

Statistique/Probabilités

# Erreur quadratique de l'estimateur à noyau de la densité conditionnelle à variable explicative fonctionnelle

Ali Laksaci

*Département de mathématiques, université Djilali-Liabes BP 89, Sidi Bel Abbès 22000, Algérie*

Reçu le 2 décembre 2006 ; accepté après révision le 31 mai 2007

Disponible sur Internet le 19 juillet 2007

Présenté par Paul Deheuvels

## Résumé

L'objet de cette Note est l'étude de l'estimateur à noyau de la densité conditionnelle d'une variable réelle  $Y$  conditionnée par une variable  $X$  à valeurs dans un espace vectoriel semi-normé. La vitesse de convergence en moyenne quadratique de cet estimateur est établie, en donnant l'expression exacte des termes dominant. *Pour citer cet article : A. Laksaci, C. R. Acad. Sci. Paris, Ser. I 345 (2007).*

© 2007 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

## Abstract

**Quadratic error of the kernel estimate of the conditional density when the regressor is functional.** This Note investigates a kernel estimator of the conditional density of a scalar response variable  $Y$  given a random variable  $X$  taking values in a semi-normed vector space. Asymptotic properties of this estimator are stated through the exact expression involved in the leading terms of the quadratic error. *To cite this article : A. Laksaci, C. R. Acad. Sci. Paris, Ser. I 345 (2007).*

© 2007 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

## Abridged English version

Let  $(X, Y)$  be a couple of random variables in  $\mathcal{F} \times \mathbb{R}$  where  $\mathcal{F}$  is a semi-metric space; we denote by  $d$  the semi-metric on  $\mathcal{F}$ . We assume that the regular version of the conditional probability of  $Y$  given  $X$  exists and has a bounded density with respect to Lebesgue measure over  $\mathbb{R}$ , denoted by  $f_Y^X$ . In the following,  $(x, y)$  will be a fixed point in  $\mathcal{F} \times \mathbb{R}$  and  $N_x \times N_y$  will denote a fixed neighborhood of  $(x, y)$ .

Given i.i.d. observations  $(X_1, Y_1), \dots, (X_n, Y_n)$  of  $(X, Y)$ , the kernel estimate of the conditional density  $f_Y^X(x, y)$  denoted  $\hat{f}_Y^X(x, y)$ , is defined by

$$\hat{f}_Y^X(x, y) = \frac{h_H^{-1} \sum_{i=1}^n K(h_K^{-1} \|x - X_i\|) H(h_H^{-1} (y - Y_i))}{\sum_{i=1}^n K(h_K^{-1} \|x - X_i\|)}$$

Adresse e-mail : [alilak@yahoo.fr](mailto:alilak@yahoo.fr).

with the convention  $\frac{0}{0} = 0$ . The functions  $K$  and  $H$  are kernels and  $h_K = h_{K,n}$  (resp.  $h_H = h_{H,n}$ ) is a sequence of positive real numbers.

The main goal of this Note is to study the nonparametric estimate  $\hat{f}_Y^X(x, y)$  of  $f_Y^X(x, y)$  when the explanatory variable  $X$  is valued in a space of eventually infinite dimension. We give precise asymptotic evaluations of the quadratic error of this estimator. Our result generalizes to a infinity dimension the asymptotic results given by Vieu [7] in the multivariate case. It is given in the following theorem (see the French version for the notations and assumptions):

**Theorem 0.1.** *Under the hypotheses (H1)–(H4) and if  $f_Y^X \in C_B^2(\mathcal{F} \times \mathbb{R})$  then*

$$E[\hat{f}_Y^X(x, y) - f_Y^X(x, y)]^2 = B_H^2(x, y)h_H^4 + B_K^2(x, y)h_K^2 + \frac{V_{HK}(x, y)}{nh_H\phi(h_K)} + o(h_H^4) + o(h_K^2) + o\left(\frac{1}{nh_H\phi(h_K)}\right).$$

## 1. Introduction

Le problème de prévision est une question très fréquente en statistique. En statistique non paramétrique, un de ces outils de prévision est le mode conditionnel qui repose sur l'estimation préalable de la densité conditionnelle. L'objectif de cette note est d'étudier cette caractéristique fonctionnelle dans le cas où la variable explicative est à valeurs dans un espace de dimension éventuellement infinie.

La littérature sur l'estimation de la densité conditionnelle est très abondante lorsque la variable explicative est à valeurs dans un espace de dimension finie. A titre d'exemple, on cite le plus ancien article sur le sujet, Roussas [5] pour la convergence en moyenne quadratique, la convergence en probabilité et la normalité asymptotique sur un processus de Markov vérifiant la condition de Doeblin. Nous renvoyons à Youndjé [8] pour une présentation assez générale sur l'estimation de la densité conditionnelle, laquelle donne des conditions nécessaires de convergence ponctuelle ou uniforme en probabilité, des conditions suffisantes à la convergence en moyenne quadratique et à la convergence uniforme presque complète pour un estimateur à noyau de la densité conditionnelle. Cette littérature est restreinte lorsque la variable explicative est fonctionnelle. Dans ce contexte, les premiers résultats ont été obtenus par Ferraty et al. [2]. Ils ont étudié la convergence presque complète d'un estimateur à noyau de la densité conditionnelle en considérant des observations i.i.d. On trouvera aussi dans cet article une application traitant de la prévision via le mode conditionnel dans le cas i.i.d. La monographie de Ferraty et Vieu [3] présente une collection importante d'outils statistiques pour la prévision, à partir de variables fonctionnelles, tels la régression, le mode conditionnel et la médiane conditionnelle.

Nous définissons dans le paragraphe 2 l'espace fonctionnel auquel est supposé appartenir la densité conditionnelle puis son estimateur à noyau. Dans le paragraphe 3 nous étudions la convergence en moyenne quadratique de l'estimateur. Le paragraphe 4 est consacré à quelques commentaires.

## 2. Modèle

Soit  $(X, Y)$  un couple de variables aléatoires à valeurs dans  $\mathcal{F} \times \mathbb{R}$ , où  $\mathcal{F}$  est un espace vectoriel semi-normé muni d'une mesure positive,  $\sigma$ -finie, notée  $\mu$ . On suppose que la version régulière de la probabilité conditionnelle de  $Y$  sachant  $X$  existe et admet une densité par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ , notée  $f_Y^X$ . Par la suite, on fixe un point  $(x, y)$  dans  $\mathcal{F} \times \mathbb{R}$  et on note  $N_x \times N_y$  un voisinage de ce point. Sans perdre de généralité, on suppose que notre modèle non paramétrique est défini par l'appartenance de  $f_Y^X$  à l'espace fonctionnel :

$$C_B^2(\mathcal{F} \times \mathbb{R}) = \left\{ \begin{array}{l} f : \mathcal{F} \times \mathbb{R} \rightarrow \mathbb{R} \\ (x, y) \rightarrow f(x, y) \text{ telle que :} \\ \forall z \in N_x, \quad f(z, \cdot) \in C^2(N_y) \quad \text{et} \quad \left( f(\cdot, y), \frac{\partial^2 f(\cdot, y)}{\partial y^2} \right) \in C_B^1(x) \times C_B^1(x) \end{array} \right\} \quad (1)$$

où  $C_B^1(x)$  est l'ensemble des fonctions continûment différentiable au sens de Gâteaux sur  $N_x$ , dont l'opérateur de dérivée d'ordre 1 au point  $x$  est borné sur la boule unité  $B(0, 1)$  de l'espace fonctionnel  $\mathcal{F}$ .

Étant donné  $(X_1, Y_1), \dots, (X_n, Y_n)$  des observations i.i.d. de  $(X, Y)$ , l'estimateur à noyau de la densité conditionnelle  $f_Y^X(x, y)$  noté  $\hat{f}_Y^X(x, y)$ , est défini par

$$\hat{f}_Y^X(x, y) = \frac{h_H^{-1} \sum_{i=1}^n K(h_K^{-1}\|x - X_i\|)H(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}\|x - X_i\|)}$$

avec la convention  $\frac{0}{0} = 0$ .  $K$  et  $H$  sont des noyaux et  $h_K = h_{K,n}$  (resp.  $h_H = h_{H,n}$ ) est une suite de nombres réels positifs. Notons que cet estimateur a été utilisé par Rosenblatt [6] dans le cas réel et a été largement étudié dans ce contexte par Youndjé [8].

### 3. Propriétés asymptotiques

Afin d'établir la convergence en moyenne quadratique de l'estimateur  $\hat{f}_Y^X(x, y)$  vers  $f_Y^X(x, y)$ , on introduit les hypothèses suivantes :

(H1) Pour tout  $r > 0$ , la variable aléatoire  $Z = r^{-1}(x - X)$  est absolument continue par rapport à la mesure  $\mu$ . Sa densité  $g(r, x, v)$  est strictement positive sur  $B(0, 1)$  et peut être écrite sous la forme :

$$g(r, x, v) = \phi(r)h(x, v) + o(\phi(r)) \quad \text{pour tout } v \in B(0, 1), \tag{2}$$

où  $\phi$  est une fonction croissante à valeurs dans  $\mathbb{R}^+$  et  $h$ , définie sur  $\mathcal{F} \times \mathcal{F}$ , à valeurs dans  $\mathbb{R}^+$ , est telle que  $0 < \int_{B(0,1)} h(x, v) d\mu(v) < \infty$ ,

(H2)  $K$  est un noyau de support  $(0, 1)$  tel que  $0 < C_3 < K(t) < C_4 < \infty$ ,

(H3)  $H$  est un noyau borné, intégrable, positif, symétrique et tel que :  $\int H(t) dt = 1$  et  $\int t^2 H(t) dt < \infty$ ,

(H4)  $\lim_{n \rightarrow \infty} h_K = 0$ ,  $\lim_{n \rightarrow \infty} h_H = 0$  et  $\lim_{n \rightarrow \infty} nh_H \phi(h_K) = \infty$ .

On établit le résultat suivant :

**Théorème 3.1.** *Sous les hypothèses (H1)–(H4) et si  $f_Y^X \in C_B^2(\mathcal{F} \times \mathbb{R})$  alors*

$$E[\hat{f}_Y^X(x, y) - f_Y^X(x, y)]^2 = B_H^2(x, y)h_H^4 + B_K^2(x, y)h_K^2 + \frac{V_{HK}(x, y)}{nh_H \phi(h_K)} + o(h_H^4) + o(h_K^2) + o\left(\frac{1}{nh_H \phi(h_K)}\right), \tag{3}$$

avec

$$B_H(x, y) = \frac{1}{2} \frac{\partial^2 f_Y^X(x, y)}{\partial y^2} \int t^2 H(t) dt, \quad B_K(x, y) = \frac{\int_{B(0,1)} K(\|v\|) D_x f_Y^X(x, y)[v] h(x, v) d\mu(v)}{\int_{B(0,1)} K(\|v\|) h(x, v) d\mu(v)}$$

et

$$V_{HK}(x, y) = (f_Y^X(x, y)) \left( \int_{B(0,1)} K(\|v\|) h(x, v) d\mu(v) \right)^{-2} \left( \int_{B(0,1)} K^2(\|v\|) h(x, v) d\mu(v) \right) \int H^2(t) dt,$$

où  $D_x$  désigne la dérivée par rapport à  $x$ .

**Schéma de la preuve.** Pour tout  $i = 1, \dots, n$ , on définit les quantités  $K_i = K(h_K^{-1}\|x - X_i\|)$ ,  $H_i(y) = h_H^{-1}H(h_H^{-1}(y - Y_i))$  et on pose

$$\hat{f}_N(x, y) = \frac{1}{n\phi(h_K)} \sum_{i=1}^n K_i H_i(y), \quad \hat{f}_D(x) = \frac{1}{n\phi(h_K)} \sum_{i=1}^n K_i.$$

On a

$$\begin{aligned} & \hat{f}_Y^X(x, y) - f_Y^X(x, y) \\ &= ((E \hat{f}_D(x))^{-1} \hat{f}_N(x, y) - f_Y^X(x, y)) - (E \hat{f}_D(x))^{-2} (\hat{f}_N(x, y) - E \hat{f}_N(x, y)) (\hat{f}_D(x) - E \hat{f}_D(x)) \\ &\quad - (E \hat{f}_D(x))^{-2} (E \hat{f}_N(x, y)) (\hat{f}_D(x) - E \hat{f}_D(x)) + (E \hat{f}_D(x))^{-2} (\hat{f}_D(x) - E \hat{f}_D(x))^2 \hat{f}_Y^X(x, y). \end{aligned}$$

Finalement, le Théorème (3.1) est une conséquence des lemmes suivants :

**Lemme 3.2.** *Sous les conditions du Théorème 3.1 on a*

$$(E \hat{f}_D(x))^{-1} E[\hat{f}_N(x, y)] - f_Y^X(x, y) = B_H(x, y) h_H^2 + B_K(x, y) h_K + o(h_H^2) + o(h_K).$$

**Lemme 3.3.** *Sous les conditions du Théorème 3.1 on a*

$$\text{Var}[\hat{f}_N(x, y)] = \frac{\int_{B(0,1)} K^2(\|v\|) h(x, v) d\mu(v)}{nh_H \phi(h_K)} \left( f_Y^X(x, y) \int H^2(t) dt \right) + o\left(\frac{1}{nh_H \phi(h_K)}\right).$$

**Lemme 3.4.** *Sous les conditions du Théorème 3.1 on a*

$$\text{Cov}(\hat{f}_N(x, y), \hat{f}_D(x)) = \frac{1}{n\phi(h_K)} (f_Y^X(x, y)) \int_{B(0,1)} K^2(\|v\|) h(x, v) d\mu(v) + o\left(\frac{1}{n\phi(h_K)}\right).$$

**Lemme 3.5.** *Sous les hypothèses (H1)–(H2) et (H4) on a,*

$$\text{Var}[\hat{f}_D(x)] = \frac{\int_{B(0,1)} K^2(\|v\|) h(x, v) d\mu(v)}{n\phi(h_K)} + o\left(\frac{1}{n\phi(h_K)}\right).$$

#### 4. Commentaires

- (i) *Remarques sur la variable fonctionnelle :* L'hypothèse (H1) n'est pas très restrictive. En effet, la première partie de cette hypothèse est vérifiée, lorsque, par exemple  $X$  est un processus de diffusion vérifiant des conditions standards (voir Niang [4]). La deuxième partie (2) est moins restrictive que la condition suivante, donnée pour tout  $(r, v) \in \mathbb{R}_*^+ \times B(0, 1)$  ( $x$  est fixé) :

$$\exists C_1, C_2 > 0, \quad 0 < C_1 \phi(r) h(x, v) \leq g(r, x, v) \leq C_2 \phi(r) h(x, v),$$

qui est une propriété classique en analyse fonctionnelle. Notons que, cette hypothèse permet de décrire le phénomène de concentration de la mesure de probabilité de la variable explicative  $X$ , car on a :

$$P(X \in B(x, r)) = \int_{B(0,1)} g(r, x, v) d\mu(v) = \phi(r) \int_{B(0,1)} h(x, v) d\mu(v) + o(\phi(r)) > 0.$$

Dans le cas de dimension finie, cette hypothèse est vérifiée quand la densité de la variable explicative  $X$  est de classe  $C^1$  et strictement positive. En effet, la densité de  $Z = r^{-1}(x - X)$  est  $g(r, x, v) = r^p f(x - rv)$  où  $f$  est la densité de  $X$  et  $p$  sa dimension, donc  $g(r, x, v) = r^p f(x) + o(r^p)$ .

- (ii) *Remarques sur le modèle non paramétrique :* Dans cette Note, nous avons opté pour une condition de dérivabilité car notre objectif est de trouver une expression de la vitesse de convergence explicite, asymptotiquement exacte et qui garde la forme usuelle de l'erreur quadratique (voir Vieu [7]). Cependant, si on procède par une condition Lipschitzienne de type :

$$\forall (y_1, y_2) \in N_y \times N_y \quad \forall (x_1, x_2) \in N_x \times N_x, \quad |f^{x_1}(y_1) - f^{x_2}(y_2)| \leq C(d(x_1, x_2)^2 + |y_1 - y_2|^2),$$

qui est moins restrictive que la condition (1), on obtient un résultat de type

$$E[\hat{f}_Y^X(x, y) - f_Y^X(x, y)]^2 = O(h_H^4 + h_K^4) + O\left(\frac{1}{nh_H \phi(h_K)}\right).$$

Mais, une telle expression (implicite) de la vitesse de convergence ne nous permettra pas de déterminer convenablement le paramètre de lissage. Autrement dit, cette condition de dérivabilité est le bon compromis pour

l'obtention d'une expression explicite de la vitesse de convergence. Notons que cette condition est souvent prise dans le cas de dimension finie.

- (iii) *Remarques sur l'erreur quadratique* : La « dimensionnalité » des observations (resp. du modèle) est bien exploitée dans l'expression de la vitesse de convergence (3). On trouvera la « dimensionnalité » du modèle dans la partie biais, tandis que, la « dimensionnalité » de la variable fonctionnelle est dans la partie dispersion à travers la propriété de concentration de la mesure de probabilité de la variable fonctionnelle qui est étroitement liée à la structure topologique de l'espace fonctionnel de la variable explicative. Ceci met en évidence le rôle de la semi-métrique dans la qualité de notre estimation. Un choix convenable de ce paramètre nous permet d'établir une solution originale pour le problème de fléau de la dimension (voir Ferraty et al. [2]). Un autre argument a un effet très important dans notre estimation. Il s'agit du paramètre de lissage  $h_K$  (resp.  $h_H$ ). L'expression de notre vitesse de convergence, décomposée principalement en deux parties : partie biais proportionnel à  $h_K$  (resp.  $h_H$ ) et partie dispersion inversement proportionnel à  $h_K$  (resp.  $h_H$ ) ( $\phi$  étant une fonction croissante en fonction du  $h_K$ ) rend ce choix relativement aisé en minimisant la partie principale de cette expression pour déterminer ce paramètre. Dans le cas où  $X$  est de dimension finie ( $X \in \mathbb{R}^p$ ), l'expression (3) est comparable à celle obtenue par Vieu [7], car, si la densité  $f$  de  $X$  est de classe  $C^1$  et strictement positive, les expressions  $B_K(x, y)$  et  $V_{HK}(x, y)$  seront :

$$B_K(x, y) = \left( \int_{B(0,1)} K(\|v\|) dv_1 \cdots dv_p \right)^{-1} \sum_{i=1}^p \frac{\partial f_Y^X(x, y)}{\partial x_i} \int_{B(0,1)} v_i K(\|v\|) dv_1 \cdots dv_p$$

et

$$V_{HK}(x, y) = f^{-1}(x)(f_Y^X(x, y)) \left( \int_{B(0,1)} K(\|v\|) dv_1 \cdots dv_p \right)^{-2} \left( \int_{B(0,1)} K^2(\|v\|) dv_1 \cdots dv_p \right) \int H^2(t) dt.$$

Mentionnons, également, que dans le cas fonctionnel notre terme de dispersion  $V_{HK}(x, y)$  est similaire à celui de l'estimateur à noyau de la dérivée de la densité conditionnelle dans le Lemme 5 de Ezzahrioui et Ould-Saïd [1]. En effet, la quantité

$$\left( \int_{B(0,1)} K(\|v\|) h(x, v) d\mu(v) \right)^{-2} \left( \int_{B(0,1)} K^2(\|v\|) h(x, v) d\mu(v) \right) = \lim_{n \rightarrow \infty} \frac{E[K_1^2]/\phi(h_K)}{(E[K_1]/\phi(h_K))^2}$$

et si on se place dans les mêmes conditions, on montre que cette limite tend vers leur paramètre normalisateur noté  $\alpha$ .

## Remerciements

Les rapporteurs ainsi que les professeurs F. Ferraty et Ph. Vieu du LSP (Université Paul Sabatier) sont vivement remerciés pour leurs commentaires pertinents.

## Références

- [1] M. Ezzahrioui, E. Ould-Saïd, Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data, Technical Report, no. 249, LMPA, Univ. Littoral Côte d'Opale, 2005, in revision.
- [2] F. Ferraty, A. Laksaci, P. Vieu, Estimating some characteristics of the conditional distribution in nonparametric functional models, *Statist. Inference Stoch. Process.* 9 (2006) 47–76.
- [3] F. Ferraty, Ph. Vieu, *Nonparametric Functional Data Analysis*, Springer Series in Statistics, Springer, New York, 2006.
- [4] S. Niang, Sur l'estimation fonctionnelle en dimension infinie : Application aux diffusions, Thèse de Doctorat, Université de Paris 6, 2002.
- [5] G.G. Roussas, On some properties of nonparametric estimates of probability density functions, *Bull. Soc. Math. Grèce (N.S.)* 9 (1) (1968) 29–43.
- [6] M. Rosenblatt, Conditional probability density and regression estimators, in: P.R. Krishnaiah (Ed.), *Multivariate Analysis II*, Academic Press, New York and London, 1969.
- [7] Ph. Vieu, Quadratic errors for nonparametric estimates under dependence, *J. Multivariate Anal.* 39 (2) (1991) 324–347.
- [8] E. Youndjé, Estimation non paramétrique de la densité conditionnelle par la méthode du noyau, Thèse de Doctorat, Université de Rouen, 1993.