Available online at www.sciencedirect.com
SCIENCE DIRECT
C. R. Acad. Sci. Paris, Ser. I 339 (2004) 713–716
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistique/Probabilités

Estimation nonparamétrique multidimensionnelle des dérivées de la régression

David Blondin

L.S.T.A., université Paris VI, 175, rue du Chevaleret, 75013 Paris, France

Reçu le 20 décembre 2003 ; accepté après révision le 27 septembre 2004

Disponible sur Internet le 27 octobre 2004

Présenté par Paul Deheuvels

Résumé

Nous présentons des estimateurs à noyau de type Nadaraya–Watson des dérivées de la régression dans un cadre multidimensionnel. En s'inspirant d'une méthode originale basée sur la théorie moderne des processus empiriques [Deheuvels et Mason, Stat. Inference Stoch. Process. 7 (2004)], nous établissons des lois limites concernant la déviation maximale de ces estimateurs. ***Pour citer cet article : D. Blondin, C. R. Acad. Sci. Paris, Ser. I 339 (2004).***

Abstract

Nonparametric, multidimensional estimation of regression derivatives. We establish uniform consistency rates for Nadaraya–Watson kernel-type estimators of the regression derivatives in a multidimensional framework. Our methods are based upon modern empirical process theory in the spirit of Deheuvels and Mason [Stat. Inference Stoch. Process. 7 (2004)] with respect to uniform deviations of nonparametric estimators. ***To cite this article: D. Blondin, C. R. Acad. Sci. Paris, Ser. I 339 (2004).***

1. Introduction

Soient $(X,Y), (X_1,Y_1), (X_2,Y_2), \ldots,$ des couples aléatoires indépendants et identiquement distribués à valeurs dans $\mathbb{R}^p \times \mathbb{R}^d$. Le couple de variables aléatoires (X,Y) est supposé admettre une densité jointe sur $\mathbb{R}^p \times \mathbb{R}^d$ notée $f_{X,Y}(\cdot,\cdot)$ et une densité marginale $f_X(\cdot)$ associée au vecteur X. Soit $\psi(\cdot)$ une fonction borélienne bornée de $\mathbb{R}^d$ dans $\mathbb{R}^q$. Nous nous intéressons à l'estimation de la fonction de régression (ou espérance conditionnelle) de $\psi(Y)$ sachant $X = \mathbf{x}$, notée

Adresse e-mail : blondin@ccr.jussieu.fr (D. Blondin).

1631-073X/$ – see front matter © 2004 Académie des sciences. Publié par Elsevier SAS. Tous droits réservés.
doi:10.1016/j.crma.2004.09.024

$$m_\psi(\mathbf{x}) = \mathbb{E}\big[\psi(Y)|X=\mathbf{x}\big] = \frac{1}{f_X(\mathbf{x})}\int_{\mathbb{R}^d}\psi(\mathbf{y})f_{X,Y}(\mathbf{x},\mathbf{y})\,\mathrm{d}\mathbf{y} := \frac{r_\psi(\mathbf{x})}{f_X(\mathbf{x})}, \tag{1}$$

lorsque cette expression est bien définie (cf. $(F.1)$–$(F.3)$ ci-dessous).

Notons $I = \prod_{i=1}^p [a_i, b_i]$ et $J = \prod_{i=1}^p [a_i', b_i'] \supset I$ deux hyper-rectangles compacts dans $\mathbb{R}^p$ tels que

$$-\infty < a_i' < a_i < b_i < b_i' < \infty \quad \text{pour } i = 1, \ldots, p.$$

La distribution du couple (X, Y) satisfait :

$(F.1)$ $f_{X,Y}(\cdot,\cdot)$ est continue sur $J \times \mathbb{R}^d$;
$(F.2)$ $f_X(\cdot)$ est continue et strictement positive sur J ;
$(F.3)$ $Y\mathbb{1}_{\{X\in J\}}$ est bornée.

Soit $g = (g_1, \ldots, g_q)$ avec $g_j : \mathbb{R}^p \to \mathbb{R}$, $j = 1, \ldots, q$. Soit $\mathbf{x} = (x_1, \ldots, x_p)$ un vecteur arbitraire de $\mathbb{R}^p$. Pour tout p-uplet $\mathbf{k} = (k_1, \ldots, k_p)$ d'éléments de $\mathbb{N}$, tel que $k := |\mathbf{k}| = k_1 + \cdots + k_p$, on définit $D^{(\mathbf{k})}g = g^{(\mathbf{k})} = (D^{(\mathbf{k})}g_1, \ldots, D^{(\mathbf{k})}g_q)$, où pour tout $j = 1, \ldots, q$:

$$D^{(\mathbf{k})}g_j = g_j^{(\mathbf{k})} = \big(\tfrac{\partial}{\partial x_1}\big)^{k_1}\cdots\big(\tfrac{\partial}{\partial x_p}\big)^{k_p} g_j.$$

Remarque 1. Pour les p-uplets de $\mathbb{N}$, $\mathbf{k}_0$ tel que $|\mathbf{k}_0| = 0$, et $\mathbf{k}_j$, $1 \leqslant j \leqslant p$, tels que $|\mathbf{k}_j| = k_j = 1$, on a :

$$g^{(\mathbf{k}_0)} = g = (g_1, \ldots, g_q) \quad \text{et} \quad g^{(\mathbf{k}_j)} = \big(\tfrac{\partial g_1}{\partial x_j}, \ldots, \tfrac{\partial g_q}{\partial x_j}\big).$$

Nos estimateurs des dérivées de la régression sont fondés sur l'estimateur à noyau de la régression introduit par Nadaraya [8] et Watson [11]. Par *noyau* nous entendons une fonction mesurable $K : \mathbb{R}^p \to \mathbb{R}$ vérifiant les conditions $(K.1)$–$(K.4)$ précisées ci-dessous. On désigne par $|K|_v$ la variation totale de $K(\cdot)$ sur $\mathbb{R}^p$ (cf. [7]), i.e. $|K|_v := \int_{\mathbb{R}^p} |\mathrm{d}K(\mathbf{v})|$.

$(K.1)$ $K(\cdot)$ est à variation bornée sur $\mathbb{R}^p$;
$(K.2)$ $K(\cdot)$ est à support compact ;
$(K.3)$ $\int_{\mathbb{R}^p} K(\mathbf{v})\,\mathrm{d}\mathbf{v} = 1$;
$(K.4)$ $K(\cdot)$ est k-fois différentiable, avec des dérivées partielles L telles que $|L|_v < \infty$.

Remarque 2. Les hypothèses $(K.2)$ et $(K.4)$ entraînent $(K.1)$ lorsque $k \geqslant 2$.

Nous travaillerons avec une fenêtre $h_n > 0$, indexée par $n = 1, 2, \ldots,$ et vérifiant les conditions :

$(H.1)$ $h_n \to 0$, lorsque $n \to \infty$;
$(H.2)$ $nh_n^{2k+p}/\log n \to \infty$, lorsque $n \to \infty$.

Remarque 3. Les hypothèses $(H.1)$ et $(H.2)$ sont souvent nécessaires et suffisantes pour la convergence uniforme des estimateurs présentés ci-dessous (voir [3]).

Nous nous limiterons à l'exposé des propriétés de nos estimateurs pour $k \in \{0, 1\}$, c'est à dire pour les p-uplets de $\mathbb{N}$ définis à la Remarque 1. Le cas général ($k \geqslant 2$) peut être traité sans difficulté additionnelle par des méthodes analogues. Introduisons les estimateurs à noyaux de $f_X(\mathbf{x})$, $r_\psi(\mathbf{x})$, $m_\psi(\mathbf{x})$ et de leurs dérivées partielles d'ordre 1 par rapport à x_j :

$$\hat f_X(\mathbf{x}) = \frac{1}{nh_n^p}\sum_{i=1}^n K\left(\frac{\mathbf{x}-X_i}{h_n}\right), \qquad \hat r_\psi(\mathbf{x}) = \frac{1}{nh_n^p}\sum_{i=1}^n \psi(Y_i)K\left(\frac{\mathbf{x}-X_i}{h_n}\right),$$

$$\hat{m}_\psi(\mathbf{x}) = \frac{\hat{r}_\psi(\mathbf{x})}{\hat{f}_X(\mathbf{x})} \quad \text{lorsque } \hat{f}_X(\mathbf{x}) \neq 0,$$

$$\hat{f}_X^{(\mathbf{k}_j)}(\mathbf{x}) = \frac{1}{nh_n^{p+1}} \sum_{i=1}^n K^{(\mathbf{k}_j)}\left(\frac{\mathbf{x}-X_i}{h_n}\right), \qquad \hat{r}_\psi^{(\mathbf{k}_j)}(\mathbf{x}) = \frac{1}{nh_n^{p+1}} \sum_{i=1}^n \psi(Y_i) K^{(\mathbf{k}_j)}\left(\frac{\mathbf{x}-X_i}{h_n}\right),$$

$$\hat{m}_\psi^{(\mathbf{k}_j)}(\mathbf{x}) = \frac{\hat{r}_\psi^{(\mathbf{k}_j)}(\mathbf{x})}{\hat{f}_X(\mathbf{x})} - \frac{\hat{r}_\psi(\mathbf{x})\hat{f}_X^{(\mathbf{k}_j)}(\mathbf{x})}{\hat{f}_X^2(\mathbf{x})} \quad \text{lorsque } \hat{f}_X(\mathbf{x}) \neq 0.$$

D'une manière générale, on définit $\hat{r}_\psi^{(\mathbf{k})}(\mathbf{x})$ via $\hat{r}_\psi^{(\mathbf{k})}(\mathbf{x}) = D^{(\mathbf{k})}\hat{r}_\psi(\mathbf{x})$. Pour l'estimateur de la régression $\hat{m}_\psi(\mathbf{x})$ et sa dérivée partielle d'ordre 1 par rapport à x_j, les termes de centrage sont définis par

$$\widetilde{\mathbb{E}}\big[\hat{m}_\psi(\mathbf{x})\big] := \frac{\mathbb{E}[\hat{r}_\psi(\mathbf{x})]}{\mathbb{E}[\hat{f}_X(\mathbf{x})]} \quad \text{et} \quad \widetilde{\mathbb{E}}\big[\hat{m}_\psi^{(\mathbf{k}_j)}(\mathbf{x})\big] := \frac{\mathbb{E}[\hat{r}_\psi^{(\mathbf{k}_j)}(\mathbf{x})]}{\mathbb{E}[\hat{f}_X(\mathbf{x})]} - \frac{\mathbb{E}[\hat{r}_\psi(\mathbf{x})]\mathbb{E}[\hat{f}_X^{(\mathbf{k}_j)}(\mathbf{x})]}{\mathbb{E}^2[\hat{f}_X(\mathbf{x})]}. \tag{2}$$

Le même principe est utilisé pour les opérateurs $D^{(\mathbf{k})}$ d'ordre supérieur.

Remarque 4. L'approximation de l'espérance en (2) permet la linéarisation de $\hat{m}_\psi^{(\mathbf{k})}(\mathbf{x}) - \widetilde{\mathbb{E}}[\hat{m}_\psi^{(\mathbf{k})}(\mathbf{x})]$ et son expression comme fonctionnelle du processus empirique. L'existence et la continuité des dérivées partielles d'ordre k de $f_{X,Y}$ et f_X, combinées aux hypothèses $(F.1)$–$(F.3)$, $(K.2)$–$(K.4)$ et $(H.1)$, impliquent

$$\widetilde{\mathbb{E}}\big[\hat{m}_\psi^{(\mathbf{k})}(\mathbf{x})\big] - m_\psi^{(\mathbf{k})}(\mathbf{x}) \to \mathbf{0}, \quad \text{uniformément en } \mathbf{x} \in I,$$

via une application du lemme de Bochner (cf. [1]) dans le cadre multidimensionnel.

Le but principal de cet article est d'établir des lois limites concernant la déviation maximale de nos estimateurs de la régression et de ses dérivées partielles. De nombreux auteurs ([2,6] et [9]) ont travaillé sur l'estimation des derivées de la régression par la méthode du noyau dans le cadre univarié. Dans ces articles, les différentes méthodologies permettent l'obtention de propriétés asymptotiques classiques telle la consistance ou la normalité asymptotique. Dans la prochaine section, lorsque $\psi(Y) \in \mathbb{R}$, nous obtenons une nouvelle loi limite exacte pour la convergence en probabilité de la déviation maximale associée à notre estimateur des dérivées partielles de la régression. La technique de démonstration est fondée essentiellement sur la théorie des processus empiriques indexés par certaines classes de fonctions à nombre de recouvrement uniformément polynomial. Ces classes sont de type Vapnik–Chervonenkis ou « *VC subgraph class* » (cf. [10]). Il est possible également d'étendre ces résultats lorsque $\psi(Y) \in \mathbb{R}^q$, via une normalisation adéquate. Dans ce cas précis, l'ensemble de points limites associé à la déviation ponctuelle est contenu dans une ellipsoïde liée à la structure de covariance asymptotique de nos estimateurs.

2. Résultats

2.1. Cadre unidimensionnel : $q = 1$

En utilisant des arguments similaires à Deheuvels et Mason [4], nous obtenons le résultat suivant :

Théorème 2.1. *Supposons que les hypothèses $(F.1)$–$(F.3)$, $(H.1)$ et $(H.2)$, $(K.1)$–$(K.4)$ soient vérifiées. Alors, nous obtenons,*

$$\lim_{n\to\infty}\left\{\frac{nh_n^{2k+p}}{2\log(h_n^{-p})}\right\}^{1/2} \sup_{\mathbf{x}\in I} \pm\left\{\hat{m}_\psi^{(\mathbf{k})}(\mathbf{x}) - \widetilde{\mathbb{E}}\big[\hat{m}_\psi^{(\mathbf{k})}(\mathbf{x})\big]\right\} = \sigma_\psi(I), \quad \textit{en probabilité}$$

où

$$\sigma_\psi(I) = \sup_{\mathbf{x}\in I}\left\{\frac{\mathrm{Var}[\psi(Y)|X=\mathbf{x}]}{f_X(\mathbf{x})}\int_{\mathbb{R}^p}\big[K^{(\mathbf{k})}(\mathbf{v})\big]^2\,\mathrm{d}\mathbf{v}\right\}^{1/2}. \tag{3}$$

A présent, nous présentons un exemple d'utilisation de la fonction $\psi(\cdot)$ lorsque $k=0$. Soit $\mathbf{t}$ un vecteur dans $\mathbb{R}^d$ fixé. Nous remarquons que pour le choix $\psi(\mathbf{y})=\mathbb{1}_{\{\mathbf{y}\leqslant\mathbf{t}\}}$ dans (1), nous obtenons la fonction de répartition conditionnelle $F(\mathbf{t}|\mathbf{x}):=\mathbb{P}\{Y\leqslant\mathbf{t}|X=\mathbf{x}\}$. Une application directe du Théorème 2.1 nous donne alors la loi exacte concernant la déviation maximale de l'estimateur à noyau de la fonction de répartition conditionnelle. D'après (3), la constante limite est

$$\sigma_F(I)=\sup_{\mathbf{x}\in I}\left\{\frac{F(\mathbf{t}|\mathbf{x})\{1-F(\mathbf{t}|\mathbf{x})\}}{f_X(\mathbf{x})}\int_{\mathbb{R}^p}K^2(\mathbf{v})\,\mathrm{d}\mathbf{v}\right\}^{1/2}.$$

2.2. *Généralisation multidimensionnelle : $q\geqslant 2$*

Soit $\Sigma_\psi(\mathbf{x})$ la matrice, supposée définie positive, de variance–covariance de $\psi(Y)$ conditionnelle à $X=\mathbf{x}$. Pour tout $\mathbf{x}\in I$, introduisons l'ellipsoïde $\mathcal{E}(\mathbf{x})$ définie par

$$\mathcal{E}(\mathbf{x})=\left\{\mathbf{u}\in\mathbb{R}^q\colon\ \mathbf{u}'V(\mathbf{x})^{-1}\mathbf{u}=1\right\}\quad\text{où}\quad V(\mathbf{x}):=\left\{\frac{1}{f_X(\mathbf{x})}\int_{\mathbb{R}^p}\left[K^{(\mathbf{k})}(\mathbf{v})\right]^2\mathrm{d}\mathbf{v}\right\}\times\Sigma_\psi(\mathbf{x}),$$

en remarquant que sous $(F.1)$–$(F.3)$ la matrice $V(\mathbf{x})$ est bien définie positive. Nous désignons par $\mathcal{L}(\{x_n\})$ l'ensemble limite, composé de toutes les valeurs d'adhérence de la suite $\{x_n\}$.

Nous obtenons le résultat suivant, par une modification appropriée du Lemme 2 de [5] :

Théorème 2.2. *Sous $(F.1)$–$(F.3)$, $(H.1)$ et $(H.2)$, $(K.1)$–$(K.4)$, l'ensemble limite satisfait, pour chaque $\mathbf{x}\in I$,*

$$\mathcal{L}\left(\left\{\frac{nh_n^{2k+p}}{2\log(h_n^{-p})}\right\}^{1/2}\pm\left\{\hat m_\psi^{(\mathbf{k})}(\mathbf{x})-\widetilde{\mathbb{E}}\left[\hat m_\psi^{(\mathbf{k})}(\mathbf{x})\right]\right\}\right)=\mathcal{E}(\mathbf{x}),\quad\textit{en probabilité.}$$

Soit $\mathcal{S}_q=\{\mathbf{u}\in\mathbb{R}^q\colon\ \|\mathbf{u}\|_q=1\}$ la sphère unité de dimension q, où $\|\cdot\|_q$ désigne la norme euclidienne sur $\mathbb{R}^q$. On considère la notation suivante pour définir le supremum sur le pavé I d'un vecteur $\mathbf{u}(\mathbf{x})$ à valeurs dans $\mathbb{R}^q$,

$$\sup_{\mathbf{x}\in I}\mathbf{u}(\mathbf{x}):=\mathbf{u}\left(\arg\max_{\mathbf{x}\in I}\left\{\left\|\mathbf{u}(\mathbf{x})\right\|_q\right\}\right).$$

Corollaire 2.3. *Sous les hypothèses du Théorème 2.2, nous avons,*

$$\mathcal{L}\left(\left\{\frac{nh_n^{2k+p}}{2\log(h_n^{-p})}\right\}^{1/2}\sup_{\mathbf{x}\in I}\pm V(\mathbf{x})^{-1/2}\left\{\hat m_\psi^{(\mathbf{k})}(\mathbf{x})-\widetilde{\mathbb{E}}\left[\hat m_\psi^{(\mathbf{k})}(\mathbf{x})\right]\right\}\right)=\mathcal{S}_q,\quad\textit{en probabilité.}$$

Références

[1] D. Bosq, J.P. Lecoutre, Théorie de l'Estimation Fonctionnelle, Economica, Paris, 1987.
[2] G. Collomb, Conditions nécessaires et suffisantes de convergence uniforme d'un estimateur de la régression, estimation des dérivées de la regression, C. R. Acad. Sci. Paris, Ser. I 288 (1979) 161–163.
[3] P. Deheuvels, D.M. Mason, Functional laws of the iterated logarithm for the increments of empirical and quantile processes, Ann. Probab. 20 (1992) 1248–1287.
[4] P. Deheuvels, D.M. Mason, General asymptotic confidence bands based on kernel-type function estimators, Stat. Inference Stoch. Process. 7 (2004), in press.
[5] H. Finkelstein, The law of the iterated logarithm for empirical distributions, Ann. Math. Statist. 42 (1971) 607–615.
[6] T. Gasser, H.G. Müller, Estimating regression functions and their derivatives by the kernel method, Scand. J. Statist. 11 (1984) 171–185.
[7] E.W. Hobson, The Theory of Functions of a Real Variable and the Theory of Fourier Series, vol. 1, 3rd ed., Cambridge University Press, 1927.
[8] E.A. Nadaraya, On estimating regression, Theory Probab. Appl. 9 (1964) 141–142.
[9] E.F. Schuster, S. Yakowitz, Contributions to the theory of nonparametric regression with application to system identification, Ann. Statist. 7 (1979) 139–149.
[10] A.W. Van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes with Applications to Statistics, Springer-Verlag, New York, 1996.
[11] G.S. Watson, Smooth regression analysis, Sankhyà Ser. A 26 (1964) 359–372.