Z. Wahrscheinlichkeitstheorie verw. Gebiete
66, 441–460 (1984)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Propriétés de convergence presque complète du prédicteur à noyau

Gérard Collomb

Université Paul Sabatier, Laboratoire de Statistique et Probabilité,
(E.R.A.-C.N.R.S. n° 591), 118, route de Narbonne F-31062 Toulouse, France

Summary. Let $(Z_n)_{\mathbb{N}}$ be a ϕ-mixing process which is valued in $\mathbb{E} \subset \mathbb{R}^q$, g be a real measurable function defined on $\mathbb{E}$, s and k be two positive integers. We suppose the existence of a function R satisfying $R(.) = E(g(Z_{n+s})/[Z_{n-k+1}, \ldots, Z_n] = .)$, $\forall n \in \mathbb{N}$, $n \geqq k$, and estimate the function R from a sequence $\{Z_i, i=1, \ldots, n\}$ by R_n with

$$R_n(u) = \frac{\sum_{i=k}^{n-s} g(Z_{i+s}) K([u-(Z_{i-k+1}, \ldots, Z_i)]\, h_n^{-1})}{\sum_{i=k}^{n-s} K([u-(Z_{i-k+1}, \ldots, Z_i)]\, h_n^{-1})}, \qquad \forall u \in \mathbb{E}^k,$$

where K is a kernel of $\mathbb{R}^{kq}$ and $h_n \in \mathbb{R}$, $h_n > 0$. We give various conditions, on both the sequences $(h_n)_{\mathbb{N}}$ and the sequence $(\phi_n)_{\mathbb{N}}$ which is associated with $(Z_n)_{\mathbb{N}}$, for the two following properties: uniform complete convergence to R for the functional estimator R_n and complete convergence to 0 for the real random variable $R_n(Z_{n-k+1}, \ldots, Z_n) - R(Z_{n-k+1}, \ldots, Z_n)$. This last result concerns the predictor $R_n(Z_{n-k+1}, \ldots, Z_n)$ of $g(Z_{n+s})$ from $\{Z_i, i=1, \ldots, n\}$ when the process $(Z_n)_{\mathbb{N}}$ is stationary and markovian of order k.

These results are proved here for a more general problem: estimation of a regression $E(Y/X)$ from $\{(X_i, Y_i), i=1, \ldots, n\}$ when these couples are not independent. We also give a lemma which is an extension of the Bernstein inequality to the case of ϕ-mixing r.r.v.

1. Introduction

On désigne par $(\xi_n)_{\mathbb{N}}$ un processus dont l'espace d'états est une partie mesurable $\mathbb{E}_0$ de $\mathbb{R}$ et on considère le problème de la prédiction de ξ_{N+1} à partir de la suite $\{\xi_i, i=1, \ldots, N\}$. On suppose que ce processus $(\xi_n)_{\mathbb{N}}$ n'est assujetti à aucune hypothèse permettant d'utiliser des méthodes paramétriques: les hypothèses que nous introduirons ultérieurement seront uniquement des hypothèses de régularité très générales.

On cherche alors à «estimer», à partir de la suite $\{\xi_i, i=1, \dots, N\}$, la v.a.r. $E(\xi_{N+1}/\xi_1, \dots, \xi_N)$ qui, lorsque $(\xi_n)_{\mathbb{N}}$ est markovien d'ordre k, est égale à $\rho(\xi_{\mathbb{N}-k+1}, \dots, \xi_N)$ où ρ est la fonction définie sur $\mathbb{E}_0^k$ par

$$\rho(u_1, \dots, u_k) = E(\xi_{N+1}/[\xi_{N-k+1}=u_1, \dots, \xi_N = u_k]), \quad \forall u \in \mathbb{E}_0^k. \tag{1.1}$$

Une prédiction de ξ_{N+1} est alors donnée par la v.a.r. $\rho_N(\xi_{N-k+1}, \dots, \xi_N)$ appelée prédicteur et définie à l'aide d'un estimateur ρ_N de la fonction ρ: l'estimateur fonctionnel ρ_N que nous considérons ici est défini, pour tout u de $\mathbb{E}_0^k$, par [dans toutes les définitions de ce type nous convenons que $0/0=0$]

$$\rho_N(u) = \sum_{i=k}^{N-1} \xi_{i+1} K((u - [\xi_{i-k+1}, \dots, \xi_i])/h) \Big/ \sum_{i=k}^{N-1} K((u - [\xi_{i-k+1}, \dots, \xi_i])/h) \tag{1.2}$$

où h est un nombre strictement positif dépendant uniquement de N et K est un noyau - sa définition figure plus bas - de $\mathbb{R}^k$.

Nous donnons, pour un problème de prédiction un peu plus général (par exemple $\mathbb{E}_0 \subset \mathbb{R}^q$, $q \geqq 1$) que celui qui vient d'être introduit, les propriétés de convergence presque complète suivantes:

$$\sup_{u \in G} |\rho_{\mathbb{N}}(u) - \rho(u)| \xrightarrow[N\to\infty]{\text{P. Co.}} 0 \quad \text{et} \quad |\rho_N(\xi_{N-k+1}, \dots, \xi_N) - \rho(\xi_{N-k+1}, \dots, \xi_N)| \xrightarrow[N\to\infty]{\text{P.Co.}} 0, \tag{1.3}$$

où G est un compact de $\mathbb{E}_0^k$, avec $h = h_N$ vérifiant des propriétés asymptotiques peu restrictives.

Ces résultats sont donnés dans le paragraphe 3, dont la lecture peut être effectuée indépendamment de celle des autres paragraphes. Ces propriétés de convergence sont obtenues presque directement à partir de théorèmes généraux énoncés dans le paragraphe 2 et donnant les propriétés semblables à (1.3) pour un estimateur analogue de la régression $r(.) = E(Y_i/X_i = .)$, lorsque les couples aléatoires (X_i, Y_i), $i=1, \dots, n$ ne sont pas nécessairement indépendants.

Il suffit en effet de poser

$$n = N-k \quad \text{et} \quad (X_j, Y_j) = ([\xi_j, \dots, \xi_{j+k-1}], \xi_{j+k}), \quad j=1, \dots, n, \tag{1.4}$$

pour voir que l'estimateur (1.2) s'écrit alors

$$\rho_N(u) = \sum_{i=1}^{n} Y_i K((u-X_i)/h) \Big/ \sum_{i=1}^{n} K((u-X_i)/h). \tag{1.5}$$

On retrouve là la définition de l'estimateur à noyau de la régression: cet estimateur non paramétrique a été abondamment étudié - nous renvoyons à Collomb, 1981, pour une revue bibliographique - lorsque les couples (X_i, Y_i), $i = 1, 2, \dots$, sont indépendants et équidistribués. Il est clair que dans bon nombre de problèmes de Statistique Appliquée ces hypothèses d'indépendance et d'équidistribution ne sont pas vérifiées, alors que peuvent être plus acceptables des hypothèses d'indépendance asymptotique et de distributions non identiques mais possédant la même fonction de régression. Cette remarque justifie le fait que les résultats obtenus pour l'estimateur de régression (1.5) – notation simplifiée par (1.4) de (1.2) – sont énoncés (dans le paragraphe 2) indépendamment du problème de prédiction qui est ici à l'origine de son étude.

Les résultats des paragraphes 2 et 3 sont prouvés dans le paragraphe 4, au début duquel est donné un lemme généralisant l'inégalité de Bernstein au cas de v.a.r. ϕ-mélangeantes.

Remarques préliminaires et bibliographie sur la prédiction non paramétrique

Les travaux relatifs à la prédiction non paramétrique concernent en général le cas d'un processus $(\xi_n)_{\mathbb{N}}$ stationnaire et markovien: (1.1) est alors satisfaite pour tout N de $\mathbb{N}$ et les ξ_i, $i \in \mathbb{N}$, sont équidistribuées.

L'estimateur ρ_N défini par (1.2) a été proposé par Watson (1964, p. 369–370) à propos d'un problème de prévision météorologique (ξ_i = «température au jour i»). Roussas (1969), pour $k=1$ et u fixé, montre la convergence en probabilité de $\rho_N(u)$. Doukhan et Ghindès (1980, 1983 avec une définition voisine de 1.2) et Collomb et Doukhan (1983, avec $\mathbb{E}_0 = \mathbb{R}^q$, $q \geqq 1$) par une technique de démonstration différente de celle utilisée ici, obtiennent une majoration de limite nulle pour $E|\rho_N(u)-\rho(u)|^2$, u fixé, ainsi que de $\int E|\rho_N(u)-\rho(u)|^2 m(u)\,du$ où m est une fonction positive donnée a priori: ces résultats apportent une information sur la vitesse de convergence du prédicteur ρ_N. Robinson (1983) s'intéresse à la loi limite de v.a.r. associées à $\rho_N(u)$; des résultats semblables à ceux des deux derniers articles cités sont obtenus par Yakowitz (1983). Les propriétés en un point fixé d'estimateurs analogues à ρ_N sont données par Georgiev (1983a, 1983b) et Yakowitz (1979a, 1979b), nous mentionnons enfin les travaux de Banon (1978), Nguyen et Pham (1981) et Pham (1981) qui, pour des problèmes d'estimation différents mais voisins de celui abordé ici, donnent les propriétés de convergence ponctuelle de divers estimateurs récursifs.

Il est évident que ρ_N, en tant qu'estimateur de la fonction d'autorégression ρ, constitue un outil pour la connaissance du processus $(\xi_n)_{\mathbb{N}}$ et donc que l'obtention de propriétés de convergence (ponctuelle ou en norme) pour l'estimateur fonctionnel ρ_N présente un intérêt certain. Cependant une telle approche du problème de la prédiction n'est pas entièrement satisfaisante. Il est clair en effet, pour $k=1$ par exemple, que le problème «naturel» de la prédiction n'est pas celui de l'estimation de $\rho(u)$ pour u fixé ou de $\rho(.)$ sur un sous-ensemble de $\mathbb{E}_0$, mais que le problème «naturel» est celui de «l'estimation» de la v.a.r. $\rho(\xi_N)$ où ξ_N est par nature une v.a.r. non indépendante de ξ_i, $i=1,\ldots,N-1$.

Bosq (1983a, 1983b) étudie la quantité $E|\tilde{\rho}_N(\xi_N)-\rho(\xi_N)|^2$ pour une classe très vaste – contenant en particulier le prédicteur à noyau $\rho_N(\xi_N)$ – de prédicteurs non paramétriques $\tilde{\rho}_N$ mais suppose connue la loi de ξ_1. Collomb (1980, 1983a) étudie cette même quantité lorsque $\tilde{\rho}_N$ est le prédictogramme, dont la définition est analogue à celle de l'histogramme ou du régressogramme: le problème traité dans ce dernier article est identique à celui qui est introduit ici dans le paragraphe 3. Notre présent travail donne des résultats concernant simultanément l'estimateur fonctionnel ρ_N et le prédicteur $\rho_N(\xi_{N-k+1},\ldots,\xi_N)$.

Bien que les résultats obtenus concernent le cas d'un processus $(\xi_n)_N$ non nécessairement markovien, il est clair que le cas markovien constitue un domaine privilégié d'application de ces résultats. Le modèle non paramétrique ainsi introduit s'oppose alors aux modèles paramétriques classiques (notamment autorégressifs, du type ARMA par exemple): ce problème du choix du modèle en analyse de séries temporelles n'est pas abordé ici. Un exemple de discussion sur ce problème est donné par Yakowitz (1979a) à propos de phénomènes hydrologiques, pour lesquels la validité d'un modèle non paramétrique markovien est mise en évidence. Enfin nous terminons cette brève revue bibliographique en mentionnant les études par simulation effectuées par Kalaidjian (1981), Bosq (1983a) et Doukhan (1983).

2. Principaux résultats

Soit (X_i, Y_i), $i=1,2,\ldots$, un processus dont l'espace d'états est $(\mathbb{E}\times\mathbb{R})$ où $\mathbb{E}$ est un borélien non vide de $\mathbb{R}^p$, $p<\infty$. On suppose l'existence d'une fonction r définie sur $\mathbb{E}$ telle que

$$E(Y_i/X_i)=r(X_i), \qquad \forall\, i\in\mathbb{N}, \tag{2.1}$$

en notant que cette condition est satisfaite dès que (X_i, Y_i), $i\in\mathbb{N}$ est un processus stationnaire.

On désigne par G un compact de $\mathbb{E}$ et on suppose que

$$\exists\, \gamma>0: P(X_i\in B)\geqq\gamma\mu(B), \qquad \forall\, i\in\mathbb{N},\ \forall\, B\in\mathscr{B}_{\hat{G}}, \tag{2.2}$$

où $\hat{G}$ est un ε-voisinage de G dans $\mathbb{E}$ et μ est la mesure de Lebesgue sur $\mathbb{R}^p$. En outre les v.a. X_i et Y_i, $i=1,2,\ldots$, sont assujetties aux hypothèses

$$\exists\, \Gamma<\infty: P(X_i\in B)\leqq\Gamma\mu(B), \qquad \forall\, i\in\mathbb{N},\ \forall\, B\in\mathscr{B}_{\mathbb{E}} \tag{2.3}$$

i.e. la loi P_{X_i} de X_i est uniformément (pour tout i de $\mathbb{N}$) absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{E}$, et

$$\exists\, M<\infty: |Y_i|\leqq M, \qquad \forall\, i\in\mathbb{N}. \tag{2.4}$$

Enfin, on désigne par $(\phi_n)_{\mathbb{N}}$ une suite réelle telle que le processus $\{U_i=(X_i, Y_i), i\in\mathbb{N}\}$ vérifie

(M) pour tous entiers n et k positifs on $a|P(A\cap B)-P(A)P(B)|\leqq\phi_n P(A)$ pour tout événement A [resp. B] appartenant à la tribu engendrée par la suite U_i, $i=1,\ldots,k$ [resp. U_i, $i=n+k, n+k+1,\ldots$], la suite $(\phi_n)_{\mathbb{N}}$ étant prise décroissante.

et on suppose que cette suite satisfait

$$\phi_n \xrightarrow[n\to\infty]{} 0 \tag{2.5}$$

[(X_i, Y_i), $i\in\mathbb{N}$, est ϕ-mélangeant au sens de Billingsley, 1968, p. 164].

On considère l'estimateur r_n de r défini, pour tout x de $\mathbb{E}$, par

$$r_n(x)=\sum_{i=1}^{n} Y_i K((x-X_i)/h_n) \Big/ \sum_{i=1}^{n} K((x-X_i)/h_n), \qquad \forall n\in\mathbb{N}, \tag{2.6}$$

où $(h_n)_{\mathbb{N}}$ est une suite réelle strictement positive de limite nulle et K est un noyau de $\mathbb{R}^p$, c.à.d. une fonction de $\mathbb{R}^p$ satisfaisant

$$\begin{gathered}|K(.)|\leqq\tilde{K}<\infty, \qquad \int|K(z)|\,dz=\bar{K}<\infty,\\ |z|^p K(z)\xrightarrow[|z|\to\infty]{}0 \quad \text{et} \quad \int K(z)\,dz>0.\end{gathered} \tag{2.7}$$

Nous donnons d'abord un résultat de convergence ponctuelle.

Théorème 1. *Si la suite* $(h_n)_{\mathbb{N}}$ *vérifie conjointement avec* $(\phi_n)_{\mathbb{N}}$ *l'hypothèse*

(H) $\quad n h_n^p/(m_n \log n)\xrightarrow[n\to\infty]{}\infty$ (2.8)

où $(m_n)_{\mathbb{N}}$ *est une suite entière croissante vérifiant*

$$\exists A<\infty: n\phi_{m_n}/m_n\leqq A, \ 1\leqq m_n\leqq n, \qquad \forall n\in\mathbb{N}, \tag{2.9}$$

alors on a

$$r_n(x)\xrightarrow[n\to\infty]{\text{P.Co.}}r(x) \tag{2.10}$$

dès que x *fixé dans* G *est un point de continuité pour* r.

Nous énonçons maintenant une propriété de convergence uniforme.

Théorème 2. *Si le noyau* K *est lipschitzien et si la condition* (H) *est satisfaite, alors on a*

$$\sup_{x\in G}|r_n(x)-r(x)|\xrightarrow[n\to\infty]{\text{P.Co.}}0 \tag{2.11}$$

dès que r *est continue sur un* ε*-voisinage de* G *dans* $\mathbb{E}$.

Enfin on donne une première application du théorème 2.

Théorème 3. *Lorsque* $\mathbb{E}$ *est un compact de* $\mathbb{R}^p$ *et* (2.2) *est satisfaite pour* $\hat{G}=\mathbb{E}$, *alors, si* (H) *est vérifiée,* K *lipschitzien et* r *continue sur* $\mathbb{E}$ *on a*

$$r_n(T)-r(T)\xrightarrow[n\to\infty]{\text{P.Co.}}0 \tag{2.12}$$

pour toute v.a. T *- par exemple* $T=X_{n+1}$ *- à valeurs dans* $\mathbb{E}$.

Ces résultats appellent les remarques suivantes.

Remarque 2.1 (l'hypothèse (H) sur $(h_n)_{\mathbb{N}}$ et $(\phi_n)_{\mathbb{N}}$). L'hypothèse conjointe sur $(h_n)_{\mathbb{N}}$ et $(\phi_n)_{\mathbb{N}}$ qui est définie par (H) est satisfaite dès que l'une des trois hypothèses ci-dessous est vérifiée

(H1) le processus (X_i, Y_i), $i\in\mathbb{N}$, est m-dépendant (i.e. $\phi_i=0$, $\forall i\geqq m$) et $n h_n^p/\mathrm{Log}\, n\xrightarrow[n\to\infty]{}\infty$ (i.e. on peut choisir $(m_n)_{\mathbb{N}}$ constante).

(H2) le processus (X_i, Y_i), $i \in \mathbb{N}$ est géométriquement mélangeant (i.e. $\exists\, \rho > 0$, $a > 0$, $\rho < 1$, $a < \infty$: $\phi_n \leqq a\rho^n$, $\forall\, n \in \mathbb{N}$) et $n h_n^p/(\mathrm{Log}\, n)^2 \xrightarrow[n\to\infty]{} \infty$,

(H3) $\exists\, w > 1$, $a > 0$: $\phi_n \leqq a n^{-w}$, $\forall\, n \in \mathbb{N}$ et $n h_n^p/(n^{1/(1+w)} \mathrm{Log}\, n) \xrightarrow[n\to\infty]{} \infty$.

Par ailleurs on note que les résultats considérés ici donnent un support mathématique aux considérations intuitives suivantes: le nombre h_n doit être pris d'autant plus grand que la dépendance entre des couples éloignés est grande.

Remarque 2.2 (v.a. indépendantes équidistribuées). Lorsque les couples (X_i, Y_i), $i = 1, \ldots, n$ sont indépendants (ou 0-dépendants) et équidistribués, (2.1) est satisfaite (par 2.4) et le problème de l'estimation de r est alors le problème classique de l'estimation de la régression. Dans ce cas particulier:

. les propriétés (2.10) et (2.11) de l'estimateur r_n ont été étudiées par de nombreux auteurs, dont les résultats sont améliorés par les théorèmes 1 et 2 ci-dessus où l'hypothèse (H) sur $(h_n)_{\mathbb{N}}$ est remplacée par (H_1): Nadaraya (1970), Schuster et Yakowitz (1979), Cheng et Taylor (1980) Mack, et Silverman (1982) obtiennent (2.10) ou (2.11) avec sur $(h_n)_N$ des hypothèses plus restrictives (du type $n^{1-\alpha} h_n^p \xrightarrow[n\to\infty]{} \infty$, avec $\alpha > 0$) que (H1).

. l'hypothèse $n h_n^p/\mathrm{Log}\, n \xrightarrow[n\to\infty]{} \infty$ aètè introduite par Devroye (1979; 1981, p. 1317) qui donne un résultat très voisin du théorème 1 et Collomb (1976, 1979) qui obtient les résultats (2.10) et (2.11) sans supposer que $|Y_1|$ est bornée (seulement une hypothèse «locale» relative aux fonctions $E(e^{\alpha Y_1}/X_1)$, $\alpha \in \mathbb{R}$, sur un ε-voisinage de G) et qui montre que l'hypothèse (H1) sur $(h_n)_N$ est une condition nécessaire pout (2.11). A propos de ce dernier résultat de convergence uniforme nous mentionnons également les travaux de Révész (1979), Johnston (1982) et Liero (1982) relatifs à la loi limite de v.a.r. analogues à (2.11), en notant que ces travaux mettent en oeuvre des techniques de démonstration (assez différentes de celles utilisées ici) développées dans le récent ouvrage de Csörgö et Révész (1981).

. le résultat (2.12) appliqué à une v.a. T indépendante de (X_i, Y_i), $i = 1, \ldots, n$ et ayant même distribution que X_1 est une extension de la propriété de convergence universelle (i.e. 2.12 avec «P.Co» remplacé par «en norme L_p, $p \geqq 1$») établie par Devroye et Wagner (1980) et Spiegelman et Sacks (1980).

Remarque 2.3 (cas d'un processus markovien). Un problème «naturel» d'estimation de régression dans lequel les v.a. (X_i, Y_i), $i = 1, \ldots, n$ sont équidistribuées mais non indépendantes est le suivant: estimer $E(Y_1/X_1)$ lorsque le processus (X_n, Y_n), $n \in \mathbb{N}$, est stationnaire et markovien.

Une question «naturelle» est alors la suivante: pour quelles suites $(m_n)_{\mathbb{N}}$ la condition (2.9) est-elle alors satisfaite? Rosenblatt (1971) montre que si le processus (X_n, Y_n), $n \in \mathbb{N}$, vérifie la condition de L_p norme (–, p. 206) pour $p = 1$ ou ∞ ou la condition de Doeblin (–, p. 209 ou Doob, 1953, p. 192), alors (2.9) est satisfaite pour toute suite entière $(m_n)_{\mathbb{N}}$ vérifiant $m_n \geqq \log n$, $\forall\, n \in \mathbb{N}$ (cas H2 de la remarque 1).

3. Applications à la prédiction

Soit $(Z_n)_{n\in\mathbb{N}}$ un processus (non nécessairement stationnaire ou markovien) dont l'espace d'états est $(\mathbb{E}_1, B_{\mathbb{E}_1})$ où $\mathbb{E}_1$ est un borélien non vide de $\mathbb{R}^q$, $q<\infty$. On désigne par k et s deux entiers positifs et par g une fonction mesurable bornée définie sur $\mathbb{E}_1$.

On suppose que

$$\begin{aligned}&\text{il existe une fonction } R \text{ définie sur } \mathbb{E}=\mathbb{E}_1^k \text{ et telle que}\\ &R(Z_{i-k+1},\dots,Z_i)=E(g(Z_{i+s})/[Z_{i-k+1},\dots,Z_i]), \qquad \forall i\in\mathbb{N},\ i=k,\end{aligned} \tag{3.1}$$

et on considère le problème de l'estimation, à partir de la suite $\{Z_i,\ i=1,\dots,n\}$, de la fonction R et de la v.a.r. $R(Z_{n-k+1},\dots,Z_n)$.

Il est clair que ce problème est étroitement lié à celui de la prédiction: on sait en effet que le meilleur (pour une fonction de perte quadratique) prédicteur de $g(Z_{n+s})$ à partir des Z_i, $i=1,\dots,n$ est $E(g(Z_{n+s}/[Z_1,\dots,Z_n])$ $=R(Z_{n-k+1},\dots,Z_n)$ lorsque $(Z_n)_{\mathbb{N}}$ est markovien d'ordre k.

On suppose en outre que le processus (Z_n), $n\in\mathbb{N}$, est ϕ-mélangeant, en ce sens que $\{U_i=Z_i,\ i\in\mathbb{N}\}$ vérifie (M) avec

$$\phi_n \xrightarrow[n\to\infty]{} 0. \tag{3.2}$$

On remarque tout de suite que lorsque (Z_n), $n\in\mathbb{N}$, est strictement stationnaire et mélangeant au sens de la théorie ergodique (voir Bradley, 1980, théorème 1) cette condition est réalisée dès que

$$\exists\, n\in\mathbb{N}:\ \phi_n<1.$$

Enfin, ce processus $(Z_n)_{\mathbb{N}}$ est également assujetti aux hypothèses

$$\exists\,\Gamma<\infty:\ P([Z_{i-k+1},\dots,Z_i]\in B)\leqq\Gamma\mu(B), \qquad \forall i\in\mathbb{N}, i\geqq k \quad \forall B\in\mathscr{B}_{\mathbb{E}} \tag{3.3}$$

et, pour un compact G de $\mathbb{E}$,

$$\exists\,\gamma>0:\ P([Z_{i-k+1},\dots,Z_i]\in B)\geqq\gamma\mu(B), \qquad \forall i\in\mathbb{N}, i\geqq k \quad \forall B\in\mathscr{B}_{\hat G} \tag{3.4}$$

où $\hat G$ est un ε-voisinage de G dans $\mathbb{E}$.

On considère l'estimateur R_n de R défini, pour tout t de $\mathbb{E}$, par

$$R_n(t)=\frac{\sum_{i=k}^{n-s} g(Z_{i+s})K((t-[Z_{i-k+1},\dots,Z_i])/h_n)}{\sum_{i=k}^{n-s} K((t-[Z_{i-k+1},\dots,Z_i])/h_n)} \tag{3.5}$$

où K est un noyau - fonction de $\mathbb{R}^{kq}$ satisfaisant (2.7) - et $(h_n)_{\mathbb{N}}$ est une suite réelle strictement positive de limite nulle vérifiant conjointement avec la suite $(\phi_n)_{\mathbb{N}}$

$$n\,h_n^{kq}/(m_n \operatorname{Log} n)\xrightarrow[n\to\infty]{}\infty \tag{3.6}$$

où $(m_n)_{\mathbb{N}}$ est une suite entière croissante vérifiant

$$\exists A<\infty: n\phi_{m_n}/m_n<A, \quad m_n\in[1,\dots,n], \qquad \forall n\in\mathbb{N}. \tag{3.7}$$

On remarque que lorsque le processus $(Z_n)_{\mathbb{N}}$ est

. strictement stationnaire, markovien et vérifie la condition de norme L_p de Rosenblatt (1971, p. 206-211) pour $p=1$ ou ∞, où la condition de Doeblin (–, p. 209 ou Doob, 1953, p. 192),

. de la forme «$Z_{n+1}=f(Z_n)+\varepsilon_n$», f bornée et les ε_n, $n\in\mathbb{N}$, étant des v.a. indépendantes équidistribuées et de loi équivalente à μ (Doukhan et Ghindés, 1980), alors ce processus est géométiquement mélangeant et la condition (3.6) peut s'écrire

$$n h_n^{kq}/(\log n)^2 \xrightarrow[n\to\infty]{} \infty. \tag{3.8}$$

Les théorèmes 1, 2 et 3 donnent immédiatement les résultats suivants.

Proposition 1. *Si u fixé dans G est un point de continuité pour R on a*

$$R_n(u) \xrightarrow[n\to\infty]{\text{P. Co.}} R(u). \tag{3.9}$$

Proposition 2. *Si R est continue sur un ε-voisinage de G et si K est lipschitzien, alors on a*

$$\sup_{u\in G}|R_n(u)-R(u)| \xrightarrow[n\to\infty]{\text{P. Co.}} 0. \tag{3.10}$$

Proposition 3. *On suppose que $\mathbb{E}$ est un compact de $\mathbb{R}^{kq}$ et que l'hypothèse* (3.4) *est satisfaite pour $\hat{G}=\mathbb{E}$. Si R est continue sur $\mathbb{E}$ et si K est lipschitzien alors on a*

$$R_n(Z_{n-k+1},\dots,Z_n)-R(Z_{n-k+1},\dots,Z_n) \xrightarrow[n\to\infty]{\text{P. Co.}} 0. \tag{3.11}$$

Remarque. Il est clair que l'ensemble des résultats asymptotiques ci-dessus ne permet pas de répondre à la question du praticien: comment choisir au mieux h_n à partir de la suite $Z_1,\dots,Z_n$. Une méthode du type «cross-validation» est proposée par Collomb (1983a, Appendix) pour ce dernier problème. Nous mentionnons enfin le travail de Collomb (1983b) qui donne une extension de ces résultats au cas d'un noyau K non nécessairement lipschitzien et d'une suite $(h_n)_N$ aléatoire définie par la méthode des k points les plus proches.

4. Démonstrations

Les démonstrations des théorèmes 2.1., 2.3. sont données dans le sous-paragraphe 4.2. et les résultats du paragraphe 3 sont prouvés dans le sous-paragraphe 4.3. Nous donnons auparavant un résultat préliminaire dont la formulation est indépendante des notations précédemment introduites.

4.1. *Inégalité de Bernstein pour un processus ϕ-mélangeant*

Le lemme ci-dessous constitue une extension de l'inégalité de Bernstein [cf. Rényi, 1966, p. 362] au cas de v.a.r. non indépendantes.

Lemme 1. *Soit, pour tout n de $\mathbb{N}$, une suite de v.a.r. $\Delta_i = \Delta_{ni}$, $i \in \mathbb{N}$, vérifiant*

$$E\Delta_i = 0, \quad |\Delta_i| \leqq d, \; E|\Delta_i| \leqq \delta \quad \text{et} \quad E\Delta_i^2 \leqq D, \tag{4.1}$$

et ϕ mélangeantes, la suite $\{\phi_k, k \in \mathbb{N}\}$ étant indépendante de n; on pose

$$\tilde{\phi}_m = \sum_{i=1}^{m} \phi_i, \qquad \forall m \in \mathbb{N}. \tag{4.2}$$

Pour tout nombre $\varepsilon > 0$ et tout entier n on a

$$P\left(\left|\sum_{i=1}^{n} \Delta_i\right| > \varepsilon\right) \leqq c\, e^{-\alpha\varepsilon + \alpha^2 nC} \tag{4.3}$$

où

$$C = 6(D + 4\delta d \tilde{\phi}_m) \quad \text{et} \quad c = 2\, e^{3\sqrt{e}n\phi_m/m} \tag{4.4}$$

et où et m sont respectivement un nombre positif et un entier inférieur à n vérifiant

$$\alpha m d \leqq 1/4. \tag{4.5}$$

(les nombres α et m, ainsi que d, δ et D, pouvant dépendre de n).

Remarque. Ce résultat améliore le résultat similaire donné par Bosq (1975) qui obtient une inégalité de forme analogue à (4.3) mais avec $c = a\,\phi_m e^{bn/m}$, $0 < a, b < \infty$. Nous notons seulement ici que le lemme 1 donne immédiatement la loi des grands nombres pour des variables bornées et ϕ-mélangeantes, sous l'unique condition que $(\phi_n)_N$ soit de limite nulle.

Démonstration. On pose

$$S = \sum_{i=1}^{n} \Delta_i.$$

La démonstration du résultat (4.3) relatif à $P(|S| > \varepsilon)$ repose sur

. l'application de l'inégalité de Markov à la v.a.r. $e^{\alpha S}$

. une majoration de $Ee^{\alpha S}$ qui est obtenue à l'aide d'un raisonnement par récurrence portant sur les éléments d'une partition de la suite $\{\Delta_i, i = 1, \ldots, n\}$ en deux groupes de sous-ensembles suffisamment peu dépendants entre eux.

La démonstration débute par ce second point.

On désigne par N un entier vérifiant conjointement avec n et m

$$(*1) \qquad 2m(N-1) \leqq n \leqq 2mN.$$

Pour tout $j = 1, 2$ et $k = 1, \ldots, N$ on pose

$$(*2) \qquad A_{j,k} = \sum_{i=i_1}^{i_2} \Delta_i$$

où

$$i_1 = \inf[(2k+j-3)m+1, n] \quad \text{et} \quad i_2 = \inf[i_1+m-1, n]$$

ainsi que

$$B_{j,k} = \sum_{t=1}^{k} A_{j,t} \quad \text{avec } B_{j,0} = 0.$$

On a

$$S = B_{1,N} + B_{2,N}$$

d'où il vient

$$(*3) \qquad E e^{\alpha S} \leqq 2^{-1}(E e^{\beta B_{1,N}} + E e^{\beta B_{2,N}}), \quad \text{avec } \beta = 2\alpha.$$

Nous considérons maintenant, pour tout $j=1, 2$ et $k=1, \ldots, N$ la quantité

$$E e^{\beta B_{j,k}} = E(e^{\beta B_{j,k-1}} e^{\beta A_{j,k}}).$$

En appliquant un résultat bien connu (voir par exemple Billingsley, 1968, p. 171, f. 20–28) sur l'espérance du produit de deux éléments d'une suite de v.a.r. ϕ-mélangeantes on obtient

$$(*4) \qquad E e^{\beta B_{j,k}} \leqq E e^{\beta B_{j,k-1}} E e^{\beta A_{j,k}} + 2 E e^{\beta B_{j,k-1}} \|e^{\beta A_{j,k}}\|_\infty \phi_m.$$

L'hypothèse (4.5) sur α, l'égalité $\beta = 2\alpha$ et l'inégalité

$$|A_{j,k}| \leqq m d$$

amènent

$$(*5) \qquad |\beta A_{j,k}| \leqq \tfrac{1}{2}.$$

L'implication «$|x| \leqq 1/2 \Rightarrow e^x \leqq 1 + x + x^2$» permet alors d'écrire

$$e^{\beta A_{j,k}} \leqq 1 + \beta A_{j,k} + \beta^2 A_{j,k}^2$$

et l'inégalité «$1 + x \leqq e^x$» donne

$$1 + \beta^2 E A_{j,k}^2 \leqq e^{\beta^2 E A_{j,k}^2}$$

d'où il vient, puisque $E\Delta_i = 0$, $\forall i \in \mathbb{N}$,

$$(*6) \qquad E e^{\beta A_{j,k}} \leqq e^{\beta^2 E A_{j,k}^2}.$$

On a d'après $(*2)$

$$E A_{j,k}^2 = \sum_{i=i_1}^{i_2} E \Delta_i^2 + \sum_{i=i_1}^{i_2} \sum_{\substack{i'=i_1 \\ i' \neq i}}^{i_2} E \Delta_i \Delta_{i'}.$$

En appliquant aux variables centrées Δ_i et $\Delta_{i'}$ le résultat général qui a donné $(*4)$, on obtient

$$|E \Delta_i \Delta_{i'}| \leqq 2 \delta d \phi_{|i-i'|}$$

d'où il vient, en tenant compte de (4.2),

$$EA_{j,k}^2 \leqq m[D + 4d\delta\tilde{\phi}_m]$$

qui avec (∗6) et (4.4) amène

$$E\,e^{\beta A_{j,k}} \leqq e^{\beta^2 mC/6}.$$

Cette dernière inégalité et l'inégalité (∗4) donnent, en tenant compte de (∗5)

$$(*7) \qquad E\,e^{\beta B_{j,k}} \leqq w E\,e^{\beta B_{j,k-1}}$$

où

$$w = e^{\beta^2 mC/6} + 2\,e^{1/2}\,\phi_m = e^{\beta^2 mC/6}(1 + \gamma\,\phi_m)$$

avec

$$\gamma = 2\,e^{1/2 - \beta^2 mC/6} \leqq 2\,e^{1/2}.$$

On a donc

$$w \leqq e^{\beta^2 mC/6}\,(1 + 2\sqrt{e}\;\phi_m)$$

qui avec (∗7) amène

$$E\,e^{\beta B_{j,N}} \leqq e^{\beta^2 NmC/6}\,(1 + 2\sqrt{e}\;\phi_m)^N.$$

En tenant compte de la partie gauche de (∗1) et de la majoration grossière «$m \leqq n$» qui donnent $N \leqq 3n/(2m)$, on obtient par (∗3)

$$E\,e^{\alpha S} \leqq c_0\,e^{\alpha^2 nC}$$

où

$$c_0 = (1 + 2\sqrt{e}\;\phi_m)^{\frac{3n}{2m}} = e^{\frac{3n}{2m}\log(1 + 2\sqrt{e}\phi_m)}$$

vérifie, d'après l'inégalité «$\log(1+x) \leqq x \;\forall x \geqq 0$»,

$$c_0 \leqq e^{3\sqrt{e}\,n\phi_m/m}.$$

On a donc, en considérant la seconde partie de (4.4),

$$(*8) \qquad E\,e^{\alpha S} \leqq (c/2)\,e^{\alpha^2 nC}$$

Pour achever la démonstration il suffit alors de remarquer que

$$P(|S| > \varepsilon) \leqq P(S > \varepsilon) + P(-S > \varepsilon)$$

et que, d'après l'inégalité de Markov,

$$P(S > \varepsilon) \leqq e^{-\alpha\varepsilon}\,E\,e^{\alpha S} \quad \text{et} \quad P(-S > \varepsilon) \leqq e^{-\alpha\varepsilon}\,E\,e^{-\alpha S}.$$

Puisque les hypothèses du lemme 1 sur les v.a.r. Δ_i, $i \in \mathbb{N}$, sont également vérifiées par les v.a.r. $-\Delta_i$, $i \in \mathbb{N}$, nous obtenons alors grâce à l'inégalité (∗8) la majoration (4-3).

4.2. Démonstration des résultats du paragraphe 2

La définition (2-6) peut aussi s'écrire

$$r_n(x)=R_{1,n}(x)/R_{0,n}(x),$$

où pour $j=0, 1$

$$R_{j,n}(x)=n^{-1}\sum_{i=1}^{n} Y_i^j K_n(x-X_i) \quad \text{avec } K_n(u)=h^{-p}K(u/h), \qquad \forall u\in R^p$$

(à l'intérieur des démonstrations et en l'absence d'ambiguité le nombre h_n sera noté h).

Il suffit d'écrire

$$\begin{aligned} r_n(x)-r(x)=&[(R_{1,n}(x)-ER_{1,n}(x))-r(x)(R_{0,n}(x)-ER_{0,n}(x)) \\ &+(ER_{1,n}(x)-r(x)ER_{0,n}(x))]/R_{0,n}(x) \end{aligned}$$

pour obtenir, en tenant compte de (2.1) et (2.4),

$$\begin{aligned} \sup_{x\in G_0}|r_n(x)-r(x)|\leqq&[\sup_{x\in G_0}|R_{1,n}(x)-ER_{1,n}(x)|+M\sup_{x\in G_0}|R_{0,n}(x)-ER_{0,n}(x)| \\ &+\sup_{x\in G_0}|ER_{1,n}(x)-r(x)ER_{0,n}(x)|]/\inf_{x\in G_0}R_{0,n}(x). \end{aligned}$$

Les propriétés de l'estimateur r_n qui sont données par les théorèmes 1 et 2 sont une conséquence immédiate de l'inégalité ci-dessus, avec $G_0=G$[resp. $G_0=\{x\}$] pour le théorème 2 [resp. 1], et des propriétés de

. convergence presque complète vers 0 des deux premiers termes de la somme entre crochets, établie dans le lemme 3,

. minoration asymptotiquement presque complète de $\inf_{x\in G_0}R_{0,n}(x)$ par un nombre strictement positif, donnée par le lemme 4,

. convergence vers 0 du troisième terme (non aléatoire) de la somme entre crochets, fournie par le lemme 5.

La propriété de convergence presque complète énoncée dans le théorème 3 découle de l'inégalité

$$|r_n(T)-r(T)|\leqq\sup_{x\in\mathbb{E}}|r_n(x)-r(x)|$$

et du théorème 2 que l'on peut alors appliquer à $G=\mathbb{E}$.

Les lemmes 3, 4 et 5 sont des conséquences du résultat préliminaire suivant obtenu par application du lemme 1 à la v.a.r.

$$R_n(x)=n^{-1}\sum_{i=1}^{n} Z_i K_n(x-X_i) \tag{4.6}$$

avec $Z_i=Y_i$, $i=1, \dots, n$ (cas $R_n=R_{1,n}$) ou $Z_i=1$, $i=1, \dots, n$ (cas $R_n=R_{0,n}$) et donc

$$|Z_i|\leqq\tilde{M}<\infty, \qquad i=1, \dots, n \quad \text{où} \quad \tilde{M}=\sup(M, 1). \tag{4.7}$$

Lemme 2. *Soit H une partie non vide de $\mathbb{E}$. Si $(m_n)_{\mathbb{N}}$ est une suite entière vérifiant* (2.9), *on a pour tout $\varepsilon>0$*

$$\sup_{x\in H} P(|R_n(x)-ER_n(x)|>\varepsilon)\leqq a e^{-bnh_n^p/m_n} \tag{4.8}$$

pour tout entier naturel n suffisamment grand, a et b étant des constantes positives indépendantes de n, h_n et m_n. [Cette inégalité est précisée par la formule (4.9) *ci-dessous].*

Démonstration. Pour tout x fixé dans H on a

$$R_n(x)-ER_n(x)=\sum_{i=1}^{n} \Delta_i$$

avec

$$\Delta_i=n^{-1} Z_i K_n(x-X_i)-En^{-1} Z_i K_n(x-X_i), \qquad i=1, \ldots, n.$$

On prend alors comme valeurs possibles de d, δ et D intervenant dans (4.1)

$$(*1) \qquad d=n^{-1} h^{-p} 2\tilde{M}\tilde{K}, \qquad \delta=n^{-1} 2\tilde{M}\bar{K}\Gamma, \qquad D=h^{-p} n^{-2} \tilde{M}^2 \tilde{K}\bar{K}\Gamma.$$

Ces majorants sont obtenus comme suit: l'inégalité (4.7) et la première partie de (2.7) amènent d'une part

$$|n^{-1} Z_i K_n(x-X_i)|\leqq \tilde{M}\tilde{K}/(nh^p)$$

qui donne immédiatement le choix ci-dessus de d, et d'autre part, pour $k=1, 2$,

$$E|n^{-1} Z_i K_n(x-X_i)|^k\leqq(n^{-1} \tilde{M})^k(\tilde{K}/h^p)^{k-1} E|h^{-p} K(x-X_i)/h)|$$

où

$$E|h^{-p} K(x-X_i)/h)|=\int h^{-p} K((x-z)/h) P_{X_i}(dz)\leqq \Gamma\bar{K}$$

d'après (2.3) et la seconde partie de (2.7). Ces deux dernières inégalités fournissent la valeur δ pour $k=1$ et la valeur D pour $k=2$ données dans $(*1)$. On pose

$$\beta=(8\tilde{M}\tilde{K})^{-1} \quad \text{et} \quad B=6\beta \tilde{M}^2 \tilde{K}\bar{K}\Gamma.$$

En appliquant alors le lemme 1 à $\alpha=\beta nh^p/m$ on obtient (après avoir remarqué que (4.5) est vérifiée) par (4.3) et pour tout n de $\mathbb{N}$

$$\sup_{x\in H} P(|R_n(x)-ER_n(x)|>\varepsilon)\leqq c_m e^{-(nh^p/m)\,t(\varepsilon, m)} \tag{4.9}$$

où $m\in\mathbb{N}$, $1\leqq m\leqq n$, et

$$t(\varepsilon, m)=\beta[\varepsilon-B(m^{-1}+16\tilde{\phi}_m m^{-1})] \quad \text{et} \quad c_m=2e^{3\sqrt{\varepsilon n \phi_m/m}}.$$

Nous considérons maintenant séparément les deux cas possibles pour la suite croissante $(m_n)_{\mathbb{N}}$ satisfaisant (2.9)

- premier cas, «$m_n \xrightarrow[n\to\infty]{} \infty$»: l'hypothèse (2.5) montre alors qu'il existe un entier n_0 pour lequel on a

$$m_n^{-1}\leqq\varepsilon/(4B) \quad \text{et} \quad \tilde{\phi}_{m_n} m_n^{-1}\leqq\varepsilon/(64B), \qquad \forall\, n\geqq n_0,$$

d'où il vient

$$t(\varepsilon, m_n) \geqq \beta\varepsilon/2, \qquad \forall\, n \geqq n_0;$$

il suffit de choisir $m=m_n$, $\forall\, n\in\mathbb{N}$, $n\geqq n_0$, dans (4.9) pour obtenir (4.8), avec $b=\beta\varepsilon/2$, après avoir noté que l'inégalité $c_{m_n}\leqq a<\infty$, $\forall\, n>n_0$, provient de l'hypothèse (2.9).

. second cas, «$\exists\, n_1$, $m_0\in\mathbb{N}$: $m_n=m_0$, $\forall\, n\in\mathbb{N}$, $n\geqq n_1$»: l'hypothèse (2.5) montre alors qu'il existe un entier m indépendant de n et satisfaisant les inégalités

$$m\geqq m_0, \qquad m^{-1}\leqq \varepsilon/(4B), \qquad \tilde{\phi}_m m^{-1}\leqq \varepsilon/(64B).$$

Ces inégalités amènent

$$t(\varepsilon, m)\geqq \beta\varepsilon/2$$

ainsi que, de par la décroissance de la suite $(\phi_n)_{\mathbb{N}}$ et la condition (2.9),

$$n\phi_m/m\leqq n\phi_{m_0}/m_0\leqq A<\infty, \qquad \forall\, n\in\mathbb{N},\ n\geqq n_1.$$

En posant $b=(\beta\varepsilon/2)\, m_0/m$ on obtient par (4.9)

$$\sup_{x\in H} P(|R_n(x)-ER_n(x)|>\varepsilon)\leqq 2e^{3A\sqrt{e}}\, e^{-(nh^p/m_0)b}, \qquad \forall\, n\in\mathbb{N},\ n\geqq n_0,$$

avec $n_0=\max(m, n_1)$, ce qui prouve (4.8) pour la suite $m_n=m_0$, $\forall\, n\in\mathbb{N}$.

Lemme 3. *Si la suite* $(h_n)_{\mathbb{N}}$ *vérifie* (2-8) *pour une suite* $(m_n)_{\mathbb{N}}$ *satisfaisant* (2.9), *la v.a.r.* $R_n(x)$ *définie par* (4.6) *vérifie pour tout* x *fixé dans* G

$$R_n(x)-ER_n(x)\xrightarrow[n\to\infty]{\text{P.Co.}} 0. \tag{4.10}$$

Si de plus K est Lipschitzien on a

$$\sup_{x\in G}|R_n(x)-ER_n(x)|\xrightarrow[n\to\infty]{\text{P.Co.}} 0. \tag{4.11}$$

Démonstration. En appliquant le lemme 2 à $H=\{x\}$ on obtient immédiatement

$$\forall\, \varepsilon>0, \qquad \sum_{n=1}^{\infty} P(|R_n(x)-ER_n(x)|>\varepsilon)\leqq \sum_{n=1}^{\infty} a e^{-bnh_n^p/m_n}$$

et l'hypothèse (2.8) implique la convergence de la série constituant la partie droite de l'inégalité ci-dessus qui donne alors le résultat (4.10).

Nous prouvons maintenant le résultat (4.11) et précisons d'abord l'hypothèse «K lipschitzien» en disant que K est lipschitzien de rapport (ou constante) $C<\infty$ et d'ordre $\gamma>0$, ce qui permet d'écrire

(*0) $$|K_n(z)-K_n(z^*)|\leqq Ch_n^{-(p+\gamma)}|z-z^*|^{\gamma}, \qquad \forall\, z, z^*\in\mathbb{R}^p.$$

Nous nous fixons un réel α qui vérifie

(*1) $$\alpha>p/\gamma+1.$$

Pour tout n de $\mathbb{N}$ nous désignons par $\{B_k,\ k=1,\ldots,l_n\}$ un recouvrement de G par l_n boules de centres t_k, $k=1,\ldots,l_n$ et de rayons inférieurs ou égaux à h_n^α. Puisque l'ensemble G est borné, il est possible de choisir ce recouvrement de manière à ce que pour tout n de $\mathbb{N}$ on ait

$$(*2)\qquad l_n \leqq L h_n^{-\alpha p}$$

où L est une constante positive indépendante de n (ne dépendant que de G). Nous posons

$$(*3)\qquad S_n(x)=R_n(x)-ER_n(x), \qquad \forall\, x\in G.$$

Soit x un point fixé dans G. Il existe alors un point t_k tel que x appartienne à la boule B_k et on a

$$(*4)\qquad |S_n(x)| \leqq |S_n(t_k)|+|\tilde{S}_n(x)| \quad \text{où} \quad \tilde{S}_n(x)=S_n(x)-S_n(t_k).$$

Les formules $(*3)$ et (4.6) montrent que

$$\tilde{S}_n(x)$$
$$=n^{-1}\sum_{i=1}^{n} Z_i[K_n(x-X_i)-K_n(t_k-X_i)]-En^{-1}\sum_{i=1}^{n} Z_i[K_n(x-X_i)-K_n(t_k-X_i)]$$

d'où il vient, en tenant compte de (4.7)

$$|\tilde{S}_n(x)| \leqq \tilde{M} n^{-1}\sum_{i=1}^{n}(|K_n(x-X_i)-K_n(t_k-X_i)|+E|K_n(x-X_i)-K_n(t_k-X_i)|)$$

qui avec $(*0)$ donne

$$|\tilde{S}_n(x)| \leqq 2\tilde{M} C h_n^{-(p+\gamma)}|x-t_k|^\gamma.$$

Puisque t_k a été choisi de manière à ce que la boule B_k contienne x, on a $|x-t_k| \leqq h_n^\alpha$, d'où il vient

$$|\tilde{S}_n(x)| \leqq 2\tilde{M} C h_n^{\alpha\gamma-(p+\gamma)}.$$

Cette majoration indépendante de x, l'inégalité $(*1)$ et l'hypothèse $h_n \xrightarrow[n\to\infty]{} 0$ impliquent

$$\sup_{x\in G}|\tilde{S}_n(x)| \xrightarrow[n\to\infty]{} 0 \qquad \text{(convergence certaine).}$$

Ce dernier résultat et l'inégalité $(*4)$ montrent que pour prouver (4.11) il suffit de prouver

$$(*5)\qquad \max_{k=1,\ldots,l_n}|S_n(t_k)| \xrightarrow[n\to\infty]{\text{P.Co.}} 0.$$

Pour tout $\varepsilon>0$ on pose

$$\beta_n=P(\max_{k=1,\ldots,l_n}|S_n(t_k)|>\varepsilon).$$

On a

$$\beta_n \leqq \sum_{k=1}^{l_n} P(|S_n(t_k)|>\varepsilon) \leqq l_n \sup_{x\in G} P(|S_n(x)|>\varepsilon)$$

d'où il vient, en appliquant le lemme 2 à $H=G$ et en tenant compte des inégalités (*1) et (*2)

$$\beta_n \leqq aLh_n^{-p(1+p/\gamma)}\, e^{-bnh_n^p/m_n}.$$

L'hypothèse (2.8) sur h_n implique d'abord

$$e^{-bnh_n^p/m_n}=n^{-b\gamma_n} \quad \text{où} \quad \gamma_n \xrightarrow[n\to\infty]{} \infty$$

et ensuite l'existence d'une constante δ pour laquell on a

$$(nh_n^p)^{-\xi}<\delta \quad \text{où} \quad m=1+p/\gamma$$

pour n assez grand. Il s'ensuit l'inégalité

$$\beta_n \leqq aL\delta n^{\xi-b\gamma_n}$$

qui donne

$$\sum_{n=1}^{\infty} \beta_n < \infty$$

et donc (*5), ce qui achève la démonstration du lemme.

Lemme 4. *Dans l'énoncé du lemme 3 on peut remplacer* (4.10) *et* (4.11) *par*

$$\exists\, \delta>0\colon \sum_{n=1}^{\infty} P(R_{0,n}(x) \leqq \delta)<\infty \tag{4.12}$$

et

$$\exists\, \delta>0\colon \sum_{n=1}^{\infty} P(\inf_{x\in G} R_{0,n}(x) \leqq \delta)<\infty. \tag{4.13}$$

Démonstration. On a pour $H=\{x\}$ ou G

(*1) $$\inf_{x\in H} R_{0,n}(x) \geqq \inf_{x\in H} ER_{0,n}(x) - \sup_{x\in H} |R_{0,n}(x)-ER_{0,n}(x)|.$$

Pour tout x de H

$$ER_{0,n}(x)=n^{-1}\sum_{i=1}^{n} EK_n(x-X_i)$$

peut s'écrire

$$ER_{0,n}(x)=n^{-1}\sum_{i=1}^{n} \int h_n^{-p} K((x-z)/h)\, P_{X_i}(dz).$$

L'hypothèse (2.2) et la dernière partie de (2.7) donnent alors

$$\exists\, \theta>0,\ \exists n_0>0\colon\ ER_{0,n}(x) \geqq \theta \int K(z)\,dz>0. \qquad \forall x\in H,\ \forall n\in\mathbb{N},\ n>n_0.$$

Cette dernière minoration, l'inégalité (*1) et le résultat (4.10) [resp. 4.11] sur $R_n=R_{0,n}$ amènent immédiatement le résultat (4.12) [resp. 4.13]: il suffit par exemple de choisir $\delta=\theta\int K(z)\,dz/2$.

Lemme 5. *Si la fonction r est continue au point x on a*

$$|ER_{1,n}(x) - r(x)\,ER_{0,n}(x)| \xrightarrow[n\to\infty]{} 0, \tag{4.14}$$

et si elle est continue sur un ε-voisinage de G on a

$$\sup_{x\in G} |ER_{1,n}(x) - r(x)\,ER_{0,n}(x)| \xrightarrow[n\to\infty]{} 0. \tag{4.15}$$

Démonstration. On remarque d'abord que la première (on pose alors $\{x\}=G$) et la seconde (r est alors uniformément continue sur un ε-voisinage de G) hypothèse sur r permettent d'écrire

(∗1) $\forall\alpha>0,\ \exists\beta_\alpha>0$: $\|u-x\|\leqq\beta_\alpha$, $|r(u)-r(x)|<\alpha$, $\forall x\in G$, β_α indépendant de x où $\|.\|$ est la norme de $\mathbb{R}^p$.

Pour tout x de G on a

$$(\ast 2)\qquad ER_{1,n}(x) - r(x)\,ER_{0,n}(x) = n^{-1}\sum_{i=1}^{n} r_{n,i}(x)$$

où

$$r_{n,i}(x) = E(Y_i - r(x))\,K_n(x-X_i) = E((E\,Y_i/X_i) - r(x))\,K_n(x-X_i)$$

vérifie d'après la définition (2.1)

$$r_{n,i}(x) = E(r(X_i) - r(x))\,K_n(x-X_i), \qquad i=1,\dots,n.$$

L'hypothèse (2.3) sur la loi de X_i amène

$$(\ast 3)\qquad |r_{n,i}(x)| \leqq \Gamma \int |r(u)-r(x)|\,h^{-p}\,|K((x-u)/h)|\,du.$$

L'hypothèse (2.4), qui implique $|r(.)-r(x)|\leqq 2M$, $\forall\,x\in G$, et la formule (∗1) montrent que pour tout $\alpha>0$ on a

$$|r(u)-r(x)| \leqq \alpha\,1_{\{\|u-x\|\leqq\beta_\alpha\}} + 2M\,1_{\{\|u-x\|>\beta_\alpha\}} \qquad \forall\,x\in G,\ \forall\,u\in\mathbb{E},$$

où β_α est indépendant de u et x. Le changement de variable $t=(x-u)/h$ dans l'intégrale figurant dans l'inégalité (∗3) et la majoration ci-dessus amènent

$$|r_{n,i}(x)| \leqq \alpha\Gamma \int 1_{\{\|ht\|\leqq\beta_\alpha\}}|K(t)|\,dt + 2M\Gamma \int 1_{\{\|t\|>\beta_\alpha/h\}}|K(t)|\,dt$$

qui avec (2.7) donne

$$(\ast 4)\qquad |r_{n,i}(x)| \leqq \alpha\Gamma\bar{K} + 2M\Gamma\gamma_\alpha(h)$$

où $\gamma_\alpha(.)$ est une fonction positive satisfaisant

$$(\ast 5)\qquad \gamma_\alpha(u) \xrightarrow[u\to 0]{} 0.$$

La majoration (∗4), indépendante de x, et l'égalité (∗2) permettent d'écrire

$$(\ast 6)\qquad \forall\,\alpha>0,\quad |ER_{1,n}(x) - r(x)\,ER_{0,n}(x)| \leqq \alpha\Gamma\bar{K} + 2M\Gamma\gamma_\alpha(h), \qquad \forall\,x\in G.$$

çe qui prouve (4.14) et (4.15) puisque on a alors

$$\forall \varepsilon>0 \ \exists n_0 \in \mathbb{N}: \forall n>n_0 \sup_{x\in G} |ER_{1,n}(x)-r(x)R_{0,n}| \leqq \varepsilon.$$

Il suffit en effet de choisir $\alpha=\varepsilon/(2\Gamma\bar{K})$ dans (*6) et d'utiliser (*5) qui montre que

$$\exists n_0 \in \mathbb{N}: \forall n>n_0, \quad \gamma_\alpha(h_n)<\varepsilon(4\Gamma M)^{-1}$$

puisque $(h_n)_{\mathbb{N}}$ est de limite nulle.

4.3. Démonstration des propositions 1, 2 et 3 du 3° paragraphe

On applique les théorèmes 1, 2 et 3 du paragraphe 2 aux couples

$$X_i=[Z_i, \ldots, Z_{i+k-1}], \quad Y_i=g(Z_{i+k+s-1}), \quad \forall i \in \mathbb{N},$$

et, pour la proposition 3 à $T=[Z_{n-k+1}, \ldots, Z_n]$, avec $p=kq$ ainsi que $R=r$, γ et Γ définis par (3.4 et 3.3), M bornant la fonction $|g(.)|$, et

$$(*1) \qquad \phi_j=\phi_{j-t} \quad \text{si } j>t, \qquad 1 \quad \text{si } j\leqq t,$$

où φ_n représente ici le terme ϕ_n introduit dans le paragraphe 3 et où

$$t=k+s-1.$$

Il suffit en effet de remarquer d'abord que la condition de φ-mélange sur $(Z_n)_{\mathbb{N}}$ implique que le processus $(U_n)_{\mathbb{N}}$ défini par

$$U_i=[Z_i, Z_{i+1}, \ldots, Z_{i+t}], \quad \forall i \in \mathbb{N}$$

est ϕ-mélangeant avec $(\phi_n)_{\mathbb{N}}$ défini par (*1), et ensuite que pour tout n de $\mathbb{N}$, les tribus engendrées par $\{U_i, i\leqq n\}$ et $\{U_i, i\geqq n\}$ sont respectivement identiques aux tribus engendrées par $\{(X_i, Y_i), i\leqq n\}$ et $\{(X_i, Y_i), i\geqq n\}$.

On constate alors que la majoration (3.2) donne immédiatement (2.5) et que l'hypothèse (3.6) concernant les suites $(\varphi_n)_{\mathbb{N}}$ et $(h_n)_{\mathbb{N}}$ implique que l'hypothèse (H) est vérifiée par cette même suite $(h_n)_{\mathbb{N}}$, conjointement avec la suite $(\phi_n)_{\mathbb{N}}$.

Références

Banon, G.: Nonparametric identification for diffusion processes. Siam J. Control Optim. **16**, 380–395 (1978)

Billingsley, P.: Convergence of Probability Measures. New York: Wiley 1968

Bosq, D.: Inégalité de Bernstein pour les processus stationnaires et mélangeants. Applications. C.R. Acad. Sci. Paris, **281**, série A, 1095–1098 (1975)

Bosq, D.: Non Parametric Prediction for a stationary Process. Lecture Notes in Statistics. No. **16**, 69–84. Berlin-Heidelberg-New York: Springer 1983

Bosq, D.: Sur la prédiction non paramétrique des variables aléatoires et de mesures aléatoires. Z. Wahrscheinlichkeitstheorie verw. Gebiete **64**, 541–553 (1983b)

Bradley, R.C.: On the ϕ-mixing condition for stationary random sequences. Duke Math. J. **47**, 421–433 (1980)

Cheng, K.F., Taylor, R.L.: On the uniform complete convergence of estimates for multivariate density functions and regression curves. Ann. Inst. Statist. Math. **32**, Part. A, 187–199 (1980)

Collomb, G.: Estimation non paramétrique de la régression par la méthode du noyau. Thèse, Université Paul Sabatier, Toulouse (1976)

Collomb, G.: Conditions nécessaires et suffisantes de convergence uniforme d'un estimateur de la régression, estimation des dérivées de la régression. C.R. Acad. Sci. Paris, **288**, Série A, 161–164 (1979)

Collomb, G.: Prédiction non paramétrique: étude de l'erreur quadratique du prédictogramme. Pub. Int., Laboratoire de Statistique et Probabilités, Université Paul Sabatier, Toulouse, France (1980)

Collomb, G.: Estimation non paramétrique de la régression: revue bibliographique. Int. Stat. Rev., **49**, 75–93 (1981)

Collomb, G.: From non parametric regression to non parametric prediction: survey of the mean square error and original results on the predictogram. Lecture Notes in Statistics. No. **16**, 182–204. Berlin-Heidelberg-New York: Springer 1983

Collomb, G.: Analyse d'une série temporelle et prédiction non paramétriques: méthodes de la fenêtre mobile et des k points les plus proches en estimation de régression pour des observations dépendantes. Prépublication (1983b)

Collomb, G., Doukhan, P.: Estimation non paramétrique de la fonction d'autorégression d'un processus stationnaire et ϕ-mélangeant: risques quadratiques pour la méthode du noyau. C.R. Acad. Sci. Paris, **296**, Série I, 859–862 (1983)

Csörgő, M., Révész, P.: Strong approximations in probability and statistics. New-York and Budapest: Academic Press 1981

Devroye, L.: The uniform convergence of the Nadaraya-Watson regression function estimate. Canad. J. Statist. **6**, 179–191 (1979)

Devroye, L., Wagner, T.S.: Distribution Free Consistency Results in Nonparametric Discrimination and Regression Function Estimation. Ann. Statist. **8**, 231–239 (1980)

Devroye, L.: On the almost everywhere convergence of nonparametric regression function estimates. Ann. Statist. **9**, 1310–1319 (1981)

Doob, J.: Stochastic Processes. New-York: Wiley 1953

Doukhan, P.: Simulations in the General First Order Autoregressive Process (Unidimensional Normal Case). Lecture Notes in Statistics, **16**, 50–68. Berlin-Heidelberg-New York: Springer 1983

Doukhan, P., Ghindès, M.: Estimation dans le processus «$X_{n+1}=f(X_n)+\varepsilon_n$». C.R. Acad. Sci. Paris, **297**, série A, 61–64 (1980)

Doukhan, P., Ghindès, M.: Estimation dans le processus autorégressif d'ordre 1. Stochastic Processes, à paraître (1983)

Georgiev, A.A.: Nonparametric system identification by kernel method. I.E.E.E. trans. Automat. Control, à paraître [adresse de l'auteur: Technical University of Wroclaw/Wypianskiego 27, I-6/50370/Wroclaw/Pologne] (1983a)

Georgiev, A.A.: A nonparametric algorithm for identification of linear dynamic systems of unknown order. Prépublication [adresse de l'auteur: voir ci-dessus] (1983b)

Johnston, G.J.: Probabilities of Maximal Deviations for Nonparametric Regression Function Estimates. J. Multivariate Anal. **12**, 402–414 (1982)

Kalaidjian, R.: Efficacités comparées de certaines méthodes de prédiction pour un arma perturbé. Statistique et Analyse des Données, **6**, 47–64 (1981)

Liero, H.: On the Maximal Deviation of the Kernel Regression Function Estimate. Math. Operationsforsch. Statist. **13**, 171–182 (1982)

Mack, Y.P., Silverman, B.W.: Weak and Strong Uniform Consistency of Kernel Regression Estimates. Z. Wahrscheinlichkeitstheorie verw. Gebiete **61**, 405–415 (1982)

Nadaraya, E.A.: Remarks on nonparametric estimates for density functions and regressions curves. Theor. Probability Appl. **15**, 134–137 (1970)

Nguyen, H.T., Pham, D.T.: Nonparametric estimation in diffusion model by discrete sampling. Publications de l'Institut de Statistique de l'Université de Paris, XXVI, **2**, 89–109 (1981)

Pham, D.T.: Nonparametric Estimation of the Drift Coefficient in the Diffusion Equation. Math. Operationsforsch. Statist, **12**, 61–74 (1981)

Rényi, A.: Calcul des Probabilités. Paris: Dunold 1966

Révész, P.: On the nonparametric estimation of the regression function. Problems Control Inform. Theory **8**, 297–302 (1979)

Robinson, P.M.: Non parametric estimators for times series. Journal of Time series Analysis, à paraître [Adresse de l'auteur: University of Surrey, Department of Mathematics/Guilford Surrey/GU2 5XH/Grande-Bretagne] (1983)

Rosenblatt, M.: Markov Processes. Structure and Asymptotic Behavior. Berlin-Heidelberg-New York: Springer 1971

Roussas, G.: Nonparametric Estimation of the Transition Distribution Function of a Markov Process. Ann. Math. Statist. **40**, 1386–1400 (1969)

Schuster, E.F., Yakowitz, S.: Contributions to the theory of nonparametric regression, with application to system identification. Ann. Statist. **7**, 139–149 (1979)

Spiegelman, G., Sachs, S.: Consistent window estimation in nonparametric regression estimation. Ann. Statist. **8**, 240–246 (1980)

Yakowitz, S.J.: A nonparametric Markov model for daily river flow. Water Resources Research, **15**, 1035–1043 (1979a)

Yakowitz, S.J.: Nonparametric estimation of Markov transition functions. Ann. Statist. **7**, 671–679 (1979b)

Yakowitz, S.J.: Nonparametric density estimation and prediction for Markov sequences. Prepublication (1983)

Watson, G.S.: Smooth regression analysis. Sankhya, **26**, ser. A, 359–372 (1964)

Reçu le 29 Novembre 1982; en forme revisée le 5 Décembre 1983