C. R. Acad. Sci. Paris, Ser. I 348 (2010) 931–934

Contents lists available at ScienceDirect

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

www.sciencedirect.com

Statistique

Estimation locale linéaire de la densité conditionnelle pour des données fonctionnelles

Local linear estimation of the conditional density for functional data

Jacques Demongeot [a], Ali Laksaci [b], Fethi Madani [a], Mustapha Rachdi [c]

[a] *Laboratoire TIMC-IMAG, UMR CNRS 5525, équipe AGIM³, faculté de médecine de Grenoble, université J. Fourier, 38700, La Tronche, France*
[b] *Université Djillali-Liabès, BP. 89, Sidi Bel-Abbès 22000, Algérie*
[c] *Laboratoire TIMC-IMAG, UMR CNRS 5525, équipe AGIM³, université P. Mendès France de Grenoble, UFR SHS, BP. 47, 38040 Grenoble cedex 09, France*

I N F O   A R T I C L E

Historique de l'article :
Reçu le 14 mai 2010
Accepté après révision le 21 juin 2010
Disponible sur Internet le 21 juillet 2010

Présenté par Paul Deheuvels

R É S U M É

Dans cette Note, nous introduisons l'estimation locale linéaire de la densité conditionnelle, quand la réponse est scalaire et la variable explicative est à valeurs dans un espace semi-métrique. Sous certaines conditions, nous établissons les convergences presque-complètes ponctuelle et uniforme, et nous donnons également les vitesses de convergence correspondantes. De plus, comme application, nous utilisons les résultats obtenus pour donner quelques propriétés asymptotiques de l'estimateur local linéaire du mode conditionnel.

A B S T R A C T

In this Note, we introduce the local linear estimation of the conditional density of a scalar response variable given a random variable taking values in a semi-metric space. Under some general conditions, we establish the pointwise and uniform almost complete convergences with rates of this estimator. Moreover, as an application, we use the obtained results to derive some asymptotic properties for the local linear estimator of the conditional mode.

1. Introduction

Dans cette Note, nous nous intéressons à l'estimation par polynômes locaux de la densité conditionnelle quand la variable explicative est fonctionnelle. Il est bien connu, en dimension finie, que cette estimation admet plusieurs avantages que la méthode à noyau de Nadaraya–Watson ne présente pas (cf. [2] et [6] pour une comparaison entre les deux méthodes). Notons que, lorsque les données appartiennent à un espace de dimension infinie, ces questions deviennent particulièrement intéressantes. D'abord pour le problème fondamental qu'elles formulent, et ensuite pour l'efficacité de la méthode d'estimation sur le plan pratique. De plus, la méthode à noyau de Nadaraya–Watson peut être considérée comme un cas particulier de la méthode étudiée dans cette note.

Hormis le fait qu'il existe plusieurs outils de prévision en statistique non-paramétrique, dont la médiane conditionnelle ou les quantiles conditionnels, il est bien connu que le mode conditionnel en est un, et dont l'estimation non-paramétrique dépend de celle de la densité conditionnelle. Dans la statistique non-paramétrique pour données fonctionnelles, [7] et [8]

Adresses e-mail : Jacques.Demongeot@imag.fr (J. Demongeot), alilak@yahoo.fr (A. Laksaci), fethi_madani@yahoo.fr, Fethi.Madani@imag.fr (F. Madani), Mustapha.Rachdi@upmf-grenoble.fr (M. Rachdi).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2010.06.013

peuvent être considérés comme les premiers travaux sur ce sujet. Dans [7], on a établi la convergence presque-complète de l'estimateur à noyau de Nadaraya–Watson de la densité et de la fonction de répartition conditionnelles quand les covariables sont fonctionnelles indépendantes. Par ailleurs, le cas de données fonctionnelles fortement mélangeantes a été étudié dans [7] et [8]. D'autre part, la convergence L^p de l'estimateur à noyau du mode conditionnel a été étudiée dans [3], alors que des résultats asymptotiques des estimateurs à noyau du quantile et du mode conditionnels ont été étudiés dans [5]. Pour d'autres références bibliographiques concernant des sujets proches de celui-ci, nous référons à [10] et aux références qui s'y trouvent.

2. Modèle

Soit $(X_1, Y_1), \ldots, (X_n, Y_n)$ un échantillon aléatoire du couple (X, Y) qui est à valeurs dans $\mathcal{F} \times \mathbb{R}$, où $\mathcal{F}$ est un espace semi-métrique muni de la semi-métrique d.

Nous supposons qu'il existe une densité conditionnelle de Y sachant X, qui est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$, et qui admet une densité bornée notée f^x. Le lissage polynomial est basé sur l'hypothèse que le paramètre fonctionnel est assez régulier pour qu'il soit localement approché par un polynôme. En statistique fonctionnelle, il y a plusieurs moyens permettant d'adapter l'estimation par polynômes locaux (cf. [1] et [4]).

Dans cette note, nous sommes intéressés par l'estimation non-paramétrique locale de f^x, pour $x \in \mathcal{F}$. Pour ceci, nous supposons que f^x est différentiable au voisinage de y (cf. [9]). Nous adaptons donc la modélisation fonctionnelle rapide locale qui a été utilisée pour l'estimation de l'opérateur de régression dans [1] (cf. aussi [4] pour les degrés supérieurs), où l'estimateur est défini par $\hat{f}^x(y) = \hat{a}$, obtenu par la minimisation de l'expression suivante :

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} \bigl(h_H^{-1} H\bigl(h_H^{-1}(y - Y_i)\bigr) - a - b\beta(X_i, x)\bigr)^2 K\bigl(h_K^{-1}\delta(x, X_i)\bigr)$$

où K et H sont des noyaux, $h_K = h_{K,n}$ (resp. $h_H = h_{H,n}$) est une suite de nombres réels positifs et $\beta(.,.)$ est une fonction réelle connue, définie sur $\mathcal{F}^2$ et telle que, $\forall \xi \in \mathcal{F}$, $\beta(\xi, \xi) = 0$. De plus, $\delta(.,.)$ est une fonction réelle définie sur $\mathcal{F}^2$, telle que $d(.,.) = |\delta(.,.)|$. Clairement, par des calculs simples, on obtient explicitement la définition suivante de $\hat{f}^x$:

$$\hat{f}^x(y) = \frac{\sum_{i,j=1}^{n} W_{ij}(x) H(h_H^{-1}(y - Y_i))}{h_H \sum_{i,j=1}^{n} W_{ij}(x)}$$

où $W_{ij}(x) = \beta(X_i, x)(\beta(X_i, x) - \beta(X_j, x)) K(h_K^{-1}\delta(x, X_i)) K(h_K^{-1}\delta(x, X_j))$, avec la convention $0/0 = 0$.

3. Principaux résultats

3.1. Convergence ponctuelle presque-complète

Dans ce qui suit, nous fixons x dans $\mathcal{F}$, et nous désignons par N_x un voisinage de x, $S_{\mathbb{R}}$ un sous-ensemble compact de $\mathbb{R}$, et par $\phi_x(r_1, r_2) = \mathbb{P}(r_2 \leqslant \delta(X, x) \leqslant r_1)$.

Notons que notre modèle non-paramétrique sera plus général, dans le sens où nous n'aurons besoin que des hypothèses suivantes :

(H1) Pour tout $r > 0$, $\phi_x(r) := \phi_x(-r, r) > 0$.

(H2) La densité conditionnelle f^x est telle que : $\exists b_1 > 0$, $b_2 > 0$, $\forall (y_1, y_2) \in S_{\mathbb{R}}^2$ et $\forall (x_1, x_2) \in N_x \times N_x$

$$\bigl|f^{x_1}(y_1) - f^{x_2}(y_2)\bigr| \leqslant C_x\bigl(d^{b_1}(x_1, x_2) + |y_1 - y_2|^{b_2}\bigr), \quad \text{où } C_x > 0 \text{ (dépend de } x\text{)}.$$

(H3) La fonction $\beta(.,.)$ est telle que :

$$\forall x' \in \mathcal{F}, \quad C_1 d\bigl(x, x'\bigr) \leqslant \bigl|\beta\bigl(x, x'\bigr)\bigr| \leqslant C_2 d\bigl(x, x'\bigr), \quad \text{où } C_1 > 0,\ C_2 > 0.$$

(H4) K est une fonction positive, différentiable et de support $[-1, 1]$.

(H5) H est une fonction positive, bornée, Lipschitzienne, et telle que :

$$\int |t|^{b_2} H(t)\,\mathrm{d}t < \infty \quad \text{et} \quad \int H^2(t)\,\mathrm{d}t < \infty.$$

(H6) La largeur de fenêtre h_K satisfait :

$$\exists n_0 \in \mathbb{N}, \quad \forall n > n_0, \quad -\frac{1}{\phi_x(h_K)} \int_{-1}^{1} \phi_x(z h_K, h_K) \frac{\mathrm{d}}{\mathrm{d}z}\bigl(z^2 K(z)\bigr)\,\mathrm{d}z > C_3 > 0$$

et

$$h_K \int\limits_{B(x,h_K)} \beta(u,x)\,\mathrm{d}P(u) = o\Bigg(\int\limits_{B(x,h_K)} \beta^2(u,x)\,\mathrm{d}P(u)\Bigg)$$

où $B(x,r) = \{x' \in \mathcal{F}/d(x',x) \leqslant r\}$ désigne une boule fermée de centre x et de rayon r, et $\mathrm{d}P(x)$ est la fonction cumulative.

(H7) La largeur de fenêtre h_H satisfait : $\lim_{n\to\infty} n^{\gamma} h_H = \infty$ pour tout $\gamma > 0$ et $\lim_{n\to\infty} \frac{\ln n}{nh_H\phi_x(h_K)} = 0$.

Remarquons que les conditions (H1), (H3) et (H6) sont les mêmes que celles utilisées dans [1]. L'hypothèse (H2) est une condition de régularité qui caractérise l'espace fonctionnel de notre modèle et elle sert pour évaluer le terme du biais dans les résultats asymptotiques de cette note. Les hypothèses (H5) et (H7) sont des conditions techniques et sont similaires à celles considérées dans [8].

Le théorème suivant donne la convergence presque-complète (p.co.) de $\hat{f}^x$.

Théorème 3.1. *Sous les hypothèses* (H1)–(H7)*, nous avons* :

$$\sup_{y\in S_{\mathbb{R}}} \big|\hat{f}^x(y) - f^x(y)\big| = O\big(h_K^{b_1} + h_H^{b_2}\big) + O\Bigg(\sqrt{\frac{\ln n}{nh_H\phi_x(h_K)}}\Bigg), \quad p.co.$$

Sketch de la démonstration du Théorème 3.1. Pour tout $y \in S_{\mathbb{R}}$, on a la décomposition suivante :

$$\hat{f}^x(y) - f^x(y) = \big\{\big(\hat{f}_N^x(y) - \mathbb{E}\big[\hat{f}_N^x(y)\big]\big) - \big(f^x(y) - \mathbb{E}\big[\hat{f}_N^x(y)\big]\big)\big\}/\hat{f}_D^x + \big(1 - \hat{f}_D^x\big) f^x(y)/\hat{f}_D^x := I_1 + I_2 + I_3$$

où $\hat{f}_N^x(y) = \sum_{i\neq j} W_{ij}(x) H(h_H^{-1}(y - Y_i))/(n(n-1)h_H\mathbb{E}[W_{12}(x)])$ et $\hat{f}_D^x = \sum_{i\neq j} W_{ij}(x)/(n(n-1)\mathbb{E}[W_{12}(x)])$. Le résultat s'obtient en démontrant la convergence des termes I_1, I_2 et I_3 et en contrôlant le terme $\hat{f}_D^x$. □

3.2. Convergence uniforme presque-complète

Dans ce paragraphe, nous établissons la convergence uniforme presque-complète de $\hat{f}^x$ sur un sous-ensemble $S_{\mathcal{F}}$ de $\mathcal{F}$ tel que : $S_{\mathcal{F}} \subset \bigcup_{k=1}^{d_n} B(x_k, r_n)$, où $x_k \in \mathcal{F}$ et r_n (resp. d_n) est une suite de nombres réels positifs. Il est important de noter que, dans le cadre multivarié, la consistance uniforme est une extension standard de la convergence ponctuelle, cependant, dans notre cadre fonctionnel, nous aurons besoin (en plus de (H1)–(H7)) de quelques outils et hypothèses topologiques supplémentaires :

(U1) Il existe une fonction différentiable $\phi(.)$ de dérivée première $\phi'(.)$, telle que :

$$\forall x \in S_{\mathcal{F}}, \quad 0 < C\phi(h) \leqslant \phi_x(h) \leqslant C'\phi(h) < \infty \quad \text{et} \quad \exists \eta_0 > 0, \quad \forall \eta < \eta_0, \quad \phi'(\eta) < C.$$

(U2) La densité conditionnelle f^x est telle que : $\forall (y_1, y_2) \in S_{\mathbb{R}} \times S_{\mathbb{R}}$, $\forall (x_1, x_2) \in S_{\mathcal{F}} \times S_{\mathcal{F}}$:

$$\big|f^{x_1}(y_1) - f^{x_2}(y_2)\big| \leqslant C\,\big(d^{b_1}(x_1,x_2) + |y_1 - y_2|^{b_2}\big).$$

(U3) La fonction $\beta(.,.)$ vérifie l'hypothèse (H3) et la condition de Lipschitz suivante :

$$\forall (x_1, x_2) \in S_{\mathcal{F}} \times S_{\mathcal{F}}, \quad \big|\beta\big(x_1, x'\big) - \beta\big(x_2, x'\big)\big| \leqslant C' d(x_1, x_2).$$

(U4) Le noyau K vérifie l'hypothèse (H4) et la condition de Lipschitz suivante :

$$\big|K(x) - K(y)\big| \leqslant C\big||x| - |y|\big|.$$

(U5) Pour tout $\gamma \in (0,1)$, $\lim_{n\to+\infty} n^{\gamma} h_H = \infty$, et pour $r_n = O(\ln n/n)$, la suite d_n vérifie :

$$\frac{(\ln n)^2}{n^{1-\gamma}\phi(h_K)} < \ln d_n < \frac{n^{1-\gamma}\phi(h_K)}{\ln n} \quad \text{et} \quad \sum_{n=1}^{\infty} n^{(3\gamma+1)/2} d_n^{1-\beta} < \infty, \quad \text{pour tout } \beta > 1.$$

Notons que les conditions (U1) et (U2) sont, respectivement, les versions uniformes de (H1) et (H2). De plus, les conditions (U1) et (U5) sont reliées à la structure topologique de la variable fonctionnelle. C'est pourquoi, comme dans le cas ponctuel, le choix de la topologie, qui est contrôlée ici par la fonction $\delta(.,.)$, joue un rôle crucial. Donc, un bon choix de cette fonction améliore la vitesse de convergence de l'estimateur. Plus précisément, un bon choix de la semi-métrique est celui qui accroît la concentration de la mesure de probabilité de la variable fonctionnelle X tout en minimisant d_n. Il faut noter que les deux hypothèses (U1) et (U2) sont satisfaites par plusieurs processus à temps continu (cf. par exemple [9] pour des processus connus).

Théorème 3.2. *Sous les hypothèses* (U1)–(U5) *et* (H5)–(H6)*, nous avons* :

$$\sup_{x\in S_{\mathcal{F}}}\sup_{y\in S_{\mathbb{R}}}\left|\hat{f}^x(y)-f^x(y)\right|=O\left(h_K^{b_1}\right)+O\left(h_H^{b_2}\right)+O\left(\sqrt{\frac{\ln d_n}{n^{1-\gamma}\phi(h_K)}}\right),\quad p.co.$$

4. Application : estimation du mode conditionnel

Notons par $\theta(x)=\arg\sup_{y\in S_{\mathbb{R}}} f^x(y)$ le mode conditionnel de Y sachant X. Soit $S_{\mathcal{F}}$ un sous-ensemble «compact» de $\mathcal{F}$. Afin de prouver la consistance forte de cet estimateur, nous aurons besoin des hypothèses suivantes :

(U6) $\forall\epsilon_0>0$, $\exists\eta>0$, $\forall r: S\to S_{\mathbb{R}}$, nous avons (cf. [11]) :

$$\sup_{x\in S_{\mathcal{F}}}\left|\theta(x)-r(x)\right|\geqslant\epsilon_0\quad\Rightarrow\quad\sup_{x\in S_{\mathcal{F}}}\left|f^x\big(r(x)\big)-f^x\big(\theta(x)\big)\right|\geqslant\eta.$$

De plus, on suppose qu'il existe un entier $j>1$ tel que $\forall x\in S_{\mathcal{F}}$, la densité f^x est de classe C^j sur l'intérieur topologique de $S_{\mathbb{R}}$ relativement à y, et que :

$$\text{(U7)}\qquad\begin{cases} f^{x(l)}\big(\theta(x)\big)=0, & \text{si } 1\leqslant l<j\\ \text{et } f^{x(j)}(\cdot) \text{ est uniformément continue sur } S_{\mathbb{R}}, & \text{telle que : } \left|f^{x(j)}\big(\theta(x)\big)\right|>C>0\end{cases}$$

où $f^{x(j)}$ désigne la dérivée d'ordre j de la densité conditionnelle f^x.

Nous estimons le mode conditionnel $\theta(x)$ par la variable aléatoire $\widehat{\theta(x)}$ qui est telle que :

$$\widehat{\theta(x)}=\arg\sup_{y\in S_{\mathbb{R}}}\hat{f}^x(y).$$

Comme conséquence du Théorème 3.2, nous déduisons le corollaire suivant.

Corollaire 4.1. *Sous les hypothèses du Théorème* 3.2*, et si la densité conditionnelle f^x satisfait les hypothèses* (H9) *et* (H10)*, alors nous obtenons* :

$$\sup_{x\in S_{\mathcal{F}}}\left|\widehat{\theta(x)}-\theta(x)\right|^j=O\left(h_K^{b_1}\right)+O\left(h_H^{b_2}\right)+O\left(\sqrt{\frac{\ln d_n}{n^{1-\gamma}\phi(h_K)}}\right),\quad p.co.$$

Références

[1] J. Barrientos-Marin, Some practical problems of recent nonparametric procedures: Testing, estimation, and application, PhD thesis from the Alicante University (Spain), 2007.
[2] C.-K. Chu, J.-S. Marron, Choosing a kernel regression estimator, With comments and a rejoinder by the authors, Statist. Sci. 6 (1991) 404–436.
[3] S. Dabo-Niang, A. Laksaci, Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle, Pub. Inst. Stat. Univ. Paris 3 (2007) 27–42.
[4] M. El Methni, M. Rachdi, Local weighted average estimation of the regression operator for functional data, Comm. Statist. Theory Methods (2010), sous presse.
[5] M. Ezzahrioui, E. Ould-Saïd, Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data, J. Nonparametr. Stat. 20 (2008) 3–18.
[6] J. Fan, I. Gijbels, Local Polynomial Modelling and Its Applications, Chapman & Hall, London, 1996.
[7] F. Ferraty, A. Laksaci, Ph. Vieu, Estimating some characteristics of the conditional distribution in nonparametric functional models, Stat. Inference Stoch. Process. 9 (2006) 47–76.
[8] F. Ferraty, Ph. Vieu, Nonparametric Functional Data Analysis. Theory and Practice, Springer Ser. Statist., Springer, New York, 2006.
[9] A. Laksaci, Contribution aux modèles non paramétriques conditionnels pour variables explicatives fonctionnelles, Thèse de doctorat de l'Université Paul Sabatier de Toulouse (France), 2005.
[10] I. Ouassou, M. Rachdi, Stein type estimation of the regression operator for functional data, Adv. Appl. Statist. Sci. 1 (2) (2010) 233–250.
[11] M. Rachdi, R. Sabre, Consistent estimates of the mode of the probability density function in nonparametric deconvolution problems, Statist. Probab. Lett. 47 (2000) 105–114.