Available online at www.sciencedirect.com
SCIENCE DIRECT®
C. R. Acad. Sci. Paris, Ser. I 342 (2006) 207–210
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistique/Probabilités

Vitesse de convergence presque sûre de l'estimateur à noyau du maximum de vraisemblance local

David Blondin

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

Reçu le 10 juin 2005 ; accepté après révision le 29 novembre 2005

Présenté par Paul Deheuvels

Résumé

Nous considérons l'estimation par la méthode du maximum de vraisemblance local de $\theta(x)$, paramètre inconnu de la distribution conditionnelle de Y sachant $X = x$. Le but de cette Note est d'étudier la vitesse de convergence uniforme presque sûre de l'estimateur à noyau du maximum de vraisemblance local. En s'appuyant sur la théorie moderne des processus empiriques indexés par des classes de fonctions, nous établissons une loi uniforme du logarithme concernant la déviation maximale de cet estimateur. ***Pour citer cet article : D. Blondin, C. R. Acad. Sci. Paris, Ser. I 342 (2006).***

Abstract

Almost sure rate of uniform consistency for the local maximum likelihood kernel estimator. We consider the local maximum likelihood estimation of $\theta(x)$, unknown parameter of the conditional distribution of Y given $X = x$. The aim of this Note is the study of strong uniform consistency rates of the local maximum likelihood kernel estimator. Under suitable regularity conditions, we establish a uniform law of the logarithm for the maximal deviation of this estimator. The method of proof is based upon functional limit laws derived by modern empirical process theory. ***To cite this article: D. Blondin, C. R. Acad. Sci. Paris, Ser. I 342 (2006).***

1. Introduction

Soient (X, Y), (X_1, Y_1), (X_2, Y_2), $\ldots$, des couples de variables aléatoires à valeurs réelles, indépendants et identiquement distribués. Le couple aléatoire (X, Y) est supposé admettre une densité jointe $f_{X,Y}$ par rapport à la mesure de Lebesgue sur $\mathbb{R}^2$ et une densité marginale f_X associée à X. Dans le cadre de notre étude, pour chaque couple $(x, y) \in \mathbb{R}^2$, la loi conditionnelle de Y sachant $X = x$ est définie par sa densité, notée $f_{Y|X}(y, x)$, et supposée de la forme :

$$f_{Y|X}(y, x) := g\big(y; \theta(x)\big), \tag{1}$$

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

1631-073X/$ – see front matter
doi:10.1016/j.crma.2005.12.011

où g est une fonction supposée connue et $\theta(x)$ désigne le paramètre fonctionnel à estimer. A titre d'exemple, si la distribution de Y sachant $X=x$ est une Exponentielle de paramètre $\theta(x)$, l'estimation du paramètre fonctionnel $\theta(x)$ se réduit alors à celle de la courbe classique de régression $\mathbb{E}[Y|X=x]$. Nous nous placerons dans un cadre non-paramétrique dans le sens où la fonction $\theta(x)$ sera seulement supposée deux fois continûment dérivable (voir (F.5), ci-après).

L'objet central de cette note sera de caractériser la vitesse exacte de convergence uniforme presque sûre pour une suite d'estimateurs à noyau de $\theta(x)$, obtenue en maximisant localement la fonction de vraisemblance conditionnelle empirique associée aux $\{(X_i,Y_i)\}_{i=1}^n$. Nos résultats seront établis uniformément en $x\in I\subseteq J$, où I et J désignent deux intervalles compacts de $\mathbb{R}$ contenus dans le support de la densité f_X. Nous introduisons T, le domaine de variation de $\theta(x)$ lorsque $x\in J$, qui constitue également un intervalle compact de $\mathbb{R}$. Sous certaines conditions de régularité (cf. (F.3–4) ci-dessous) et sous réserve d'identifiabilité de notre modèle, $\theta(x)$ est défini comme la solution du problème de maximisation suivant

$$\theta(x):=\arg\max_{t\in T}\big\{\mathbb{E}_x\big[\log g(Y;t)\big]\big\}\quad \text{ou } \mathbb{E}_x\big[\psi\big(Y;\theta(x)\big)\big]=0, \tag{2}$$

lorsque $\mathbb{E}_x[\cdot]:=\mathbb{E}[\cdot\,|X=x]$ et $\psi(y;t):=\frac{\partial \log g(y;t)}{\partial t}$.

Soit $I_\theta(x)$ l'information de Fisher locale ou variance conditionnelle de la variable $\psi(Y;\theta(x))$, définie par

$$I_\theta(x):=\mathbb{E}_x\big[\big\{\psi(Y;t)\big\}^2\big]_{t=\theta(x)}=\mathbb{E}_x\left[-\left\{\frac{\partial\psi(Y;t)}{\partial t}\right\}\right]_{t=\theta(x)}=-\mathbb{E}_x\big[\psi'(Y;t)\big]_{t=\theta(x)}\quad \text{cf. (F.3).} \tag{3}$$

Nous imposons certaines conditions sur la distribution du couple (X,Y), parmi les hypothèses (F.1–5), présentées ci-dessous.

(F.1) $f_X(\cdot)$ et $I_\theta(\cdot)$ sont continues et strictement positives sur J.
(F.2) $Y\mathbb{1}_{\{X\in J\}}$ est bornée.
(F.3) Les dérivées partielles d'ordre $1,2,3$ (par rapport à t) de $\log g(y,t)$ existent et sont continues sur $\mathbb{R}\times T$. De plus, il existe des fonctions $H_i(y)$ intégrables telles que

$$\left|\frac{\partial^i \log g(y;t)}{\partial t^i}\right|\leqslant H_i(y),\quad \text{pour } i=1,2.$$

(F.4) Il existe des constantes strictement positives C_1 et C_2 telles que

$$\inf_{x\in I}\mathbb{E}_x\big[-\psi'\big(Y;\theta(x)+\epsilon\big)\big]>C_2>0,\quad \text{lorsque } |\epsilon|\leqslant C_1.$$

(F.5) Les dérivées $f_X'(x)$, $I_\theta'(x)$, $\theta'(x)$ et $\theta''(x)$ sont continues et bornées.

L'hypothèse (F.1) est essentielle car si la densité marginale $f_X(x)$ ou l'information de Fisher locale $I_\theta(x)$ sont nulles, $\theta(x)$ ne peut pas être estimé convenablement par la méthode du maximum de vraisemblance. L'hypothèse (F.2) est classique en régression non-paramétrique, elle peut être relaxée via une condition de moment appropriée (cf. Deheuvels et Mason [2]). L'hypothèse (F.3) constitue l'extension naturelle des conditions nécessaires à la théorie classique du maximum de vraisemblance (voir, par exemple, Serfling [6], §4.2.2, pp. 144–149). La condition (F.4) justifie notre approche en (2) et permettra de démontrer la consistance forte de notre estimateur, étape première et fondamentale de la démonstration de notre loi limite. Enfin, les conditions de régularité en (F.5) sont liées au traitement du biais de notre estimateur (voir ci-dessous) construit avec un noyau K d'ordre 2 (cf. (K.2)).

D'après (2), pour chaque $n\geqslant 1$, l'estimateur $\hat\theta_n(x)$ du maximum de vraisemblance local, construit avec la méthode du noyau, est défini comme la solution par rapport à t de l'équation suivante

$$\hat r_n(x,t)=\frac{1}{nh}\sum_{i=1}^n\psi(Y_i;t)K\left(\frac{x-X_i}{h}\right)=0\quad \big(\text{i.e. } \hat r_n\big(x,\hat\theta_n(x)\big)=0\big), \tag{4}$$

où $K:\mathbb{R}\to\mathbb{R}$ désigne le noyau et h le paramètre de lissage ou fenêtre. Nous obtenons alors une suite d'estimateurs $\{\hat\theta_n(x);\ n\geqslant 1\}$ du maximum de vraisemblance local. Cette technique d'estimation, dénommée '*local likelihood estimation*', a pour origine une idée développée par Tibshirani et Hastie [8]. La normalité asymptotique de cet estimateur a été démontrée par Staniswalis [7] dans le cadre du plan fixe. En s'appuyant sur les travaux de Härdle, Janssen et Serfling [5], Zhao [9] a obtenu la vitesse optimale de convergence presque sûre du supremum sur un intervalle compact de la déviation $\hat{\theta}_n(x) - \theta(x)$. Notre but principal sera de compléter ce résultat en énonçant une loi limite pour la déviation maximale de l'estimateur du maximum de vraisemblance local. Cette loi limite uniforme du logarithme est proche dans la forme de certains travaux célèbres en statistique fonctionnelle, parmi lesquels nous pouvons citer les articles de Einmahl et Mason [3] ainsi que Deheuvels et Mason [2] concernant l'estimation non-paramétrique des fonctions de densité et de régression.

Par convenance, on désigne par $[u^j K]$ le j-ième moment associé au noyau K, $j \in \mathbb{N}$. Le noyau K est supposé vérifier les hypothèses suivantes.

(K.1) K est une fonction de densité continue par morceaux et à variation bornée sur $\mathbb{R}$;
(K.2) K est un noyau d'ordre 2 (i.e. $[uK] = 0$ et $[u^2 K] < \infty$) à support compact.

Afin d'éviter les valeurs négatives du logarithme, nous considérons la notation $L(u) = \log(u \vee e)$. Nous travaillerons avec une fenêtre $(h_n)_{n \geqslant 1}$, suite de constantes strictement positives, vérifiant les conditions de monotonie et de convergence énoncées ci-dessous.

(H.1) $h_n \searrow 0$, $nh_n \nearrow \infty$, $nh_n/L(h_n^{-1}) \to \infty$, $L(h_n^{-1})/LLn \to \infty$ et $L(h_n^{-1})/nh_n^5 \to \infty$, lorsque $n \to \infty$.

L'hypothèse (H.1) regroupe des conditions classiques sur le paramètre de lissage en estimation non-paramétrique par la méthode du noyau.

2. Résultats

Nos hypothèses garantissent l'existence d'une suite de solutions à l'Éq. (4) mais pas leur unicité. Par la suite, nous supposerons que notre estimateur vérifie

$$\sup_{x \in I} \left| \hat{\theta}_n(x) - \theta(x) \right| < \tau, \quad \text{presque sûrement,}$$

où τ désigne une constante choisie suffisamment petite. Ainsi, toutes les racines de l'Éq. (4) sont suffisamment proches les unes des autres. Pour plus de détails concernant cet argument classique, nous renvoyons à Zhao, pp. 82–83, [9] ou Serfling [6].

Théorème 2.1. *Supposons que les hypothèses* (F.1–5), (H.1) *et* (K.1–2) *soient vérifiées. Alors, nous avons, lorsque* $n \to \infty$,

$$\left| \left\{ \frac{nh_n}{2L(1/h_n)} \right\}^{1/2} \sup_{x \in I} \pm \left\{ \hat{\theta}_n(x) - \theta(x) \right\} - \sigma_\theta(I) \right| \overset{p.s.}{=} \mathrm{o}(1), \tag{5}$$

où

$$\sigma_\theta(I) = \sup_{x \in I} \left\{ \frac{1}{f_X(x) I_\theta(x)} \int_{\mathbb{R}} \left[K^2(u) \right] \mathrm{d}u \right\}^{1/2}.$$

Contrairement aux résultats de Zhao [9] ou Härdle et al. [5], cette loi limite peut servir de point de départ à la construction d'intervalles de confiance pour le paramètre $\theta(x)$, uniformément en $x \in I$ (cf. [2]). Précisons que pour la plupart des applications statistiques, la convergence en probabilité est une notion suffisante et alors nous pouvons réduire les hypothèses sur la fenêtre h_n. Il est également possible d'obtenir la consistance forte de $\hat{\theta}_n(x)$ lorsque la fenêtre est aléatoire, de la forme $\hat{h}_n$, telle que

$$a_n \leqslant \hat{h}_n \leqslant b_n, \quad \text{presque sûrement,}$$

avec $\{a_n;\ n \geqslant 1\}$ et $\{b_n;\ n \geqslant 1\}$ deux suites telles que $0 < a_n < b_n \leqslant 1$, $n \geqslant 1$, satisfaisant $b_n \to 0$ et $na_n/Ln \to \infty$, lorsque $n \to \infty$ (cf. Einmahl et Mason [4]). Précisons enfin que le Théorème 2.1 se généralise aisément au cadre multivarié en suivant, par exemple, les extensions multidimensionnelles concernant l'estimation non-paramétrique de la fonction de régression (cf. [1]).

3. Résumé de la démonstration

Posons, pour $x \in I$ et $t \in T$,

$$\hat{r}_n'(x,t) = \frac{\partial}{\partial t}\big[\hat{r}_n(x,t)\big] = \frac{1}{nh_n}\sum_{i=1}^{n}\psi'(Y_i;t)K\left(\frac{x-X_i}{h_n}\right).$$

La démonstration se décompose en trois étapes principales. Dans un premier temps, nous établissons la consistance forte de l'estimateur du maximum de vraisemblance (avec une vitesse préliminaire convenable)

$$\sup_{x\in I}\big|\hat{\theta}_n(x)-\theta(x)\big| \overset{p.s.}{=} \mathrm{O}\left(\left\{\frac{L(h_n^{-1})}{nh_n}\right\}^{1/2} + h_n\right) = \mathrm{o}(1).$$

Ce premier résultat, combiné au developpement de Taylor de la fonction $\hat{r}_n(x,\cdot)$ au voisinage de $\theta(x)$, nous conduit à réduire l'étude asymptotique de la déviation $\hat{\theta}_n(x)-\theta(x)$ à celle d'un terme de régression, i.e.

$$\left\{\frac{nh_n}{L(h_n^{-1})}\right\}^{1/2}\big\{\hat{\theta}_n(x)-\theta(x)\big\} = \left\{\frac{nh_n}{L(h_n^{-1})}\right\}^{1/2}\left\{\frac{\hat{r}_n(x,\theta(x))}{-\hat{r}_n'(x,\theta(x))}\right\} + \mathrm{o}(1).$$

Le reste de la démonstration consiste alors à établir la vitesse de convergence presque sûre et une loi limite pour le terme au numérateur, puis, en utilisant un argument du type Slutsky en combinaison avec l'étude du terme aléatoire au numérateur, nous obtenons (5). Les outils de démonstration sont principalement fondés sur la théorie moderne des processus empiriques et plus particulièrement les travaux développés par Einmahl et Mason [3] et Deheuvels et Mason [2].

Références

[1] D. Blondin, Estimation nonparamétrique multidimensionnelle des dérivées de la régression, C. R. Math. Acad. Sci. Paris, Ser. I 339 (10) (2004) 713–716.
[2] P. Deheuvels, D.M. Mason, General asymptotic confidence bands based on kernel-type function estimators, Stat. Inference Stoch. Process. 7 (3) (2004) 225–277.
[3] U. Einmahl, D.M. Mason, An empirical process approach to the uniform consistency of kernel-type function estimators, J. Theor. Probab. 13 (1) (2000) 1–37.
[4] U. Einmahl, D.M. Mason, Uniform in bandwidth consistency of kernel-type functions estimators, Ann. Statist. 33 (3) (2005), à paraître.
[5] W. Härdle, P. Janssen, R. Serfling, Strong uniform consistency rates of estimators of conditional functionals, Ann. Statist. 16 (4) (1988) 1428–1449.
[6] R. Serfling, Approximations Theorems of Mathematical Statistics, Wiley, New York, 1980.
[7] J.G. Staniswalis, The kernel estimate of a regression function in likehood-based models, J. Amer. Statist. Assoc. 84 (405) (1989) 276–283.
[8] R. Tibshirani, T. Hastie, Local likelihood estimation, J. Amer. Statist. Assoc. 82 (398) (1987) 559–567.
[9] P.-L. Zhao, Asymptotics of kernel estimators based on local maximum likelihood, Nonparametric Statist. 4 (1994) 79–90.