Normalité asymptotique des erreurs quadratiques intégrées d'estimateurs par noyau sous des conditions de mélangeance

Carlos TENREIRO*
Departamento de Matemática, Universidade de Coimbra

RÉSUMÉ

Dans cet article nous obtenons des théorèmes de limite centrale pour les erreurs quadratiques intégrées des estimateurs non paramétriques par noyau de la densité et de la régression, sous des conditions d'indépendance asymptotique, comme application de théorèmes de limite centrale pour des U-statistiques dégénérées.

Classification AMS 1991: 62G20, 62G07.

1 Introduction

Soit $(X_i, Y_i), i \in \mathbb{Z}$ un processus stochastique fortement stationnaire et géométriquement absolument régulier (cf. [2]), à valeurs dans $\mathbb{R}^d \times \mathbb{R}$. On suppose que X_0 admet une densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Soient, pour $x \in \mathbb{R}^d$,

$$f_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

l'estimateur par noyau de la densité $f(x)$ (cf. Rosenblatt [11] et Parzen [10]), et

$$\mu_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h_n}\right) / f_n(x),$$

l'estimateur par noyau de la fonction de régression $\mu(x) = E[Y_0|X_0 = x]$ (cf. Nadaraya [7] et Watson [15]). K est un noyau sur $\mathbb{R}^d$, c'est-à-dire, une application de $\mathbb{R}^d$ dans $\mathbb{R}$ telle que $\int_{\mathbb{R}^d} K(x)dx = 1$, et (h_n) est une suite de nombres réels strictement positifs tendant vers zéro lorsque n tend vers l'infini.

Dans ce papier nous nous intéressons au comportement asymptotique des erreurs quadratiques intégrées (pondérées) correspondantes à chacun des estimateurs précédents. Elles sont définies par

$$I_n = \int_A \{f_n(x) - f(x)\}^2 \omega(x)dx \text{ et } J_n = \int_A \{\mu_n(x) - \mu(x)\}^2 f_n^2(x)\omega(x)dx,$$

où A est un borélien de $\mathbb{R}^d$ dont la frontière est de mesure de Lebesgue nulle et ω est une fonction de poids.

* Ce travail a bénéficié de l'appui de la JNICT et du Gouvernement Français.

Des théorèmes de limite centrale pour ces variables ont été obtenus par divers auteurs quand les variables $(X_i, Y_i), i \in \mathbb{Z}$, sont indépendantes. Ce sont les cas de Bickel et Rosenblatt [1], Rosenblatt [12] et Hall [4], pour la variable I_n, et Nadaraya [8], Konakov [6], Nadaraya [9] et Hall [5] pour la variable J_n. Dans le cas mélangeant, la normalité asymptotique de la variable I_n avec $\omega = 1$ et $A = \mathbb{R}^d$, a été obtenue par Takahata et Yoshihara [13].

Dans cet article, nous présentons une approche qui permet d'unifier le problème de la détermination des lois limite pour les erreurs quadratiques intégrées des estimateurs non paramétriques par noyau de la densité et de la régression au cas multivarié, dans un cadre de dépendance. De la même façon que dans le cas d'échantillonnage (cf. [4] et [5]), les variables $I_n - EI_n$ et $J_n - EJ_n$ peuvent asymptotiquement s'écrire comme la somme de deux U-statistiques dégénérées: une d'ordre 1 et l'autre d'ordre 2. En utilisant un théorème limite pour des U-statistiques dégénérées engendrées par un processus fortement stationnaire et géométriquement absolument régulier, que nous énonçons dans la Section 2, nous dérivons la normalité asymptotique des écarts centrés $I_n - EI_n$ et $J_n - EJ_n$. Ces résultats, que nous présentons dans la Section 3, généralisent au cas mélangeant, ceux obtenus par Hall [4] et [5].

Toutes les démonstrations sont données dans la Section 4.

2 U-statistiques dégénérées. Loi asymptotique

Soient $(Z_i, i \in \mathbb{Z})$ un processus stochastique fortement stationnaire et géométriquement absolument régulier, à valeurs dans $\mathbb{R}^p$, et $(g_n(\cdot), n \in \mathbb{N}^*)$ et $(h_n(\cdot,\cdot), n \in \mathbb{N}^*)$, des suites d'applications mesurables de $\mathbb{R}^p$ et $\mathbb{R}^p \times \mathbb{R}^p$, respectivement, dans $\mathbb{R}$. Dans cette section nous établissons le comportement asymptotique du couple:

$$(\mathcal{G}_n, \mathcal{H}_n),$$

où

$$\mathcal{G}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g_n(Z_i), \text{ et}$$

$$\mathcal{H}_n = \frac{1}{n} \sum_{1 \le j < i \le n} \{h_n(Z_i, Z_j) - Eh_n(Z_i, Z_j)\},$$

où on suppose que $E[g_n(Z_0)] = 0$, $E[h_n(Z_0, y)] = 0$ et $h_n(x, y) = h_n(y, x)$, pour tout $n \in \mathbb{N}^*$ et $x, y \in \mathbb{R}^p$. $\mathcal{G}_n$ et $\mathcal{H}_n$ sont donc des U-statistiques dégénérées, respectivement d'ordres 1 et 2.

Pour $r > 0$ et $n \in \mathbb{N}^*$, introduisons les notations suivantes:

$$\begin{aligned}
u_n(r) &= \max\Big\{\max_{1 \le i \le n} \|h_n(Z_i, Z_0)\|_r, \|h_n(Z_0, \overline{Z}_0)\|_r\Big\}, \\
v_n(r) &= \max\Big\{\max_{1 \le i \le n} \|G_{n0}(Z_i, Z_0)\|_r, \|G_{n0}(Z_0, \overline{Z}_0)\|_r\Big\}, \\
w_n(r) &= \|G_{n0}(Z_0, Z_0)\|_r, \\
z_n(r) &= \max_{\substack{0 \le i \le n \\ 1 \le j \le n}} \max\Big\{\|G_{nj}(Z_i, Z_0)\|_r, \|G_{nj}(Z_0, Z_i)\|_r, \|G_{nj}(Z_0, \overline{Z}_0)\|_r\Big\},
\end{aligned} \tag{2.1}$$

où $G_{nj}(u,v) = E[h_n(Z_j,u)h_n(Z_0,v)]$, pour $j \in \mathbb{N}$ et $u,v \in \mathbb{R}^p$, $\overline{Z}_0$ est une copie de Z_0 qui lui est indépendante, et $\|\xi\|_r = E^{\frac{1}{r}}|\xi|^r$, pour toute variable aléatoire réelle ξ telle que $E|\xi|^r < \infty$.

Théorème 2.2 *On suppose qu'il existe* $\delta_0 > 0$, $\gamma_0 < \frac{1}{2}$ *et* $\gamma_1 > 0$, *tels que:*

i) $\|g_n(Z_0)\|_4 = O(1)$;
ii) $E[g_n(Z_j)g_n(Z_0)] = c_j + o(1)$, *pour tout* $j = 0,1,2,\ldots$;
iii) $u_n(4+\delta_0) = O(n^{\gamma_0})$;
iv) $v_n(2+\frac{\delta_0}{2}) = o(1)$;
v) $w_n(2+\frac{\delta_0}{2}) = o(n^{\frac{1}{2}})$;
vi) $z_n(2)n^{\gamma_1} = O(1)$;
vii) $E[h_n(Z_0,\overline{Z}_0)]^2 = 2\sigma_2^2 + o(1)$.

Alors le vecteur aléatoire $(\mathcal{G}_n, \mathcal{H}_n)$ *est asymptotiquement normal de moyenne zéro et de matrice de covariance donnée par* $\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$, *où* $\sigma_1^2 = c_0 + 2\sum_{j=1}^{\infty} c_j$.

3 Hypothèses et résultats

Les hypothèses introduites dans les points suivants concernent le processus, le noyau, la fenêtre h_n et la fonction de poids.

H$_1$) Les lois des vecteurs (X_i, X_0), $i \geq 1$, admettent des versions des densités $f_{(X_i,X_0)}$ qui satisfont les contraintes: il existe $\epsilon > 0$ tel que,

$$\sup_{x \in A^\epsilon} f(x) < \infty \text{ et}$$

$$\sup_{i \in \mathbb{N}^*} \sup_{\substack{x \in A^\epsilon \cap \{u \in \mathbb{R}^d | f(u) > 0\} \\ y \in A^\epsilon}} \frac{f_{(X_i,X_0)}(y,x)}{f(x)} < \infty,$$

où A^ϵ désigne l'ensemble des points de $\mathbb{R}^d$ dont les distances à A sont inférieures à ϵ.

La deuxième condition est, dans un cadre d'échantillonnage, conséquence de la première. S'il y a dépendance, elle est par exemple satisfaite dans le cas d'un processus gaussien stationnaire, parce que la variance conditionnelle ne dépend pas de la variable conditionante.

H$_2$) Le noyau K supposé borné à support borné est tel que

$$\int_{\mathbb{R}^d} K(z) z_i dz = 0,\ i = 1,\ldots,d, \text{ et } z = (z_1,\ldots,z_d).$$

H$_3$) La suite (h_n) satisfait les conditions classiques

$$h_n \to 0 \text{ et } nh_n^d \to +\infty, \text{ lorsque } n \to +\infty,$$

et de plus on admet qu'il existe $\gamma \in]0,1[$ tel que

$$\limsup_n n^\gamma h_n^d < \infty.$$

Cette dernière condition est peu restrictive et est vérifiée par exemple si $h_n = O(n^{-\delta})$, $0<\delta<\frac{1}{d}$, dès qu'on choisit $\gamma \in]0, \delta d]$.

H_4) La fonction de poids ω est une fonction bornée sur A^ϵ et continue sur l'intérieur de A.

Dans les théorèmes suivants, nous établissons la normalité asymptotique des variables aléatoires $I_n - EI_n$ et $J_n - EJ_n$.

On notera par $\mathbb{I}_A$ la fonction indicatricè de l'ensemble A e par $\xrightarrow[n\to+\infty]{d} N(0,\sigma^2)$ la convergence vers la loi normale de moyenne zéro et de variance σ^2.

Théorème 3.1 *Sous les hypothèses* H_1),...,H_4), *on suppose que* f *admet des dérivées partielles d'ordre 2 continues et bornées sur* A^ϵ. *Notons*

$$\nu_1^2 = var\big((\gamma\,\omega)(X_0)\big) + 2\sum_{j=1}^{+\infty} cov\big((\gamma\,\omega)(X_j), (\gamma\,\omega)(X_0)\big)\ , \quad et$$

$$\nu_2^2 = \int_A f^2(x)\omega^2(x)dx \int_{\mathbb{R}^d}\Big(\int_{\mathbb{R}^d} K(u+z)K(u)du\Big)^2 dz,$$

où

$$\gamma(x) = \frac{1}{2}\sum_{i,j=1}^{d}\int_{\mathbb{R}^d} u_i u_j K(u)du \frac{\partial^2 f}{\partial x_i \partial x_j}(x)\mathbb{I}_A(x).$$

(1) *Si* $nh_n^{d+4} \to \lambda \in [0, +\infty[$, $n\to+\infty$, *alors*

$$nh_n^{d/2}\{I_n - EI_n\} \xrightarrow[n\to+\infty]{d} N(0, 4\lambda\nu_1^2 + 2\nu_2^2).$$

(2) *Si* $nh_n^{d+4} \to \lambda = +\infty$, $n\to+\infty$, *alors*

$$\sqrt{n}h_n^{-2}\{I_n - EI_n\} \xrightarrow[n\to+\infty]{d} N(0, 4\nu_1^2).$$

Théorème 3.2 *Sous les hypothèses* H_1),...,H_4), *on suppose que* f *et* μ *admettent des dérivées partielles, respectivement d'ordre 1 et d'ordre 1 et 2, continues et bornées sur* A^ϵ *et que* $\sigma^2 f$ *est continue sur l'intérieur de* A *et bornée sur* A^ϵ, *où* $\sigma^2(x)$ *est la variance de la loi de* Y_0 *conditionnée par* $X_0 = x$. *On admet aussi que* $E|Y_0|^{8+\delta} < \infty$, *pour un* $\delta > 0$. *Notons*

$$\begin{aligned}\eta_1^2 \;=\;& var\big(\{Y_0 - \mu(X_0)\}(\phi\,\omega)(X_0)\big)\\ &+2\sum_{j=1}^{+\infty} cov\big(\{Y_j - \mu(X_j)\}(\phi\,\omega)(X_j), \{Y_0 - \mu(X_0)\}(\phi\,\omega)(X_0)\big)\ , \quad et\end{aligned}$$

$$\eta_2^2 = \int_A \sigma^4(x)f^2(x)\omega^2(x)dx \int_{\mathbb{R}^d}\Big(\int_{\mathbb{R}^d} K(u+z)K(u)du\Big)^2 dz,$$

où

$$\phi(x) = \frac{1}{2}\sum_{i,j=1}^{d}\int_{\mathbb{R}^d} u_i u_j K(u)du\Big\{\frac{\partial^2 \mu}{\partial x_i\partial x_j}(x)f(x) + 2\frac{\partial \mu}{\partial x_i}(x)\frac{\partial f}{\partial x_j}(x)\Big\}\,\mathbb{I}_A(x).$$

(1) *Si* $nh_n^{d+4} \to \lambda \in [0,+\infty[$, $n \to +\infty$, *alors*

$$nh_n^{d/2}\{J_n - EJ_n\} \underset{n\to+\infty}{\xrightarrow{d}} N(0, 4\lambda\eta_1^2 + 2\eta_2^2).$$

(2) *Si* $nh_n^{d+4} \to \lambda = +\infty$, $n \to +\infty$, *alors*

$$\sqrt{n}h_n^{-2}\{J_n - EJ_n\} \underset{n\to+\infty}{\xrightarrow{d}} N(0, 4\eta_1^2).$$

4 Démonstrations

Démonstration du Théorème 2.2: Par la suite nous ne donons que les lignes générales de la démonstration. Pour les détailles voir Tenreiro [14].

Dans un premier temps, les hypothèses i), iii), vi) et vii) et la structure de dépendance considérée pour le processus $(Z_i, i \in \mathbb{Z})$, nous permettent de conclure que pour tout a et b réels, la variable aléatoire $a\mathcal{G}_n + b\mathcal{H}_n$ est, en probabilité, asymptotiquement équivalente à la martingale

$$\sum_{\alpha=1}^{k}\{U_{\alpha,n} - E[U_{\alpha,n} \mid F_{\alpha-1,n}]\} \tag{4.1}$$

où pour $\alpha = 1, \ldots, k$

$$U_{\alpha,n} = \sum_{i=\alpha(r+m)-m+1}^{\alpha(r+m)} \left(\frac{1}{\sqrt{n}} g_n(Z_i) + \frac{1}{n} \sum_{j=1}^{(\alpha-1)(r+m)+1} h_n(Z_i, Z_j)\right), \text{ et}$$

$$F_{\alpha,n} = \sigma(Z_1, \ldots, Z_{\alpha(r+m)+1}),$$

avec $k = k(n) = [\frac{n}{m+r}]$, $m = m(n) = [n^{\delta_1}]$, $r = r(n) = [n^{\delta_2}]$ et $0 < \delta_2 < \delta_1 < \min(\gamma_1/2\,,\,(1-2\gamma_0)/3)$ ($[x]$ désigne la partie entière du réel x).

Deuxièmement, l'ensemble d'hypothèses considéré, le choix de δ_1 et δ_2, et la convergence exponentielle vers zéro des coefficients de mélange de type β, permettent d'obtenir les égalités

$$\sum_{\alpha=1}^{k} var[U_{\alpha,n}|F_{\alpha-1,n}] = a^2\sigma_1^2 + b^2\sigma_2^2 + o_p(1), \text{ et}$$

$$\sum_{\alpha=1}^{k} E|U_{\alpha,n}|^4 = o(1).$$

Ainsi, d'après le théorème de limite centrale de Dvoretzky [3], on conclut que la variable aléatoire (4.1) et donc $a\mathcal{G}_n + b\mathcal{H}_n$ sont asymptotiquement normales de moyenne zéro et de variance $a^2\sigma_1^2 + b^2\sigma_2^2$. □

Démonstration du Théorème 3.1: On a:

$$\begin{aligned} I_n &= \int_A \{f_n(x) - Ef_n(x)\}^2\omega(x)dx \\ &\quad +2\int_A \{f_n(x) - Ef_n(x)\}\{Ef_n(x) - f(x)\}\omega(x)dx \\ &\quad +\int_A \{Ef_n(x) - f(x)\}^2\omega(x)dx. \end{aligned}$$

Ainsi:

$$
\begin{aligned}
(4.2) \qquad I_n - EI_n &= \frac{1}{\sqrt{n}h_n^{-2}}\frac{2}{\sqrt{n}}\sum_{i=1}^{n} G_n(X_i) \\
&+\frac{1}{nh_n^{d/2}}\frac{2}{n}\sum_{1\le j<i\le n}\{H_n(X_i,X_j)-EH_n(X_i,X_j)\} \\
&+\frac{1}{nh_n^{d/2}}\frac{1}{n}\sum_{i=1}^{n}\{H_n(X_i,X_i)-EH_n(X_i,X_i)\} \\
&= \frac{1}{\sqrt{n}h_n^{-2}}U_n+\frac{1}{nh_n^{d/2}}V_n+\frac{1}{nh_n^{d/2}}W_n,
\end{aligned}
$$

où pour $u, v \in \mathbb{R}^d$:

$$
G_n(u) = \frac{1}{h_n^d}\int_A\left\{K\left(\frac{x-u}{h_n}\right) - EK\left(\frac{x-X_0}{h_n}\right)\right\} h_n^{-2}\{Ef_n(x)-f(x)\}\omega(x)dx, \text{ et}
$$

$$
\begin{aligned}
H_n(u,v) = \frac{1}{h_n^{3d/2}}\int_A &\left\{K\left(\frac{x-u}{h_n}\right) - EK\left(\frac{x-X_0}{h_n}\right)\right\} \\
&\times\left\{K\left(\frac{x-v}{h_n}\right) - EK\left(\frac{x-X_0}{h_n}\right)\right\}\omega(x)dx.
\end{aligned}
$$

Normalité asymptotique de U_n: D'après les hypothèses sur le noyau K et sur f, un développement de Taylor nous permet d'écrire pour $x \in A$:

$$
\begin{aligned}
h_n^{-2}\{Ef_n(x)-f(x)\} &= h_n^{-2}\int_{\mathbb{R}^d} K(u)[f(x-uh_n)-f(x)]du \\
&= \sum_{i,j=1}^{d}\int_{\mathbb{R}^d} K(u)u_iu_j\int_0^1 \frac{\partial^2 f}{\partial x_i\partial x_j}(x-uh_nt)(1-t)dtdu.
\end{aligned}
$$

Ainsi, comme K est à support borné et les dérivées partielles d'ordre 2 de f sont continues et bornées sur A^ϵ, on conclut d'après le théorème de la convergence dominée que pour $x \in A$:

$$
\lim_{n\to+\infty} h_n^{-2}\{Ef_n(x)-f(x)\} = \frac{1}{2}\sum_{i,j=1}^{d}\int_{\mathbb{R}^d} u_iu_jK(u)du\,\frac{\partial^2 f}{\partial x_i\partial x_j}(x) = \gamma(x).
$$

La normalité asymptotique de U_n sera établie à l'aide du Théorème 2.2. Comme ω est borné sur A^ϵ ainsi que les dérivées partielles d'ordre 2 de f, nous avons:

$$
\sup_{n\in\mathbb{N}^*}\sup_{u\in\mathbb{R}^d}|G_n(u)| < \infty,
$$

et donc la condition i) du Théorème 2.2 est satisfaite. De plus, comme ω est continue sur l'intérieur de A et la frontière de A est de mesure de Lebesgue nulle, nous avons:

$$
\lim_{n\to+\infty} E[G_n(X_j)G_n(X_0)] = cov\big((\gamma\omega)(X_j),(\gamma\omega)(X_0)\big), \text{ pour } j = 0,1,2,\ldots.
$$

D'après le Théorème 2.2, on déduit alors que U_n converge en loi, lorsque n tend vers l'infini, vers la loi normale de moyenne zéro et de variance $4\nu_1^2$.

Normalité asymptotique de V_n: Soient pour $r > 0$ et $n \in \mathbb{N}^*$, $u_n(r)$, $v_n(r)$, $w_n(r)$ et $z_n(r)$ définis par (2.1), pour le processus $(X_i, i \in \mathbb{Z})$, avec $h_n = H_n$. D'après les hypothèses sur le processus, le noyau et la fonction de poids, il existe $C > 0$ tel que, pour $r \geq 1$ et $n \in \mathbb{N}^*$, on a:

$$\begin{array}{rcl} u_n(r) & \leq & C(h_n^d)^{\frac{1}{r}-\frac{1}{2}}, \\ v_n(r) & \leq & C(h_n^d)^{\frac{1}{r}}, \\ w_n(r) & \leq & C, \text{ et} \\ z_n(r) & \leq & Ch_n^d. \end{array} \tag{4.3}$$

On ne montrera que la première inégalité. Les autres s'obtiennent avec une technique analogue. En tenant en compte qu'uniformément par rapport à $u, v \in \mathbb{R}^d$, on a:

$$H_n(u,v) = \frac{1}{h_n^{3d/2}} \int_A K\left(\frac{x-u}{h_n}\right) K\left(\frac{x-v}{h_n}\right) \omega(x)dx + O(h_n^{d/2}), \tag{4.4}$$

on conclut que pour $r \geq 1$ et $i \geq 1$:

$$h_n^{3d/2} \| H_n(X_i, X_0)\|_r \leq E^{\frac{1}{r}} \left| \int_A K\left(\frac{x-X_i}{h_n}\right) K\left(\frac{x-X_0}{h_n}\right) \omega(x)dx \right|^r + O(h_n^{2d}),$$

où, K étant borné à support borné, on a:

$$\begin{aligned} & E\left| \int_A K\left(\frac{x-X_i}{h_n}\right) K\left(\frac{x-X_0}{h_n}\right) \omega(x)dx \right|^r \\ & = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left| \int_A K\left(\frac{x-v}{h_n}\right) K\left(\frac{x-u}{h_n}\right) \omega(x)dx \right|^r \frac{f_{(X_i,X_0)}(v,u)}{f(u)} f(u)dvdu \\ & \leq (h_n^d)^{r+1} \sup_{x\in A} |\omega(x)|^r \sup_{i\in\mathbb{N}^*} \sup_{\substack{x\in A^\epsilon \cap \{u\in\mathbb{R}^d | f(u)>0\} \\ y\in A^\epsilon}} \frac{f_{(X_i,X_0)}(y,x)}{f(x)} \\ & \qquad \times \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |K|^r(y)|K|^r(y+z)dydz. \end{aligned}$$

Nous déduisons alors que $h_n^{3d/2} \| H_n(X_i, X_0)\|_r \leq C(h_n^d)^{1+\frac{1}{r}}$, avec $C > 0$. De même, si on remplace dans les développements précédents $f_i(v|u)$ par $f(v)$, on trouve pour $h_n^{3d/2} \| H_n(X_0, \overline{X}_0)\|_r$ une majoration du même ordre. Nous en déduisons alors que:

$$u_n(r) \leq C(h_n^d)^{\frac{1}{r}-\frac{1}{2}}, \text{ avec } C > 0.$$

D'après les inégalités (4.3), les conditions iii),...,vi) du Théorème 2.2 sont alors satisfaites avec $\delta_0 > 0$, fixé, $\gamma_0 = \frac{2+\delta_0}{8+2\delta_0} < \frac{1}{2}$ et $\gamma_1 \in]0, \gamma]$, où γ est tel que $\limsup_n n^\gamma h_n^d < \infty$ (hypothèse H_3). Il nous reste à déterminer la forme de la variance asymptotique. On a:

$$\begin{aligned} & E\left[\frac{1}{h_n^{3d/2}} \int_A K\left(\frac{x-X_0}{h_n}\right) K\left(\frac{x-\overline{X}_0}{h_n}\right) \omega(x)dx \right]^2 \\ & = \frac{1}{h_n^{3d}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left(\int_A K\left(\frac{x-u}{h_n}\right) K\left(\frac{x-v}{h_n}\right) \omega(x)dx \right)^2 f(u)f(v)dudv \\ & = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} K(y)K(y+z)(\omega \cdot \mathbb{I}_A)(v+(y+z)h_n)dy \right)^2 f(v+zh_n)f(v)dzdv \\ & = \int_A f^2(x)\omega^2(x)dx \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} K(u+z)K(u)du \right)^2 dz + o(1), \end{aligned}$$

d'après le théorème de la convergence dominée, en tenant en compte la continuité des fonctions ω et f sur l'intérieur de A, et le fait qu'elles sont bornées sur A^ϵ. Ainsi, d'après (4.4):

$$\lim_{n\to+\infty} E[H_n(X_0,\overline{X}_0)]^2 = \nu_2^2 = \int_A f^2(x)\omega^2(x)dx \int_{\mathbb{R}^d}\Big(\int_{\mathbb{R}^d} K(u+z)K(u)du\Big)^2 dz.$$

D'après le Théorème 2.2 on déduit que V_n converge en loi, lorsque n tend vers l'infini, vers la loi normale de moyenne zéro et de variance $2\nu_2^2$.

Normalité asymptotique de $I_n \dot{-} EI_n$: Comme le dernier terme de la décomposition (4.2) est tel que $W_n = o_p(1)$, elle peut donc se reécrire:

$$I_n - EI_n = \frac{1}{\sqrt{n}h_n^{-2}}U_n + \frac{1}{nh_n^{d/2}}V_n + o_p\left(\frac{1}{nh_n^{d/2}}\right).$$

La conclusion découle maintenant des points précédents et du Théorème 2.2. □

Démonstration du Théorème 3.2: Soient, pour $x\in\mathbb{R}^d$ et $u=(u_1,u_2)\in\mathbb{R}^d\times\mathbb{R}$:

$$r_n(x,u) = (u_2-\mu(u_1))K\left(\frac{x-u_1}{h_n}\right), \text{ et } s_n(x,u_1) = (\mu(u_1)-\mu(x))K\left(\frac{x-u_1}{h_n}\right).$$

Nous avons:

$$\begin{aligned} J_n &= \frac{1}{n^2h_n^{2d}}\sum_{i,j=1}^n \int_A \{r_n(x,Z_i)+\bar{s}_n(x,X_i)\}\{r_n(x,Z_j)+\bar{s}_n(x,X_j)\}\omega(x)dx \\ &\quad +\frac{2}{nh_n^{2d}}\sum_{i=1}^n \int_A \{r_n(x,Z_i)+\bar{s}_n(x,X_i)\}E[s_n(x,X_0)]\omega(x)dx \\ &\quad +\frac{1}{nh_n^{2d}}\int_A E^2[s_n(x,X_0)]\omega(x)dx, \end{aligned}$$

où $\bar{s}_n(x,u_1) = s_n(x,u_1) - E[s_n(x,X_0)]$ et $Z_i = (X_i,Y_i), i\in\mathbb{Z}$. L'écart centré $J_n - EJ_n$ peut être donc décomposé dans la somme de U-statistiques dégénérées d'ordre 1 et 2. En négligeant les termes d'ordre supérieur, on déduit:

$$\begin{aligned} J_n - EJ_n &= \frac{1}{\sqrt{n}h_n^{-2}}\frac{2}{\sqrt{n}}\sum_{i=1}^n G_n^*(Z_i) \\ &\quad +\frac{1}{nh_n^{d/2}}\frac{2}{n}\sum_{1\le j<i\le n}\{H_n^*(Z_i,Z_j) - EH_n^*(Z_i,Z_j)\} + o_p\left(\frac{1}{nh_n^{d/2}}\right) \qquad (4.5)\\ &= \frac{1}{\sqrt{n}h_n^{-2}}U_n^* + \frac{1}{nh_n^{d/2}}V_n^* + o_p\left(\frac{1}{nh_n^{d/2}}\right), \end{aligned}$$

où pour $u=(u_1,u_2), v=(v_1,v_2)\in\mathbb{R}^d\times\mathbb{R}$:

$$\begin{aligned} G_n^*(u) &= \frac{1}{h_n^{2d}}\int_A r_n(x,u)h_n^{-2}E[s_n(x,X_0)]\omega(x)dx \\ &= \frac{1}{h_n^d}(u_2-\mu(u_1))\int_A K\left(\frac{x-u_1}{h_n}\right)h_n^{-2}E[(\mu_n(x)-\mu(x))f_n(x)]\omega(x)dx, \end{aligned}$$

et

$$\begin{aligned} H_n^*(u,v) &= \frac{1}{h_n^{3d/2}} \int_A r_n(x,u) r_n(x,v) \omega(x) dx, \\ &= \frac{1}{h_n^{3d/2}} (u_2 - \mu(u_1))(v_2 - \mu(v_1)) \int_A K\left(\frac{x-u_1}{h_n}\right) K\left(\frac{x-v_1}{h_n}\right) \omega(x) dx. \end{aligned}$$

Normalité asymptotique de U_n^*: D'après les hypothèses sur le noyau K, sur f et μ, nous avons pour $x \in A$:

$$\begin{aligned} & h_n^{-2} E[(\mu_n(x) - \mu(x)) f_n(x)] \\ &= \sum_{i,j=1}^{d} \int_{\mathbb{R}^d} K(u) u_i u_j \left\{ \frac{\partial \mu}{\partial x_i}(x) \int_0^1 \frac{\partial f}{\partial x_j}(x - h_n u t) dt \right. \\ & \qquad \left. + f(x - h_n u) \int_0^1 \frac{\partial^2 \mu}{\partial x_i \partial x_j}(x - h_n u t)(1-t) dt \right\} du. \end{aligned}$$

Le théorème de la convergence dominée permet d'écrire que pour $x \in A$:

$$\begin{aligned} & \lim_{n \to +\infty} h_n^{-2} E[(\mu_n(x) - \mu(x)) f_n(x)] \\ &= \frac{1}{2} \sum_{i,j=1}^{d} \int_{\mathbb{R}^d} u_i u_j K(u) du \left\{ \frac{\partial^2 \mu}{\partial x_i \partial x_j}(x) f(x) + 2 \frac{\partial \mu}{\partial x_i}(x) \frac{\partial f}{\partial x_j}(x) \right\} = \phi(x). \end{aligned}$$

De plus comme les dérivées partielles d'ordre 1 de f et d'ordre 1 et 2 de μ sont continues et bornées sur A^ϵ, ω est continue sur l'intérieur de A, et la frontière de A est de mesure de Lebesgue nulle, nous avons:

$$\sup_{n \in \mathbb{N}^*} E|G_n^*(Z_0)|^4 < \infty, \text{ et pour } j = 0, 1, 2, \ldots,$$

$$\lim_{n \to +\infty} E[G_n^*(Z_j) G_n^*(Z_0)] = cov\big(\{Y_j - \mu(X_j)\}(\phi\omega)(X_j), \{Y_0 - \mu(X_0)\}(\phi\omega)(X_0)\big).$$

D'après le Théorème 2.2, on déduit que U_n^* converge en loi, lorsque n tend vers l'infini, vers la loi normale de moyenne zéro et de variance $4\eta_1^2$.

Normalité asymptotique de V_n^*: Soient $u_n(r)$, $v_n(r)$, $w_n(r)$ et $z_n(r)$ les quantités définies pour $r > 0$ et $n \in \mathbb{N}^*$, par (2.1) avec $h_n = H_n^*$. De la même façon que dans la démonstration du Théorème 3.1, et en tenant en compte que $E|Y_0|^{8+\delta} < \infty$, il existe $C > 0$ tel que pour $r \geq 1 - \frac{2}{10+\delta}$ et $n \in \mathbb{N}^*$, on a:

$$\begin{aligned} u_n(r) &\leq C(h_n^d)^{\frac{1}{r} - \frac{2}{8+\delta} - \frac{1}{2}}, \text{ si } r < 4 + \tfrac{\delta}{2}, \\ v_n(r) &\leq C(h_n^d)^{\frac{1}{r} - \frac{2}{8+\delta}}, \text{ si } r < 2 + \tfrac{\delta}{4}, \\ w_n(r) &\leq C, \text{ si } r < 2 + \tfrac{\delta}{4}, \text{ et} \\ z_n(r) &\leq C(h_n^d)^{\frac{4+\delta}{8+\delta}}, \text{ si } r < 2 + \tfrac{\delta}{4}. \end{aligned}$$

Les conditions iii),...,vi) du Théorème 2.2 sont alors satisfaites avec $\delta_0 \in]0, \frac{\delta}{2}[$, $\gamma_0 = \frac{1}{2} + \frac{2}{8+\delta} - \frac{1}{4+\delta_0}$, $\gamma_1 \in]0, \frac{4+\delta}{8+\delta}\gamma]$, où γ est tel que $\limsup_n n^\gamma h_n^d < \infty$. Précisons maintenant la forme de la variance asymptotique. On a:

$$
\begin{aligned}
&E[H_n^*(Z_0,\overline{Z}_0)]^2 \\
&= \frac{1}{h_n^{3d}} E\left[(Y_0-\mu(X_0))(\overline{Y}_0-\mu(\overline{X}_0))\int_A K\left(\frac{x-X_0}{h_n}\right)K\left(\frac{x-\overline{X}_0}{h_n}\right)\omega(x)dx\right]^2 \\
&= \frac{1}{h_n^{3d}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\sigma^2(u)\sigma^2(v)\left[\int_A K\left(\frac{x-u}{h_n}\right)K\left(\frac{x-v}{h_n}\right)\omega(x)dx\right]^2 f(u)f(v)dudv \\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}(\sigma^2 f)(v+uh_n)(\sigma^2 f)(v) \\
&\qquad\qquad \times\left[\int_{\mathbb{R}^d}K(y)K(y+u)(\omega\cdot\mathbb{I}_A)(v+(u+y)h_n)dy\right]^2 dudv.
\end{aligned}
$$

Par hypothèse les fonctions $\sigma^2 f$ et ω sont continues sur l'intérieur de A et bornées sur A^ϵ, ce qui permet de conclure, d'après le théorème de la convergence dominée que:

$$
\lim_{n\to+\infty} E[H_n^*(Z_0,\overline{Z}_0)]^2 = \eta_2^2 = \int_A \sigma^4(x)f^2(x)\omega^2(x)dx\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}K(u+z)K(u)du\right)^2 dz.
$$

On déduit d'après le Théorème 2.2 que V_n^* converge en loi, lorsque n tend vers l'infini, vers la loi normale de moyenne zéro et de variance $2\eta_2^2$.

Normalité asymptotique de $J_n - EJ_n$: Le résultat annoncé découle maintenant des points précédents, de l'égalité (4.5) et du Théorème 2.2. □

Bibliographie

[1] Bickel, P.J., Rosenblatt, M.: On some global measures of the deviations of density function estimates. Ann. Statist., **1**, 1071-1095 (1973).

[2] Bradley, R.C.: Basic properties of strong mixing conditions. In: Eberlein, E. et al.: Dependence in Probability and Statistics, a survey of recent results, pp. 165-192, Boston: Birkhäuser 1986.

[3] Dvoretzky, A.: Asymptotic normality for sums of dependent random variables. In: Proc. 6th Berkeley Sympo. Math. Statist. Probab., Univ. Calif. 1970, vol. 2, 513-535 (1972).

[4] Hall, P.: Central limit theorem for integrated square error properties of multivariate nonparametric density estimators. J. Multivariate Anal., **14**, 1-16 (1984).

[5] Hall, P.: Integrated square error properties of kernel estimators of regression functions. Ann. Statist., **12**, 241-260 (1984).

[6] Konakov, V.D.: On a global measure of deviation for an estimate of the regression line. Theory Probab. Appl., **22**, 858-868 (1977).

[7] Nadaraya, E.A.: On estimating regression. Theory Probab. Appl., **9**, 141-142 (1964).

[8] Nadaraya, E.A.: The limit distribution of the quadratic deviation of nonparametric estimates of the regression function. Soobshch. Akad. Nauk Gruz. SSR, **74**, 33-36 (1974). (En Russe).

[9] Nadaraya, E.A.: A limit distribution of the square error deviation of nonparametric estimators of the regression function. Z. Wahrsch. Verw. Gebiete, **64**, 37-48 (1983).

[10] Parzen, E.: On estimation of a probability density function and mode. Ann. Math. Statist., **33**, 1065-1076 (1962).

[11] Rosenblatt, M.: Remarks on some non-parametric estimates of a density function. Ann. Math. Statist., **27**, 832-837 (1956).

[12] Rosenblatt, M.: A quadratic measure of deviation of two-dimensional density estimates and a test of independence. Ann. Statist., **3**, 1-14 (1975) [correction **10**, 646 (1982)].

[13] Takahata, H., Yoshihara, K.: Central limit theorems for integrated square error of non-parametric density estimators based on absolutely regular random sequences. Yokohama Math. J., **35**, 95-111 (1987).

[14] Tenreiro, C.: Théorèmes de limite centrale pour des U-Statistiques dégénérées. Manuscrit non publié (1993).

[15] Watson, G.S.: Smooth regression analysis. Sankhya A, **26**, 359-372 (1964).

Departamento de Matemática, Universidade de Coimbra, Apartado 3008, 3000 Coimbra, Portugal.