ANNALES DE L'I. H. P., SECTION B

P. DOUKHAN

J. LEON

F. PORTAL

# Une mesure de la déviation quadratique d'estimateurs non paramétriques

*Annales de l'I. H. P., section B*, tome 22, nº 1 (1986), p. 37-66

<http://www.numdam.org/item?id=AIHPB_1986__22_1_37_0>

*Ann. Inst. Henri Poincaré,*
Vol. 22, n° 1, 1986, p. 37-66. *Probabilités et Statistiques*

# Une mesure de la déviation quadratique d'estimateurs non paramétriques (*)

par

**P. DOUKHAN (**), J. LEON (***), F. PORTAL (**)**

RÉSUMÉ. — Soit $X_n = (U_n, V_n)$ une suite équidistribuée de variables aléatoires indépendantes ou $\phi$-mélangeante à valeurs dans $\mathbb{R}^{d+e}$, nous désignons par $\theta$, soit la densité, $f$, de la loi de $U_1$, soit la fonction $r(x) = E(V_1/U_1 = x)$, soit la fonction $rf$ et par $\hat{\theta}_n$ une estimation non paramétrique de $\theta$ construite sur l'échantillon $X_1, \ldots, X_n$. Nous montrons dans ce travail la convergence des statistiques $a_n \int | \hat{\theta}_n - \theta |^2 d\mu - b_n$ vers une loi normale $N(0, \sigma^2)$, dans les cas où $\hat{\theta}_n$ désigne un estimateur à noyau ou de projection. Notons que les constantes $\sigma^2$, $a_n$ et $b_n$ ne dépendent pas du mélange, seules les fenêtres d'estimation varient entre le cas indépendant et le cas $\phi$-mélangeant.

ABSTRACT. — Let $X_n = (U_n, V_n)$ be a stationary $\mathbb{R}^{d+e}$-valued sequence of independent or $\phi$-mixing random variables, and $\theta$ being the density, $f$, of the law of $U_1$ or the regression function $r(x) = E(V_1/U_1 = x)$ or the function $rf$, we consider non parametric estimates $\hat{\theta}_n$ of $\theta$ based on the sample $\{X_1, \ldots, X_n\}$. We show here the convergence of the statistics

$$a_n \int | \hat{\theta}_n - \theta |^2 d\mu - b_n$$

(*) Au moment de la frappe de ce texte nous apprenons la publication d'un article de Peter Hall traitant du risque quadratique concernant la fonction de régression dans le cas de variables indépendantes pour des estimateurs à noyau (*Annals of Statistics* 1984, t. **12**, n° 1, p. 241-260). Il donne dans cet article des valeurs optimales pour la fenêtre d'estimation toutefois, utilisant des méthodes de martingales, son travail ne permet pas de traiter de variables mélangeantes.

(**) E. R. A. 532, Bât. 425, Université de Paris-Sud, Centre d'Orsay, 91405 Orsay Cedex, France.

(***) Université Centrale du Vénézuela, Faculté des Sciences, Département de Mathématiques, Caracas Vénézuela (actuellement à l'E. R. A. 532, Orsay, France).

*Annales de l'Institut Henri Poincaré* - Probabilités et Statistiques - 0246-0203
Vol. 22/86/01/37/30/\$ 5,00/© Gauthier-Villars

to a normal $N(0, \sigma^2)$ law, where $\hat{\theta}_n$ is a kernel estimator or a projection estimator. We remark that the constants $\sigma^2$, $a_n$ and $b_n$ do not depend on the mixing, only window estimates vary between the independent and the $\phi$-mixing case.

---

## 0. INTRODUCTION

Soit $(X_n)_{n \geqslant 1}$, une suite équidistribuée de variables aléatoires $X_n = (U_n, V_n) \in \mathbb{R}^{d+e}$, nous considérons ici le comportement de statistiques $\gamma_n$ de la forme :

$$\gamma_n = a_n \int (\hat{\theta}_n - \theta)^2 d\mu - b_n,$$

où les suites $a_n$ et $b_n$ sont déterministes et à ajuster, la mesure $\mu$ est positive $\sigma$-finie et absolument continue par rapport à la mesure de Lebesgue, $dx$, et à support dans $M \subset \mathbb{R}^d$, $\theta$ et $\hat{\theta}_n$ désignent une fonction à préciser et un estimateur non paramétrique.

Le but de ce travail est de déterminer, $a_n$, $b_n$ et $\sigma^2$ de sorte que :

$$\mathscr{L}(\gamma_n) \underset{n \to \infty}{\longrightarrow} N(0, \sigma^2).$$

Nous traitons ici le cas de variables aléatoires indépendantes et le cas de variables aléatoires $\phi$-mélangeantes ([*22*]). Il est intéressant de noter que, dans le second cas, les constantes ne dépendent pas du mélange, mais que seules les fenêtres d'estimation changent.

Notant $f$, la densité de $U_1$ et $r(x) = E(V_1/U_1 = x)$ une version régulière de l'espérance conditionnelle de $V_1$ sachant $U_1$, nous considérons les cas :

$\theta = f$, $\theta = rf$ et $\theta = r$, $\hat{\theta}_n$ désigne alors une estimation non paramétrique de la fonction $\theta$ considérée qui se présente sous les formes :

$$\hat{f}_n(x) = \frac{1}{n} \sum_{k=1}^{n} K_n(x, U_k) \qquad \text{est une estimation de } f$$

$$\hat{g}_n(x) = \frac{1}{n} \sum_{k=1}^{n} V_k K_n(x, U_k) \quad \text{est une estimation de la fonction } g = rf$$

$$\hat{r}_n(x) = \begin{cases} 0 \quad \text{si} \quad \hat{f}_n(x) = 0 \\ \dfrac{\hat{g}_n(x)}{\hat{f}_n(x)} \quad \text{sinon} \end{cases}$$

Les fonctions $K_n$ envisagées correspondent à des estimateurs à noyaux et se présentent donc sous les deux formes :

1. *Cas des noyaux :* Soit K, une fonction continue et bornée telle que $\int K(z)dz = 1$, nous posons $K_n(x, y) = h_n^{-d}K((x - y)/h_n)$, où $h_n > 0$ vérifie $\lim_{n\to\infty} h_n = 0$ ([*16*]).

2. *Cas des projections :* Soit $\{ \psi_i ; i \geqslant 1 \}$, une base orthonormale de l'espace de Hilbert $L^2(M, d\mu)$, formée de fonctions continues, nous posons

$$K_n(x, y) = \rho(y) \sum_{i=1}^{q(n)} \psi_i(x)\psi_i(y), \text{ où}$$

$$q(n) \in \mathbb{N}, \quad \text{vérifie} \quad \lim_{n\to\infty} q(n) = +\infty \ ([2])$$

Nous appelons fenêtre de l'estimation la suite $h_n$ (resp. $q(n)$).

Dans le cas 1., le problème est traité par Rosenblatt dans le cas d'estimation de la densité d'une suite de variables indépendantes à l'aide d'une approximation poissonnienne [*18*], d'autre part Csörgo et Révész rassemblent dans ([*5*], chapitre 6) les résultats concernant ces problèmes dans le cas $d = e = 1$ en utilisant des approximations fortes du type de celle de Komlos Major et Tusnady. Une approche de ce type peut être trouvée dans [*10*], toutefois les fenêtres d'estimation y sont moins satisfaisantes du fait qu'aucun principe d'invariance n'admet une vitesse qui semble optimale dans le cas multidimensionnel et mélangeant ; c'est d'ailleurs la raison pour laquelle nous étudions, dans [*10*], plus particulièrement le risque uniforme inaccessible autrement que par des principes d'invariance. Nous utiliserons ici des approximations gaussiennes dans $L^2(M, d\mu)$ ([*14*] dans le cas indépendant, [*8*] et [*9*] dans le cas mélangeant) d'une manière analogue à León [*15*] dans le cas d'une estimation de densités pour des variables indépendantes, le traitement des processus gaussiens qui interviennent utilise le développement de Karhunen-Loeve [*5*].

Enfin, notre problème étant étroitement lié à celui de l'évaluation de de risques intégrés, de la forme $E\int(\hat{\theta}_n - \theta)^p d\mu$, citons les travaux de Collomb ([*3*]) qui traite ce problème dans le cas indépendant, Collomb et Doukhan ([*4*]) le traitent pour $p = 2$ dans le cas mélangeant et Doukhan et Portal ([*11*]) le considèrent lorsque $p$ est un entier pair. Enfin Doukhan ([*7*]) donne une étude par simulation du comportement du risque quadratique de ces estimateurs dans le cas particulier d'un processus autorégressif non linéaire d'ordre 1.

Faisons une première réduction du problème. Posons

$$\Gamma_n = a_n \int (\hat{\theta}_n - \mathrm{E}\,\hat{\theta}_n)^2 d\mu - b_n .$$

Nous remarquerons que dans le cas $\theta = f$ et $\theta = rf$, la convergence de $\Gamma_n$ entraîne celle de $\gamma_n$ sous l'hypothèse additionnelle $\lim_{n\to\infty} a_n \int (\mathrm{E}\,\hat{\theta}_n - \theta)^2 d\mu = 0$ qui est réalisée sous des conditions de différentiabilité sur $\theta$ et sous certaines hypothèses précisées plus loin sur $\mathrm{K}_n$. Le cas de $r_n$ est considéré d'une façon différente ; écrivant $\hat{r}_n - r = \hat{u}_n + \mathrm{R}_n$ avec $\hat{u}_n = \frac{1}{f}(\hat{g}_n - \hat{f}_n r)$ et $\mathrm{R}_n = \frac{(\hat{f}_n - f)}{f^2}((\hat{g}_n - g) + \hat{r}_n(f - \hat{f}_n))$, nous verrons que l'hypothèse

$$\lim_n a_n b_n \mathrm{E} \int \mathrm{R}_n^2 d\mu = 0$$

qui est réalisée sous des conditions de dimension et de différentiabilité entraîne que la convergence de $\Gamma'_n = a_n \int \hat{u}_n^2 d\mu - b_n$ implique celle de $\gamma'_n = a_n \int (\hat{r}_n - r)^2 d\mu - b_n$ vers une normale $\mathrm{N}(0, \Sigma^2)$.

Nous notons de plus $\| \psi \| = \left( \int | \psi |^2 d\mu \right)^{1/2}$, pour $\psi$ dans $\mathrm{L}^2(\mathrm{M}, d\mu)$ et $(.,.)$ désigne le produit scalaire associé de plus $| \,.\, |$ désigne la norme euclidienne de $\mathbb{R}^k$ ($k = d$ ou $e$) et si F est $l$ fois différentiable nous posons $\| \mathrm{D}^l \mathrm{F}(x) \| = \mathrm{Sup} \left\{ \frac{| \mathrm{D}^l \mathrm{F}(x)(h_1, \ldots, h_l) |}{| h_1 | \ldots | h_l |} ; h_1, \ldots, h_l \neq 0 \right\}$. Enfin nous faisons dans tout ce travail les hypothèses suivantes sur $\mu$ et sur la loi de la suite $(\mathrm{X}_n)_{n \geq 1}$ :

*i*) $\mu$ est absolument continue et s'écrit $\mu(dx) = \rho(x)dx$ avec $\| \rho \|_\infty < \infty$ et $\rho \geqslant 0$ ; M désigne le support de la loi $\mu$.

*ii*) Les couples $\mathrm{X}_1 = (\mathrm{U}_1, \mathrm{V}_1)$ et $(\mathrm{X}_1, \mathrm{X}_k)$ admettent pour tout $k > 1$, une densité par rapport à la mesure de Lebesgue de $\mathbb{R}^{d+e}$ (resp. $\mathbb{R}^{2d+2e}$) notée $f_1$ (resp. $f_k$).

Ainsi : $g(x) = r(x)f(x) = \int y f_1(x, y) dy$ et, si nous posons

$$g_s(x) = \int | y |^s f_1(x, y) dy, \qquad g_s(x) = \mathrm{E}( | \mathrm{V}_1 |^s / \mathrm{U}_1 = x) f(x) .$$

*Remarque.* — Dans le cas indépendant l'existence d'une densité pour le couple $(X_1, X_k)$ résulte de l'existence de la densité $f_1$ de $X_1$. Dans le cas non indépendant, nous faisons l'hypothèse que la suite $X_n$ est $\phi$-mélangeante (resp. $\Psi$-mélangeante), c'est-à-dire que $\lim_{n\to\infty} \phi_n = 0$ (resp. $\lim_{n\to\infty} \Psi_n = 0$) où, si $M_i^j$ désigne la tribu engendrée par $X_i, \ldots, X_j$ :

$$\phi_n = \operatorname{Sup}\{\,|\mathbb{P}(A\cap B)/\mathbb{P}(B) - \mathbb{P}(A)|\,;\ A\in M_0^i,\ B\in M_{i+n}^\infty,\ i\in\mathbb{N}\,\},$$
$$\Psi_n = \operatorname{Sup}\{\,|\mathbb{P}(A\cap B)/(\mathbb{P}(A)\mathbb{P}(B)) - 1|\,;\ A\in M_0^i,\ B\in M_{i+n}^\infty,\ i\in\mathbb{N}\,\}$$

## 1. ESTIMATEURS A NOYAUX

Nous faisons ici deux types d'hypothèses notées (H) et (H'), utilisées dans la suite, respectivement pour les estimateurs $\hat{f}_n$ et $\hat{g}_n$ et $\hat{r}_n$. Cette dernière estimation est en effet spécifique du fait qu'elle fait intervenir le rapport de deux estimateurs. Dans la suite $M_1$ désigne un voisinage strict de M :

HYPOTHÈSES (H).

*i)* K *est une fonction continue et à support compact et paire par rapport à chaque coordonnée vérifiant, si m est un entier fixé et k est défini par* $\frac{d}{2} < k \leqslant \frac{d}{2} + 1$ ($k = 2$ *lorsque* $d \leqslant 3$) *et* $m_1 = \operatorname{Max}\{k, m\}$ :

$\int p(z)K(z)dz = 0$ *pour tout monôme homogène de degré strictement inférieur à* $m_1$ *et, de plus* $\int K(z)dz = 1$.

*ii)* $f,\ g \in L^\infty(M_1, dx)$, $D^l f$, $D^l g \in L^2(M_1, d\mu)$, *si* $l = 0$ *ou* $m$ *et* $D^l f$, $D^l g_2 \in L^1(M_1, d\mu)$, $l = 0$ *ou* $k$.

*iii)* $E|V_1|^h < \infty$ *avec* $h = 3$ *dans le cas indépendant et* $h = 4 + \delta$ *dans le cas mélangeant, pour un* $0 < \delta \leqslant 1$.

HYPOTHÈSES (H') $d \leqslant 3$.

*i)* K *est une fonction continue positive, à support compact et paire par rapport à chaque coordonnée et, de plus* $\int K(z)dz = 1$.

*ii)* $\int \frac{d\mu}{f^4} < \infty$, *et les fonctions* $\frac{\rho}{f^2}$, $r$, $D^l f$, $D^l g$ *et* $D^l g_2$ *sont bornées sur* $M_1$ *si* $l = 0$ *ou* 2.

*iii)* $\exists \alpha > 0$, $Ee^{\alpha|V_1|} < \infty$.

*Remarques.*

. De tels noyaux K existent comme le montrent Bretagnolle et Huber ([*1*]), on peut en trouver des expressions explicites dans Gässer et Müller ([*12*]), une construction explicite peut être donnée sous la forme $P(t^2)\exp(-t^2/2)$.

. Les conditions $(H'-ii)$ peuvent être simplifiées en supposant $M_1$ compact et les fonctions qui interviennent bornées sur $M_1$, dans les hypothèses (H') ; il faut aussi supposer $f$ minorée sur M. Ces hypothèses sont celles habituellement utilisées dans ce cas.

## 1.1. **Étude du cas indépendant.**

Écrivons $\hat{g}_n(x) - \mathrm{E}\,\hat{g}_n(x) = \dfrac{1}{nh_n^d}\displaystyle\sum_{k=1}^{n} Y_{h_n,k}(x)$, où

$$Y_{b,k} = V_k K\left(\frac{x-U_k}{b}\right) - EV_k K\left(\frac{x-U_k}{b}\right).$$

Nous approximons $n^{-1/2}\displaystyle\sum_{k=1}^{n} Y_{b,k}$ par une variable aléatoire gaussienne $z_b$, sur $L^2(M, d\mu)$, centrée de la covariance $\Gamma_b(x, y) = EY_{b,1}(x)Y_{b,1}(y)$. Cette approximation basée sur le résultat de ([*14*]) sera faite de manière plus précise lorsque nous aurons mis en évidence le comportement asymptotique de $\| z_b \|_2^2$ qui fera apparaître de façon naturelle les suites $a_n$ et $b_n$ ainsi que la variance asymptotique $\sigma^2$ de $\Gamma_n$. Nous aurons besoin dans toute la suite de considérer avec Dehling ([*6*]) la norme de la trace d'un opérateur autoadjoint et compact de $L^2(M, d\mu)$, A définie par :

$$\| A \|_1 = \operatorname{Sup}\left\{ \sum_{i=1}^{\infty} |(Ae_i, e_i)| ; (e_i)_{i\geqslant 1} \text{ base orthonormale de } L^2(M, d\mu) \right\}.$$

Alors $\|A\|_1 = \displaystyle\sum_{i=1}^{+\infty} |\lambda_i|$ si les $\{\lambda_i ; i \geqslant 1\}$ désignent les valeurs propres de A, nous aurons de plus besoin du résultat classique suivant (Lemme XI.31, p. 65 [*17*]).

LEMME 1. — *Soit* A *un opérateur compact et autoadjoint de* $L^2(M, d\mu)$

*associé à un noyau intégral* $G(x, y)$ *symétrique et continu, alors si* A *est positif :*

$$\| A \|_1 = \int K(x, x)\mu(dx)$$

Évaluons à présent $E \| z_b \|^2 = E \| Y_{b,1} \|^2$, nous écrivons :

$$E \| z_b \|^2 = \alpha - \beta \qquad \text{avec} \qquad \alpha = \iint g_2(u) K^2\left(\frac{x-u}{b}\right) du \mu(dx)$$

et
$$\beta = \int \left| \int g(u) K\left(\frac{x-u}{b}\right) du \right|^2 \mu(dx).$$

Ainsi $\alpha = b^d \iint g_2'(x - bz) K^2(z) \mu(dx) dz$ et $\beta = b^{2d} \int \left| \int g(x - bz) K(z) dz \right|^2 \mu(dx)$.

La continuité des translations dans $L^2(M \times \mathbb{R}^d, d\mu \otimes dx)$ montre donc que $E \| z_b \|^2 = 0(b^d)$ ; pour poursuivre le développement de $E \| z_b \|^2$ nous faisons sous les hypothèses de différentiabilité (H) un développement limité à l'aide de la formule de Taylor avec reste intégral conduisant au :

LEMME 2. — *Supposons la fonction* $g_2$ *différentiable à l'ordre* $k$, *et* $D^k g_2 \in L^1(M, d\mu)$ *alors :*

$$E \| z_b \|^2 = b^d \int g_2 d\mu \int K^2(z) dz + \sum_{l=2}^{k-1} \frac{b^{d+l}}{l!} \iint D^l g_2(x)(z, \ldots, z) K^2(z) dz \mu(dx)/l! + 0(b^{d+k}),$$

*En particulier, sous les hypothèses* (H),

$$E \| z_b \|^2 = b^d \int g_2 d\mu \int K^2(z) dz + 0(b^{d+2})$$

Posons $t_b = b^{-3d/2}(\| z_b \|^2 - E \| z_b \|^2)$, le lemme 2 permet d'évaluer le terme de biais $b^{-3d/2} E \| z_b \|^2$, sous les hypothèses (H) si $d \leqslant 3$ et si $D^k g_2 \in L^1(M, d\mu)$ pour un $k > \frac{d}{2}$ sinon. Nous étudions à présent le comportement de $t_b$ lorsque $b$ tend vers 0. Suivant la méthode indiquée par Csörgo et Révész ( [*5* ], chapitre 6), nous diagonalisons la covariance $\Gamma_b$ de $z_b$ dans une base orthonormale de $L^2(M, d\mu)$, si $\lambda_{i,b}$ désigne la valeur propre associée au vecteur propre $e_{i,b}$, le développement de Karhunen-Loeve de $z_b$ s'écrit, si $\{ \gamma_i, i \geqslant 1 \}$ désigne une suite de variables N(0, 1) indépendantes :

$$z_b = \sum_{i=1}^{\infty} \sqrt{\lambda_{i,b}}\, e_{i,b} \gamma_i \qquad \text{et} \qquad t_b = b^{-3d/2} \sum_{i=1}^{\infty} \lambda_{i,b}(\gamma_i^2 - 1),$$

Alors, $\lim_{b \to 0} \mathscr{L}(t_b) = \mathrm{N}(0, \sigma^2)$ lorsque :

*a*) $2b^{-3d} \sum_{i=1}^{+\infty} \lambda_{i,b}^2 \underset{b \to 0}{\longrightarrow} \sigma^2$ (Convergence de la variance)

*b*) $\mathrm{Max} \{ \lambda_{i,b} ; i \geqslant 1 \}^2 b^{-3d} \underset{b \to 0}{\longrightarrow} 0$ (Condition de Lindeberg).

Le lemme 1 permet d'évaluer $\sum_{i=1}^{\infty} \lambda_{i,b}^2$ qui est la trace du carré de l'opérateur de covariance $\Gamma_b$ : $\sum_{i=1}^{\infty} \lambda_{i,b}^2 = \iint \Gamma_b^2(x, y)\mu(dx)\mu(dy)$.

Posons $\mathrm{R}_b(x, y) = \mathrm{E}\,|\,\mathrm{V}_1\,|^2 \mathrm{K}\left(\frac{x - \mathrm{U}_1}{b}\right)\mathrm{K}\left(\frac{y - \mathrm{U}_1}{b}\right)$ et $s_b(x) = \int g(x - bz)\mathrm{K}(z)dz$, nous avons $\Gamma_b(x, y) = \mathrm{R}_b(x, y) - b^{2d} s_b(x) s_b(y)$ et $\int s_b^2(x)\mu(dx) = 0(1)$ quand $b \to 0$.

Ainsi, montrer que $\sigma_b^2 = \iint \mathrm{R}_b^2(x, y)\mu(dx)\mu(dy) b^{-3d}$ converge quand $b \to 0$, entraînera la condition *a*.

$$\sigma_b^2 = b^{-d} \iint \mu(dx)\mu(dy)\left(\int g_2(x - bu)\mathrm{K}(u)\mathrm{K}\left(u + \frac{y - x}{b}\right)du\right)^2$$

$$\sigma_b^2 = \iint \rho(x + by)\rho(x)dxdy\left(\int g_2(x - bu)\mathrm{K}(u)\mathrm{K}(u + y)dy\right)^2$$

La continuité de la translation dans $\mathrm{L}^2(\mathbb{R}^k, dt)$ entraîne que :

$$\lim \sigma_b^2 = \int g_2^2(x)\rho^2(x)dx \int \mathrm{H}^2(y)dy \quad \text{où} \quad \mathrm{H}(y) = \int \mathrm{K}(u)\mathrm{K}(u + y)du\,.$$

Pour prouver le point *b*, nous remarquons que :

$$\mathrm{Max} \{ \lambda_{i,b} ; i \geqslant 1 \}^2 = \mathrm{Sup} \{ \| \Gamma_b \psi \|^2 ; \| \psi \| \leqslant 1 \}.$$

Le théorème de Banach-Steinhaus montre qu'il suffit de montrer que :

$$\forall \psi \in \mathrm{L}^2(\mathrm{M}, d\mu),\ \lim_{b \to 0} \| \Gamma_b \psi \|^2 b^{-3d} = 0\,.$$

De plus, un raisonnement analogue au précédent montre que l'on peut remplacer $\Gamma_b$ par $R_b$ :

$$\| R_b\psi \|^2 = \int \mu(dx)\left(\int R_b(x, y)\psi(y)\mu(dy)\right)^2$$

$$\| R_b\psi \|^2 = b^{4d}\int \mu(dx)\left(\iint g_2(x-ub)K(u)K(v)\rho(x+b(v-u))\psi(x+b(v-u))dudv\right)^2$$

L'inégalité de Schwarz dans $L^2(\mathbb{R}^{2d}, | K(u)K(v) | dudv)$ entraîne

$$\| R_b\psi \|^2 \leqslant b^{4d}\left[\iint g_2^2(x - ub) | K(u)K(v) | \rho^2(x + b(v - u))dudv \right.$$
$$\left. \iint \psi^2(x + b(\beta - \alpha)) | K(\alpha)K(\beta) | d\alpha d\beta \right] \mu(dx)$$

$$\| R_b\psi \|^2 \leqslant b^{4d} \| \psi \|^2 \| g \|_\infty^2 \| \rho \|_\infty^2 \left(\int | K(u) | du\right)^4 (1 + \varepsilon_b) \quad \text{où} \quad \lim_{b\to 0} \varepsilon_b = 0 .$$

Par conséquent les conditions $a$ et $b$ sont réalisées.

Ces résultats étant établis, l'approximation gaussienne de $\hat{g}_n - E\hat{g}_n$ devient claire ; nous considérons $\Delta_b$ défini à l'aide du lemme 2 par

$$b^{-3d/2}E \| z_b \|^2 = \Delta_b + 0(1) \quad \text{quand} \quad b \to 0$$

Nous posons :

$$G_{n,b}(t) = \mathbb{P}\left( \left\| n^{-1/2}b^{-3d/4} \sum_{n=1}^{n} Y_{b,k} \right\|^2 - \Delta_b \leqslant t \right)$$

$$G_b(t) = \mathbb{P}( \| b^{-3d/4}z_b \|^2 - \Delta_b \leqslant t )$$

Le résultat de [*14*] montre que si $v_b^3 = E \| b^{-3d/4}Y_{b,1} \|^3$ et $\varepsilon > 0$ :

$$| G_{n,b}(t) - G_b(t) | \leqslant v_b^3 n^{-1/2}\varepsilon^{-3} + \mathbb{P}( | \| b^{-3d/4}z_b \| - (t + \Delta_b)^{1/2} | < \varepsilon)$$

Or le lemme 2 montre, sachant que $\| g \|_\infty < \infty$, que $v_b = 0(b^{-d/4})$. Ainsi

$$\lim_{n\to\infty} | G_{n,h_n}(t) - G_{h_n}(t) | = 0 \quad \text{lorsque} \quad \lim_{n\to\infty} nh_n^{3d} = +\infty .$$

Ce calcul est analogue à celui de Léon ([*15*]), lorsque $V_n = 1$ p. s. il fournit un résultat concernant l'estimation d'une densité :

THÉORÈME 3. — *Sous les hypothèses* (H), *lorsque la suite* $X_n = (U_n, V_n)$

*est indépendante et équidistribuée nous avons* $\lim\limits_{n\to\infty} \mathscr{L}(\Gamma_n^0) = N(0, \sigma_0^2)$ *et* $\lim\limits_{n\to\infty} \mathscr{L}(\Gamma_n^1) = N(0, \sigma_1^2)$, *si* $\lim nh_n^{3d} = +\infty$, *où*

$$\Gamma_n^0 = nh_n^{d/2}\int |\hat{g}_n - E\hat{g}_n|^2 d\mu - h_n^{-d/2}\Delta_{h_n}^0$$

$$\Gamma_n^1 = nh_n^{d/2}\int (f_n - Ef_n)^2 d\mu - h_n^{-d/2}\Delta_{h_n}^1$$

*où*

$$\Delta_b^0 = \int f d\mu \int K^2(z)dz \quad si \quad d < 3 \quad et, sinon$$

$$\Delta_b^0 = \int f d\mu \int K^2(z)dz + \sum_{l=2}^{k-1} b^l \iint D^l f(x)(z, \ldots, z)K^2(z)dz\mu(dx)/l\,!$$

$$\Delta_b^1 = \int g_2 d\mu \int K^2(z)dz \quad si \quad d < 3 \quad et, sinon$$

$$\Delta_b^1 = \int g_2 d\mu \int K^2(z)dz + \sum_{l=2}^{k-1} b^l \iint D^l g_2(x)(z, \ldots, z)K^2(z)dz\mu(dx)/l\,!$$

*et* $\sigma_0^2 = 2\int f^2(x)\rho^2(x)dx \int H^2(z)dz$ *et* $\sigma_1^2 = 2\int g_2^2(x)\rho^2(x)dx \int H^2(z)dz$

*avec* $H(z) = \int K(u+z)K(u)du$.

COROLLAIRE 4. — *Si* $D^l f$, $D^l g \in L^2(M, d\mu)$ *pour* $l \leqslant m$, *et* K *vérifie l'hypothèse* (H) $\lim \mathscr{L}(\gamma_n^0) = N(0, \sigma_0^2)$ *et* $\lim \mathscr{L}(\gamma_n^1) = N(0, \sigma_1^2)$, *lorsque* $\lim nh_n^{d/(2+2m)} = 0$ *et* $\lim nh_n^{3d} = +\infty$ *où* $\gamma_n^1 = nh_n^{d/2}\int |\hat{g}_n - g|^2 d\mu - h_n^{-d/2}\Delta_{h_n}^1$, $\gamma_n^0 = nh_n^{d/2}\int (\hat{f}_n - f)^2 d\mu - h_n^{-d/2}\Delta_{h_n}^0$.

*Démonstration.* — Comme on l'a déjà vu il suffit de montrer le résultat pour $\hat{g}_n : \gamma_n^1 = \Gamma_n^1 + 2nh_n^{d/2}\int (\hat{g}_n - E\hat{g}_n)(E\hat{g}_n - g)d\mu + nh_n^{d/2}\int |E\hat{g}_n - g|^2 d\mu$, où $\int |E\hat{g}_n - g|^2 d\mu = 0(h_n^{2m})$, le terme correspondant tend vers 0 si $\lim\limits_{n\to\infty} nh_n^{d/(2+2m)} = 0$

$$\alpha = E\left|\int (\hat{g}_n - E\hat{g}_n)(E\hat{g}_n - g)d\mu\right|$$

$$\leqslant \left|E\left|\int (\hat{g}_n(x) - E\hat{g}_n(x))(E\hat{g}_n(x) - g(x))\mu(dx)\right|^2\right|^{1/2}$$

$$\alpha \leqslant E\iint \mathrm{Cov}(\hat{g}_n(x), \hat{g}_n(y))(\bar{g}_n(x) - g(x))(\bar{g}_n(y) - g(y))\mu(dx)\mu(dy)$$

$$\text{Or} \quad \operatorname{Cov}(\hat{g}_n(x), \hat{g}_n(y)) = \frac{1}{nh_n^{2d}}\left\{ \mathrm{E}\,|\,\mathrm{V}_1\,|^2 \mathrm{K}\left(\frac{x-\mathrm{U}_1}{h_n}\right)\mathrm{K}\left(\frac{y-\mathrm{U}_1}{h_n}\right) - \mathrm{EV}_1\mathrm{K}\left(\frac{x-\mathrm{U}_1}{h_n}\right)\mathrm{EV}_1\mathrm{K}\left(\frac{y-\mathrm{U}_1}{h_n}\right)\right\}$$

$$\operatorname{Cov}(\hat{g}_n(x), \hat{g}_n(y)) \leqslant \frac{1}{nh_n^{2d}}\left\{ \mathrm{E}\,|\mathrm{V}_1|^2\mathrm{K}^2\left(\frac{x-\mathrm{U}_1}{k_n}\right)\mathrm{E}\,|\mathrm{V}_1|^2\mathrm{K}^2\left(\frac{y-\mathrm{U}_1}{h_n}\right)\right\}^{1/2}$$

Par conséquent $\lim\limits_{n\to\infty} nh_n^{2d} = 0$ entraîne $nh_n^{d/2}\alpha \underset{n\to\infty}{\longrightarrow} 0$ en probabilité.

*Remarque.* — Si $d > 3$, la forme du résultat est peu manipulable car elle implique la connaissance d'une fonctionnelle de $\mathrm{D}^l f$ ou de $\mathrm{D}^l g_2$ pour tout $l$ tel que : $2 \leqslant l \leqslant \dfrac{d}{2}$.

Les statistiques $\gamma_n^0$ et $\gamma_n^1$ convergent si $h_n = n^{-\alpha}$, lorsque $\dfrac{2}{d+4m} < \alpha < \dfrac{1}{3d}$ ; ainsi $m > \dfrac{5d}{4}$ est la condition qui assure l'existence d'une fenêtre de convergence. Pour $d = 1$ (resp. $d = 2, 3, 4$ ou $5$) l'ordre minimal de différentiabilité est donc 2 (resp. 3, 4, 6, 7). Toutefois les fenêtres ne sont pas optimales vis-à-vis du risque quadratique intégré (cela se traduirait par $\alpha = \dfrac{1}{d+2m}$ or $\left.\dfrac{1}{d+2m} < \dfrac{2}{d+4m}\right)$, ce qui est le cas pour tous les résultats de ce type (cf. Csörgo et Revesz [5]).

Traitons à présent le cas de l'estimateur $\hat{r}_n$ de $r$. Pour commencer étudions le comportement de $\Gamma'_n = nh_n^{d/2}\displaystyle\int \hat{u}_n^2 - h_n^{-d/2}\Delta_n$ où $\hat{u}_n = \dfrac{1}{f}(\hat{g}_n - r\hat{f}_n)$ et $\Delta_n$ est une suite à déterminer. La méthode déjà employée est utilisée à nouveau, nous ferons toutefois dès à présent l'hypothèse que $\mathrm{K} \geqslant 0$ ce qui entraîne $d < 3$ étant données les hypothèses (H), les conditions de différentiabilité seront alors toutes d'ordre 2, en effet si K n'était pas positif il serait exclu d'évaluer des moments de $\hat{r}_n$.

Écrivant $\hat{u}_n(x) = \dfrac{1}{nf(x)h_n^d}\displaystyle\sum_{k=1}^{n} \Lambda_{k_n,k}(x)$ où $\Lambda_{b,k}(x) = (\mathrm{V}_k - r(x))\mathrm{K}\left(\dfrac{x-\mathrm{U}_k}{b}\right)$, nous considérons le gaussien $y_b$, associé à $\sqrt{nh_n^d}\,[\hat{u}_n(x) - \mathrm{E}\hat{u}_n(x)]$; centré et de covariance $\Gamma'_b : \Gamma'_b(x, y) = \mathrm{E}\Lambda_{b,1}(x)\Lambda_{b,1}(y) - \mathrm{E}\Lambda_{b,1}(x)\mathrm{E}\Lambda_{b,1}(y)$. Alors $\mathrm{E}\,\|\,y_b\,\|^2 = \alpha - \beta$, où $\alpha = \displaystyle\int \mathrm{E}\,|\,\mathrm{V}_1 - r(x)\,|^2\mathrm{K}^2\left(\frac{x-\mathrm{U}_1}{b}\right)\frac{\mu(dx)}{f^2(x)}$, et

$$\beta = \int \left|\mathrm{E}(\mathrm{V}_1 - r(x))\mathrm{K}\left(\frac{x-\mathrm{U}_1}{b}\right)\right|^2 \frac{\mu(dx)}{f^2(x)}.$$

Nous obtenons alors $E\|y_b\|^2 = b^d \int K^2(z)dz \int \frac{g_2(x) - r^2(x)f(x)}{f^2(x)} + 0(b^{d+2})$ lorsque $\frac{1}{f^2} D^l g_2, \frac{r}{f^2} D^l g, \frac{r^2}{f^2} D^l f \in L^1(M, d\mu)$ pour $l = 0, 1, 2$.

Le comportement du gaussien sera normal si :

*a)* $\lim\limits_{b\to\infty} 2 \int \Gamma_b'^2(x, y)\mu(dx)\mu(dy)b^{-3d} = \Sigma^2$

*b)* $\forall\psi \in L^2(M, d\mu), \lim\limits_{b\to 0} b^{-3d} \| \Gamma\psi \|^2 = 0$.

Soit $R_b'(x, y) = E\Lambda_{b,1}(x)\Lambda_{b,1}(y)$, $r_b'(x, y) = E\Lambda_{b,1}(x)E\Lambda_{b,1}(y)$, le fait que $\iint r_b'^2 d\mu^{\otimes 2} b^{-3d} \underset{b\to 0}{\longrightarrow} 0$ entraîne que l'on doit montrer que $2\Sigma_b^2 \to \Sigma^2$ où

$$\Sigma_b^2 = \iint R_b'^2(x, y)\mu(dx)\mu(dy)b^{-3d}$$

$$\Sigma_b^2 = b^{-3d} \iint \frac{\mu(dx)\mu(dy)}{f^2(x)f^2(y)}$$

$$\left[\iint \left[g_2(u) - (r(x) + r(y))g(u) + r(x)r(y)f(u)\right] K\left(\frac{x-u}{b}\right) K\left(\frac{y-u}{b}\right) du\right]^2$$

Posons $h(x) = \frac{\rho(x)}{f^2(x)}$, nous obtenons :

$$\Sigma_b^2 = \iint h(x+bt)h(x)dxdt \left(\int (g_2(x-bz) - (r(x) + r(x+bt))g(x-bz) + r(x)r(x+bt)f(x-bz)K(z)K(z+t)dz\right)^2.$$

La continuité de la translation dans $L^2$ montre ainsi :

$$\lim_{b\to 0} \Sigma_b^2 = \int (g_2(x) - r^2(x)f(x))^2 \frac{\rho^2(x)}{f^4(x)} dx \int H^2(t)dt = \frac{1}{2}\Sigma^2.$$

De même que précédemment d'autre part :

$$\| R_b\psi \|^2 = b^{4d} \int \frac{\mu(dx)}{f^2(x)}$$

$$\left(\iint \left[g_2(x-bu) - (r(x) + r(x+b(v-u))g(x-bu) + r(x)r(x+b(v-u))f(x-bu)\right] K(u)K(v)\psi(x + b(v - u))h(x + b(v - u))dudy\right)^2$$

Ainsi $\left\|\frac{\rho}{f^2}\right\|_\infty < \infty$, $\| g_2 \|_\infty + \| r \|_\infty + \| f \|_\infty < \infty$ entraîne le résultat.

THÉORÈME 5. — *Sous les hypothèses* (H'), *nous avons :*

$\lim \Gamma'_n = N(0, \Sigma^2)$ *si* $\lim nh_n^{3d} = \infty$ *où*

$$\Gamma'_n = nh_n^{d/2} \int | h_n |^2 d\mu - h_n^{-d/2}\Delta \quad avec \quad \Delta = \int \frac{g_2 - r^2 f}{f^2} d\mu \int K^2(z)dz$$

$$et \quad \Sigma^2 = 2 \int (g_2 - r^2 f)^2 \frac{\rho^2}{f^4} dx \int H^2(t)dt, \qquad H(t) = \int K(t)K(t+z)dz.$$

*Remarque.* — Dans le cas particulier où $V_1 = r(U_1) + W_1$ avec $W_1$ indépendante de $U_1$, les constantes $\Delta$ et $\Sigma^2$ prennent la forme :

$$\Delta = w^2 \int \frac{d\mu}{f} \quad \text{et} \quad \Sigma^2 = w^4 \int \frac{\rho^2(x)}{f^2(x)} dx \int H^2(t)dt \qquad \text{où} \qquad w^2 = E\,|\,W_1\,|^2$$

*Démonstration.* — Le terme de centrage $c_b = \|\, E\Lambda_{b,1} \,\|^2$ contenu dans le gaussien

$$c_b = \int \mu(dx) \left[ \int \left[ g(u) - r(x)f(u) \right] K\left(\frac{x-u}{b}\right) du \right]^2$$

peut en outre être évalué à l'aide de la formule de Taylor :

$$c_b \leqslant b^{2d+4} \int | z |^2 | K(z) | dz (\| r \|_\infty^2 \| D^2 f \|_2^2 + \| D^2 g \|_2^2)$$

Ainsi $b^{-3d/2}(\|z_b\|^2 - E\|z_n\|^2) = b^{-3d/2}(\|z_b\|^2 + c_b - E\|z_b\|^2) + o_P(1)$ quand $b \to 0$.

Les calculs précédents permettent d'obtenir la convergence de

$$nh_n^{d/2} \int | h_n - Eh_n |^2 d\mu - h_n^{-d/2}\Delta.$$

Cette dernière évaluation permet de conclure au théorème 5.

Pour obtenir à la convergence de la fonctionnelle

$$\gamma'_n = nh_n^{d/2} \int | \hat{r}_n - r |^2 d\mu - h_n^{-d/2}\Delta,$$

il reste à montrer que $\lim\limits_{n\to\infty} nE \int | R_n(x) |^2 \mu(dx) = 0$, ce qui sera le cas si

$$i) \lim_{n\to\infty} nE \int (\hat{f}_n - f) | \hat{g}_n - g |^2 \frac{d\mu}{f^4} = 0$$

$$ii) \lim_{n\to\infty} nE \int (\hat{f}_n - f) | \hat{r}_n |^2 \frac{d\mu}{f^4} = 0$$

Afin d'éviter de recommencer plus loin le même calcul considérons dès à présent des variables dépendantes.

Pour évaluer $i$), nous utilisons l'inégalité de Schwarz :

$$\mathrm{E}\int(\hat{f}_n-f)^2\mid \hat{g}_n-g\mid^2\frac{d\mu}{f^4}\leqslant\left|\mathrm{E}\int(\hat{f}_n-f)^4\frac{d\mu}{f^4}\mathrm{E}\int\mid \hat{g}_n-g\mid^4\frac{d\mu}{f^4}\right|^{1/2}$$

$$\leqslant \mathrm{C}\left[\frac{1}{n^2h_n^{2d}}+h_n^8\right],\ \text{d'après [4] si}\ \sum_{n=1}^{\infty}n^2\phi_n^{1/2}<\infty$$

Le second terme utilise l'inégalité de Hölder :

$$\int\mathrm{E}(\hat{f}_n-f)^4\mid \hat{r}_n\mid^2\frac{d\mu}{f^4}\leqslant\left(\mathrm{E}\int(\hat{f}_n-f)^6\frac{d\mu}{f^4}\right)^{2/3}\left(\mathrm{E}\int\mid \hat{r}_n\mid^6\frac{d\mu}{f^4}\right)^{1/3}$$

$$\leqslant \mathrm{C}\left(\frac{1}{n^2h_n^2d}+h_n^8\right)(\log n)^2$$

Car $\mid \hat{r}_n\mid\leqslant \mathrm{Max}\{\mid \mathrm{V}_1\mid,\dots,\mid \mathrm{V}_n\mid\}$ et $\mathrm{E}e^{\alpha|\mathrm{V}_1|}<\infty$ entraînent $\mathrm{E}\mid \hat{r}_n\mid^6\leqslant \mathrm{C}(\log n)^6$.

Ainsi $\gamma_n'$ converge si, $\lim\limits_{n\to\infty}\dfrac{nh_n^{2d}}{(\log n)^2}=\infty$, $\lim nh_n^8(\log n)^2=0$.

THÉORÈME 6. — $\lim \mathscr{L}(\gamma_n')=\mathrm{N}(0,\Sigma^2)$ *sous les hypothèses* (H') *lorsque* $h_n=n^{-\alpha}$, $\dfrac{1}{8}<\alpha<\dfrac{1}{3d}$.

*Remarque.* — $d=1$ ou $d=2$.

### 1.2. Cas mélangeant.

Nous considérons de manière naturelle, dans ce paragraphe, le processus gaussien $\mathrm{Z}_b$ sur $\mathrm{L}^2(\mathrm{M}, d\mu)$ de covariance $\Gamma_b^{\phi}(x,y)=\Gamma_b(x,y)+\tau_b(x,y)+\tau_b(y,x)$ avec $\tau_b(x,y)=\sum_{k=2}^{\infty}\mathrm{E}\mathrm{Y}_{b,1}(x)\mathrm{Y}_{b,k}(y)$, associé au cas mélangeant.

Nous évaluons les mêmes expressions que dans le cas indépendant :

$$\mathrm{E}\parallel \mathrm{Z}_b\parallel^2=\int\Gamma_b^{\phi}(x,x)\mu(dx)=\mathrm{E}\parallel z_b\parallel^2+2\int\tau_b(x,x)\mu(dx).$$

Pour évaluer le second terme nous aurons besoin de l'une des hypothèses de mélange suivantes :

($\mathrm{M}_1$) $\left\{\begin{array}{l}(\mathrm{X}_n)_{n\geqslant 1}\ \text{suit est une suite}\ \psi\text{-mélangeante, telle que}\ \sum_{k=1}^{\infty}\psi_k<\infty\\ \text{(cf. [22]).}\end{array}\right.$

Nous en déduisons l'inégalité $|f_k - f_1^{\otimes 2}| \leqslant \psi_k f_1^{\otimes 2}$

$$(\mathrm{M}_2)\ \left\{\begin{array}{l}(\mathrm{X}_n)_{n\geqslant 1} \text{ est une suite } \phi\text{-mélangeante telle que :}\\ f_k(u_1, v_1, u_2, v_2) \leqslant \mathrm{C} f_1(u_1, v_1) f_1(u_2, v_2) \text{ pour une constante C indé-}\\ \text{pendante de } k \text{ et } (u_1, u_2) \in \mathrm{M}_1^2,\ (v_1, v_2) \in \mathbb{R}^{2l} \text{ et il existe } 0 < s < \dfrac{1}{4}\\ \text{tel que } \Sigma\phi_n^s < \infty.\end{array}\right.$$

$$(\mathrm{M}_3)\ \left\{\begin{array}{l}(\mathrm{X}_n)_{n\geqslant 1} \text{ est une suite } \phi\text{-mélangeante, } \mu(\mathrm{M}) < \infty \text{ et il existe}\\ s \in \left]0, \dfrac{1}{2}\right[ \text{ tel que } \Sigma\phi_k^{s/2}(c_k^{1-s} + 1) < \infty \text{ avec}\\ c_k = \operatorname{Sup}\left\{ \displaystyle\iint \left[ |v_1|^2 + |v_2|^2 + 1 \right] f_k(u_1, v_1, u_2, v_2) dv_1 dv_2 \,;\right.\\ \hfill \left.(u_1, u_2) \in \mathrm{M}_1^2 \right\}.\end{array}\right.$$

Le calcul peut donc être poursuivi,

$$s_k = \int |\mathrm{S}_{b,k}(x, x)| \mu(dx) = b^{2d} \iiiint\!\!\int |v_1 v_2| \, |\mathrm{K}(u_1)\mathrm{K}(u_2)|$$
$$|f_k(x - u_1 b, v_1, x - u_2 b, v_2) - f_1(x - u_1 b, v_1) f_1(x - u_2 b, v_2)| \mu(dx) du_1 du_2 dv_1 dv_2$$

l'hypothèse $(\mathrm{M}_1)$, $s_k \leqslant b^{2d}\psi_{k-1} \left| \displaystyle\int g_1(x - ub) |\mathrm{K}(u)| du\mu(dx) \right|^2$.

Sous l'hypothèse de $\phi$-mélange $s_k < 2\phi_{k-1}^{1/2}\, b^d \displaystyle\int g_2(x - bu)\mathrm{K}^2(u) du\mu(dx)$ par l'inégalité de mélange. Les hypothèses $(\mathrm{M}_2)$ et $(\mathrm{M}_3)$ conduisent à une autre évaluation

$$s_k \leqslant b^{2d}(\mathrm{C} + 1)\left(\int g_1(x - ub) |\mathrm{K}(z)| dz\right)^2 \quad \text{sous l'hypothèse } (\mathrm{M}_2)$$

$$s_k \leqslant b^{2b}\left[ \mu(\mathrm{M})\left(\int |\mathrm{K}(u)| du\right)^2 c_k + \left(\int g_1(x - bu) |\mathrm{K}(u)| du\mu(dx)\right)^2 \right]$$
sous l'hypothèse $(\mathrm{M}_3)$ ;

ainsi,

$$s_k \leqslant b^{2d}\left[ \mu(\mathrm{M})\left[ \left(\int |\mathrm{K}(u)| du\right)^2 c_k + \int g_2(x - bu)\mathrm{K}^2(u) du\mu(dx) \right]\right].$$

Sous l'hypothèse $(\mathrm{M}_1)$, $\displaystyle\int \tau_b(x, x)\mu(dx) = 0(b^{2d}) \sum_{k=1}^{\infty} \psi_k$.

Sous l'hypothèse $(M_2)$ $\left|\int \tau_b(x, x)\mu(dx)\right| = \Sigma \phi_k^{s/2} 0(b^{d(2-s)})$ et sous l'hypothèse $(M_3)$ $\left|\int \tau_b(x, x)\mu(dx)\right| = \sum_{k=2}^{\infty} \phi_{k-1}^{s/2}[c_k^{1-s} + 1] 0(b^{d(2-s)})$.

LEMME 7. — *Supposons les hypothèses* (H) *et* $(M_i)$ (*pour* $i = 1, 2$ *ou* 3) *réalisées :*

$$E \| Z_b \|^2 = b^d \int g_2 d\mu \int K^2(z)dz + \sum_{k=2}^{k=1} \iint \frac{b^{d+l}}{l!} D^l g_2(x)(z, \ldots, z) K^2(z) dz \mu(dx) + 0(b^{3d/2})$$

*quand* $b \to 0$ *en particulier, lorsque* $d \leqslant 3$,

$$E \| Z_b \|^2 = b^d \int g_2 d\mu \int K^2(z)dz + 0(b^{3d/2}), \qquad \text{quand} \quad b \to 0.$$

Posons $T_b = b^{-3d/2}(\| Z_b \|^2 - E \| Z_b \|^2)$, nous aurons montré que $\lim_{b\to 0} \mathscr{L}(T_b) = N(0, \sigma^2)$ où $\sigma^2$ a été défini dans le théorème 3 quand nous aurons vérifié :

*a*) $$\lim 2b^{-3d} \iint (\Gamma_b^\phi(x, y))^2 \mu(dx)\mu(dy) = \sigma^2$$

*b*) $$\forall \psi \in L^2(M, d\mu), \lim_{b\to 0} \| \Gamma_b^\phi \psi \|^2 b^{-3d} = 0$$

Nous remarquons qu'il suffit de montrer, en utilisant l'inégalité de Schwarz, que :

$$\lim_{b\to 0} \iint \tau_b^2(x, y)\mu(dx)\mu(dy) b^{-3d} = 0$$

or $$\beta_b = \iint \tau_b^2(x, y)\mu(dx)\mu(dy) = \iint [\Sigma a_k(x, y)]^2 \mu(dx)\mu(dy), \quad \text{où}$$

$$a_k(x, y) = E \mid Y_{b,1}(x) Y_{b,k}(y) \mid$$

$$a_k(x, y) \leqslant 2\phi_k^{1/2} \mid E Y_{b,1}^2(x) E Y_{b,1}^2(y) \mid^{1/2} \quad \text{par l'inégalité du mélange}$$

$$a_k(x, y) \leqslant b^{2d} \iint \mid K(u_1)K(u_2) \mid t_k(x - u_1 b, y - u_2 b) du_1 du_2 \quad \text{avec}$$

$$t_k(x, y) = \iint \mid v_1 \mid \mid v_2 \mid \mid f_k(x_1, v_1, y, v_2) - f_1(x, v_1) f_1(y, v_2) \mid dv_1 dv_2 .$$

Sous $(M_1)$, $| t_k(x, y) | \leqslant \psi_k g_1(x) g_1(y)$; sous $(M_2)$, $| t_k(x, y) | < (C+1) g_1(x) g_1(y)$ et, sous l'hypothèse $(M_3)$, $| t_k(x, y) | \leqslant 2c_k + \| g_2 \|_\infty$ si $(x, y) \in M_1^2$ et

$a_k(x, y) \leqslant b^{2d}[2c_k + \| g_2 \|_\infty^2]\left(\int | K(u) | \, du\right)^2$, des inégalités parallèles suivent pour les hypothèses $(M_1)$ et $(M_2)$. Soient $s, t \geqslant 0$ $s + t = 1$

$$\beta_b \leqslant b^{4d}\left(\sum_{k=2}^{\infty} (2c_k + \| g_2 \|_\infty^2)^t \phi_{k=1}^{s/2}\right)^2 \left(\int [EY_{b,1}^2(x)]^s \mu(dx)\right)^2 \mu(M)^{2t} = 0(b^{2d(2-s)})$$

sous l'hypothèse $(M_3)$ (de même $\beta_b = 0(b^{2d(2-s)})$ sous $(M_2)$ et $\beta_b = 0(b^{4d})$ sous $(M_1)$).

La condition $(b)$ résulte alors de l'inégalité :

$$\int \mu(dx)\left[\int \tau_b(x, y)\psi(y)\mu(dy)\right]^2 \leqslant \iint \tau_b^2(x, y)\mu(dx)\mu(dy) \int \psi^2(y)\mu(dy),$$

donc le comportement de $T_b$ est le même que celui associé au cas indépendant $t_b$.

Nous pouvons à présent effectuer l'approximation gaussienne, posons :

$$G_{n,b}(t) = \mathbb{P}\left(\left\| n^{-1/2} b^{-3d/4} \sum_{k=1}^{n} Y_{b,k} \right\|^2 - \Delta_b \leqslant t\right)$$
$$G_b(t) = \mathbb{P}(\| b^{-3/4} Z_b \|^2 - \Delta_b \leqslant t)$$

Le résultat de [9] montre si $v_b^{4+\delta} = E \| b^{-3d/4} Y_{b,1} \|^{4+\delta}$ et $\varepsilon > 0$ :

$$| G_{n,b}(t) - G_b(t) | \leqslant \frac{v_b}{\varepsilon n^{(1-b)/6}} + \frac{v_b^2}{\varepsilon^2 n^{2(1-b)/3}} + \frac{v_b^3}{\varepsilon^3 n^{(1-b)/3}} + \mathbb{P}(| \| b^{-3d/4} Z_b \| - (t + \Delta_b)^{\frac{1}{2}} | < \varepsilon)$$

sous les hypothèses de mélange :

$$\sum_{n=1}^{\infty} n^2 \phi_n^{\delta/(4+\delta)} < \infty, \qquad \phi_n = 0(n^{-3/4(1/b-1)}) \qquad \text{quand} \quad n \to \infty .$$

De plus $v_b \leqslant C b^{-d/4}$ d'après le lemme 7, sachant que $E | V_1 |^{4+\delta} < \infty$. Ainsi $\lim_{n\to\infty} | G_{n,h_n}(t) - G_{h_n}(t) | = 0$ lorsque $\lim n^{1-b} h_n^{9d/2} = \infty$.

THÉORÈME 8. — *Sous les hypothèses* (H), *en utilisant les notations du théorème* 3 : $\lim_{n\to\infty} \mathscr{L}(\Gamma_n^0) = N(0, \sigma_0^2)$, $\lim_{n\to\infty} \mathscr{L}(\Gamma_n^1) = N(0, \sigma_1^2)$ lorsque $\lim n^{1-b} h_n^{9d/2} = \infty$

*et sous les hypothèses de mélange* $(M_i)$ $(i=1, 2$ *ou* $3)$, *lorsque de plus :*

$$b \in \left]0, \frac{1}{4}\right[, \ \delta \in [0, 1] : \sum_{n=1}^{\infty} n^2 \phi_n^{\delta/(4+\delta)} < \infty \qquad et \qquad \phi_n = 0\left(n^{-\frac{3}{4}\left(\frac{1}{b}-1\right)}\right)$$

*quand* $n \to \infty$.

*Lorsque, de plus*

$$\lim_{n\to\infty} nh_n^{d/2+2m} = 0, \qquad \lim_{n\to\infty} \mathscr{L}(\gamma_n^0) = N(0, \sigma_0^2) \qquad et \qquad \lim_{n\to\infty} \mathscr{L}(\gamma_n^1) = N(0, \sigma_1^2).$$

*Remarque.* — L'hypothèse de différentiabilité assurant l'existence d'une fenêtre de convergence est $m > d\dfrac{7+2b}{4(1-b)}$, c'est-à-dire $m > \dfrac{7d}{4}$ sous des hypothèses de mélange complémentaires (c'est-à-dire que $d=1, 2, 3, 4$ ou $5$ donne lieu à $m \geqslant 2, 4, 6, 8$ ou $9$).

Pour terminer la démonstration du théorème, il reste à remarquer que l'évaluation de $\mathrm{Cov}^2(\hat{g}_n(x), \hat{g}_n(y))$ faite dans le cas indépendant s'étend dans le cas mélangeant en utilisant l'inégalité de mélange, sous l'hypothèse $\sum_{n=1}^{\infty} n\phi_n^{1/2} < \infty$.

Traitons à présent le cas de la régression pour $d \leqslant 3$ ; il apparaît naturel de manière analogue à ce qui précède de considérer le noyau non positif $\theta_b$ défini par :

$$\theta_b(x, y) = \sum_{k=2}^{\infty} \{ \zeta_{b,k}(x, y) + \zeta_{b,k}(y, x) \}$$

où
$$\zeta_{b,k} = E\Lambda_{b,k}(x)\Lambda_{b,k}(y) - E\Lambda_{b,1}(x)E\Lambda_{b,k}(y)$$

et $\Lambda_{b,k}$ est la fonction définie en fin de paragraphe 1.1.

Nous montrons, de la même manière que dans le cas de $\hat{g}_n$, que

$$E \| Y_b \|^2 = b^d\Delta + 0(b^{d+2})$$

où
$$\Delta = \int K^2(z)dz \int (g_2(x) - r^2(x)f(x)) \frac{\mu(dx)}{f^2(x)}$$

si $Y_b$ est le gaussien centré sur $L^2(M, d\mu)$ de covariance $\Gamma_b'^{\phi}$, $\Gamma_b'^{\phi} = \Gamma_b' + \theta_b$ ; $Y_b$ est le gaussien associé à $\sqrt{nh_n^d}[h_n(x) - Eh_n(x)]$ dans le cas mélangeant. L'hypothèse de mélange requise est encore ici $(M_1)$, $(M_2)$ ou $(M_3)$.

*Remarques.* — L'hypothèse $(M_3)$ peut être allégée en supprimant $\mu(M) < \infty$, en effet dans le cas de la régression la mesure de base $\mu(dx)$ est remplacée par $\frac{\mu(dx)}{f^2(x)}$ qui est de masse finie sous les hypothèses (H').

L'hypothèse ($M_2$) peut être remplacée par l'existence de deux fonctions $h_1$ et $h_2$ telles que $\int(|u|^2 + 1)(h_1(x) + h_2(x))dx < \infty$ si $x = (u, v)$, $f_k(u_1, v_1, u_2, v_2) \leqslant h_1(u_1, v_1)h_2(u_2, v_2)$, $\forall(u_1, u_2) \in M^2$, $\forall v_1, v_2$.

Nous obtenons ainsi le :

THÉORÈME 9. — *Sous les hypothèses* (H') *et* ($M_i$) ($i = 1, 2$ *ou* 3), *nous avons :* $\lim\limits_{n\to\infty} \mathscr{L}(\Gamma'_n) = N(0, \Sigma^2)$ *lorsque* $\lim n^{(1-b)}h_n^{9d/2} = \infty$ *sous les hypothèses de mélange* $0 < b < \frac{1}{4}$ *et* $\phi_n = 0(n^{-3/4(1/b-1)})$ *et* $\exists\delta \in\, ]0, 1]$, $\sum_{n=1}^{\infty} n^2 \phi_n^{\delta(4+\delta)} < \infty$, *les notations sont celles du théorème 5. De plus si* $\lim nh_n^{\delta}(\log n)^2 = 0$ ; $\lim \mathscr{L}(\gamma'_n) = N(0, \Sigma^2)$.

*Remarques.* — Dans ce cadre seule la condition $d = 1$ assure l'existence d'une fenêtre de convergence pour l'estimateur $\hat{r}_n$ de $r$.

Par contre, ce qui concerne la statistique $\Gamma'_n$ peut être étendu à une dimension quelconque $h_n^{-d/2}\Delta$ étant dans ce cas remplacé comme c'était le cas dans le lemme 2 par un développement limité obtenu à l'aide de la formule de Taylor à l'ordre $k\,(d/2 < k \leq d/2 + 1)$ de l'expression $b^{-3d/2}E\,\|\,Y_b\,\|^2$. Il faut pour cela adjoindre des hypothèses additionnelles concernant les dérivées d'ordre 4 de $f$, $r$ et $g_2$.

## 2. ESTIMATEURS DE PROJECTION

Nous énonçons d'abord les hypothèses utilisées dans la suite. Dans le cas indépendant nous supposerons :

*i)* $f, g \in L^2(M, d\mu)$

*ii)* $\|\,f\rho\,\|_\infty < \infty$ (pour l'estimation de $f$)

*iii)* $\|\,g_3\rho^2\,\|_\infty < \infty$ et $\|\,g_2\rho\,\|_\infty < \infty$ (pour l'estimation de $g$).

Dans le cas $\phi$-mélangeant nous supposons, de plus l'existence de $0 < b < \frac{1}{4}$ et $0 < \delta \leq 1$ tels que :

*i)* $\|\,g_{4+\delta}\rho^{3+\delta}\,\|_\infty < \infty$

$$ii) \quad \sum_{n=1}^{\infty} n^2 \phi_n^{\delta/(4+\delta)} < \infty, \qquad \phi_n = 0\left(n^{-\frac{3}{4}\left(\frac{1}{b}-1\right)}\right) \quad \text{quand} \quad n \to \infty$$

$$iii) \lim_{q \to \infty} q^{-1/2} \sum_{i=1}^{q} c_{ii}^k = 0, \qquad \text{si} \qquad c_{ii}^k = \iint t_k(x, y)\psi_i(x)\psi_i(y)\mu(dx)\mu(dy)$$

avec $$t_k(x, y) = \iint u \, . \, v(f_k(x, u, y, v) - f_1(x, u) f_1(y, v)) du dv \, .$$

Nous supposons de plus réalisée l'une des hypothèses du mélange suivantes :

($M_1$) La suite $(X_n)_{n \geq 1}$ est $\psi$-mélangeante et $\Sigma\psi_k < \infty$.

($M_2$) La suite est $\phi$-mélangeante et vérifie :

$$f_k(u_1, v_1, u_2, v_2) \leqslant C f_1(u_1, v_1) f_1(u_2, v_2) \, .$$

($M_3$) La suite est $\phi$-mélangeante et vérifie : $\displaystyle\sum_{k=0}^{\infty} \phi_k^{1/2} C_k^{1/2} < \infty$ où

$$C_k = \mathrm{Sup} \left\{ \iint (\,|\, v_1 \,|^2 + |\, v_2 \,|^2 + 1) f_k(u_1, v_1, u_2, v_2) dv_1 dv_2 \; ; (u_1, u_2) \in M^2 \right\}$$

Fixons maintenant les notations.

Les fonctions $f$ et $g$ étant des éléments de $L^2(M, d\mu)$, elles s'écrivent :

$$f = \sum_{j=1}^{\infty} c_j \psi_j \quad \text{et} \quad g = \sum_{j=1}^{\infty} a_j \psi_j \quad \text{avec} \quad a_j = \int r f \psi_j d\mu \quad \text{et} \quad c_j = \int f \psi_j d\mu \, .$$

Des estimations sans biais des coefficients $a_j$ et $c_j$ sont données par

$$\hat{a}_{j,n} = \frac{1}{n} \sum_{k=1}^{n} V_k \psi_j(U_k) \rho(U_k) \quad \text{et} \quad \hat{c}_{j,n} = \frac{1}{n} \sum_{k=1}^{n} \psi_j(U_k) \rho(U_k) \quad \text{(cf. [13])} \, .$$

Cela donne des estimations $\hat{f}_n = \displaystyle\sum_{j=1}^{q(n)} \hat{c}_{j,n} \psi_j$ et $\hat{g}_n = \displaystyle\sum_{j=1}^{q(n)} \hat{a}_{j,n} \psi_j$ sans biais pour les fonctions $f_n = \displaystyle\sum_{j=1}^{q(n)} c_j \psi_j$ et $g_n = \displaystyle\sum_{j=1}^{q(n)} a_j \psi_j$.

Ainsi $\hat{f}_n(x) = \frac{1}{n}\sum_{k=1}^{n} \rho(U_k)L_{q(n)}(x, U_k)$ et $\hat{g}_n(x) = \frac{1}{n}\sum_{k=1}^{n} V_k\rho(U_k)L_{q(n)}(x, U_k)$

avec $L_q(x, y) = \sum_{j=1}^{q} \psi_j(x)\psi_j(y)$.

## 2.1. Étude du cas indépendant.

Écrivons $\sqrt{n}(\hat{g}_n - g_n) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} Y_{q(n),k}$ avec

$$Y_{q,k}(x) = V_k\rho(U_k)L_q(x, U_k) - EV_k\rho(U_k)L_q(x, U_k).$$

Soit $\Gamma'_q$ l'opérateur de covariance de $Y_{q,1}$, il vérifie :

$$\Gamma'_q(\psi_i, \psi_j) = \int g_2(x)\psi_i(x)\psi_j(x)\rho^2(x)dx - a_ia_j \qquad \text{pour} \quad i, j = 1, \ldots, q.$$

Pour éliminer le terme $a_ia_j$, procédons comme León [15], nous considérons le gaussien $z'_q$ centré, à valeurs dans $E_q = \text{Vect}\,[\psi_1, \ldots, \psi_q]$ et de covariance $\Gamma'_q$ et le gaussien $z_q$ centré à valeurs dans $E_q$ et de covariance $\Gamma_q$ :

$$\Gamma_q(\psi_i, \psi_j) = \int g_2\psi_i\psi_j\rho^2dx.$$

La variable $z'_q$ s'écrit $z'_q = \sum_{j=1}^{q} \xi_j\psi_j$ où les $\xi_j$ sont gaussiennes réelles ; soit $\xi_0$ une variable gaussienne de variance 1 indépendante de $\xi_1, \ldots, \xi_q$, une réalisation de $z_b$ s'écrit : $z_q = z'_q + \xi_0 \sum_{j=1}^{q} a_j\psi_j$. Procédant comme [15], nous montrons que $\| g_2\rho \|_\infty < \infty$ entraîne $E\,|\,\| z'_q \|^2 - \| z_q \|^2\,| = 0(1)$ quand $q \to \infty$.

Nous notons à présent $E\| z_q \|^2 = \sum_{j=1}^{q} \int g_2(x)\psi_j^2(x)\rho^2(x)dx = q\Delta_q$.

Cette expression ne peut être évaluée de façon générale comme c'est le cas pour les noyaux.

De même nous posons : $\sigma_q^2 = \frac{2}{q}\sum_{i=1}^{q}\sum_{j=1}^{q}\left(\int \psi_i(x)\psi_j(x)\rho^2(x)g_2(x)dx\right)^2$, nous avons :

PROPOSITION 10. — *Supposons* $\lim_{q\to\infty} \sigma_q^2 = \sigma^2$, *alors si*

$$t_q = q^{-1/2}(\| z_q \|^2 - \mathrm{E}\| z_q \|^2),$$

*nous obtenons :* $\lim_{q\to\infty} \mathscr{L}(t_q) = \mathrm{N}(0, \sigma^2)$.

*Démonstration.* — Considérons une base orthonormale de $\mathrm{E}_q$ qui diagonalise la covariance $\Gamma_q$, nous écrivons, $z_q = \sum_{j=1}^{q} \lambda_{i,q}^{1/2}\gamma_i e_{i,q}$ où $\gamma_1, \ldots, \gamma_q$ sont des réalisations indépendantes de loi N(0, 1). Ainsi $t_q = q^{-1/2}\sum_{i=1}^{q}\lambda_{i,q}(\gamma_i^2 - 1)$, de manière analogue au développement de Karhunen-Loeve ; toutefois ici nous obtenons un tableau triangulaire fini, ainsi appliquer le théorème de Lindeberg se traduit par :

*a*)
$$\lim_{q\to\infty} \frac{2}{q}\sum_{i=1}^{q}\lambda_{i,q}^2 = \sigma^2$$

*b*)
$$\lim_{q\to\infty} \mathrm{Max}\,\{\lambda_{i,q}\,;\, i = 1, \ldots, q\}^2 \Big/ \left(\sum_{i=1}^{q}\lambda_{i,q}^2\right) = 0\,.$$

La condition *a*) est l'hypothèse de la proposition 10, la condition *b*) s'écrit : $\lim_{q\to\infty} q^{-1}(\Gamma_q\psi, \psi)^2 = 0$.

$$\text{Or } (\Gamma_q\psi, \psi) = \sum_{i=1}^{q}\sum_{j=1}^{q}\int g_2(x)\psi_i(x)\psi_j(x)\rho^2(x)dx\,\alpha_i\alpha_j \qquad \text{avec} \qquad \alpha_i = (\psi_i, \psi)$$

$$= \int g_2(x)\left(\sum_{i=1}^{q}\psi_i(x)\alpha_i\right)^2\rho^2(x)dx$$

$$= \| g_2\rho \|_\infty \| \psi \|^2.$$

*Remarque.* — Posant $t_q' = q^{-1/2}\| z_q' \|^2 - q^{1/2}\Delta_q$, les calculs précédents montrent que $\lim_{q\to\infty} \mathscr{L}(t_q') = \mathrm{N}(0, \sigma^2)$.

THÉORÈME 11. — *Supposons* $E \| Y_{q(n),1} \|^3 = o(n^{1/2})$ quand $n \to \infty$ *et* $\lim_{n\to\infty} q(n) = \infty$, *alors :*

$$\lim_{n\to\infty} \mathscr{L}[nq(n)^{-1/2} \| \hat{g}_n - g_n \|^2 - q(n)^{1/2}\Delta_{q(n)}] = N(0, \sigma)^2,$$

*si* $$\| g_2\rho \|_\infty < \infty \qquad et \qquad \lim \sigma_q^2 = \sigma^2$$

*où* $$\sigma_q^2 = \frac{2}{q}\sum_{i=1}^{q}\sum_{j=1}^{q}\left(\int \psi_i(x)\psi_j(x)\rho^2(x)g_2(x)dx\right)^2$$

*et* $$\Delta_q = q^{-1}\sum_{j=1}^{q}\int g_2(x)\psi_j^2(x)\rho^2(x)dx.$$

*Démonstration.* — La méthode est celle de Léon [*15*], en ce qui concerne l'approximation gaussienne, dans le cas de l'estimateur $\hat{f}_n$.

Notons qu'un résultat analogue peut être écrit pour $\hat{\hat{f}}_n$. La seule différence qui existe entre les deux calculs est ici l'utilisation du théorème de Lindeberg sans employer d'évaluation de la trace de $\Gamma_q^m$ pour un $m > 2$. Sous l'hypothèse additionnelle $\| g_n - g \|^2 = 0(q(n)^{1/2}n^{-1})$ qui s'exprime comme une condition de régularité de $g$, nous obtenons

$$\lim \mathscr{L}(nq(n)^{-1/2} \| \hat{g}_n - g \|^2 - q(n)^{-1/2}\Delta_{q(n)}) = N(0, \sigma^2).$$

L'expression $\| g_n - g \|^2$ s'est en effet que le reste du développement en série de $g$ sur la base $\{ \psi_j, j \geq 1 \}$ de $L^2(M, d\mu)$.

La condition $E \| Y_{q(n),1} \|^3 = o(n^{1/2})$ est satisfaite lorsque $\sum_{i=1}^{q(n)} \| \psi_i \|_\infty^2 = 0(q(n)^{-2}n)$ quand $n \to \infty$. En effet $E \| Y_{q(n),1} \|^3 \leq 2^3 \| g_3\rho^2 \|_\infty q(n)\left(\sum_{i=1}^{q(n)} \| \psi_i \|_\infty^2\right)^{1/2}$ (*cf.* [*11*]).

Nous allons à présent étudier des applications de ces résultats :

2.1.*a*) *Séries trigonométriques :* f, $g_2 \in L^2([-\pi, \pi]^d, dx)$.

Nous considérons non plus l'espace euclidien $L^2$ mais l'espace hermitien qui lui est associé, nous notons

$$\hat{f}_n(x) = \sum_{-J_n \leq \vec{j} \leq J_n} \hat{c}_{j,n}\psi_{\vec{j}}(x) \qquad \text{et} \qquad \hat{g}_n(x) = \sum_{-J_n \leq \vec{j} \leq J_n} \hat{a}_{\vec{j},n}\psi_{\vec{j}}(x);$$

ici, de plus, nous avons $q(n) = (2J_n + 1)^d$ et $\{ \vec{j}; -J_n \leq \vec{j} \leq J_n \} = \{ (j_1, \ldots, j_d);$ $-J_n \leq j_\alpha \leq J_n,\ \alpha = 1, \ldots, d \}$ et les fonctions $\psi_{\vec{j}}$ sont définies par $\psi_{\vec{j}}(x) = \exp(\vec{j}\,.\,x)(2\pi)^{-d/2}$. Ainsi $\Delta_q$ et $\sigma_q^2$ peuvent être calculés :

Dans le cas de l'estimateur de densité $\hat{f}_n$ :

$$\lim_q \Delta_q = \Delta = (2\pi)^{-d} \int_{[-\pi,\pi]^d} f(x)dx$$

$(= (2\pi)^{-d}$ lorsque $f$ est à support dans $[-\pi, \pi]^d)$ et $\sigma^2 = 2(2\pi)^{-d} \| f \|^2$.

Dans le cas de l'estimation $\hat{g}_n : \lim_q \Delta_q = \Delta = (2\pi)^{-d} \int_{[-\pi,\pi]^d} g_2(x)dx$ $(= \mathrm{E} \mid \mathrm{V}_1 \mid^2 (2\pi)^{-d}$ si $f$ est à support dans $[-\pi, \pi]^d)$ et $\sigma^2 = 2(2\pi)^{-d} \| g_2 \|^2$. Ainsi, les hypothèses du théorème 11 sont vérifiées si $\lim q(n)^3 n^{-1} = 0$ ou encore $\lim \mathrm{J}_n^{3d} n^{-1} = 0$. Si $\mathrm{J}_n = n^\alpha$, $\alpha < \frac{1}{3d}$ ; nous notons l'analogie avec le cas des noyaux.

Le terme de biais $\| g_n - g \|^2 = 0(q(n)^{1/2} n^{-1})$ lorsqu'il existe $l = l_1 + \ldots + l_d$, $\frac{\partial^l g_2}{\partial^{l_1} x_1 \ldots \partial^d x_d} \in \mathrm{L}^2([-\pi, \pi]^d, dx)$, et $l$ vérifie $\lim_{n\to\infty} n \mathrm{J}_n^{-d-2l} = 0$ ; par exemple lorsque $\mathrm{J}_n = n^\alpha$, $\frac{2}{d+4l} < \alpha < \frac{1}{3d}$ est là une condition assurant la convergence vers une loi $\mathrm{N}(0, \sigma^2)$ de la statistique

$$\gamma_n = nq(n)^{-1/2} \int_{[-\pi,\pi]^d} \mid \hat{g}_n - g \mid^2 dx - q(n)^{1/2} \Delta$$

L'ordre de dérivabilité vérifie nécessairement $l > \frac{5d}{4}$, condition identique à celle imposée dans le cas des estimateurs à noyaux. Un résultat parallèle suit aussi pour le cas de l'estimateur $\hat{f}_n$.

2.1.*b*) *Polynômes de Legendre normalisés* (cf. [*19*]).

Par soucis de simplicité, nous nous restreignons au cas $d = 1$, $\mathrm{M} = [-1, 1]$ et $d\mu(x) = dx$. Dans ce cas, si $g_2$ (resp. $f$) est continue nous voyons avec [*15*] que

$$\lim_{q\to\infty} \sigma_q^2 = \frac{2}{\pi} \int_{-1} g_2^2(x) \frac{dx}{\sqrt{1-x^2}} = \sigma^2 \left( \text{resp.} \lim_{q\to\infty} \sigma_q^2 = \frac{2}{\pi} \int_{-1}^{1} f^2(x) \frac{dx}{\sqrt{1-x^2}} \right)$$

et
$$\lim \Delta_q = \frac{1}{\pi} \int_{-1}^{1} g_2(x) \frac{dx}{\sqrt{1-x^2}}.$$

La condition pour que $\lim_{n\to\infty} \mathscr{L}(nq(n)^{-1/2} \| \hat{g}_n - g_n \|^2 - q(n)^{1/2} \Delta_{q(n)}) = \mathrm{N}(0, \sigma^2)$ est ici $\lim_{n\to\infty} q(n)^4 n^{-1} = 0$.

La condition additionnelle pour que

$$\lim_{n\to\infty} \mathscr{L}(nq(n)^{-1/2} \| \hat{g}_n - g \|^2 - q(n)^{1/2} \Delta_{q(n)}) = \mathrm{N}(0, \sigma^2)$$

est définie comme suit ; soit $\Delta$ l'opérateur différentiel défini par :

$\Delta u(x) = \dfrac{d}{dx}\left[(x^2 - 1)\dfrac{d}{dx}u(x)\right]$, $u \in L^2([-1, 1])$, et les dérivées sont prises au sens des distributions.

La condition $\Delta^s g \in L^2([-1, 1])$ (resp. $\Delta^s f \in L^2([-1, 1])$) entraîne $\| g_n - g \|^2 = 0(q(n)^{-4s})$ comme le montre l'équation différentielle qui régit les polynômes de Legendre (Cette condition est assurée lorsque $f^{(2s)} \in L^2([-1, 1])$) ; lorsque $\Delta^s g \in L^2([-1, 1])$ la statistique converge si, de plus, $\lim_{n\to\infty} n^{-2}q(n)^{8s+1} = \infty$ et $s \geqslant 1$.

Une extension au cadre multidimensionnel est aisément envisageable.

## 2.2. Cas mélangeant.

Considérons la variable aléatoire gaussienne $Z_q$ centrée, à valeurs dans $E_q = \text{Vect}\,[\psi_1, \dots, \psi_q]$ et de covariance

$$\Gamma_q^\phi(\psi_i, \psi_j) = \Gamma_q(\psi_i, \psi_j) - a_i a_j + \tau_q(\psi_i, \psi_j) + \tau_q(\psi_j, \psi_i)$$

avec
$$\tau_q(\psi_i, \psi_j) = \sum_{k=0}^{\infty} E(Y_{q,k}, \psi_i)(Y_{q,k}, \psi_j)$$

De plus :

$$E(Y_{q,1}, \psi_i)(Y_{q,k}, \psi_j) = E\,[V_1 \rho(U_1)\psi_i(U_1) - EV_1 \rho(U_1)\psi_i(U_1)]\,[V_k \rho(U_k)\psi_j(U_k) - EV_k \rho(U_k)\psi_j(U_k)]$$

avec $t_k(x, y) = \displaystyle\iint v_1 . v_2\, [f_k(x, v_1, y, v_2) - f_1(x, v_1) f_1(y, v_2)]\, dv_1 dv_2$.

Nous évaluons d'abord : $E \| Z_q \|^2 = \displaystyle\sum_{i=1}^{q} \tau_q(\psi_i, \psi_i) + E \| z'_q \|^2$.

L'évaluation précédente permet d'écrire :

$$\sum_{i=1}^{q} \tau_q(\psi_i, \psi_i) = \sum_{k=2}^{\infty} \iint L_q(x, y) t_k(x, y) \mu(dx)\mu(dy).$$

Nous remarquons d'autre part que les coefficients $c_{ij}^k$ du développement de la fonction $t_k$ dans la base $(\psi_i \otimes \psi_j)_{i,j \geqslant 1}$ de $L^2(M^2, d\mu^{\otimes 2})$,

$$c_{ij}^k = \iint t_k(x, y)\psi_i(x)\psi_j(y)\mu(dx)\mu(dy),$$

vérifient :

$$\iint L_q(x, y)t_k(x, y)\mu(dx)\mu(dy) = \sum_{i=1}^{q} c_{ij}^k .$$

Ainsi, sous l'hypothèse : $q^{-1/2} \sum_{i=1}^{q} c_{ii}^k \underset{q\to\infty}{\longrightarrow} 0$, nous obtenons :

$$\lim_{q\to\infty} q^{-1/2} \iint L_q(x, y)t_k(x, y)\mu(dx)\mu(dy) = 0$$

De plus $\left| \iint L_q(x, y)t_k(x, y)\mu(dx)\mu(dy) \right| \leqslant q^{1/2} \| t_k \|_2$, l'hypothèse $(M_3)$ permet d'appliquer le théorème de convergence dominée car $\| t_k \|_2 \leqslant 2c_k^{1/2}\phi_{k-1}^{1/2}$, ainsi $\lim q^{-1/2} \sum_{i=1}^{q} \tau_q(\psi_i, \psi_i) = 0$, les hypothèses $(M_1)$ et $(M_2)$ aboutissent à des évaluations analogues.

Nous avons donc montré que le terme de centrage $E \| Z_q \|^2$ est asymptotiquement équivalent à celui correspondant au cas indépendant.

Posant $a_k(i, j) = \iint \psi_i(x)\psi_j(y)t_k(x, y)\mu(dx)\mu(dy)$

$$A_q = \sum_{i,j=1}^{q} \left( \sum_{k=2}^{\infty} a_k(i, j) \right)^2 = 0(1)$$

entraînera comme précédemment les conditions du théorème central limite en utilisant, comme au paragraphe 1.2, l'inégalité de Schwarz

$$A_q = \sum_{k,l=2}^{\infty} \iiiint \mu(dx)\mu(dy)\mu(du)\mu(dv)t_k(x, y)t_l(u, v)L_q(x, u)L_q(y, v)$$

$$A_q = \sum_{k,l=2}^{\infty} (t_k, t_l^q)_{L^2(M^2, d\mu^{\otimes 2})}$$

où $t_l^q$ est le développement de $t_l$ dans la base $(\psi_i \otimes \psi_j)_{i,j\geqslant 1}$ de l'espace produit $L^2(M^2, d\mu^{\otimes 2})$

$$A_q \leqslant \left( \sum_{k=2}^{\infty} \| t_k \|_2 \right)^2 \leqslant \left( \sum_{k=2}^{\infty} c_k^{1/2}\phi_{k-1}^{1/2} \right)^2 \quad \text{sous l'hypothèse } (M_3).$$

Cette expression est aussi bornée sous les autres hypothèses.

Faisons maintenant l'approximation gaussienne ; nous posons :

$$G_{n,q}(t) = P\left( \left\| n^{-1/2} q^{-1/4} \sum_{k=1}^{n} Y_{q,k} \right\|^2 - q^{1/2}\Delta_q \leqslant t \right)$$

$$G_q(t) = P(\| q^{-1/4} Z_q \|^2 - q^{1/2}\Delta_q \leqslant t),$$

Soit, à présent $v_q = \| q^{-1/4} Y_{q,1} \|_{4+\delta}$, nous obtenons comme précédemment grâce à [9] :

$$| G_{n,q}(t) - G_q(t) | \leqslant \frac{v_q}{\varepsilon n^{(1-b)/6}} + \frac{v_q^2}{\varepsilon^2 n^{2(1-b)/3}} + \frac{v_q^3}{\varepsilon^3 n^{(1-b)/3}}$$
$$+ P(| \| q^{-1/4} Z_q \| - (t + q^{1/2}\Delta_q)^{1/2} | < \varepsilon)$$

Sous les hypothèses :

$$\sum_{n=1}^{\infty} n^2 \phi_n^{\delta/(4+\delta)} < \infty \qquad \text{et} \qquad \phi_n = 0\left(n^{-\frac{1}{4}\left(\frac{1}{b}-1\right)}\right) \quad \text{quand} \quad n \to \infty$$

pour un $b$, $0 < b < \frac{1}{4}$.

THÉORÈME 12. — *Supposons* $(E \| Y_{q(n),1} \|^{4+\delta})^{1/(4+\delta)} = 0(n^{(1-b)/9})$ *quand* $n \to \infty$ *pour une suite* $q(n)$ *telle que* $\lim_{n\to\infty} q(n) = \infty$, *sous les hypothèses de mélange* $(M_i)$ ($i = 1, 2$ *ou* $3$) *et si* $b \in \left]1, \frac{1}{4}\right[$, $\delta \in ]0, 1]$ *vérifient*

$$\sum_{n=1}^{\infty} n^2 \phi^{\delta/(4+\delta)} < \infty \qquad \text{et} \qquad \phi_n = 0\left(n^{\frac{3}{4}\left(\frac{1}{b}-1\right)}\right) \qquad \textit{quand} \quad n \to \infty,$$

*alors* $\lim_{n\to\infty} \mathscr{L}(nq(n)^{-1/2} \| \hat{g}_n - g_n \|^2 - q(n)^{1/2}\Delta_{q(n)}) = N(0, \sigma^2)$ *si* $\| g_n \|_\infty < \infty$, *et* $\lim \sigma_q^2 = \sigma^2$ *et* $\Delta_q = q^{-1} \sum_{j=1}^{q} \int g_2(x)\psi_j^2(x)\rho^2(x)dx$ *(l'expression de* $\sigma_q^2$ *est donnée dans l'énoncé du théorème 11). Nous supposons de plus l'hypothèse* $\lim_{q\to\infty} q^{-1/2} \sum_{i=1}^{q} c_{ii}^k = 0$ *réalisée pour tout* $k$.

Notons que le résultat analogue vaut encore dans le cas de l'estimation $\hat{f}_n$ de la densité qui correspond à une suite $V_n \equiv 1$, déterministe. Sous l'hypothèse additionnelle de régularité de $g$ : $\| g_n - g \|^2 = o(q(n)^{1/2} n^{-1})$, le résultat $\lim_{n\to\infty} \mathscr{L}(nq(n)^{-1/2} \| \hat{g}_n - g \|^2 - q(n)^{1/2}\Delta_{q(n)}) = N(0, \sigma^2)$ découle du précédent.

La condition $(E \| Y_{q(n),1} \|^{4+\delta})^{1/(4+\delta)} = o(n^{(1-b)/9})$ est réalisée lorsque $\left(\sum_{i=1}^{q(n)} \| \psi_i \|_\infty^2\right)^{1+\delta/2} = o(q(n)^{-1} n^\alpha)$ quand $n \to \infty$, si $\alpha = (1-b)(4+\delta)/9$, sous l'hypothèse $\| g_{4+\delta} \rho^{3+\delta} \|_\infty < \infty$. Cela découle de l'inégalité :

$$E \| Y_{q,1} \|^{4+\delta} \leqslant 2^5 q \| g_{4+\delta} \rho^{3+\delta} \|_\infty \left(\sum_{i=1}^{q} \| \psi_i \|_\infty^2\right)^{1+\delta/2}$$

2.2.*a*) *Séries trigonométriques.*

Les résultats sont les mêmes que dans le cas indépendant (2.1.*a*). Lorsque $q(n) = (2J_n + 1)^d$, les conditions du théorème 12 sont vérifiées si $\lim_{n\to\infty} J_n^d n^{-2/3(1-b)} = 0$. Le terme de biais est contrôlé de la même manière que dans le cas indépendant ; enfin l'inégalité de Schwarz montre que l'hypothèse $q^{-1/2} \sum_{i=1}^{q} c_{ii}^k \xrightarrow[q\to\infty]{} 0$ est réalisée lorsque $\sum_{i \in \mathbb{Z}^d} | i |^\varepsilon (c_{ii}^k)^2 < \infty$ pour un $\varepsilon > 0$, condition réalisée sous l'hypothèse $Dt_k \in L^2([-\pi, \pi]^d, dx)$.

2.2.*b*) *Polynômes de Legendre normalisés.*

La condition qui assure la convergence de la statistique

$$nq(n)^{1/2} \| \hat{g}_n - g_n \|^2 - q(n)^{1/2} \Delta_{q(n)}$$

devient $\lim_{n\to\infty} q(n)^{27+9\delta} n^{-(1-b)(4+\delta)} = 0$ et

$$\frac{\partial^2}{\partial x \partial y}\left((x^2-1)(y^2-1)\frac{\partial^2}{\partial x \partial y} t_k(x, y)\right) \in L^2([-1, 1]^2)$$

assure l'hypothèse $q^{-1/2} \sum_{i=1}^{q} c_{ii}^k \xrightarrow[q\to\infty]{} 0$ (cette condition est en fait beaucoup plus forte).

Pour que, $nq(n)^{1/2} \| \hat{g}_n - g \|^2 - q(n)^{1/2} \Delta_{q(n)}$ converge il faut, de plus que $\Delta^s g \in L^2([-1, 1])$, $\lim_{n\to\infty} n^{-2} q(n)^{8s+1} = \infty$ si $8s+1 > \dfrac{18(3+\delta)}{(1-b)(4+\delta)}$ $\left(s = 3\right.$ convient toujours et $s = 2$ convient si $\left. b < \dfrac{13}{85}\right)$.

## 3. EXEMPLES

Des exemples importants de suites mélangeantes vérifiant les hypothèses $(M_i)$ peuvent être énoncés. Tout d'abord, une suite $m$-dépendante et stationnaire dont le $m$-uplet $(X_1, \ldots, X_m)$ admet une densité régulière vérifie les hypothèses $(M_1)$ ; c'est le cas d'une suite de type MA($q$) très utilisée en statistiques.

De plus une chaîne de Markov Doeblin récurrente, dont la transition admet une densité $\pi(x, u, y, v)$ vérifie :

$$c_k(x, v) = \int (|u|^2 + |v|^2 + 1) f_k(x, u, y, v) du dv \\ \leqslant \| \pi \|_\infty + \| g_2 \|_\infty + \| f \|_\infty \} + C \| f \|_\infty$$

avec $C = \text{Sup} \{ E(|V_2|^2 / U_1 = u' ; V_1 = v) ; (u, u') \in M_1^2, v \in R^2 \} < \infty$ (dans le cas des estimateurs de projection $M_1$ est remplacé par M).

Un cas particulier intéressant est celui d'un processus autorégressif non linéaire d'ordre $p$ : $X_n = r(X_{n-1}, \ldots, X_{n-p}) + \varepsilon_n$ avec $\| r \|_\infty < \infty$, $(\varepsilon_n)$ est une suite indépendante ; équidistribuée, centrée et possédant des moments. Alors lorsque $U_n = (X_{n-1}, \ldots, X_{n-p})$ et $V_n = X_n$, la constante C est majorée par $\| r \|_\infty^2 + E |\varepsilon_1|^2$. De plus la fonction $r$ est ici la fonction d'autorégression (cf. [*4*] [*7*]).

## BIBLIOGRAPHIE


[*1*] J. Bretagnolle, C. Huber, Estimation des densités : risques minimaux. *Z. W.*, t. **47**, 1979, p. 119-137.

[*2*] N. N. Cencov, Evaluations of unknown distribution density from observation. *Soviet. Math.*, t. **3**, 1962, p. 1559-1562.

[*3*] G. Collomb, *Thèse*. Université Paul Sabatier, Toulouse, 1976.

[*4*] G. Collomb, P. Doukhan, *C. R. A. S.* Série I, t. **296**, 1983, p. 859-862.

[*5*] M. Csörgö, P. Revesz, *Strong approximations in Probability and Statistics*, 1981, Academic Press.

[*6*] H. Dehling, Limit theorems for sums of weakly dependant Banach space valued random variables. *Z. W.*, t. **63**, 1983, p. 393-432.

[*7*] P. Doukhan, Simulations in the first order autogression process. Specifying Statistical Models. *Lecture Notes in Statistics*, t. **16**, 1983, p. 50-68, Spinger-Verlag.

[*8*] P. Doukhan, J. León and F. Portal, *C. R. A. S.* Série I., 1984.

[*9*] P. Doukhan, J. León and F. Portal, Calcul de la vitesse de convergence dans le théorème central limite, vis-à-vis des distances de Levy, Dudley et Prohorov, dans le cas de variables aléatoires dépendantes (*Probability and Mathematical Statistics*, Vol. 6, Fasc. 1), 1985.

[*10*] P. Doukhan, J. León and F. Portal, *Applications Statistiques des principes d'Invariance dans un cadre mélangeant*, 1984 (à paraître).

[*11*] P. Doukhan, F. Portal, *C. R. A. S.* Série I, t. **297**, 1983, p. 129-132.

[*12*] T. Gasser and H. G. Muller, Kernel estimations of regression functions, Smoothing technics for curve estimation. *Lecture Notes in Mathematics*, t. **757**, 1979, p. 23-68, Springer-Verlag.

[*13*] W. Greblicki, D. Rutowska and L. Rutowski, An orthogonal series estimate of time varying regression. *Ann. Inst. Stat. Math. A*, t. **35**, 1983, p. 215-228.

[14] J. KUELBS and T. KURTZ, Berry Essen estimates in Hilbert space an application to L. I. L. *Ann. Prob.*, t. **2**, 1974, p. 387-407.
[15] J. LEÓN, Asymptotic behaviour of the quadratic deviation of multivariante density estimates. *Ann. Inst. Henry Poincaré B*, t. **19**, 1983, p. 297-309.
[16] E. J. PARZEN, On estimation of a Probability Density and Mode. *Ann. Math.*, t. **35**, 1962, p. 1065-1076.
[17] M. REED and B. SIMON, *Methods of modern Mathematical Physics*, t. **III**, 1970, Academic Press.
[18] M. ROSENBLATT, A quadratic measure of deviation of two dimensional density estimates and a test of independence. *Ann. of Stat.*, t. **3**, 1975, p. 1-14.
[19] G. SANSONE, *Orthogonal functions*, Interscience Publisher Inc., New York, 1959.
[20] S. C. SCHWARTZ, Estimation of a probability density by an orthogonal serie. *Ann. Math. Stat.*, t. **38**, 1967, p. 1261-1265.
[21] G. SZEGÖ, Orthogonal Polynomials. *Amer. Math. Soc. Coll. Publ.*, t. **23**, 1959.
[22] H. TAKAHATA, $L_\infty$-bound for asymptotic normality of weakly dependent summands using Stein's result. *Ann. Prob.*, t. **9**, 1981, p. 676.

*(Manuscrit reçu le 5 juillet 1984)*
*(révisé le 14 mai 1985)*