Estimation non paramétrique de la densité par histogrammes généralisés (II)

Paul Deheuvels

Pub. Inst. Stat. Univ. Paris
1977, XXII - 1 - 23

ESTIMATION NON PARAMETRIQUE DE LA DENSITE PAR HISTOGRAMMES GENERALISES (II)

Paul DEHEUVELS, Maître de Conférences à l'Université Paris VI

Résumé: On détermine les estimations de la densité par la méthode du noyau, optimales pour le critère du M.I.S.E., pour des densités dans $\mathbb{R}^p$, $p \geq 2$, et pour de petits échantillons. Le cas de la loi multinormale fait l'objet d'une étude particulière.

Cette note fait suite à [4], où a été étudié en détail le cas de densités dans $\mathbb{R}$.

I) Introduction:

L'estimation de la densité dans $\mathbb{R}^p$, par des méthodes non paramétriques, est en général beaucoup plus délicate pour $p \geq 2$, que pour $p=1$, bon nombre de méthodes permettant d'obtenir des résultats pour les densités réelles se généralisant mal, ou pas du tout, pour des densités dans des espaces de dimension supérieure.

Ce type d'estimation est cependant très précieux pour traiter des problèmes d'estimation du support, ou de reconnaissance de formes. Il peut être aisément associé à des algorithmes de classification permettant de grouper par affinité des nuages de points dans l'espace.

Or, comme dans le cas réel, la plupart des résultats disponibles (citons notamment [1],[3],[5],[8], et renvoyons à [4] pour une bibliographie du cas réel) ne donnent que des propriétés asymptotiques, qui rendent hasardeux l'usage des estimations pour des échantillons de taille petite et moyenne.

Le but de ce qui suit est d'exposer des résultats permettant d'effectuer valablement de telles estimations pour des densités dans $\mathbb{R}^p$, $p \geq 2$. Le cas $p=1$ a été étudié plus en détail dans [4].

On supposera par la suite, que $\{X_N, N\geq 1\}$ est une suite de v.a. indépendantes, de même loi dans $\mathbb{R}^p$. On posera:

$$X_N = \begin{bmatrix} X_{1N} \\ \vdots \\ X_{pN} \end{bmatrix}.$$

On supposera que X possède une densité continue f, relativement à la mesure de Lebesgue dans $\mathbb{R}^p$.

On considèrera les estimations de f, de la forme suivante:

$$(1,1) \qquad f_N(x) = \frac{|\operatorname{Det}(A_N)|}{N} \sum_{i=1}^{N} K[A_N(X_i - x)] \ ,$$

où K est une fonction bornée de $\mathbb{R}^p$ dans $\mathbb{R}$, telle que $\int_{\mathbb{R}^p} K(v)dv = 1$, et A_N est un automorphisme linéaire de $\mathbb{R}^p$.

Deux cas particuliers de (1,1) sont fréquemment utilisés:

$$(1,2) \qquad f_N(x) = \frac{1}{N}\sum_{i=1}^{N} \prod_{j=1}^{p} \frac{1}{\delta_{j,N}} k\left[\frac{X_{j,i} - x_j}{\delta_{j,N}}\right] ,$$

$$(1,3) \qquad f_N(x) = \frac{1}{N\delta_N^p} \sum_{i=1}^{N} K\left[\frac{X_i - x}{\delta_N}\right] .$$

Historiquement, Rosenblatt (1956),[10], et Parzen (1962),[11], ont introduit pour p=1 l'estimation (1,3), Nadayara (1964),[9] a introduit (1,3) pour p=2, Maniya (1961),[6] a utilisé des estimations analogues dans $\mathbb{R}^2$, pour des noyaux k uniformes Epanechnikov (1969),[5] a introduit et étudié les estimations (1,2).

Citons également, pour l'étude de la convergence des estimations, Cacoullos (1966), [1], Van Ryzin (1970) , Deheuvels (1974),[2],[3].

Nous nous intéresserons ici particulièrement à la mesure d'efficacité des estimations fournie par l'erreur quadratique intégrée, ou critère du M.I.S.E. (minimum integrated square error):

$$(1,4) \qquad \begin{aligned} M^2 &= \int_{\mathbb{R}^p} E((f_N(x) - f(x))^2)dx = B_1^2 + B_2^2 \quad \text{, avec:} \\ B_1^2 &= \int_{\mathbb{R}^p} (E(f_N(x)) - f(x))^2 dx \quad , \quad B_2^2 = \int_{\mathbb{R}^p} V(f_n(x))dx . \end{aligned}$$

II) Moments des estimations:

Les développements suivants sont une généralisation à $\mathbb{R}^p$ des résultats obtenus dans [4] pour $\mathbb{R}$. On notera $||.||$, une norme quelconque dans $\mathbb{R}^p$, et sa norme associée dans $\mathcal{L}(\mathbb{R}^p,\mathbb{R}^p)$.

On pose $A^{-1} = B = \left[b_{ij}, 1\leq i,j\leq p\right]$, et, sous réserve que ces quantités soient définies:

$$[[B, y_1^{r_1}..y_p^{r_p}K^s]] = \int_{\mathbb{R}^p} \prod_{i=1}^{p} \left[\sum_{j=1}^{p} b_{ij}y_j\right]^r K^s(y_1,\ldots,y_p)dy_1..dy_p ,$$

$$[[y_1^{r_1}..y_p^{r_p}K^s]] = [[Id, y_1^{r_1}..y_p^{r_p}K^s]].$$

Par changement de variable, on obtient l'identité:

$$E(K^s(A(X - x))) = \frac{1}{|\operatorname{Det}(A)|} \int_{\mathbb{R}^p} K^s(v)f(x + A^{-1}v)dv ;$$

Supposons maintenant que f soit n fois continûment différentiable en x; on obtient formellement:

$$|\mathrm{Det}(A)|\, E\left[K^r(A(X-x))\right] = f(x) + \sum_{i=1}^{n} S_i(x,A,K^r) + R_n(x,A,K^r) \text{ , avec:}$$

$$(2,1) \qquad S_i(x,A,K^r) = \sum_{\substack{r_1+\dots+r_p = i \\ r_j\geq 0,\, 1\leq j\leq p}} \frac{1}{r_1!\dots r_p!} \frac{\partial^i f(x)}{\partial x_1^{r_1}\dots\partial x_p^{r_p}} [[A^{-1}, y_1^{r_1}\dots y_p^{r_p} K^r]]$$

$$R_n(x,A,K^r) = \int_{\mathbb{R}^p} \epsilon_n(x,A^{-1}v)\, ||A^{-1}v||^n K^r(v)dv \text{ , où } \epsilon_n(x,v)\, ||v||^n \text{ est le}$$

reste du développement de Taylor à l'ordre n de f(x+v)-f(v).

Il est clair, par (2,1), que, sous réserve que les moments de K^r soient définis à un ordre suffisant:

$$[[B, y_1^{r_1}\dots y_p^{r_p} K^r]] = O\left(||B||^{r_1+\dots+r_p}\right) \text{, et } S_i(x,A,K^r) = O\left(||A||^{-i}\right).$$

Pour que (2,1) soit utilisable, il faut que $R_n(x,A,K^r)$ soit $o\left(||A||^{-n}\right)$; on obtient ce résultat aux conditions suivantes:

<u>Proposition</u> (2,1): Si f est n fois différentiable en x, et si l'une des conditions suivantes est vérifiée:

a) K est nul en dehors d'un compact;

b) $[[y_1^{r_1}\dots y_p^{r_p} K^r]]$ est défini $\forall\, r_1,\dots,r_p \geq 0$, et f est définie sur $\mathbb{R}^p$, et y admet une majoration de la forme: $|f(x)| \leq C(1 + ||x||^R)$, $R \geq 0$;

c) f vérifie la condition b, sauf éventuellement au voisinage d'un nombre fini de points, où elle admet des discontinuités infinies, et K vérifie l'hypothèse:

$$(2,2) \qquad \lim_{x\to\infty} ||x||^R K^r(x) = 0 \text{ , } \forall\, R \geq 0;$$

Alors (2,1) est vérifié, avec $R_n(x,A,K) = o\left(||A||^{-n}\right)$.

Preuve: Même démonstration que dans $\mathbb{R}$, voir [4].

Dans la suite, on supposera toujours que K est borné, et vérifie (2,2). On supposera que f vérifie les hypothèses de la proposition (2,1),c.

<u>Corollaire</u> (2,2): Avec les hypothèses et notations de la proposition (2,1),

$$E(f_N(x)) = f(x) + \sum_{i=1}^{n} S_i(x,A_N,K) + R_n(x,A_N,K) \text{ , avec } S_i(x,A_N,K) = O\left(||A_N||^{-i}\right),$$

$$R_n(x,A_N,K) = o\left(||A_N||^{-n}\right);$$

$$(2,3) \qquad \left\{E(f_N(x)) - f(x)\right\}^2 = \sum_{i=2}^{n+1}\left\{\sum_{j=1}^{i-1} S_j(x,A_N,K)S_{i-j}(x,A_N,K)\right\} + T_{n+1}(x,A_N,K) \text{ , avec:}$$

$$T_{n+1}(x,A_N,K) = o\left(||A_N||^{-n-1}\right);$$

4.
$$V(f_N(x)) = \frac{|Det(A_N)|}{N}\sum_{i=0}^{n} S_i(x,A_N,K^2) - \frac{1}{N}\sum_{i=0}^{n-p}\Big\{\sum_{j=0}^{i} S_j(x,A_N,K)S_{i-j}(x,A_N,K)\Big\}$$
$$+ \frac{1}{N}U_n(x,A_N,K) \ , \text{ avec } U_n(x,A_N,K) = o\left(||A_N||^{-n+p}\right).$$

Remarques: a) $Det(A_N) = O\left(||A_N||^p\right)$, ce qui explique la présence de la sommation de 0 à n-p dans (2,3), $V(f_N(x))$; si $n < p$, les résultats restent valables en supprimant cette sommation devenue impossible.

b) Ces résultats restent valables pour n=0, en supprimant les sommations impossibles. On obtient alors:

$$E(f_N(x)) = f(x) + R_0(x,A_N,K) \to 0 \ , \text{ si } ||A_N|| \to \infty \ ,$$
$$V(f_N(x)) = \frac{|Det(A_N)|}{N} f(x)\,[[K^2]] + \frac{U_0(x,A_N,K)}{N} \ , \text{ avec: }$$
$$U_0(x,A_N,K) \to 0 \ , \text{ si } ||A_N|| \to \infty .$$

La condition nécessaire et suffisante pour que $f_N(x) \to f(x)$ p.s., est que:

(2,4) $\qquad ||A_N|| \to \infty$, et $|Det(A_N)| = \dfrac{N\,o(1)}{LogLog(N)}$;(Deheuvels (1974),[2],[3])

La condition nécessaire et suffisante pour que $\sup_{x\in\mathbb{R}^p} |f_N(x) - f(x)| \to 0$ p.s., est que:

(2,5) $\qquad ||A_N|| \to \infty$, et $|Det(A_N)| = \dfrac{N\,o(1)}{Log(N)}$;(mêmes références)

Nous utiliserons ici des estimations asymptotiquement optimales pour le critère du M.I.S.E., et telles que:

$$||A_N|| \sim (Cte)\ N^{1/(4+p)} \ , \text{ et } \ |Det(A_N)| \sim (Cte)\ N^{p/(4+p)} \ ;$$

Compte tenu de (2,4) et (2,5), les estimations seront toujours uniformément convergentes p.s., avec des hypothèses minimales sur f et K (f uniformément continue), qui seront toujours vérifiées en pratique.

III) Moyenne quadratique intégrée (développements asymptotiques):

On estime $M^2 = B_1^2 + B_2^2$, défini par (1,4), en utilisant (2,1) et (2,3).

Proposition (3,1): Si f est n fois continûment différentiable dans un ouvert contenant I, domaîne de $\mathbb{R}^p$, et dans l'un ou l'autre des cas suivants:

a) I est un compact, et K est un noyau nul en dehors d'un compact;

b) Les dérivées partielles de f à l'ordre p, $0 \le p \le n+1$, continues, appartiennent à $L^1(V(I))\cap L^2(V(I))$, V(I) étant un voisinage ouvert de I, et, ou bien K est nul en dehors d'un compact, ou bien $I = \mathbb{R}^p$;

On a alors:

$$B_1^2 = \int_I (E(f_N(x))-f(x))^2 dx = \sum_{i=2}^{n+1} V_i(A_N,K) + \zeta_{n+1}(A_N,K) \text{ , avec:}$$

$$V_i(A_N,K) = \left(\| A_N \|^{-i}\right), \quad \zeta_{n+1}(A_N,K) = \left(\| A_N \|^{-n-1}\right) ;$$

$$B_2^2 = \int_I V(f_N(x))dx = \frac{|\mathrm{Det}(A_N)|}{N} \sum_{i=0}^{n} W_i(A_N,K) - \frac{1}{N} \sum_{i=0}^{n-p} V_i^*(A_N,K) + \frac{1}{N} \xi_n(A_N,K) ,$$

avec: $V_i^*(A_N,K)$ et $W_i(A_N,K) = \left(\| A_N \|^{-i}\right)$, $\xi_n(A_N,K) = \left(\| A_N \|^{p-n}\right)$;

et:

(3,1) $$V_i(A,K) = \sum_{j=1}^{i-1} \left\{ \sum_{\substack{r_1+\dots+r_p=j \\ s_1+\dots+s_p=i-j \\ r_j,s_j \geq 0, 1\leq j\leq p}} \left[\frac{[[A^{-1},y_1^{r_1}\dots y_p^{r_p}K]][[A^{-1},y_1^{s_1}\dots y_p^{s_p}K]]}{r_1!\dots r_p!s_1!\dots s_p!} \int_I \left\{\frac{\partial^j f(x)}{\partial x_1^{r_1}\dots\partial x_p^{r_p}}\right\}\left\{\frac{\partial^{i-j}f(x)}{\partial x_1^{s_1}\dots\partial x_p^{s_p}}\right\} dx \right] \right\} = \sum_{j=1}^{i-1} \Omega_{ji} ;$$

$$V_i^*(A,K) = \sum_{j=0}^{i} \Omega_{ji} ;$$

$$W_i(A,K) = \sum_{\substack{r_1+\dots+r_p=i \\ r_j\geq 0, 1\leq j\leq p}} \left[\frac{[[A^{-1},y_1^{r_1}\dots y_p^{r_p}K^2]]}{r_1!\dots r_p!} \int_I \frac{\partial^i f(x)}{\partial x_1^{r_1}\dots\partial x_p^{r_p}} dx \right]$$

Preuve: Même démonstration que dans $\mathbb{R}$, voir [4].

<u>Remarques</u>: Ces expressions se simplifient considérablement, si on fait des hypothèses supplémentaires convenables sur f, ou K:

1°) <u>Hypothèse</u> (A): On suppose que f et ses dérivées partielles d'ordre n s'annulent au bord de I, supposé être un ensemble produit; en se ramenant au cas où $I = \mathbb{R}^p$, on obtient:

$$\int_{\mathbb{R}^p} \frac{\partial^{r_1+\dots+r_p} f(x)}{\partial x_1^{r_1}\dots\partial x_p^{r_p}} dx = 0 \text{ , si } r_1+\dots+r_p \geq 1.$$ De même, en notant:

$$J(r_1,\dots,r_p,s_1,\dots,s_p) = \int_{\mathbb{R}^p} \left\{\frac{\partial^{r_1+\dots+r_p} f(x)}{\partial x_1^{r_1}\dots\partial x_p^{r_p}}\right\}\left\{\frac{\partial^{s_1+\dots+s_p} f(x)}{\partial x_1^{s_1}\dots\partial x_p^{s_p}}\right\} dx ,$$

on a les relations:

$$J(r_1,\dots,r_i,\dots,r_p,s_1,\dots,s_i,\dots,s_p) = -J(r_1,\dots,r_i-1,\dots,r_p,s_1,\dots,s_i+1,\dots,s_p) ,$$

$$J(r_1,\dots,r_p,s_1,\dots,s_p) = J(s_1,\dots,s_p,r_1,\dots,r_p) ,$$

6.

$$J(r_1,\dots,r_p,s_1,\dots,s_p) = 0, \text{ si } r_1+\dots+r_p+s_1+\dots+s_p \text{ est impair.}$$

On pourra toujours alors se ramener à:

(3,2) $$K(q_1,\dots,q_p) = (-1)^{i_1+\dots+i_p} J(i_1,\dots,i_p,q_1-i_1,\dots,q_p-i_p), \text{ avec } q_1+\dots+q_p \text{ pair.}$$

On en déduit que si (A) est vrai,

(3,3) $$W_i(A,K) = 0,\ \forall i\geq 1\ ,\ \text{et } V_{2i+1}(A,K) = V^*_{2i-1}(A,K) = 0,\ \forall i\geq 1.$$

2°) <u>Hypothèse</u> (B): On suppose que $f(x_1,\dots,x_p) = \prod_{i=1}^{p} f_i(x_i)$, et que les dérivées des f_i s'annulent à l'infini; dans ce cas, (A) est vrai, et:

(3,4) $$K(q_1,\dots,q_p) = 0, \text{ si l'un quelconque des } q_i \text{ est impair, } 1 \leq i \leq p.$$

3°) <u>Hypothèse</u> (C): On suppose que (A) est vérifié, et que, à une translation d'origine près, f admet les plans de coordonnées comme plans de symétrie, c'est à dire:

$$f(\pm x_1,\dots,\pm x_p) = f(x_1,\dots,x_p)\ ; \text{ dans ce cas, (3,4) reste vrai.}$$

4°) <u>Hypothèse</u> (D): On suppose que K est symétrique à l'ordre n, si:

$$[[y_1^{r_1}\dots y_p^{r_p}K]] = 0, \text{ si l'un des } r_i \text{ est impair}, 1\leq i\leq p, 0\leq r_i, r_1+\dots+r_p\leq n.$$

a) Si K est symétrique d'ordre 1 (on dira que K est <u>centré</u>), $\forall B$, $\forall i, 1 \leq i \leq p$,

$$[[B,y_iK]] = 0.$$

b) Si K est symétrique d'ordre 2, pour que $[[B,y_iy_jK]] = 0$, $\forall i\neq j, 1\leq i,j\leq p$, il faut et il suffit que B soit telle que:

$$^tB \begin{bmatrix} [[y_1^2K]] & & 0 \\ & \ddots & \\ 0 & & [[y_p^2K]] \end{bmatrix} B = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{bmatrix} \quad ; \text{ si } [[y_i^2K]] = \text{Cte},$$

cette condition équivaut au fait que B est une matrice orthogonale. Si K est centré, on peut toujours se ramener à ce cas, et, en particulier:

(3,5) $$[[y_i^2K]] = 1,\quad [[y_iy_jK]] = 0\ ,\ \forall i\neq j,\ 1 \leq i,j \leq p.$$

Si (3,5) est vérifié, K étant centré, on dira que K est <u>sous forme réduite</u>.

c) Il n'est pas toujours possible de supposer que K est symétrique d'ordre 3, à un changement d'axes près; si c'est le cas, on construit aisément des exemples, où les seules matrices B telles que $[[B,y_1^{r_1}\dots y_p^{r_p}K]] = 0$, pour $r_1+\dots+r_p \leq 3$, et l'un des r_i impair, sont diagonales.

Cependant, la plupart des noyaux couramment utilisés sont symétriques, c'est à dire, tels que:

(3,6) $$K(\pm x_1,\dots,\pm x_p) = K(x_1,\dots,x_p)\ .$$

Si (3,6) est vérifié, K est symétrique d'ordre quelconque.

4°) Par la suite, nous supposerons toujours que (A) est vérifié, et que K est centré, sous forme réduite (3,5). Dans (3,1), par (3,3), on obtient les simplifications suivantes:

a) Si A_N est quelconque:

$$B_1^2 = \sum_{i=2}^{[\frac{1}{2}(n+1)]} V_{2i}(A_N,K) + \zeta_{n+1}(A_N,K) ,$$

$$B_2^2 = \frac{|\mathrm{Det}(A_N)|}{N}[[K^2]] - \frac{1}{N}\sum_{i=0}^{[\frac{1}{2}(n-p)]} V_{2i}^*(A_N,K) + \frac{1}{N}\xi_n(A_N,K) \quad \text{, avec:}$$

(3,7)
$$V_{2i}(A,K) = \sum_{\substack{q_1+\dots+q_p=2i \\ 0\le q_i,\, 1\le i\le p}} K(q_1,\dots,q_p) \sum_{\substack{r_j+s_j=q_j,\, r_j,s_j\ge 0,\, 1\le j\le p \\ r_1+\dots+r_p\ge 2,\; s_1+\dots+s_p\ge 2}} \left[\frac{(-1)^{r_1+\dots+r_p}}{r_1!\dots r_p!\,s_1!\dots s_p!}\right.$$
$$\left.[[A^{-1},y_1^{r_1}\dots y_p^{r_p}K]][[A^{-1},y_1^{s_1}\dots y_p^{s_p}K]]\right]$$

$V_{2i}^*(A,K)$ est définie comme $V_{2i}(A,K)$, mais sans les conditions restrictives: $r_1+\dots+r_p\ge 2, s_1+\dots+s_p\ge 2$.

b) Si A est une matrice orthogonale, on a, dans (3,7):

(3,8) $$[[A^{-1},y_i^2K]]=1\text{ , }[[A^{-1},y_iy_jK]] = 0,\ \forall\ i\neq j\ ,\ 1\le i,j\le p.$$

En particulier, $V_4(A,K)$ est indépendant de A.

c) Remarquons enfin que, dans les formules précédentes:

(3,9) $$K(2q_1,\dots,2q_p)(-1)^{q_1+\dots+q_p} \ge 0,\ \forall\ q_1,\dots,q_p \ge 0.$$

IV) Optimisation asymptotique du M.I.S.E. (premier ordre)

On applique ici les résultats du §III. On étudiera tout d'abord le cas d'une densité dans $\mathbb{R}^2$, particulièrement important en pratique, et permettant des développements plus simples que dans le cas de $\mathbb{R}^p$, $p\ge 3$. Le cas d'une densité dans $\mathbb{R}$ est étudié dans [ ].

A) Cas d'une densité dans $\mathbb{R}^2$:

On suppose que (A) est vérifiée, et que K est sous forme réduite (3,5). En se limitant aux premiers termes des développements (3,7), on obtient:

$$B_1^2 \sim \tfrac{1}{4}K(4,0)[[A_N^{-1},y_1^2K]]^2 + \tfrac{1}{4}K(0,4)[[A_N^{-1},y_2^2K]]^2 + \tfrac{1}{2}K(2,2)[[A_N^{-1},y_1^2K]][[A_N^{-1},y_2^2K]]$$
$$+K(2,2)[[A_N^{-1},y_1y_2K]]^2 + K(1,3)[[A_N^{-1},y_1y_2K]][[A_N^{-1},y_2^2K]]$$

(4,1) $$+ K(3,1)[[A_N^{-1},y_1y_2K]][[A_N^{-1},y_1^2K]] ;$$

$$B_2^2 \sim \frac{|\mathrm{Det}(A_N)|}{N}[[K^2]].$$

Posons maintenant: $A_N^{-1} = \begin{bmatrix} \lambda\cos(\alpha) & \lambda\sin(\alpha) \\ \mu\cos(\beta) & \mu\sin(\beta) \end{bmatrix}$; on obtient, compte tenu de (3,5):

$$(4,2) \quad \begin{aligned} M^2 = B_1^2+B_2^2 &\sim \tfrac{1}{4}K(4,0)\lambda^4+\tfrac{1}{4}K(0,4)\mu^4+\tfrac{1}{2}K(2,2)\lambda^2\mu^2+K(2,2)\lambda^2\mu^2\cos^2(\alpha-\beta) \\ &+K(1,3)\lambda\mu^3\cos(\alpha-\beta)+K(3,1)\lambda^3\mu\cos(\alpha-\beta)+\frac{[[K^2]]}{N\lambda\mu\,|\sin(\alpha-\beta)|}\;. \end{aligned}$$

Jusqu'ici, nous n'avons fait aucun choix particulier du repère, relativement à f. On utilise alors le lemme suivant:

Lemme (4,1): Il existe un repère Oy_1y_2, dans lequel $K(1,3) = K(3,1) = 0$.

Preuve: On effectue le changement de repère $Oy_1y_2 \to Oy_1'y_2'$, en notant $K'(r,s)$ les quantités correspondant à (3,2) dans le nouveau repère. On peut toujours poser:

$$\frac{\partial f}{\partial y_1} = a\frac{\partial f}{\partial Y_1} + b\frac{\partial f}{\partial Y_2}\;,\quad \frac{\partial f}{\partial y_2} = c\frac{\partial f}{\partial Y_1} + d\frac{\partial f}{\partial Y_2}\quad ;\ \text{on en déduit:}$$

a) En prenant $c=1, d=0$, on peut toujours se ramener au cas où $K(1,3) = 0$.

b) On pose $a=c=1$, et on écrit que $K'(1,3) = K'(3,1) = 0$. On en déduit deux relations équivalentes, de la forme:

$$b = \frac{P(d)}{Q(d)}\;,\quad d = \frac{R(b)}{S(b)}\;,\ \text{où } P,Q,R,S \text{ sont des polynômes, avec } \overset{\circ}{d}P = \overset{\circ}{d}R = 2,$$

et $\overset{\circ}{d}S = \overset{\circ}{d}Q = 3$. Les solutions triviales étant exclues, il existe nécessairement une solution de ce système.

Remarque: Le lemme (4,1) est trivial lorsque l'hypothèse (B) ou (C) est vérifiée.

Supposons maintenant que le repère soit choisi tel que $K(1,3) = K(3,1) = 0$; comme, d'après (3,9), $K(2,2) \geq 0$, on constate que le choix $\beta - \alpha = \frac{1}{2}\pi$ minimise M^2, indépendamment de λ et μ, dans (4,2). On se ramène donc, quitte à choisir $\alpha = 0$, à:

$A_N^{-1} = \begin{bmatrix} \lambda & 0 \\ 0 & \mu \end{bmatrix}$. On obtient aisément alors les valeurs asymptotiquement optimales de α et β pour le critère du M.I.S.E.; remarquons en passant que ce choix est intrinsèque à f, ne dépendant pas de la base initiale choisie, puisqu'un changement de base dans $\mathbb{R}^2$ ne modifie la mesure que par la multiplication par une constante.

On résume les résultats:

Théorème (4,2): Etant donnée une densité f, vérifiant (A) à l'ordre 4, et un noyau

(centré, sous forme réduite (3,5): $[[y_1^2K]]=[[y_2^2K]] = 1$, $[[y_1y_2K]] = 0$; si le repère $)y_1y_2$ est choisi tel que:

4,3) $\quad K(1,3) = K(3,1) = 0,$

l'erreur quadratique intégrée M^2 est asymptotiquement minimale pour l'estimation 1,1):

$$f_N(x) = \frac{|\mathrm{Det}(A_N)|}{N} \sum_{i=1}^{N} K[A_N(X_i - x)] \text{ , lorsque:}$$

4,4)

$$A_N = \begin{bmatrix} 1/\lambda_N & 0 \\ 0 & 1/\mu_N \end{bmatrix}, \text{ avec: } \quad \lambda_N = \frac{[[K^2]]^{\frac{1}{6}}\left(K(0,4)/K(4,0)\right)^{\frac{1}{8}}}{\left[\left(K(0,4)\ K(4,0)\right)^{\frac{1}{2}} + K(2,2)\right]^{\frac{1}{6}} N^{-1/6}},$$

$$\mu_N = \frac{[[K^2]]^{\frac{1}{6}}\left(K(4,0)/K(0,4)\right)^{\frac{1}{8}}}{\left[\left(K(0,4)\ K(4,0)\right)^{\frac{1}{2}} + K(2,2)\right]^{\frac{1}{6}} N^{-1/6}}$$

Pour le choix précédent, en notant M^2_{Opt} l'erreur quadratique intégrée minimale,

4,5)

$$M^2_{\mathrm{Opt}} \sim \frac{3\ [[K^2]]^{2/3}}{2\left[\left(K(0,4)K(4,0)\right)^{\frac{1}{2}} + K(2,2)\right]^{-1/3} N^{-2/3}} \quad \bullet$$

emarques: 1°) Quitte à effectuer éventuellement des homothéties sur les axes, on urait pu rajouter à la condition (4,3):

4,6) $\quad K(0,4) = K(4,0).$

On obtient alors, au lieu de (4,4) et (4,5):

4,7)

$$\lambda_N = \mu_N = \left[\frac{2[[K^2]]}{N\int_{\mathbb{R}^2} (\Delta f)^2 dx}\right]^{\frac{1}{6}}, \quad M^2_{\mathrm{Opt}} \sim \frac{3\ [[K^2]]^{2/3}}{2\left\{\int_{\mathbb{R}^2} (\Delta f)^2 dx\right\}^{-1/3} N^{-2/3}} \quad \bullet$$

2°) La solution fournie par (4,4) n'est pas unique. Quitte à se ramener, comme 'écédemment, au cas où (4,3) et (4,6) sont vérifiées, si λ_N est défini par (4,7), 1 constate que $A_N = \lambda_N^{-1}\mathfrak{M}$ rend M^2 asymptotiquement minimal, lorsque $\mathfrak{M}$ est une matrice 'thogonale quelconque. On peut, d'ailleurs, vérifier directement cette propriété dans 1,7), car le laplacien $\Delta f = \dfrac{\partial^2 f}{\partial y_1^2} + \dfrac{\partial^2 f}{\partial y_2^2}$ a une expression indépendante de la base 'thonormée choisie (pour le produit euclidien naturel dans $\mathbb{R}^2$).

3°) On comparera ce résultat (4,7), avec son homologue pour les densités dans $\mathbb{R}$ voir [4]); on obtient, dans ce dernier cas, et avec des notations analogues à celles le nous avons utilisées ici:

$$(4,8) \qquad \lambda_N = \left[\frac{\Gamma[K^2]]}{\int_{\mathbb{R}}(f''(x))^2dx}\right]^{\frac{1}{5}} \frac{1}{N^{1/5}} \cdot$$

On obtient dans (4,17) les développements analogues pour $\mathbb{R}^p$, p quelconque.

4°) On peut constater que, dans un remère convenable, il existe une estimation asymptotiquement optimale, de la forme (1,3). Si le repère n'est pas choisi de maniè que (4,3) et (4,6) soit vérifié, les estimations optimales de la forme (1,3) seront obtenues en minimisant M^2 dans (4,2), pour $\lambda = \mu$, et $\alpha = \beta + \frac{1}{2}\pi$. On obtient alors, dans le repère choisi, les formules (4,7) pour les coefficients optimaux.

Ces formules n'étant liées au repère que par le coefficient:

$$(4,9) \qquad \int_{\mathbb{R}^2} (\Delta f)^2 dx = \Theta \text{ , on en déduit le:}$$

<u>Corollaire</u> (4,3): Etant donnée une densité f, vérifiant (A) à l'ordre 4, et un noyau K centré, sous forme réduite (3,5), l'erreur quadratique intégrée M^2 est asymptotiqu ment minimale pour l'estimation (1,1), lorsqu'elle est sous la forme (1,3), avec un coefficient déterminé par (4,7), et pour un repère Oy_1y_2 tel que dans (4,9), Θ soit minimum.

Signalons que Epanechnikov (1969),[5] obtient, dans le cas des noyaux (1,2), le critère du Θ minimum donné ici en général, pour les estimations (1,1).

5°) (4,5), ou (4,7) montrent que, asymptotiquement, M^2_{Opt} est, indépendamment de f, proportionnelle à $\Gamma[K^2]]^{2/3}$. On en déduit donc que les noyaux qui minimisent $\Gamma[K^2]]$, tout en restant sous forme réduite (3,5), sont asymptotiquement uniformément en f les meilleurs pour le critère du M.I.S.E.; le même raisonnement de calcul des variations que dans le cas unidimentionnel (voir [4]) peut être utilisé ici (voir Epanecl nikov (1969),[5], Lehmann et Hodges (1956)) pour obtenir de tels noyaux.

On obtient ainsi le noyau optimal sous forme réduite:

$$(4,10) \qquad \begin{aligned} K(y_1,y_2) &= \frac{1}{18\pi}[6 - (y_1^2+y_2^2)] \text{ , pour } y_1^2+y_2^2 \leq 6 \text{ ,} \\ &= 0 \text{ , si } y_1^2+y_2^2 > 6. \end{aligned}$$

Il est intéressant de donner un tableau mesurant l'efficacité des estimations (1,1 pour les noyaux les plus usuels dans $\mathbb{R}^2$. On comparera ce tableau à celui obtenu dans [4] pour $\mathbb{R}$. Les noyaux sont mis sous forme réduite (3,5), et on posera éventuelleme

$$\rho = (y_1^2+y_2^2)^{\frac{1}{2}}.$$

Noyau	$[[K^2]]$	$\sim M^2_{Opt} / M^2$
$K = \frac{1}{18\pi}(6-\rho^2),\ \rho^2 \le 6$	$\frac{2}{9\pi}$	1,00
$K = \frac{1}{4\pi},\ \rho \le 2$	$\frac{1}{4\pi}$	92,44%
(4,11) $K = \frac{1}{12},\ \lvert y_1\rvert$ et $\lvert y_2\rvert \le 3^{\frac{1}{2}}$	$\frac{1}{12}$	89,65%
$K = \frac{1}{2\pi} e^{-\frac{1}{2}\rho^2}$	$\frac{1}{4\pi}$	92,44%
$K = \frac{9}{80}\left[1-\frac{y_1^2}{5}\right]\left[1-\frac{y_2^2}{5}\right],\ \lvert y_i\rvert \le 5^{\frac{1}{2}}$, $1 \le i \le 2$	$\frac{9}{125}$	98,83%

emarque: Si K est un noyau produit (1,2), $K(y_1,y_2) = k(y_1)k(y_2)$, k étant un noyau ans $\mathbb{R}$, sous forme réduite (3,5), on a:

$$[[y^2k]] = 1\text{, et } [[K^2]] = [[k^2]]^2.$$

Le dernier noyau du tableau (4,11) est, en conséquence, le noyau produit sous orme réduite qui minimise $[[K^2]]$, ou, d'une manière équivalente $[[k^2]]$. Il réalise insi l'optimalité asymptotique de M^2 pour cette catégorie de noyaux.

6°) La recherche d'un repère tel que (4,3) et (4,6) soient vérifiés peut paraître licate, particulièrement à partir d'un échantillon. Cependant, on peut, dans la upart des cas pratiques, simplifier considérablement ce problème en le ramenant à recherche de symétries, en utilisant la remarque du lemme (4,1). Il suffit, en fet que les axes de coordonnées choisis soient axes de symétrie affine de la distri-tion, pour que (4,3) soit vérifié.

B) Cas d'une densité dans $\mathbb{R}^p$:

On suppose ici, comme pour p=2, que (A) est vérifiée, et que K est sous forme duite (3,5). On utilisera les notations simplificatrices:

$$D_{ii} = K(0,..,0,4,0,..,0) \quad , \quad D_{ij} = K(0,..,2,..,2,..,0) \text{ (2 en i et j)}$$

$$E_{ij} = K(0,..,1,..,3,..,0) \quad , \quad E_{ji} = K(0,..,3,..,1,..,0) \text{ (1 en i, 3 en j)}$$

En se limitant aux premiers termes des développements (3,7), on obtient:

$$B_1^2 \sim \tfrac{1}{4}\sum_{i=1}^{p}\left\{\sum_{j=1}^{p} D_{ij}[[A_N^{-1},y_i^2K]][[A_N^{-1},y_j^2K]]\right\} + \sum_{1\le i<j\le p}\left\{D_{ij}[[A_N^{-1},y_iy_jK]]^2\right.$$

$$(4,12) \qquad \left.+E_{ij}[[A_N^{-1},y_iy_jK]][[A_N^{-1},y_j^2K]] + E_{ji}[[A_N^{-1},y_iy_jK]][[A_N^{-1},y_i^2K]]\right\}$$

$$B_2^2 \sim \frac{|\mathrm{Det}(A_N)|}{N} [[K^2]] .$$

De même que dans le cas p=2, on obtient:

Théorème (4,4): Etant donnée une densité f vérifiant (A) à l'ordre 4, et un noyau centré K, sous forme réduite (3,5); si le repère $Oy_1,\dots,y_p$ est choisi tel que:

(4,13) $\forall\, i \neq j\ ,\ 1 \le i,j \le p\ ,\ E_{ij} = E_{ji} = 0,$

l'erreur quadratique intégrée M^2 est asymptotiquement minimale pour l'estimation (1,1):

$$f_N(x) = \frac{|\mathrm{Det}(A_N)|}{N} \sum_{i=1}^{N} K[A_N(X_i - x)] \text{ , lorsque:}$$

$$A_N = \begin{bmatrix} 1/\lambda_{1,N} & & 0 \\ & \ddots & \\ 0 & & 1/\lambda_{p,N} \end{bmatrix} \text{ , avec:}$$

(4,14)
$$\lambda_{i,N} = \sqrt{\mu_i} \left[\frac{p\,[[K^2]]}{N\sqrt{\prod_{j=1}^{p}\mu_j}\ \int_{\mathbb{R}^p}\left\{\sum_{j=1}^{p}\mu_j \frac{\partial^2 f}{\partial x_j^2}\right\}^2 dx} \right]^{\frac{1}{4+p}} \text{ , où}$$

$\mu_1,\dots,\mu_p$ sont solutions du système: $\mu_i \ge 0,\ 1 \le i \le p,$

$$\forall\, i,\ 1 \le i \le p,\ \sum_{j=1}^{p} D_{ij}\mu_i\mu_j = C \Leftrightarrow \int_{\mathbb{R}^p}\left\{\mu_i \frac{\partial^2 f}{\partial x_i^2}\right\}\left\{\sum_{j=1}^{p}\mu_j \frac{\partial^2 f}{\partial x_j^2}\right\} dx = C.$$

Pour le choix précédent, en notant M^2_{Opt} l'erreur quadratique intégrée minimale,

(4,15)
$$M^2_{Opt} \sim \frac{(p+4)\,[[K^2]]^{\frac{4}{4+p}} \left[\int_{\mathbb{R}^p}\left\{\sum_{j=1}^{p}\mu_j \frac{\partial^2 f}{\partial x_j^2}\right\}^2 dx\right]^{\frac{p}{4+p}} p^{\frac{4}{4+p}}}{4p\left[N\sqrt{\prod_{j=1}^{p}\mu_j}\right]^{\frac{4}{4+p}}}$$

Remarques: 1°) Une forme plus simple de (4,14) peut être obtenue pour $\lambda_{i,N}$, en constatant que:

(4,16)
$$\int_{\mathbb{R}^p}\left\{\sum_{j=1}^{p}\mu_j \frac{\partial^2 f}{\partial x_j^2}\right\}^2 dx = pC \text{ , et que: } \lambda_{i,N} = \sqrt{\mu_i}\left[\frac{[[K^2]]}{NC\sqrt{\prod_{j=1}^{p}\mu_j}}\right]^{\frac{1}{4+p}} .$$

Cependant, les formulations (4,14) et (4,15) de la solution sont tout aussi intéressantes, et fournissent directement, d'une manière analogue à (4,7):

Lorsque μ_i = Cte, $1 \le i \le p$, est solution de (4,14), $\forall i, 1 \le i \le p$,

$$(4,17) \qquad \lambda_{i,N} = \left[\frac{p[[K^2]]}{N\int_{\mathbb{R}^p}(\Delta f)^2 dx}\right]^{\frac{1}{4+p}}, \quad M^2_{opt} \sim \frac{(p+4)\left[p[[K^2]]\right]^{\frac{4}{4+p}}\left[\int_{\mathbb{R}^p}(\Delta f)^2 dx\right]^{\frac{p}{4+p}}}{4p\, N^{4/(4+p)}}$$

On constate que résoudre en $\{\mu_i, 1\le i\le p\}$ le système (4,14) est équivalent à effectuer des affinités convenables sur les axes, pour se ramener à (4,17). En définissant Θ par (4,9), dans $\mathbb{R}^p$, on en déduit la version dans $\mathbb{R}^p$ du corollaire (4,3):

Corollaire (4,5): Etant donnée une densité f, vérifiant (A) à l'ordre 4, et un noyau K centré, sous forme réduite (3,5), l'erreur quadratique intégrée M^2 est asymptotiquement minimale pour l'estimation (1,1), lorsque celle-ci est sous la forme (1,3), avec un coefficient déterminé par (4,17), et pour un repère $Oy_1,\dots,y_p$ tel que:

$$\Theta = \int_{\mathbb{R}^p}(\Delta f)^2 dx \text{ , y soit minimum.}$$

Remarque: Il existe toujours un tel repère, ce qui peut être prouvé en faisant varier les extrémités des vecteurs de base sur la boule unité (compacte) de $\mathbb{R}^p$, $\Theta \ge 0$, y ayant une borne inférieure atteinte. On déduit aisément que, pour un tel repère, les conditions (4,13) sont vérifiées, ce qui fournit une démonstration plus générale du lemme (4,1).

2°) La solution fournie par (4,14),(4,16), ou (4,17) n'est pas unique. Quitte à se ramener au cas (4,17), on constate que si $\lambda_N = \lambda_{i,N}$ est défini par (4,17), $A_N=\lambda_N^{-1}\mathfrak{M}$ rend M^2 asymptotiquement minimal, lorsque $\mathfrak{M}$ est une matrice orthogonale quelconque. On vérifie directement cette propriété, car Θ est invariant par changement de base orthonormé.

3°) Lorsque $N\to\infty$, M^2_{Opt} est asymptotiquement proportionnel, dans (4,15) ou (4,17), à $[[K^2]]^{\frac{4}{4+p}}$. Le même raisonnement que pour (4,10) permet d'obtenir:

Proposition (4,6): Le noyau positif K, sous forme réduite (3,5), minimisant $[[K^2]]$, et fournissant asymptotiquement l'erreur quadratique intégrée M^2 minimale, indépen--damment de f, par (4,15), dans $\mathbb{R}^p$, est, en notant $\rho = (y_1^2+\dots+y_p^2)^{\frac{1}{2}}$:

$$(4,18) \qquad K = \frac{(q+1)!}{\pi^q 2^{q+1}(q+2)^{q+1}}\left[(2q+4) - \rho^2\right] \text{ , } \rho \le \sqrt{2q+4} \text{ , si } p=2q \text{ ,}$$

$$K = \frac{(2q+3)!}{\pi^{q}2^{2q+3}(2q+5)^{\frac{1}{2}(2q+3)}(q+1)!}\left[(2q+5)-\rho^{2}\right],\ \rho \le \sqrt{2q+5}\ ,\ \text{si } p=2q+1.$$

On obtient les valeurs correspondantes de $[[K^2]]$:

$$(4,19)\qquad [[K^{2}]] = \frac{(q+1)!}{\pi^{q}(q+2)^{q+1}2^{q-1}},\ \text{si } p=2q\ ;\ [[K^{2}]] = \frac{(2q+3)!}{\pi^{q}(2q+5)^{\frac{1}{2}(2q+3)}2^{2q+1}(q+1)!},$$

si $p = 2q+1$.

<u>Exemples</u>: $p=1$, $K = \dfrac{3}{20\sqrt{5}}[5-\rho^{2}]$, résultat d'Epanechnikov (1969),[ ],

$p=2$, $K = \dfrac{1}{18\pi}[6-\rho^{2}]$, donné en (4,10),

$p=3$, $K = \dfrac{15}{392\pi\sqrt{7}}[7-\rho^{2}]$.

Nous donnons ici, comme en (4,11) pour p=2, et en [4] pour p=1, un tableau mesurant l'efficacité asymptotique de quelques estimations à noyau usuelles dans $\mathbb{R}^3$ (p=3). Les noyaux sont mis sous forme réduite (3,5):

	Noyaux	$[[K^2]]$	$\sim M_{Opt}^2 / M^2$
	$K = \dfrac{15}{392\,\pi\sqrt{7}}[7-\rho^{2}]$, $\rho \le \sqrt{7}$	$\dfrac{15}{98\pi\sqrt{7}}$	1,00
	$K = \dfrac{3}{20\,\pi\sqrt{5}}$, $\rho \le \sqrt{5}$	$\dfrac{3}{20\pi\sqrt{5}}$	91,88%
(4,20)	$K = \dfrac{1}{(2\pi)^{3/2}}e^{-\frac{1}{2}\rho^{2}}$	$\dfrac{1}{8\,\pi^{3/2}}$	89,29%
	$K = \dfrac{1}{24\sqrt{3}}$, $\lvert x_i\rvert \le \sqrt{3}$	$\dfrac{1}{24\sqrt{3}}$	85,83%
	$K = \dfrac{27}{320\sqrt{5}}\left[1-\dfrac{x_1^2}{5}\right]\left[1-\dfrac{x_2^2}{5}\right]\left[1-\dfrac{x_3^2}{5}\right]$ $\lvert x_i\rvert \le \sqrt{5}$, $1 \le i \le 3$	$\dfrac{27}{625\sqrt{5}}$	97,28%

Le tableau (4,20) montre, avec le tableau (4,11), que l'efficacité des meilleurs noyaux produits reste excellente pour $\mathbb{R}^p$, p=2,3. Pour les noyaux uniformes, cubiques, et normaux, l'efficacité est sensiblement moins bonne que dans $\mathbb{R}$ (p=1).

4°) La condition (4,13) est vérifiée si les plans de coordonnées sont plans de symétrie de la distribution (à une translation d'origine près). Dans le cas particulier de la loi multinormale, il existe une indétermination d'ordre p-1 dans le choix d'un tel système de coordonnées symétriques, puisque tout système d'axes conjugués relativement à la forme quadratique définissant la distribution est solution.

5°) L'efficacité asymptotique atteingible par les noyaux positifs (ordre de décroissance $N^{-4/(4+p)}$) est de plus en plus médiocre lorsque p croît. Pour atteindre une efficacité supérieure (ordre de décroissance $N^{-1+\epsilon}$, $\epsilon > 0$ arbitrairement petit), il est nécessaire de permettre aux noyaux de prendre des valeurs négatives.

6°) Nous n'avons pas justifié jusqu'ici le fait que les noyaux sont choisis centrés. Un raisonnement analogue à celui des théorèmes (4,4) et (4,2) montre qu'en général, lorsqu'un noyau est décentré, l'ordre de décroissance de M^2 est en $N^{-2/(p+2)}$, et est de ce fait toujours moins bon que pour un noyau centré.

V) Optimisation asymptotique du M.I.S.E. (deuxième ordre)

Nous avons constaté, dans les remarques n°2 des théorèmes (4,2) et (4,4), que le choix de A_N dans (1,1), pour obtenir asymptotiquement une erreur quadratique intégrée minimale, n'est pas unique. Ce fait peut sembler paradoxal, particulièrement pour des noyaux non isotropes, comme des noyaux produits. En fait, cela provient de l'expression asymptotique (4,1) ou (4,12), qui ne fait intervenir que des invariants ou des moments d'ordre 2 du noyau, expressions qui ne sont pas modifiées par une trans-formation orthogonale, dans le cas réduit (4,7) ou (4,17).

On peut alors se poser la question de la détermination d'une telle transformation pour minimiser les termes suivants du développement de M^2. Nous étudions ici ce problème, et tout d'abord dans le cas de densités dans $\mathbb{R}^2$:

A) Cas d'une densité dans $\mathbb{R}^2$:

Nous allons, tout d'abord, faire des hypothèses sensiblement plus précises sur f que celles du §IV. On suppose que (A) est vraie à l'ordre 6, et que:

$$(5,1)\quad \begin{aligned} &K(1,3) = K(3,1) = 0 \quad \text{(hypothèse (4,3))} \\ &K(1,5) = K(5,1) = K(3,2) = K(2,3) = 0. \text{(voir (3,2))} \end{aligned}$$

En général, si f est quelconque, il n'est pas toujours possible de déterminer un repère (voir lemme (4,1) et remarque du corollaire (4,5)), tel que (5,1) soit vérifié. Le cas standard de vérité de (5,1) est obtenu lorsque les axes sont axes de symétrie de la distribution, ou plus généralement si l'une des hypothèses (B) ou (C) est vraie, ce qui implique (3,4).

(5,1) étant vrai, par une affinité sur les axes, il est toujours possible de se ramener au cas où (4,6) est vrai:

$$(4,6)\qquad K(0,4) = K(4,0).$$

Il est plus facile de faire des hypothèses sur K que sur f; tout d'abord, on peut toujours se ramener au cas où K est sous forme réduite (3,5):

(3,5) $$[[y_1^2 K]] = [[y_2^2 K]] = 1 \ , \ [[y_1 y_2 K]] = 0 \ ;$$

En fait, la plupart des noyaux usuels sont des noyaux vérifiant l'une des hypothèses (B) ou (C), en remplaçant f par K. On pourra donc supposer qu'en plus de (3,5):

(5,2) $$[[y_1^4 K]] = [[y_2^4 K]] \ , \ [[y_1^r y_2^s K]] = 0, \text{ si r ou s est impair.}$$

Pour simplifier les notations, on supposera que $A_N = \lambda_N^{-1} B_N$, B_N étant une matrice orthogonale directe: $B_N = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$; on posera, de plus:

$$K_{rs} = [[B_N^{-1}, y_1^r y_2^s K]] \ , \ K_{rs}^0 = [[y_1^r y_2^s K]] = [[Id, y_1^r y_2^s K]].$$

Il sera fait usage du fait que:

$K_{30} = K_{03} = 0$, et que $K_{21} = K_{12} = 0$, $\forall\, B_N$, ce qui peut être prouvé aisément en développant ces expressions en fonction des K_{rs}^0.

Comte tenu de ces notations et hypothèses, en se limitant aux deux premiers termes des développements (3,7), on obtient:

(5,3)
$$\begin{aligned} B_1^2 &= \tfrac{1}{4}\lambda_N^2 (K(4,0)+K(0,4)+2K(2,2)) \\ &+ \frac{\lambda_N^6}{4!}\Big[K(0,6)K_{04}+6K(2,4)K_{22}+K(2,4)K_{04}+6K(4,2)K_{22}+K(4,2)K_{40}+K(6,0)K_{40}\Big] \\ &+ O(\lambda_N^8) \ , \end{aligned}$$
$$B_2^2 = \frac{[[K^2]]}{N\lambda_N^2} - \frac{1}{N}\Big[K(0,0) + O(\lambda_N^2)\Big] .$$

On constate par (5,3) que, lorsque $\lambda_N \to 0$, $N\lambda_N^2 \to \infty$, on minimise M^2 asymptotiquement, uniformément par rapport à λ_N , en choisissant B_N telle que:

(5,4)
$$\begin{aligned} C(\theta) &= K(0,6)K_{04}+6K(2,4)K_{22}+K(2,4)K_{04}+6K(4,2)K_{22}+K(4,2)K_{40}+K(6,0)K_{40} \\ &= a\,K_{04} + b\,K_{22} + c\,K_{40} \ , \text{ soit minimum.} \end{aligned}$$

En utilisant l'hypothèse (5,2), et en notant $K_4 = K_{04}^0 = K_{40}^0$, $K_2 = K_{22}^0$, développons (5,4):

(5,5) $$C(\theta) = (K_4(a+c)+K_2 b) + (\tfrac{1}{2}\sin^2 2\theta)\Big[2\,(b-a-c)(K_4 - 3K_2)\Big] .$$

On en déduit que $C(\theta)$ est minimum: pour $\theta = 0 + \frac{k\pi}{2}$, si $(b-a-c)(K_4-3K_2) \geq 0$,

pour $\theta = \frac{\pi}{4} + \frac{k\pi}{2}$, si $(b-a-c)(K_4-3K_2) \leq 0$.

On dira, dans le premier cas ($\theta = 0$), que le noyau est en position droite, dans le ieuxième cas ($\theta = \frac{\pi}{4}$), que le noyau est en position oblique.

Nous résumons ces résultats:

'héorème (5,1): Pour estimer une densité f, vérifiant (A) à l'ordre 6, et (4,6),(5,1), u l'une des hypothèses (B) ou (C), par l'estimation (1,1), avec un noyau K vérifiant 3,5),(5,2), on pose:

$$\Omega = 5(K(2,4)+K(4,2)) - (K(0,6)+K(6,0))$$

1°) On place le noyau en position droite, si $\Omega\left[[[y_1^4K]]-3[[y_1^2y_2^2K]]\right] \geq 0$, en position blique sinon.

2°) On pose $A_N = \lambda_N^{-1}\,\mathrm{Id}$, où :

5,6) $$\lambda_N = \left[\frac{2[[K^2]]}{N\int_{\mathbb{R}^2}(\Delta f)^2 dx}\right]^{\frac{1}{6}}\left[1 - \left[\frac{2[[K^2]]}{N}\right]^{\frac{1}{3}}\left\{\frac{C+o(1)}{4!\left[\int_{\mathbb{R}^2}(\Delta f)^2 dx\right]^{\frac{4}{3}}}\right\}\right]$$

ù $C = C(0)$ ou $C(\frac{1}{4}\pi)$ dans (5,5), suivant que le noyau soit en position droite ou blique:

$$C(0) = [[y_1^4K]]\left[K(0,6)+K(2,4)+K(4,2)+K(6,0)\right] + 6[[y_1^2y_2^2K]]\left[K(2,4)+K(4,2)\right] ,$$

$$C(\tfrac{\pi}{4}) = \tfrac{1}{2}[[y_1^4K]]\left[K(0,6)+7[K(2,4)+K(4,2)]+K(6,0)\right]$$

$$+\tfrac{1}{2}[[y_1^2y_2^2K]]\left[3K(0,6)-6[K(2,4)+K(4,2)]+3K(6,0)\right] .$$

xemples: 1°) Noyau carré; il peut se mettre sous forme symétrique sous les deux ormes suivantes:

$$K = 1/12\ ,\ |y_1|\ \text{et}\ |y_2| \leq \sqrt{3}\ ,\ \text{ou}\ K = 1/12\ ,\ |y_1| + |y_2| \leq \sqrt{6}\ .$$

our ces positions, on obtient respectivement: ($K_4=9/5$, $K_2=1$), et ($K_4=12/5$, $K_2=2/5$). i $\Omega \geq 0$, la meilleure orientation est oblique, si $\Omega \leq 0$, la meilleure orientation st la position droite.

2°) Noyau produit; ce cas englobe le précédent, avec $K(y_1,y_2)=k(y_1)k(y_2)$. En suppoant que K est sous forme réduite (3,5), $K_4 = \beta_2$, où β_2 est le coefficient d'aplatissement de k. On obtient:

En position droite: $K_4 = \beta_2$, $K_2 = 1$, , $K = k(y_1)k(y_2)$,

En position oblique: $K_4 = \frac{1}{2}(\beta_2+3)$, $K_2 = \frac{1}{2}(\beta_2-1)$, $K=k\left[\frac{x+y}{\sqrt{2}}\right]k\left[\frac{x-y}{\sqrt{2}}\right]$.

a) Si $1 < \beta_2 \leq 3$, la position oblique est meilleure si $\Omega \geq 0$, moins bonne autrement.

b) Si $\beta_2 = 3$ (cas du noyau normal), la position est indifférentiée.

c) Si $3 \leq \beta_2$, la position droite est meilleure si $\Omega \geq 0$, moins bonne autrement.

Remarques: 1°) Le théorème (5,1) montre que, sauf dans le cas où f ou K ont une symétrie de révolution pour les $K(i,j)$ et K_{ij} à un ordre convenable, les axes asymptotiquement optimaux sont définis de manière unique.

2°) Les axes optimaux précédents conservent leur optimalité asymptotique uniformé--ment en λ_N.

3°) Compte tenu de la remarque 6 du théorème (4,2), l'application pratique de ces résultats consiste essentiellement à déterminer des axes de symétrie de la distribu--tion, puis à les normaliser, en y déterminant des échelles convenables. Ceci étant fait, il n'y a que deux possibilités d'orientation optimale qui peuvent être choisie soit en effectuant une hypothèse paramétrique sur f, et en estimant Ω, ou, plus particulièrement son signe, soit empiriquement.

4°) Le terme correctif de (5,6), par rapport à (4,17), est un bon critère pour apprécier la validité des approximations utilisées pour obtenir les formules asympto--tiques. Pour une hypothèse paramétrique particulière, on peut aisément apprécier la taille de l'échantillon nécessaire pour que (5,6) donne un résultat sensiblement égal à (4,17).

B) Cas d'une densité dans $\mathbb{R}^p$ ($p \geq 3$):

On fait sur K et f des hypothèses analogues à celles du §B:

On suppose que (A) est vraie à l'ordre 6, et que, en notant:

$$F_{i,r} = K(0,..,r,..,0) \quad (r \text{ en } i),$$

(5,7) $$F_{ij,rs} = K(0,..,r,..,s,..,0) \quad (r \text{ en } i,\ s \text{ en } j),\ i\neq j,$$

$$F_{ijk,rst} = K(0,..,r,..,s,..,t,..,0) \quad (r \text{ en } i,\ s \text{ en } j,\ t \text{ en } k), \text{ pour}$$

i,j,k distincts, on ait:

$F_{i,r}$, $F_{ij,rs}$, $F_{ijk,rst} = 0$, si l'un des r, s ou t est impair, la somme des indices r, respectivement r+s, ou r+s+t étant ≤ 6. Plus généralement, on supposera que (3,4) est vrai. Ces hypothèses sont vérifiées par exemple si (B) ou (C) est vraie.

On suppose de plus, que (4,17) est vraie, ou, avec les notations (5,7):

$$\sum_{j=1}^{p} F_{ij,22} = \frac{1}{p}\int_{\mathbb{R}^p} (\Delta f)^2 dx \quad , \quad \text{avec } F_{i,r+s} = F_{ii,rs}\ ,\ \forall j,\ 1 \leq j \leq p.$$

On suppose que K est sous forme réduite (3,5), et que K est symétrique d'ordre 6:

$$[[y_1^{r_1}..y_p^{r_p}K]] = 0, \text{ si l'un des } r_i \text{ est impair}, 0\leq r_i, r_1+..+r_i\leq 6, 1\leq i\leq p.$$

On note, de même que dans le §B:

$$A_N = \lambda_N^{-1} B_N \text{ , où } B_N \text{ est une matrice orthogonale,}$$

$$K_{i,r} = [[B_N^{-1}, y_i^r K]] \ ,\ K_{ij,rs} = [[B_N^{-1}, y_i^r y_j^s K]] \ .$$

$$K^0_{i,r} = [[y_i^r K]] \ ,\ K^0_{ij,rs} = [[y_i^r y_j^s K]] \ .$$

On obtient, en se limitant aux deux premiers termes des développements (3,7):

$$B_1^2 = \tfrac{1}{4}\lambda_N^4\Big\{\sum_{i=1}^{p} F_{i,4} + \sum_{i\neq j} F_{ij,22}\Big\} + \frac{\lambda_N^6}{4!}\Big\{\sum_{i=1}^{p} F_{i,6}K_{i,4} + 6\sum_{i\neq j} F_{ij,24}K_{ij,22}$$

5,8)
$$+ \sum_{i\neq j} F_{ij,24}K_{i4} + 6 \sum_{i,j,k \text{ distincts}} F_{ijk,222}K_{ij,22}\Big\} + O(\lambda_N^8)$$

$$B_2^2 = \frac{[[K^2]]}{N\,\lambda_N^p} - \frac{1}{N}\Big[K(0,\ldots,0) + O(\lambda_N^2)\Big] \quad .$$

On constate, par (5,8) que, lorsque $\lambda_N \to 0$, $N\lambda_N^p \to \infty$, on minimise M^2, asymp-
ptiquement, uniformément par rapport à λ_N, en choisissant B_N, telle que:

$$C(B_N) = \sum_{i=1}^{p} F_{i,6}K_{i,4} + 6\sum_{i\neq j} F_{ij,24}K_{ij,22} + \sum_{i\neq j} F_{ij,24}K_{i,4}$$

5,9)
$$+ 6 \sum_{i,j,k \text{ distincts}} F_{ijk,222}K_{ij,22} \quad \text{, soit minimum.}$$

On peut alors résoudre ce système en utilisant la représentation d'Euler de B_N;
$_N$ peut être représentée comme un produit de $\frac{1}{2}n(n-1)$ matrices, représentant chacune
ne matrice de rotation dans le sous-espace défini par un couple de coordonnées.
On obtient dans $\mathbb{R}^3$: (B_N étant supposé directe)

$$B_N = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & -\sin\psi \\ 0 & \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\varphi & -\sin\varphi & 0 \\ \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

La solution existe toujours, B_N pourvant être représentée continûment comme
'image d'un compact de $\mathbb{R}^{\frac{1}{2}n(n-1)}$, le minimum de $C(B_N)$ existe et est atteint.

On peut, cependant, partiellement ramener le problème de la minimisation de (5,9)
ı cas de $\mathbb{R}^2$, traité dans le §A; supposons pour simplifier, qu'il existe une symétrie
omplète des axes:

5,10) $F_{i,6} = F_6$, $F_{ij,24} = F_{24}$, $F_{ijk,222} = F_{222}$, $K^0_{i,r} = K_r$, $K^0_{ij,rs} = K_{rs}$,

$\forall\ i,j,k$.

Supposons qu'on se limite pour B_N à une rotation dans le plan Oy_iy_j ; on constate
isément que dans (5,9), les termes contenant une seule fois i ou j ne sont pas modi-
fiés. Seuls les termes $K_{ij,22}$, $K_{i,4}$, $K_{j,4}$ sont modifiés. On est alors ramené à un
roblème résolu par le théorème (5,1).

On obtient:

<u>Théorème</u> (5,2): Pour estimer une densité f, vérifiant (A) à l'ordre 6, et (5,7), (4,17), ou l'une des hypothèses (B) ou (C), par l'estimation (1,1), avec un noyau sous forme réduite (3,5) et symétrique d'ordre 6 (hypothèse (D)), on détermine une transformation orthogonale B_N, telle que $C(B_N)$ (donné en (5,9)) soit minimum. On pose alors:

$$A_N = \lambda_N^{-1} B_N \text{ , avec:}$$

$$(5,11) \qquad \lambda_N = \left[\frac{p[[K^2]]}{N \int_{\mathbb{R}^p} (\Delta f)^2 dx}\right]^{\frac{1}{4+p}} \left[1 - \left[\frac{p[[K^2]]}{N}\right]^{\frac{2}{4+p}} \left[\frac{\operatorname{Inf}_{B_N} C(B_N) + o(1)}{4(4+p)\left[\int_{\mathbb{R}^p} (\Delta f)^2 dx\right]^{\frac{6+p}{4+p}}}\right]\right]$$

On obtient alors l'erreur quadratique intégrée minimale, à un terme du second ordre près.

Quand (5,10) est vrai, une condition nécessaire et suffisante, pour que $B_N = Id$ minimise $C(B_N)$, est que:

$$\left[K_4 - 3K_{22}\right]\left[(7-p)F_{24} - F_6 + 3(p-2)F_{222}\right] \geq 0$$

<u>Remarques</u>: Les remarques 1,2,3,4 suivant le théorème (5,1) sont valables dans $\mathbb{R}^p$.

Il est utile de signaler que si on peut trouver la forme des noyaux optimaux asymptotiquement, l'optimalité de noyaux tels que (4,18) est définie à un équivalent près. Il n'est donc pas prouvé que pour les valeurs usuelles de N, on n'obtienne pas de meilleurs résultats du point de vue du critère du M.I.S.E., avec d'autres noyaux que (4,18). On s'épargne en tout cas le problème de l'orientation du noyau, si on prend un noyau isotrope. Pour des raisons pratiques, il arrive cependant que des noyaux non isotropes soient préférables, et on peut alors utiliser ce qui précède.

VI) <u>Calcul exact du M.I.S.E.</u>

On peut, comme dans [4], obtenir des conditions suffisantes pour que, dans les développements (3,7), les restes: $\zeta_{n+1}(A_N, K)$, $\xi_n(A_N, K)$, aient pour limite 0 lorsque $N \to \infty$. Ces développements peuvent alors être considérés comme des sommes de séries, et on aura exactement:

$$(6,1) \qquad B_1^2 = \sum_{i=2}^{\infty} V_{2i}(A_N, K) \quad , \quad B_2^2 = \frac{|\operatorname{Det}(A_N)|}{N}[[K^2]] - \frac{1}{N}\sum_{i=0}^{\infty} V_{2i}^*(A_N, K) \text{ .}$$

Une démonstration généralisée dans $\mathbb{R}^p$, de [4] permet d'énoncer:

<u>Théorème</u> (6,1): Si f est analytique dans $\mathbb{R}^p$, et somme de son développement en tout point, si (A) est vérifié, et K est centré sous forme réduite (3,5), si les conditions de la proposition (3,1) sont vérifiées pour n ∞, si, de plus,

$$(6,2) \qquad ||A_N|| < \frac{1}{p \operatorname*{LimSup}_{n\,\infty} \left\{ \left[\operatorname*{Sup}_{q_1+\dots+q_p=n} K(q_1,\dots,q_p) \right] \left[\operatorname*{Sup}_{r_1+\dots+r_p\leq n} [[A_N^{-1}, y_1^{r_1} \dots y_p^{r_p} K]]^2 \right] \frac{1}{n!} \right\}^{\frac{1}{n}}}$$

Alors (6,1) est vrai.

<u>Remarques</u>: 1°) Dans le cas de la loi normale réduite, on obtient des simplifications remarquables dans (3,7) et (6,1). Plus précisément, si:

$$f = \frac{1}{(2\pi)^{\frac{1}{2}p}} \exp\left\{ -\frac{1}{2} \sum_{i=1}^{p} y_i^2 \right\} \text{ , on a :}$$

$$K(q_1,\dots,q_p) = \prod_{i=1}^{p} \left\{ \int_{\mathbb{R}} e^{-\frac{1}{2}t^2} \frac{d^{q_i}}{dt^{q_i}} \left[\frac{e^{-\frac{1}{2}t^2}}{2\pi} \right] dt \right\} \text{ , d'où :}$$

$$K(q_1,\dots,q_p) = 0, \text{ si l'un des } q_i \text{ est impair , } 1 \leq i \leq p \text{ ,}$$

$$(6,3) \qquad K(2q_1,\dots,2q_p) = \frac{(-1)^{q_1+\dots+q_p} \left(\frac{1}{2}\right)^{p+2(q_1+\dots+q_p)}}{\pi^{\frac{1}{2}p}} \prod_{i=1}^{p} \left[\frac{(2q_i)!}{q_i!} \right]$$

Sous réserve que A_n vérifie (6,2), on obtient dans $\mathbb{R}^p$ des développements analogues à ceux qui ont été obtenus pour $\mathbb{R}$ dans [4]. Compte tenu que, si K est sous forme réduite (3,5), les estimations asymptotiquement optimales sont de la forme (1,3), on se ramène, comme dans le cas de $\mathbb{R}$, à des expressions B_1^2 , B_2^2 , fonctions d'un seul paramètre réel δ_n .

La procédure suivante peut alors être appliquée:(K étant un noyau donné)

a) On fait une hypothèse ne normalité sur f, et on estime une base, dans laquelle l'estimée de f est normale centrée réduite.

b) On détermine, compte tenu des paramètres estimés, et avec une recherche de robustesse comme celle décrite dans [4], le coefficient δ_n optimal, en utilisant les formules (6,1).

c) On construit l'histogramme.

Cette procédure peut permettre, jointe à une procédure de classement, un classement robuste de nuages de points dans $\mathbb{R}^p$.

2°) La procédure décrite précédemment peut être appliquée avec n'importe quelle distribution convenablement régulière avec support dans $\mathbb{R}^p$. Elle est à conseiller systématiquement, parallèlement à toute estimation paramétrique, pour donner une représentation visuelle de la distribution, robuste lorsque s'introduit une erreur d'hypothèse sur celle-ci. La loi normale présente l'inconvénient d'avoir une infinité d'axes de symétrie conjugués. Pour faire apparaître éventuellement des axes de symétrie, il est intéressant de généraliser les hypothèses de normalité, en considérant des distributions de la forme:

$$f = (A + Bq_1(x)) \exp(-\tfrac{1}{2}q_2(x)) \text{ , où } q_1 \text{ et } q_2 \text{ sont des formes quadratiques}$$

positives.

VII) Utilisation de noyaux non nécessairement positifs:

Comme dans $\mathbb{R}$, l'usage de noyaux pouvant être négatifs permet d'améliorer considérablement l'ordre asymptotique de convergence de M^2 vers 0.

Supposons que K soit un noyau produit, de la forme:

$K = k(y_1)..k(y_p)$, où k est un noyau dans $\mathbb{R}$, tel que:

$[[y^r k]] = 0$, pour $r=1,2,..,m$; comme:

$[[y_1^{r_1}..y_p^{r_p}K]] = [[y^{r_1}k]]..[[y^{r_p}k]] = 0$, s'il existe un $r_i, 1 \leq r_i \leq m$.

On en déduit que:

$V_{2i}(Id,K) = 0$, $1 \leq i \leq m$, de même pour V^* qui se simplifie.

Pour des estimations de la forme (1,2), ou (1,3), on obtient aisément alors une décroissance asymptotique en:

$$N^{-(2(m+1))/(p+2(m+1))} .$$

Nous étudierons plus particulièrement ce type d'estimation dans une prochaine note.

_ • _ • _ • _ • _ • _

Références bibliographiques:

[1] Cacoullos (1966) Estimation of a multivariate density, Ann. Inst. Stat. Math. 18, p.179-89.

[2] Deheuvels P. (1974) Estimation de la densité, Thèse, Université Paris VI,

[3] (1974) Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre, et uniforme presque sûre des estimateurs de la densité, C.R.Acad.Sci. Paris, 278, p.1217-20,

[4] (1977) Estimation non paramétrique de la densité par histogrammes généralisés (I), à paraître dans la Revue de Statistique Appliquée, Sept. 1977.

[5] Epanechnikov (1969) Non parametric estimation of a multivariate probability density, Teor. Prob. Appl. 14, p.153-8

[6] Maniya (1961) Remarks on non parametric estimates of a bivariate probability density, Soobshch Akad. Nauk. Gruz. SSR, 27, 4, pp.385-400.

[7] Murthy (1964) Estimation of bivariate probability density, Ann. Math. Statist. 34, p.457

[8] (1965) Estimation of probability density, Ann. Math. Statist., 36, p.1027-31

[9] Nadayara (1964) Estimation of bivariate probability density, Soobshch. Akad. Nauk. Gruz. SSR., 36, 2, p.267-8

[10] Rosenblatt (1956) Remarks on some non parametric estimates of a density, Ann. Math. Statist., 27, p.832-7

[11] Parzen (1962) On the estimation of probability density and the mode, Ann. Math. Statist., 33, p.1065-76

Reçu en Juin 1977
Institut de Statistique
Université P. et M. Curie
Tour 45-55,
4 Place Jussieu
75230 Paris Cedex 05

57 bis,Avenue du Petit Chambord
92340 BOURG LA REINE