

Statistique/Probabilités

Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle

Sophie Dabo-Niang^a, Ali Laksaci^b

^a Laboratoire GREMARS, maison de la recherche, Université Lille 3, BP 60149, 59653 Villeneuve d'Ascq cedex, France

^b Laboratoire de mathématique, Université Djillali Liabes, Sidi Bel Abbès, Algérie

Reçu le 25 août 2005 ; accepté après révision le 24 octobre 2006

Disponible sur Internet le 15 décembre 2006

Présenté par Paul Deheuvels

Résumé

Nous étudions l'estimateur à noyau du mode de la distribution d'une variable réelle Y conditionnellement à une variable explicative X , à valeurs dans un espace semi-métrique. Nous établissons la convergence en norme L^p de l'estimateur. Les résultats asymptotiques établis sont liés aux probabilités de petites boules de la loi de la variable explicative et aussi à la régularité de la densité conditionnelle. Nos conditions et résultats unifient les deux cadres des variables explicatives de dimensions finie et infinie.

Pour citer cet article : S. Dabo-Niang, A. Laksaci, C. R. Acad. Sci. Paris, Ser. I 344 (2007).

© 2006 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

Abstract

Nonparametric estimation of the conditional mode when the regressor is functional. We study a kernel estimator of the conditional mode of a scalar response variable Y given a random variable X taking values in a semi-metric space. We establish the consistency in L^p norm of the estimator. The asymptotic results are closely related to the concentration properties on small balls of the probability measure of the underlying explanatory variable and the regularity of the conditional density. Our conditions and results unify both cases of finite and infinite dimensional regressors. *To cite this article: S. Dabo-Niang, A. Laksaci, C. R. Acad. Sci. Paris, Ser. I 344 (2007).*

© 2006 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

1. Introduction

Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathcal{F} \times \mathbb{R}$, où \mathcal{F} est un espace semi-métrique. On note d la semi-métrique sur \mathcal{F} . On suppose que la version régulière de la probabilité de Y conditionnellement à X existe et admet une densité bornée par rapport à la mesure de Lebesgue sur \mathbb{R} . Pour tout $x \in \mathcal{F}$, on désigne par f^x la densité conditionnelle de Y sachant $X = x$ (resp. par $f^{x(j)}$ la dérivée d'ordre j de cette densité conditionnelle). Par la suite, on fixe un point x dans \mathcal{F} , on note N_x un voisinage de ce point et on suppose qu'il existe un compact (α_x, β_x) dans \mathbb{R} tel que le mode θ de la densité conditionnelle f^x soit unique dans ce compact.

Nous nous intéressons ici au problème de l'estimation du mode conditionnel θ qui est un outil de prévision alternatif à la méthode de régression (voir Collomb et al. [2]).

Adresses e-mail : sophie.dabo@univ-lille3.fr (S. Dabo-Niang), alilak@yahoo.fr (A. Laksaci).

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^x noté \hat{f}^x , est défini par :

$$\hat{f}^x(y) = \frac{h_H^{-1} \sum_{i=1}^n K(h_K^{-1}d(x, X_i))H(h_H^{-1}(y - Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))}, \quad \forall y \in \mathbb{R}, \quad (1)$$

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 [12] dans le cas réel et a été largement étudié dans ce contexte par Youndjé [13]. Nous renvoyons à Ferraty et al. [5] pour le cas fonctionnel.

L'estimateur $\hat{\theta}$ de θ est tel que

$$\hat{f}^x(\hat{\theta}) = \sup_{y \in (\alpha_x, \beta_x)} \hat{f}^x(y). \quad (2)$$

En général, cet estimateur n'est pas unique. Ainsi, dans toute la suite de cette Note, $\hat{\theta}$ désigne n'importe quelle variable aléatoire vérifiant l'équation (2).

La littérature sur l'estimation du mode conditionnel est très abondante lorsque la variable explicative est à valeurs dans un espace de dimension finie. Citons, par exemple, Quintela del Rio et Vieu [11], Ould-Saïd [10], et Louani et Ould-Saïd [9]. Cette littérature est relativement restreinte lorsque la variable explicative est de nature fonctionnelle. Dans ce dernier cas, les premiers résultats ont été obtenus par Ferraty et al. [5]. Ils ont étudié la convergence presque complète d'un estimateur à noyau du mode conditionnel pour des observations i.i.d. Ce résultat à été généralisé par Ferraty et al. [6] au cas α -mélangeant.

Dans cette Note, on établit la vitesse de convergence en norme L^p de l'estimateur à noyau du mode conditionnel dans le cas i.i.d. Cette vitesse de convergence est étroitement liée à la propriété de concentration de la mesure de probabilité de la variable explicative sur des petites boules, voir Dabo-Niang et Rhomari [3]. Ce résultat nous permet d'établir une solution originale au problème du fléau de la dimension et de généraliser à la dimension infinie de nombreux résultats asymptotiques existant dans le cas multivarié. De plus, en utilisant les résultats récents en théorie des probabilités sur les petites boules, nous précisons notre résultat pour certains processus à temps continu.

Dans le paragraphe 2, nous donnons les hypothèses, puis nous étudions la convergence de l'estimateur. Le paragraphe 3 donne quelques commentaires.

2. Hypothèses et propriétés asymptotiques

Afin d'établir la convergence en norme L^p ($p \in [1, \infty[$) de l'estimateur $\hat{\theta}$, on introduit les hypothèses suivantes :

- (H1) $\forall (x_1, x_2) \in N_x \times N_x, \forall y \in (\alpha_x, \beta_x), |f^{x_1}(y) - f^{x_2}(y)| \leq C_1 d(x_1, x_2)^{b_1}$ et $\forall (y_1, y_2) \in (\alpha_x, \beta_x) \times (\alpha_x, \beta_x), \forall z \in N_x, |f^z(y_1) - f^z(y_2)| \leq C_2 |y_1 - y_2|^{b_2}$, où C_1, C_2, b_1, b_2 sont des constantes positives ;
 (H2) f^x est de classe C^j par rapport à y sur le compact (α_x, β_x) , $f^{x^{(l)}}(\theta) = 0$, pour $1 \leq l < j$, et $0 < |f^{x^{(j)}}(y)| < \infty$, pour tout $y \in (\alpha_x, \beta_x)$;
 (H3) K est un noyau de support $(0, 1)$ tel que $0 < C_3 < K(t) < C_4 < \infty$;
 (H4) H est un noyau vérifiant :

$$\begin{cases} \text{(i) Il existe une fonction intégrable } g \text{ telle que } |H(t) - H(s)| \leq g(|t - s|), \\ \text{(ii) et } \int |t|^{b_2} H(t) dt < \infty; \end{cases}$$

- (H5) $\lim_{n \rightarrow \infty} h_K = 0, \lim_{n \rightarrow \infty} h_H = 0$ et $\lim_{n \rightarrow \infty} n \phi_x(h_K) = \infty$, où $\phi_x(h_K) = P(X \in B(x, h_K))$, $B(x, h_K)$ est la boule centrée en x et de rayon h_K .

Remarque 1.

1. Les hypothèses (H1) et (H2) sont des conditions de régularité qui caractérisent l'espace fonctionnel de notre modèle, ce qui justifie l'emploi des méthodes non paramétriques pour le problème considéré. Tandis que (H3)–(H5) sont des hypothèses techniques relativement standards dans ce contexte, elles sont nécessaires pour démontrer le Théorème 2.1.

2. La « planéité » de la densité conditionnelle f^x autour du mode θ est un argument très important dans l'estimation du mode conditionnel. Cette « planéité » est contrôlée par le nombre de dérivées nulles au point θ (voir (H2)). Cet argument a une grande influence sur la vitesse de convergence de notre estimateur (voir l'expression (3)).

Le théorème suivant donne un résultat de convergence en norme L^p de l'estimateur du mode conditionnel :

Théorème 2.1. *Sous les hypothèses (H1)–(H5) on a, pour tout $p \in [j, \infty[$:*

$$\|\hat{\theta} - \theta\|_p = O(h_K^{b_1/j} + h_H^{b_2/j}) + O\left(\left(\frac{1}{n\phi_x(h_K)}\right)^{1/(2j)}\right), \tag{3}$$

où $\|\cdot\|_p = (E^{1/p}|\cdot|^p)$.

Schéma de la preuve. Sous l'hypothèse (H2), le développement de Taylor de la fonction f^x , au voisinage de θ , en particulier pour le point $\hat{\theta}$, est

$$f^x(\hat{\theta}) = f^x(\theta) + (\hat{\theta} - \theta)^j \frac{f^{x(j)}(\theta^*)}{j!},$$

où θ^* se situe entre $\hat{\theta}$ et θ , il est donc nécessairement dans le compact (α_x, β_x) . Ainsi à l'aide d'arguments analytiques, on montre que

$$|\hat{\theta} - \theta|^j \leq \frac{j!}{\min_{y \in (\alpha_x, \beta_x)} f^{x(j)}(y)} \sup_{y \in (\alpha_x, \beta_x)} |\hat{f}^x(y) - f^x(y)|.$$

L'inégalité de Minkowski nous permet d'écrire :

$$\left\| \sup_{y \in (\alpha_x, \beta_x)} |\hat{f}^x(y) - f^x(y)| \right\|_p \leq S_1 + S_2 + S_3 + S_4,$$

où

$$S_1 = \left\| \sum_{i=1}^n W_{ni}(x) \sup_{y \in (\alpha_x, \beta_x)} |H_i(y) - E(H_i(y)/X_i)| \right\|_p, \quad S_2 = \left\| \sum_{i=1}^n W_{ni}(x) \sup_{y \in (\alpha_x, \beta_x)} |E(H_i(y)/X_i) - f^{X_i}(y)| \right\|_p,$$

$$S_3 = \left\| \sum_{i=1}^n W_{ni}(x) \sup_{y \in (\alpha_x, \beta_x)} |f^{X_i}(y) - f^x(y)| \right\|_p, \quad S_4 = \sup_{y \in (\alpha_x, \beta_x)} |f^x(y)| \left(P \left(\sum_{i=1}^n W_{ni}(x) = 0 \right) \right)^{1/p}, \quad \forall p \geq j,$$

$$W_{ni}(x) = \frac{K(h_K^{-1}d(x, X_i))}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))} \quad \text{et} \quad H_i(y) = h_H^{-1}H(h_H^{-1}(y - Y_i)).$$

Ainsi, le Théorème 2.1 est une conséquence des lemmes suivants :

Lemme 2.2. *Sous les hypothèses (H1), (H3), (H4)(ii) et (H5), on a : $S_2 = O(h_H^{b_2})$.*

Lemme 2.3. *Sous les hypothèses (H1), (H3) et (H5), on a : $S_3 = O(h_K^{b_1})$.*

Lemme 2.4. *Sous les hypothèses (H3) et (H5), on a : $S_4 = o\left(\frac{1}{n\phi_x(h_K)}\right)$.*

Lemme 2.5. *Sous les hypothèses (H3), (H4)(i) et (H5), on a : $S_1 = O(\sqrt{1/(n\phi_x(h_K))})$.*

3. Commentaires

(1) En considérant des arguments similaires à ceux de Ferraty et al. [5] et les résultats récents sur la probabilité des petites boules (voir par exemple Bogachev [1] et Li et Shao [8]), nous obtenons une vitesse de convergence en norme

L^p de $\hat{\theta}$ du type $O((\log n)^{-b_1/2})$ pour les processus stochastiques ayant une mesure de probabilité absolument continue par rapport à la mesure de Wiener, pour $h_K \sim \eta(\log n)^{-1/2}$, $h_H \sim n^{-\epsilon}$ et une densité conditionnelle de classe C^1 . Pour le mouvement brownien fractionnaire de paramètre δ , la vitesse de convergence en norme L^p est de type $O((\log n)^{-\delta b_1/2})$, pour $h_K \sim \eta(\log n)^{-\delta/2}$, $h_H \sim n^{-\epsilon}$, et une densité f^x de classe C^1 . Ces vitesses sont obtenues pour la métrique de la convergence uniforme.

- (2) Pour le cas fractal (voir Ferraty et Vieu [4]) (resp. le cas de dimension finie ($X \in \mathbb{R}^k$) sous la condition de la positivité stricte de la densité de X) et pour $h_K \sim n^{-1/(2b_1+\alpha)}$ (resp. $h_K \sim n^{-1/(2b_1+k)}$), $h_H \sim n^{-\epsilon}$ et une densité conditionnelle de classe C^1 , on obtient une vitesse de convergence en norme L^p du type $O(n^{-b_1/(2b_1+\alpha)})$ où α est la dimension fractale (resp. de l'ordre de $O(n^{-b_1/(2b_1+k)})$).
- (3) Le fléau de la dimension dans le cas fini dimensionnel est la dégradation de la vitesse de convergence par rapport à l'augmentation de la dimension. Dans le cas infini dimensionnel, la dégradation de la vitesse est liée à la faible concentration de la mesure de probabilité de la variable explicative fonctionnelle. Dans les deux cas, il s'agit d'un problème de concentration. L'utilisation d'une semi-métrique qui donne une forte concentration à la mesure de probabilité de la variable fonctionnelle, est alors une solution originale de ce problème de dégradation de la vitesse dans le deuxième cas. Pour plus de détails sur le choix de la semi-métrique nous renvoyons à Ferraty et Vieu [7].
- (4) Le choix du paramètre de lissage a une importance capitale dans l'estimation par la méthode du noyau. La vitesse de convergence donnée dans le théorème précédent nous permet de faire un choix optimal de ce paramètre. En effet, il suffit de choisir un paramètre de lissage qui minimise l'erreur en moyenne d'ordre p donnée par le théorème ci-dessus (voir Youndjé [13], pour le cas réel). Nous pouvons également utiliser une autre méthode de sélection telle que la validation croisée.

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] V.I. Bogachev, Gaussian Measures, Math. Surveys and Monographs, vol. 62, Amer. Math. Soc., 1999.
- [2] G. Collomb, W. Härdle, S. Hassani, A note on prediction via conditional mode estimation, J. Stat. Plann. Inference 15 (1987) 227–236.
- [3] S. Dabo-Niang, N. Rhomari, Estimation non paramétrique de la régression avec variable explicative dans un espace métrique, C. R. Math. Acad. Sci. Paris, Ser. I 336 (1) (2003) 75–80.
- [4] F. Ferraty, P. Vieu, Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés, C. R. Acad. Sci. Paris, Sér. I Math. 330 (2) (2000) 139–142.
- [5] F. Ferraty, A. Laksaci, P. Vieu, Estimating some characteristics of the conditional distribution in nonparametric functional models, Stat. Inference Stoch. Process. 9 (2006) 47–76.
- [6] F. Ferraty, A. Laksaci, P. Vieu, Functional times series prediction via conditional mode, C. R. Math. Acad. Sci. Paris, Ser. I 340 (5) (2005) 389–392.
- [7] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis, Springer-Verlag, 2006.
- [8] W.V. Li, Q.M. Shao, Gaussian processes: inequalities, small ball probabilities and applications, in: C.R. Rao, D. Shanbhag (Eds.), Stochastic Processes: Theory and Methods, in: Handbook of Statistics, vol. 19, North-Holland, Amsterdam, 2001.
- [9] D. Louani, E. Ould-Saïd, Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis, J. Nonparametric Stat. 11 (4) (1999) 413–442.
- [10] E. Ould-Saïd, A note on ergodic processes prediction via estimation of the conditional mode function, Scand. J. Statist. 24 (2) (1997) 231–239.
- [11] A. Quintela del Rio, P. Vieu, A nonparametric conditional mode estimate, Nonparametric Statistics 8 (1997) 253–266.
- [12] M. Rosenblatt, Conditional probability density and regression estimators, in: P.R. Krishnaiah (Ed.), Multivariate Analysis II, Academic Press, New York and London, 1969.
- [13] E. Youndjé, Estimation non paramétrique de la densité conditionnelle par la méthode du noyau, Thèse de Doctorat, Université de Rouen, 1993.