



Convergence Uniforme Presque Sure D'une Classe De Prédicteurs A Noyau Pour Un Processus Fortement Mélangeant

Afif Hayek, Jean-Pierre Lecoutre

► To cite this version:

Afif Hayek, Jean-Pierre Lecoutre. Convergence Uniforme Presque Sure D'une Classe De Prédicteurs A Noyau Pour Un Processus Fortement Mélangeant. Annales de l'ISUP, 1990, XXXV (1), pp.23-41. hal-03666456

HAL Id: hal-03666456

<https://hal.science/hal-03666456>

Submitted on 12 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pub. Inst. Stat. Univ. Paris
XXXV, fasc. 1, 1990, 23 à 41

CONVERGENCE UNIFORME

PRESQUE SURE D'UNE CLASSE

DE PREDICTEURS A NOYAU POUR

UN PROCESSUS FORTEMENT MELANGEANT.

(Afif HAYEK, Jean-Pierre LECOUTRE)

Laboratoire de Statistique Théorique et
Appliquée,
Université Paris VI , 4 place Jussieu, 75252
Paris Cedex 05

RESUME

On établit une propriété de convergence uniforme presque sûre d'une classe d'estimateurs de la fonction d'autoregression pour un processus fortement mélangeant. On en déduit alors une classe de prédicteurs non paramétriques associés à un noyau et construits comme un M-estimateur fonctionnel.

ABSTRACT

We prove that a general nonparametric estimator of the autoregression function for a

strictly stationary, strongly mixing process is uniformly convergent in some compact set. The principal application of our results concerns a predictor defined by the kernel method and constructed as a functional M-estimate.

MOTS CLES:

autoregression non paramétrique, prédicteur non paramétrique, processus fortement mélangeant.

CLASSIFICATION AMS: 62 G 05

1) INTRODUCTION

De nombreux travaux récents (Collomb, 1984; Vieu, 1987; Sarda, 1988) ont étudié le problème de la prédiction de X_{T+1} à partir des observations $X_1, \dots, X_T$ d'une série temporelle. Ils se situent dans le cas où $(X_t ; t \in \mathbb{N})$ est un processus réel strictement stationnaire et $\emptyset$ -mélangeant, c'est à dire vérifiant pour tout entier $t > 0$:

$$\emptyset_n = \sup \{ |P(B \setminus A) - P(B)| ; A \in \mathcal{M}_0^t, B \in \mathcal{M}_{t+n}^\infty \} \xrightarrow{n \rightarrow \infty} 0,$$

où $\mathcal{M}_0^t$ (resp. $\mathcal{M}_{t+n}^\infty$) désigne la tribu engendrée par $(X_s; s \leq t)$ [resp. $(X_s; s \geq t+n)$] et avec la convention $P(B \setminus A) = P(B)$ si $P(A) = 0$.

Malheureusement, cette hypothèse est assez forte et n'est pas vérifiée notamment par les processus auto-regressifs généraux (cf Gastwirth, Rubin, 1975). Nous l'affaiblions donc en supposant seulement que le processus $(X_t; t \in \mathbb{N})$ est α -mélangeant (ou fortement mélangeant), c'est à dire tel que pour tout entier $t > 0$:

$$\alpha_n = \sup \{ |P(A \cap B) - P(A)P(B)| ; A \in \mathcal{M}_0^t, B \in \mathcal{M}_{t+n}^\infty \} \xrightarrow{n \rightarrow \infty} 0,$$

Cette condition est alors vérifiée par la plupart des processus ARMA (cf Pham, Tran, 1985). Dans Bosq (1983), les propriétés de convergence d'une classe de prédicteurs non paramétriques font intervenir un coefficient de mélangeance de Gastwirth-Rubin permettant de retrouver les deux mélangeances précédentes. Le cas d'un processus non bornée, géométriquement mélangeant, a également été traité par Bosq (1988). L'étude de la prévision non paramétrique

sous la condition minimale d'ergodicité a été effectuée par Delecroix (1987), mais elle ne permet pas alors l'étude des vitesses de convergence.

2) DEFINITION DE LA CLASSE DE PREDICTEURS

Nous considérons ici la classe générale de prédicteurs non paramétriques, introduite par Collomb, Härdle (1986), définie à partir d'un noyau K comme un M -estimateur d'une fonction d'autoregression r calculée à partir de la loi de X_{s+1} conditionnellement à $(X_1, \dots, X_s)$, où s est un entier fixé strictement positif.

Soit K un noyau de Parzen-Rosenblatt [cf Bosq, Lecoutre (1987)] positif, pair et ψ une application de $\mathbb{R}$ dans $\mathbb{R}$ strictement monotone et continûment différentiable. Pour tout x de $\mathbb{R}^s$, $r(x)$ est définie comme solution en θ de:

$$H(x; \theta) = E[\psi(X_{s+1} - \theta) | (X_1, \dots, X_s) = x] \quad f(x) = 0 \quad (2.1)$$

où f est la densité de $(X_1, \dots, X_s)$.

Si ψ est la fonction identité, $r(x)$ est la

régression usuelle de X_{s+1} sur $(X_1, \dots, X_s)$:

$$r(x) = E(X_{s+1} | [X_1, \dots, X_s] = x).$$

L'estimateur $r_n(x)$ de $r(x)$ associé à cette fonctionnelle est alors défini comme solution de l'équation en θ de:

$$H_n(x; \theta) = \sum_{i=1}^n w_{ni}(x) \psi(X_{i+s} - \theta) = 0 \quad (2.2)$$

où on a posé:

$w_{ni}(x) = [1/(nh_n^s)] K\{[x - (X_i, \dots, X_{i+s-1})]/h_n\}$
 $(h_n; n \in \mathbb{N})$ étant une suite de nombres réels strictement positifs, avec $n = T - s$ ($s < T$).

Si la fonction ψ est choisie bornée, l'estimateur ainsi obtenu possédera des propriétés de robustesse et atténuera notamment l'influence de valeurs aberrantes éventuelles.

3) RESULTAT PRINCIPAL

En suivant le même schéma de démonstration que Collomb, Härdle (1986), nous allons étendre au cas d'un processus α -mélangeant le résultat qu'ils ont obtenu pour un processus $\emptyset$ -mélangeant. Pour cela, nous aurons besoin des

hypothèses suivantes:

$H_1)$ f est uniformément bornée sur $\mathbb{R}^S$, strictement positive sur un compact C , et l'équation (2.1) admet une solution unique.

$H_2)$ Il existe des constantes positives, finies, β et M telle que:

$$\forall \theta \in \mathbb{R}, E\{\exp[\beta\psi(X_1 - \theta)]\} \leq M$$

$H_3)$ $\forall \varepsilon > 0, \exists \lambda > 0$ tel que pour tout réel θ fixé:

$$\sup_{x \in C} \sup_{\{u; |u-x| < \lambda\}} |E[\psi(X_{S+1}-r(x)-\theta) \mid (X_1, \dots, X_S)=u] f(u) - H(x, r(x)+\theta)| \leq \varepsilon$$

Nous pouvons alors énoncer le résultat suivant:

THEOREME

Sous les hypothèses H_1 et H_2 , si on choisit la suite (h_n) telle que:

$$h_n \rightarrow 0, nh_n^S / [m_n (\log n)^2] \rightarrow \infty$$

où m_n est une suite croissante d'entiers, non majorée, avec $1 \leq m_n \leq n/2$ et telle qu'il existe une constante finie A vérifiant:

$$(n/m_n) [1/m_n \sum_{i=1}^{m_n} a_i + e\sqrt{e} (a_{m_n})^{2m_n/(3n)}] < A$$

alors, en tout point x de continuité de r :

$$r_n(x) \xrightarrow{\text{p.s.}} r(x), \quad n \rightarrow \infty$$

Si de plus l'hypothèse H_3 est vérifiée, en prenant un noyau K lipschitzien, alors $r_n(x)$ existe et est p.s. unique pour tout x de C et n suffisamment grand, avec:

$$\sup_{x \in C} |r_n(x) - r(x)| \xrightarrow{\text{p.s.}} 0, \quad n \rightarrow \infty$$

4) APPLICATION A LA PREDICTION

L'équation (2.2) permet de construire une classe générale de prédicteurs $r_T(X_T, \dots, X_{T-s+1})$ de X_{T+1} à partir des observations $(X_1, \dots, X_T)$ qui, dans les conditions du théorème, vérifiera:

$$\{r_T(X_T, \dots, X_{T-s+1}) - E(X_{T+1} | [X_T, \dots, X_{T-s+1}])\}^2 \mathbf{1}_{\{(X_T, \dots, X_{T-s+1}) \in C\}} \xrightarrow{\text{p.s.}} 0 \quad \text{quand } T \rightarrow \infty$$

En choisissant une fonction ψ bornée, on construit ainsi une classe de prédicteurs non paramétriques robustes. Ceci évite l'arbitraire

du choix d'un modèle paramétrique et fournit des prédicteurs simples à construire et réalisant de très bonnes performances. Des simulations effectuées par Kalaidjian (1983), Carbon et Delecroix (1987) sur des processus ARMA, perturbés ou non, ont démontré la qualité de ce type de méthodes.

5) DEMONSTRATION

La fonction ψ étant strictement monotone, pour tout $\varepsilon > 0$ fixé:

$$\forall x \in C, H(x; r(x) + \varepsilon) < 0 < H(x; r(x) - \varepsilon) \quad (5.1)$$

Nous utiliserons alors les résultats suivants:

LEMME 1

Sous les hypothèses de la première partie du théorème:

$$\forall \theta \in \mathbb{R}, \forall x \in C,$$

$$H_n(x; \theta) - E H_n(x; \theta) \xrightarrow{p.s.} 0, n \rightarrow \infty \quad (p.s.)$$

Sous les hypothèses de la seconde partie du théorème:

$$\forall \theta \in \mathbb{R},$$

$$\sup_{x \in C} |H_n(x; \theta) - E H_n(x; \theta)| \xrightarrow{p.s.} 0, n \rightarrow \infty \quad (p.s.)$$

LEMME 2

Sous les conditions de la première partie du théorème:

$$\forall \theta \in \mathbb{R}, \forall x \in C,$$

$$E H_n(x; r(x) + \theta) - H(x; r(x) + \theta) \longrightarrow 0, n \longrightarrow \infty$$

Sous les conditions de la seconde partie du théorème:

$$\forall \theta \in \mathbb{R},$$

$$\sup_{x \in C} |E H_n(x; r(x) + \theta) - H(x; r(x) + \theta)| \longrightarrow 0, n \longrightarrow \infty$$

Ces lemmes permettent de déduire de (5.1), pour n assez grand:

$$\forall x \in C,$$

$$H_n(x; r(x) + \varepsilon) < 0 < H_n(x; r(x) - \varepsilon) \quad (\text{p.s.})$$

Il suffit alors de remarquer que $H_n(x; r_n(x)) = 0$ pour en conclure que:

$$\forall x \in C, r(x) - \varepsilon < r_n(x) < r(x) + \varepsilon \quad (\text{p.s.})$$

Enfin, l'unicité de la solution de (2.1) implique:

$$\exists \theta_0 \in \mathbb{R}, \quad H(x; r(x) + \theta_0) \neq 0 \quad (\text{p.s.})$$

on obtient donc par les lemmes 1 et 2, pour n assez grand:

$$\exists \theta_0 \in \mathbb{R}, \forall x \in \mathbb{C}, H_n(x; r(x) + \theta_0) \neq 0 \quad (\text{p.s.})$$

d'où l'existence et l'unicité de la solution $r_n(x)$ de l'équation (2.2), du fait que ψ est monotone.

6) DEMONSTRATION DES LEMMES

La démonstration du lemme 1 repose sur une extension de l'inégalité de Bernstein aux processus fortement mélangeants, due à M. Carbon (1983).

LEMME 3

Soit $(Z_i; i \in \mathbb{N})$ un processus réel fortement mélangeant tel que:

$$\forall i \in \mathbb{N}, E(Z_i) = 0;$$

$$\forall n \in \mathbb{N}, \forall i \in \mathbb{N}, 1 \leq i \leq n, |Z_i| \leq d_n, E(Z_i^2) \leq D_n$$

$$\text{Alors: } \forall \varepsilon > 0, \forall n \in \mathbb{N}, n \geq 3,$$

$$P\left\{\left|\sum_{i=1}^n Z_i\right| > \varepsilon\right\} \leq 2 \exp\{-n\varepsilon + 4n^2 \exp[D_n + 8d_n^2 \tilde{\alpha}(k)] + 2\sqrt{e} \alpha_k^{2k/(3n)} n/k\}$$

où k est un entier inférieur à $n/2$, η vérifiant

$$0 < \eta \leq 1/(4kd_n e)$$

La fonction ψ n'étant pas nécessairement supposée bornée, pour pouvoir appliquer ce lemme, nous allons tronquer les v.a. $\psi(X_i - \theta)$ au moyen d'une suite croissante M_n de nombres réels positifs, convenablement choisie. Nous écrivons alors:

$$H_n(x; \theta) - E H_n(x; \theta) = \sum_{i=1}^n T_i + \sum_{i=1}^n \Delta_i$$

où on a posé:

$$\Delta_i = w_{ni}(x) \psi(X_{i+s} - \theta) 1_{\{|\psi(X_{i+s} - \theta)| \leq M\}} - E[w_{ni}(x) \psi(X_{i+s} - \theta) 1_{\{|\psi(X_{i+s} - \theta)| \leq M\}}]$$

les v.a. T_i étant obtenues à partir de Δ_i en remplaçant l'évènement $\{|\psi(X_{i+s} - \theta)| \leq M_n\}$ par son complémentaire.

Il est facile de voir que pour tout $\varepsilon > 0$:

$$P\{|H_n(x; \theta) - E H_n(x; \theta)| > 2\varepsilon\} \leq P\left\{\left|\sum_{i=1}^n T_i\right| > \varepsilon\right\} + P\left\{\left|\sum_{i=1}^n \Delta_i\right| > \varepsilon\right\} \quad (6.1)$$

On supposera dans toute la suite que θ est fixé.

Une application de l'inégalité de Markov nous donne d'après H_2 :

$$P\{|\psi(X_{i+s}-\theta)| > M_n\} \leq M e^{-\beta M_n}$$

D'autre part, l'inégalité de Schwarz permet d'obtenir:

$$E \left[\sup_{x \in C} \sum_{i=1}^n T_i \right] \leq 2\bar{k} / (nh_n^S) \sum_{i=1}^n \frac{\sqrt{E[\psi^2(X_{i+s}-\theta)]}}{\sqrt{P\{|\psi(X_{i+s}-\theta)| > M_n\}}}$$

où $\bar{k}$ est un majorant de k .

Une nouvelle utilisation de H_2 et de l'inégalité de Markov conduit à:

$$\forall \varepsilon > 0, \exists c_1 > 0, P\left\{ \sup_{x \in C} \left| \sum_{i=1}^n T_i \right| > \varepsilon \right\} \leq c_1 h_n^{-S} e^{-\beta M_n / 2} \quad (6.2)$$

Par ailleurs, les v.a. Δ_i vérifient les hypothèses du lemme 3 avec:

$d_n = 2\bar{k}M_n / (nh_n^S)$, $D_n = \bar{k}B M_n^2 / (n^2 h_n^S)$, où B est un majorant de f .

On obtient donc, en prenant $\eta = (4kd_n e)^{-1}$:

$\forall \varepsilon > 0,$

$$P\left\{\left|\sum_{i=1}^n \Delta_i\right| > \varepsilon\right\} \leq 2 \exp\{-nh_n^S / (8ekM_n) [\varepsilon - BM_n / (2k)]\} \\ + 2n / (ek) [\tilde{\alpha}(k) / k + e\sqrt{\alpha}(\alpha_k)^{2k / (3n)}]\} \quad (6.3)$$

Choisissons la suite M_n de la forme $M_n = c \log n$, $c > 0$, et $k = m_n$ vérifiant les hypothèses du théorème et telle que, au moins pour n assez grand, $BM_n / (2k) < \varepsilon / 2$. Ainsi, il existe des constantes positives c_2 et c_3 telles que:

$$P\left\{\left|\sum_{i=1}^n \Delta_i\right| > \varepsilon\right\} \leq c_2 \exp[-c_3 nh_n^S / (m_n \log n)] \quad (6.4)$$

D'après le choix de h_n , il est facile de voir que:

$$\sum_{n=1}^{\infty} P\left(\left|\sum_{i=1}^n \Delta_i\right| > \varepsilon\right) < \infty$$

D'autre part, le majorant de (6.2) s'écrit: $c_1 / (nh_n^S n^{c\beta/2 - 1})$; il suffit donc de choisir c tel que $\beta c > 4$ pour avoir le terme général d'une série convergente. La majoration (6.1) permet alors de conclure.

Compte tenu de (6.1) et (6.2), pour démontrer la seconde partie du lemme, il suffit de montrer que:

$$\sup_{x \in \mathbb{C}} \left| \sum_{i=1}^n \Delta_i \right| \xrightarrow[n \rightarrow \infty]{} 0, \quad \text{(p.s.)}$$

Soit $(B_j ; 1 \leq j \leq l_n)$ un recouvrement de C en l_n boules de centre v_j , de rayon inférieur ou égal à $h_n^S / \text{Log } n$, avec:

$$l_n \leq L h_n^{-\delta S} (\text{Log } n)^S$$

où δ est un nombre fixé tel que $\delta > s+1$ et L une constante positive.

Soit x fixé dans C , et notons $\Delta_i(x)$ la v.a. notée Δ_i précédemment; si B_j est la boule contenant x , on peut écrire:

$$|\Delta_i(x)| \leq |\Delta_i(v_j)| + |\tilde{\Delta}_i(x)|$$

où

$$\tilde{\Delta}_i(x) = \Delta_i(x) - \Delta_i(v_j)$$

Ainsi:

$$\begin{aligned} \sum_{i=1}^n \tilde{\Delta}_i(u) &= \sum_{i=1}^n [\omega_{ni}(x) - \omega_{ni}(v_j)] \psi(x_{i+s}^{-\theta}) 1_{\{|\psi(x_{i+s}^{-\theta})| \leq M_n\}} \\ &\quad - E\left\{ \sum_{i=1}^n [\omega_{ni}(x) - \omega_{ni}(v_j)] \psi(x_{i+s}^{-\theta}) 1_{\{|\psi(x_{i+s}^{-\theta})| \leq M_n\}} \right\} \end{aligned}$$

Le noyau K étant lipschitzien, il existe une constante A finie telle que:

$$\sup_{x \in C} \left| \sum_{i=1}^n \tilde{\Delta}_i(x) \right| \leq 2AM_n / (\text{Log } n) h_n^{\delta-(s+1)}$$

le choix de M_n implique que le majorant tend vers zéro.

Il suffit donc de montrer que:

$$\max_{1 \leq j \leq 1_n} \left| \sum_{i=1}^n \bar{\Delta}_i(v_j) \right| \xrightarrow{\text{p.s.}} 0, \quad n \xrightarrow{\text{p.s.}} \infty$$

Pour tout $\varepsilon > 0$, on obtient:

$$P \left\{ \max_{1 \leq j \leq 1_n} \left| \sum_{i=1}^n \bar{\Delta}_i(v_j) \right| > \varepsilon \right\} \leq \frac{1}{n} \sup_{x \in C} P \left\{ \left| \sum_{i=1}^n \Delta_i(x) \right| > \varepsilon \right\}$$

En appliquant la majoration (6.4), on peut trouver une constante positive c_4 telle que:

$$P \left\{ \max_{1 \leq j \leq 1_n} \left| \sum_{i=1}^n \Delta_i(v_j) \right| > \varepsilon \right\} \leq c_4 h_n^{\delta S} (\text{Log } n)^S \exp[-c_3 n h_n^S / (m_n \text{Log } n)]$$

si on pose $\mathcal{V}_n = n h_n^S / [m_n (\text{Log } n)^2]$, le majorant s'écrit:

$$c_4 [1/(n h_n^S)]^S (\text{Log } n)^S n^{\delta - c_3} \mathcal{V}_n$$

c'est bien le terme général d'une série convergente puisque $\mathcal{V}_n \xrightarrow{\text{p.s.}} \infty$. Le lemme 1 est ainsi démontré.

La démonstration du lemme 2 est immédiate par application du lemme de Bochner (cf Bosq, Lecoutre, 1987), en remarquant que:

$$E H_n(x; r(x) + \theta) - H(x; r(x) + \theta) = \\ (1/h_n^S) \int \{E[\psi(X_{i+s} - r(x) - \theta) \mid (X_i, \dots, X_{i+s-1}) = t] f(t) \\ - H(x; r(x) + \theta)\} K[(x-t)/h_n] dt$$

BIBLIOGRAPHIE

- 1) D. Bosq (1983) Sur la prédiction non paramétrique de variables aléatoires et de mesures aléatoires. Z. Wahrsch. verw. Gebiete, 64, 541 - 553.
- 2) D. Bosq (1988) Prédiction non paramétrique d'un processus stationnaire non borné. Publ. du L.S.T.A., n° 81, Univ. Paris VI.
- 3) D. Bosq, J.P.Lecoutre (1987) Théorie de l'estimation fonctionnelle, Economica.
- 4) M. Carbon (1983) Inégalité de Bernstein pour les processus fortement mélangeants, non nécessairement stationnaires. Applications C.R. Acad. Sci. Paris A, 297, 303 - 306.
- 5) M. Carbon, M. Delecroix (1987) Comparaison de méthodes paramétrique et non paramétrique de prévision. Publ. IRMA Lille, vol. 9, n° 11.

- 6) G. Collomb (1984) Propriétés de convergence presque complète du prédicteur à noyau. Z. Wahrsch. verw. Gebiete, 66, 441 - 460.
- 7) G. Collomb, W. Härdle (1986) Strong uniform convergence rates in robust nonparametric time series analysis and prédiction: Kernel regression estimation from dependent observations. Stochastic Processes Appl. 23, 77 - 89.
- 8) M. Delecroix (1987) Sur l'estimation et la prévision non paramétrique des processus ergodiques. Thèse d'Etat, Univ. Lille 1.
- 9) J. Gastwirth, H. Rubin (1975) The asymptotic distribution theory of the empiric CDF for mixing stochastic processes. Ann. Statist. 3, 809 - 824.
- 10) R. Kalaidjian (1983) Efficacités comparées de certaines méthodes de prévision pour un ARMA perturbé. Statist. et Analyse des données 6, 3, 47 - 64.

11) T. Pham, L. Tran (1985) Some mixing properties of time series models. Stochastic Process. Appl. 19, 297 - 303.

12) P. Vieu (1987) Sur les contributions à l'estimation fonctionnelle. Thèse Doctorat, Math. Appli. Univ. Paul Sabatier.

13) P. Sarda (1988) Thèse à paraître.

Reçu en Septembre 1989

1. Introduction

This paper deals with two concepts such as prediction and outliers. Regarding outlier, available literature is vast. Just to cite a few, Barnett and Lewis (1994) is a treatise devoted entirely to this topic. Chen (1973) deals with the estimation problem in the presence of an outlier. Lingappan (1974) gives the estimator of the parameters of exponential and gamma populations in terms of hypergeometric functions.