-x_i)$ soit un noyau de *Parzen d'ordre s* (4.3) à *support compact*. Il suffit de remarquer que le déterminant $|\alpha_{i,j}|$ diffère de 0, si $\alpha_{i,j} = \int x^j \cdot v(x-x_i)$, $0\leqq i, j<s$.

b) Notons $\tilde{f}$ la transformée de Fourier de f: $\tilde{f}(x) = \int \exp(i\,t\,x)\cdot f(t)\,dt$. g_0 étand le cas particulier décrit plus haut, $K=\frac{1}{2\pi}\cdot \tilde{g}_0$ est un *noyau de Parzen d'ordre s*:

comme $\tilde{K}=g_0$, formellement $\int x^j \cdot K(x)\,dx=(i)^j \cdot g_0^{(j)}(0)$, or g_0 est constante et vaut 1 sur $[-\frac{1}{4},\frac{1}{4}]$. Il suffit de montrer la finitude de $\|{}_sK\|_1$. Or

$$\|{}_sK\|_1=\frac{1}{2\pi \cdot s!}\|x^{j+k+1}\cdot \tilde{1}_I \cdot \tilde{\Phi}_j \cdot \tilde{\Phi}_k\|_1.$$

Comme

$$|x\cdot \tilde{1}_I|=\left|2\cdot \sin\left(\frac{x}{2}\right)\right|\leqq 2,$$

comme

$$\|x^j \cdot \tilde{\Phi}\|_2^2=(2\pi)\cdot \|\Phi_j^{(j)}\|_2^2,$$

il vient:

$$\text{(A.6)}\quad \|{}_sK\|_1 \leqq \frac{2}{s!}\cdot \|\Phi_j^{(j)}\|_2 \cdot \|\Phi_k^{(k)}\|_2.$$

Enfin:

$$\text{(A.7)}\quad \|K\|_2^2=(2\pi)^{-1}\cdot \|g_0\|_2^2 \leqq (2\pi)^{-1}.$$

4. Démonstration du théorème 4.2

Compte-tenu de (3.9), (4.23) et des évaluations précédentes (A.5), (A.6) et (A.7), le membre de gauche de (4.24) est majoré par:

$$(2\cdot e)^{4/p}\cdot (6\cdot \sqrt{2\pi})^{\frac{2s}{2s+1}}\cdot (C_p)^{\frac{2s}{p\cdot(2s+1)}}\cdot (s!)^{\frac{-1}{2s+1}}\cdot (\|\Phi_j^{(j)}\|_2 \cdot \|\Phi_k^{(k)}\|_2)^{\frac{2}{2s+1}}.$$

Remarquons d'abord que $C_p^{1/p}=0(p^{\frac{1}{2}})$. Il faut maintenant choisir les Φ_j: on va prendre les polynômes de Legendre pour $\Phi_j^{(j)}$. Précisément: on prend

$$\Phi_j=(Q_j)_{1/8},\qquad \text{où } Q_j(x)=\frac{(2j+1)!}{2^{2j+1}\cdot (j!)^2}\cdot (1-x^2)^j \cdot 1_{|x|\leqq 1}.$$

Alors

$$\|\Phi_j^{(j)}\|_2=\frac{(2j+1)!}{j!}\cdot 2^{2j}\cdot (2j+1)^{-\frac{1}{2}}.$$

Le produit des deux derniers termes est alors $O(s^{\frac{1}{2}})$, puisque $j+k+1=s$, en prenant j et k équivalents à $\frac{1}{2}s$, ce qui démontre le résultat.

5. Démonstration du lemme 4.3

Fixons f; soit $t>0$; on pose $Z_n=Y_n$ si $n^{-t}\leqq Y_n \leqq n^t$, $Z_n=1$ sinon. Puisque $y_n(f)$ tend vers $y(f)>0$, pour n grand, $2\cdot n^{-t}\leqq y_n(f)\leqq n^t-n^{-t}$, donc $\{Y_n \neq Z_n\}$ implique $\{|Y_n-y_n(f)|\geqq n^{-t}\}$, donc $P(Y_n \neq Z_n)\leqq C(f)\cdot n^{-d+t}$, (i). Il s'ensuit que

$$E_f|Z_n-y_n(f)|\leqq C(f)\cdot n^{-d}\cdot (1+n^t\cdot (1\vee y_n(f))).$$

L'inégalité (4.23), où l'on porte $q=a$, $r=1$, $y=Z_n$, $x=y_n(f)$ donne, puisqu'ici $y^{a-1} \vee x^{a-1}$ est majoré par $n^{t \cdot |a-1|}$,

$$E_f|Z_n^a - y_n(f)^a| \leqq M \cdot C(f) \cdot n^{-d+t\cdot(M+1)} \cdot (1+n^t \cdot (1 \vee y_n(f))),$$

et donc le résultat, puisque $y_n(f)^a \to y(f)^a$, si t a été choisi >0, mais assez petit pour que $t \cdot (M+2) < d$.

6. Démonstration du lemme 4.1

Démonstration. De (4.7): si $0 \leqq r \leqq s$, $E(|X|^r) \leqq (E|X|^s))^{r/s}$.

De (4.8): on peut supposer les z_i positives. On se ramène au cas où les z_i sont des indicatrices 1_{B_i} en remarquant que toute v.a. $z(0 \leqq z \leqq 1)$ peut s'interpréter comme espérance conditionnelle d'une indicatrice (construction: prendre le carré $(0 \leqq s,\ t \leqq 1\}$ muni de la mesure de Lebesgue. Soit $F(s)=P(z \leqq s)$. Alors, si $\mathscr{B}$ est la σ-algèbre des événements ne dépendant que de y, si $B = \{s, t | 0 < s \leqq 1;\ F(s) \leqq t \leqq 1\}$, $E^{\mathscr{B}}(1_B)$ a même loi que z), puis on applique Jensen car $q \geqq 1$. Calquons l'idée de Young [18]. L'inégalité:

$$(1-2a)+2a \cdot (1+x)^q \leqq (1-a)^2 + 2a \cdot (1+x)^q + a^2 \cdot (1+2x)^q$$
$$(\text{pour } \ 0 \leqq 2a < 1 \leqq q;\ 0 \leqq x)$$

qu'on montre par dérivation, montre que si B', B'' sont indépendants, avec $P(B') = P(B'') = a$, si $P(B)=2a$, $E((t+1_B)^q) \leqq E((t+1_{B'}+1_{B''})^q)$ $(t>0)$ et donc qu'on obtient le sup sur $\mathscr{X}_u$ en divisant ad infinitum les B_i suivant la règle plus haut. Le cas extrémal est donc celui de la v.a. de Poisson.

De (4.9): posons $Y'' = \frac{1}{2}(Y-Y')$, où Y' est une copie indépendante de Y. Les Y_i'' sont bornées par 1, et peuvent se factoriser comme $\varepsilon_i \cdot A_i$, où (ε_i, A_i) sont indépendantes, $P(\varepsilon_i=1)=P(\varepsilon_i=-1)=\frac{1}{2}$. Par intégrations successives, $E(|Y''|^p) \leqq C_p \cdot E((\sum A_i^2)^{p/2})$, où C_p est la constante de Khintchine $C_p = \operatorname{Sup}\{E(|\sum a_i \varepsilon_i|^p) | \sum a_i^2 = 1\}$. On termine en remarquant que $E(|Y|^p) \leqq 2^p \cdot E(|Y''|^p)$ (Jensen), en utilisant (4.8) et $\sum E(A_i^2) \leqq \frac{1}{2} \cdot u^2$; enfin, Young, pour $p>3$, puis Haagerup ([18] et [6]) ont calculé la constante de Khintchine.

Références

1. Abou-Jaoudé, S.: Thèse, Université de Paris VI (1977)
2. Bickel, J.P., Rosenblatt, H.: On some global measures of the deviation of density functions estimates. Ann. Statist. **1**, 1071-1095 (1973)
3. Carrol, R.J.: On sequential density estimation. Z. Wahrscheinlichkeitstheorie und verw. Gebiete **36**, 137-151 (1976)
4. Farrell, R.H.: On the lack of a uniformly consistent sequence of estimators of a density function in certain cases. Ann. Math. Statist. **38**, 471-474 (1967)
5. Farrell, R.H.: On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. Ann. Math. Statist. **43**, 170-180 (1972)
6. Haagerup, U.: The best constant in the Khintchine inequality. Research announcement Odense University Denmark, Preprint n° 3 (1977)

7. Meyer, T.G.: Bounds for estimation of density functions and their derivatives. Ann. Statist. **5**, 136-142 (1977)
8. Nadaraya, E.A.: On the integral mean square error of some non parametric estimates for the density function. Theor. Probability Appl. (SIAM) **19**, 131-141 (1974)
9. Parzen, E.: On estimation of a probability density function and mode. Ann. Math. Statist. **33**, 1065-1076 (1962)
10. Rosenblatt, M.: Curve estimates. Ann. Math. Statist. **42**, 1815-1842 (1971)
11. Rosenblatt, M.: A quadratic measure of deviation of two dimensional density and a test of independance. Ann. Statist. **3**, 1-14 (1975)
12. Schwartz, S.C.: Estimation of a probability density by an orthogonal serie. Ann. Math. Statist. **38**, 1261-1265 (1967)
13. Silvermann, B.W.: Weak and strong uniform consistency of the kernel estimate of a density and its derivative. Ann. Statist. **6**, 177-184 (1978)
14. Wahba, G.: Optimal convergence properties of variable Knot, kernel and orthogonal series methods for density estimation. Ann. Statist. **3**, 15-29 (1975)
15. Walter, G.G.: Properties of Hermite series estimation of Probability density. Ann. Statist. **6**, 1258-1264 (1977)
16. Watson, G.S.: Density estimation by orthogonal series. Ann. Math. Statist. **40**, 1496-1498 (1969)
17. Weiss, L., Wolfowitz, J.: Estimation of a density at a point. Z. Wahrscheinlichkeitstheorie und verw. Gebiete **7**, 327-335 (1967)
18. Young, R.M.G.: On the best constants in the Khintchine inequality. J. London Math. Soc. **14**, 496-504 (1976)
19. Boyd, D.W., Steele, J.M.: Lower Bounds for non parametric density estimation rates. Ann. Statist. **6**, 932-935 (1978)

Reçu le 7 juillet 1977; en forme révisée le 3 juillet 1978