C. R. Acad. Sci. Paris, Ser. I 350 (2012) 717–720

Contents lists available at SciVerse ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistique

Estimation de la fonction de régression pour variable explicative et réponse fonctionnelles dépendantes

Nonparametric regression for functional response and functional regressor under dependance

Frédéric Ferraty[a], Ali Laksaci[b], Amel Tadj[b], Philippe Vieu[a]

[a] *Institut de mathématiques de Toulouse, université Paul-Sabatier, 118, route de Narbonne, 31062 Toulouse, France*
[b] *Laboratoire de mathématiques, université Djilali-Liabes, BP 89, Sidi Bel Abbes, 22000, Algérie*

INFO ARTICLE

Historique de l'article :
Reçu le 13 janvier 2011
Accepté le 30 juillet 2012
Disponible sur Internet le 22 août 2012

Présenté par le Comité de rédaction

RÉSUMÉ

Nous considérons l'estimateur à noyau de la fonction de régression lorsque la variable réponse est dans un espace de Banach et la variable explicative est à valeurs dans un espace semi-métrique. En considérant des observations β-mélangeantes, on établit la vitesse de convergence presque complète de l'estimateur construit.

ABSTRACT

We consider a kernel estimate of the regression when the response variable is in a Banach space and the explanatory variable takes its values in a semi-metric space. Our main result states the almost complete convergence (with rate) of the constructed estimate when the sample considered is a β-mixing sequence.

Abridged English version

Let us introduce n pairs of random variables $(X_i, Y_i)_{i=1,\dots,n}$ that we suppose drawn from the pair (X, Y), valued in $\mathcal{F} \times \mathcal{B}$, where $\mathcal{F}$ is a semi-metric space, d denoting the semi-metric (i.e. a metric for which $d(x, x') = 0 \not\Rightarrow x = x'$) and $\mathcal{B}$ is a Banach space of type $\tau \in]1, 2]$. For $x \in \mathcal{F}$, one assumes that the conditional mean of Y given $X = x$, denoted $m(x)$, exists. The purpose of this paper is to study the nonparametric estimate $\widehat{m}(x)$ of $m(x)$ when both variables (response and explanatory) are of functional nature (curves, images, ...) and when the observations $(Y_i, X_i)_{i\in\mathbb{N}}$ are β-mixing. We give the almost complete convergence (with rate) of our estimate.

1. Introduction

Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathcal{F} \times \mathcal{B}$, où $\mathcal{B}$ est un espace de Banach de type $\tau \in]1, 2]$ et $\mathcal{F}$ est un espace semi-métrique. On note $\|.\|$ la norme sur $\mathcal{B}$ et on note $d(.,.)$ la semi métrique sur $\mathcal{F}$. On considère une suite de n observations de même loi que le couple (X, Y). Pour $x \in \mathcal{F}$, on suppose que l'espérance conditionnelle de Y sachant $X = x$, notée $m(x) = E[Y|X = x]$, existe. L'estimateur naturel de $m(x)$ par la méthode du noyau est

Adresses e-mail : ferraty@cict.fr (F. Ferraty), alilak@yahoo.fr (A. Laksaci), ameltdz@yahoo.fr (A. Tadj), vieu@cict.fr (P. Vieu).

1631-073X/$ – see front matter
http://dx.doi.org/10.1016/j.crma.2012.07.014

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} K(h^{-1}d(x, X_i))Y_i}{\sum_{i=1}^{n} K(h^{-1}d(x, X_i))}, \quad \forall x \in \mathcal{F}, \tag{1}$$

où $K(.)$ est un noyau et $h := h_n$ est une suite de nombres réels positifs. Le but de cette Note est d'étudier l'estimateur non-paramétrique $\widehat{m}(x)$ de $m(x)$ lorsque les deux variables X et Y sont à valeurs dans un espace de dimension infinie et les observations $(Y_i, X_i)_{i=1,\dots,n}$ sont dépendantes. La modélisation statistique des données fonctionnelles dépendantes est motivée par un grand nombre de problèmes concrets liés à l'étude de phénomènes temporels (voir, pour des exemples, Ramsay et Silverman [9], Bosq [2] pour les modèles paramétriques et Ferraty et Vieu [5], Ferraty et Romain [4] dans le cas non-paramétrique). Typiquement, les séries chronologiques fonctionnelles sont issues d'un découpage d'un processus à temps continu sur des intervalles. Ceci nous permet d'étudier le processus en tenant compte des morceaux de trajectoires continus. D'une manière générale, les premiers résultats sur l'estimation non-paramétrique de la fonction de régression à variable réponse fonctionnelle ont été obtenus par Bosq et Delecroix [3]. En 1990, Lecoutre [8] a étudié la convergence uniforme presque complète d'un estimateur de type régressogramme de la fonction de régression lorsque la variable réponse est dans un espace de Banach et la variable explicative est réelle. Récemment, Ferraty et al. [6] ont obtenu la convergence uniforme presque complète d'un estimateur à noyau d'une variable réponse Banachique conditionnée par une variable explicative dans un espace semi-métrique lorsque les observations sont indépendantes et identiquement distribuées. Nous renvoyons à Ferraty et al. [7] pour l'étude de la normalité asymptotique de cet estimateur.

L'apport principal de cette Note est de généraliser, au cas dépendant, les résultats de Ferraty et al. [6]. Plus précisément, on établit la vitesse de convergence presque complète de $\hat{m}$ lorsque les observations sont β-mélangeantes. L'expression de la vitesse de convergence exploite la structure topologique de la variable explicative et de la variable réponse, ainsi que la structure de dépendance des observations.

Nos hypothèses sont présentées au paragraphe 2 alors que le résultat principal est donné au paragraphe 3.

2. Notations et hypothèses

Nous commençons par rappeler la notion de β-mélange pour une suite de variables aléatoires.

Définition 2.1. Soit $\{Z_i,\ i = 1, 2, \dots\}$ une suite stationnaire de variables aléatoires. Pour tout entier n, on note $\beta(n)$ le coefficient de mélange défini par

$$\beta(n) = \sup\left\{\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}\left|P(A_i \cap B_j) - P(A_i)P(B_j)\right|:\ A_i \in \mathcal{F}_1^k(Z) \text{ et } B_j \in \mathcal{F}_{k+n}^{\infty}(Z),\ I,\ J \text{ et } k \in \mathbb{N}^*\right\},$$

où $\mathcal{F}_i^k(Z)$ est la tribu engendrée par $\{Z_j,\ i \leqslant j \leqslant k\}$. Nous disons que la suite (Z_i) est β-mélangeante si le coefficient de mélange $\beta(n) \to 0$ lorsque $n \to \infty$.

On trouvera dans la littérature plusieurs exemples de processus stochastiques vérifiant cette condition de mélange. Par exemple, dans notre contexte fonctionnel, Allam et Mourid [1] ont donné les conditions suffisantes pour qu'un processus autorégressif Banachique soit β-mélangeant. Par la suite, on note par $B(x, h) = \{x' \in \mathcal{F}/d(x', x) < h\}$ la boule de centre x de rayon h et on introduit les hypothèses suivantes :

(H1) $P(X \in B(x, r)) = \phi_x(r) > 0$ pour tout $r > 0$,
(H2) $\exists b > 0$, $\exists C > 0$, $\exists \mu > 0$, $\forall x_1, x_2 \in B(x, \mu)$, $\|m(x_1) - m(x_2)\| \leqslant C d^b(x_1, x_2)$,
(H3) $(X_i, Y_i)_{i \geqslant 1}$ est une suite β-mélangante dont le coefficient de mélange vérifie

$$\exists a > 2, \quad \beta(n) = O\left(n^{-a}\right),$$

(H4) $\exists C > 0$, $\forall r > 1$, $E(\|Y\|^r | X) \leqslant C$, $r! < \infty$ p.s., avec $r! = r(r-1)\cdots(r-[r]+1)$ et où $[.]$ désigne la partie entière,
(H5) K est une fonction de support $[0, 1[$ vérifiant $0 < C\mathbf{1}_{[0,1[} < K(t) < C'\mathbf{1}_{[0,1[} < \infty$ et si $K(1) = 0$, la fonction $\phi_x(.)$ vérifie la condition supplémentaire

$$\exists C > 0,\ \exists \eta_0 > 0,\ \forall 0 < \eta < \eta_0, \quad \int_0^{\eta} \phi_x(u)\,\mathrm{d}u > C\eta\phi_x(\eta),$$

(H6) $\exists \delta \in (2/a, 1)$, $\lim_{n\to\infty}(n^{1-\delta}\phi_x(h)/\log n) = \infty$.

3. La convergence presque complète

Le théorème ci-dessous établit la convergence ponctuelle (ainsi que la vitesse) de notre estimateur. On pourra remarquer que dans le cas particulier d'indépendance (i.e. $\delta = 0$) on retrouve les vitesses de Ferraty et al. [6]. Qui plus est, si $\mathcal{B} = \mathbb{R}$ est muni de la topologie euclidienne (ou plus généralement si $\mathcal{B}$ est un espace de Hilbert ($\tau = 2$)) on retrouve les vitesses habituelles pour réponse scalaire (Ferraty et Vieu [5]).

Théorème 3.1. *Sous les conditions* (H1)–(H6)*, on a*

$$\left\|\widehat{m}(x)-m(x)\right\| = O\left(h^b\right) + O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\delta}\phi_x(h)}}\right) + O\left(\frac{1}{n^{1-\delta}\phi_x(h)}\right)^{1-1/\tau}.$$

Schéma de la preuve. La démonstration de ce Théorème est basée sur la décomposition suivante :

$$\widehat{m}(x)-m(x) = \frac{1}{\widehat{f}(x)}\left[\widehat{g}(x)-E\widehat{g}(x)\right] + \frac{1}{\widehat{f}(x)}\left[E\widehat{g}(x)-m(x)\right] + \left[1-\widehat{f}(x)\right]\frac{m(x)}{\widehat{f}(x)},$$

où

$$\widehat{f}(x) = \frac{1}{nE[K(h^{-1}d(x,X_1))]}\sum_{i=1}^{n} K\left(h^{-1}d(x,X_i)\right)$$

et

$$\widehat{g}(x) = \frac{1}{nE[K(h^{-1}d(x,X_1))]}\sum_{i=1}^{n} K\left(h^{-1}d(x,X_i)\right)Y_i.$$

Le Théorème 3.1 est une conséquence des résultats préliminaires suivants :

Lemme 3.2. *Sous les hypothèses* (H1), (H2), (H5) *et* (H6)*, on a*

$$\left\|E\widehat{g}(x)-m(x)\right\| = O\left(h^b\right).$$

Lemme 3.3. *Sous les hypothèses* (H1) *et* (H3)–(H6)*, on a*

$$\|\widehat{g}(x)-E\widehat{g}(x)\| = O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\delta}\phi_x(h)}}\right) + O\left(\frac{1}{n^{1-\delta}\phi_x(h)}\right)^{1-1/\tau}. \tag{2}$$

Lemme 3.4. *Sous les hypothèses* (H1), (H3), (H5) *et* (H6)*, on a*

$$\widehat{f}(x)-1 = O_{a.co.}\left(\sqrt{\frac{\log n}{n^{1-\delta}\phi_x(h)}}\right). \tag{3}$$

Une des idées essentielles de la preuve des lemmes 3.3 et 3.4 consiste à utiliser une inégalité exponentielle pour variables aléatoires Banachiques mélangeantes due à Rhomari [10]. Quant au lemme 3.2, il ne pose pas de problèmes particuliers puisqu'il concerne des termes non stochastiques de biais qui ne sont pas affectés par la dépendance des variables.

4. Perspectives

Parmi les perspectives de travaux futurs, deux pistes nous intéressent plus particulièrement. Tout d'abord, citons l'obtention de vitesse de convergence uniforme avec toutes les retombées que cela peut représenter en matière d'applications (choix automatique de h, construction de modèles semi-paramétriques, ...). Enfin, il y a l'extension au cadre α-mélangeant qui devrait être possible en utilisant des inégalités exponentielles de Bosq [2] à condition toutefois de remplacer la structure de Banach par une structure de Hilbert.

Remerciements

Les rapporteurs sont vivement remerciés pour leurs commentaires pertinents.

Références

[1] A. Allam, T. Mourid, Geometric absolute regularity of Banach space-valued autoregressive processes, Statist. Probab. Lett. 60 (2002) 241–252.
[2] D. Bosq, Linear Processes in Function Spaces: Theory and Applications, Lecture Notes in Statist., vol. 149, Springer, 2000.
[3] D. Bosq, M. Delecroix, Nonparametric prediction of a Hilbert-space valued random variable, Stochastic Process. Appl. 19 (1985) 271–280.
[4] F. Ferraty, Y. Romain, The Oxford Handbook of Functional Data Analysis, Oxford University Press, 2011.
[5] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis. Theory and Practice, Springer, New York, 2006.
[6] F. Ferraty, A. Laksaci, A. Tadj, P. Vieu, Kernel regression with functional response, Electron. J. Stat. 5 (2011) 159–171.
[7] F. Ferraty, I. Van Keilegom, P. Vieu, Regression when both response and predictor are functions, J. Multivariate Anal. 109 (2012) 10–28.

[8] J.P. Lecoutre, Uniform consistency of a class of regression function estimators for Banach-space valued random variable, Statist. Probab. Lett. 10 (1990) 145–149.
[9] J.O. Ramsay, B.W. Silverman, Applied Functional Data Analysis: Methods and Case Studies, Springer, New York, 2002.
[10] N. Rhomari, Approximation et inégalités exponentielles pour les sommes de vecteurs aléatoires dépendants, C. R. Acad. Sci. Paris, Ser. I 334 (2002) 149–154.