Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 345 (2007) 411–414
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistics

CLT and $\mathbb{L}^q$ errors in nonparametric functional regression

Laurent Delsol

L.S.P., Université Paul-Sabatier, 31062 Toulouse, France

Received 15 November 2006; accepted after revision 23 July 2007

Available online 29 September 2007

Presented by Paul Deheuvels

Abstract

We study a nonparametric functional regression model and we provide an asymptotic law with explicit constants under α-mixing assumptions. Then we establish both pointwise confidence bands for the regression operator and asymptotic $\mathbb{L}^q$ errors for its kernel estimator. ***To cite this article: L. Delsol, C. R. Acad. Sci. Paris, Ser. I 345 (2007).***

Résumé

Loi et erreurs $\mathbb{L}^q$ asymptotiques dans un modèle de régression non-paramétrique. On étudiera dans cet article la normalité asymptotique de l'estimateur à noyau pour des données α-mélangeantes fonctionnelles. L'explicitation des constantes apparaissant dans la loi asymptotique permet d'établir des intervalles de confiance ponctuels pour l'opérateur de régression ainsi que l'expression des erreurs $\mathbb{L}^q$. ***Pour citer cet article : L. Delsol, C. R. Acad. Sci. Paris, Ser. I 345 (2007).***

1. Introduction

We focus on the usual regression model: $Y = r(X) + \epsilon$, where Y is a real random variable and X a random variable which takes values on a semi-metric space (E, d). Only regularity assumptions are made on r that is why the model is called nonparametric. Then, we consider a dataset of n pairs (X_i, Y_i) identically distributed as (X, Y) which may be dependent.

One considers an element x of E and estimates $r(x)$ by the following kernel estimator:

$$\hat{r}(x) = \frac{\sum_{i=1}^{n} Y_i K(\frac{d(X_i,x)}{h_n})}{\sum_{i=1}^{n} K(\frac{d(X_i,x)}{h_n})},$$

where K is a kernel function and h_n a smoothing parameter. This estimator has been introduced by Ferraty and Vieu [3] to generalise to the functional case the classical Nadaraya–Watson estimator. The results on $\mathbb{L}^q$ errors given below are new even in the independent and multivariate case.

E-mail address: delsol@cict.fr.

1631-073X/$ – see front matter
doi:10.1016/j.crma.2007.07.021

2. Notation and main assumptions

Firstly, we introduce some key-functions which take an important place in our approach:

$$\phi(s)=\mathbb{E}\big[\big(r(X)-r(x)\big)\mid d(X,x)=s\big],\quad F(h)=\mathbb{P}\big(d(X,x)\leqslant h\big),\quad \tau_h(s)=\frac{F(hs)}{F(h)}.$$

We start with some assumptions on the model (already introduced in [2]):

$$r \text{ is bounded on a neighbourhood of } x, \tag{1}$$

$$F(0)=0 \text{ and } \phi'(0) \text{ exists}, \tag{2}$$

$$\sigma_\epsilon^2(s):=\mathbb{E}[\epsilon^2\mid X=s] \text{ is continuous on a neighbourhood of } x \text{ and } \sigma_\epsilon^2:=\sigma_\epsilon^2(x)>0, \tag{3}$$

$$\forall s\in[0;1],\quad \lim_{n\to+\infty}\tau_{h_n}(s)=\tau_0(s)\quad \text{with } \tau_0(s)\neq 1_{[0;1]}(s). \tag{4}$$

Assumption (2) enables to get the exact expression of the constants where a standard Lipschitz condition would only give upper bounds. Then, one may note that assumptions dealing with the law of X only concern small ball probabilities. Besides, many standard processes fulfill assumption (4). In order to control the sum of covariances we propose the following assumptions:

$$\exists p>2,\ \exists M>0,\quad \mathbb{E}\big[|\epsilon|^p\mid X\big]\leqslant M \text{ a.s.}, \tag{5}$$

$$\exists C,\ \forall i,j\in\mathbb{Z}\quad \max\big(\mathbb{E}\big[|\epsilon_i\epsilon_j|\mid X_i,X_j\big],\mathbb{E}\big[|\epsilon_i|\mid X_i,X_j\big]\big)\leqslant C \text{ a.s.} \tag{6}$$

Finally we make some assumptions on the functional kernel estimator:

$$h_n=\mathrm{O}\bigg(\frac{1}{\sqrt{nF(h_n)}}\bigg),\qquad \lim_{n\to+\infty}nF(h_n)=+\infty, \tag{7}$$

$$K \text{ has a compact support } [0;1], \text{ is } C^1 \text{ and non-increasing on }]0;1[,\ K(1)>0. \tag{8}$$

Our results will be expressed using the following constants:

$$M_0=\bigg(K(1)-\int_0^1\big(sK(s)\big)'\tau_0(s)\,\mathrm{d}s\bigg),\qquad M_j=\bigg(K^j(1)-\int_0^1(K^j)'(s)\tau_0(s)\,\mathrm{d}s\bigg),\quad j=1,2,$$

and the random variable Z_n defined with $B_n=h_n\phi'(0)\frac{M_0}{M_1}$ by:

$$Z_n:=\frac{M_1}{\sqrt{M_2\sigma_\epsilon^2}}\sqrt{nF(h_n)}\big(\hat r(x)-r(x)-B_n\big).$$

If (X_i,Y_i) are dependent, we assume them to be α-mixing (see [7], p. 34, Notation 2.1). The mixing coefficients are denoted by $\{\alpha(n),\ n\in\mathbb{N}\}$ and we introduce the notations:

$$\forall k\geqslant 2,\quad \Theta_k(s):=\max\Big(\max_{1\leqslant i_1<\cdots<i_k\leqslant n}P\big(d(X_{i_j},x)\leqslant s,\ 1\leqslant j\leqslant k\big),\ F^k(s)\Big),$$

$$\Gamma_i:=Y_iK\bigg(\frac{d(X_i,x)}{h_n}\bigg),\quad \Delta_i:=K\bigg(\frac{d(X_i,x)}{h_n}\bigg),\quad U_{i,n}=\frac{\Gamma_i\mathbb{E}[\Delta_i]-\Delta_i\mathbb{E}[\Gamma_i]}{F(h_n)\sqrt{nF(h_n)}}.$$

We need the following assumptions:

$$\exists (u_n)_{n\in\mathbb{N}}\in\mathbb{N}^{\mathbb{N}},\quad \mathrm{O}\bigg(\frac{n[\alpha(u_n)]^{\frac{p-2}{p}}}{F(h_n)^{\frac{p-2}{p}}}\bigg)+\mathrm{O}\bigg(u_n\frac{\Theta_2(h_n)}{F(h_n)}\bigg)\xrightarrow{n\to+\infty}0, \tag{H1}$$

$$\text{and}\quad I_n:=n\int_0^1\alpha^{-1}\Big(\frac{s}{2}\Big)Q_{U_{1,n}}^2(s)\inf\bigg(\frac{3M_1\sqrt{\sigma_\epsilon^2M_2}}{2},\alpha^{-1}\Big(\frac{s}{2}\Big)Q_{U_{1,n}}(s)\bigg)\mathrm{d}s\to 0,\quad \text{(see [6])} \tag{H2}$$

where $Q_{U_{i,n}}(x)=\inf\{t,\mathbb{P}(|U_{i,n}|>t)\geqslant x\}$, and $\alpha^{-1}(\frac{x}{2})=\inf\{t,\alpha([t])\geqslant\frac{x}{2}\}$. These assumptions will be simplified in the particular case of arithmetic α-mixing coefficients (see (9) and (10) below).

3. Asymptotic normality

Theorem 3.1. *Under assumptions* (1)–(8)*,* (H1) *and* (H2) *we get*

$$Z_n \to N(0,1).$$

Remark. If the α-mixing coefficients are arithmetic of order a: $\alpha(i) \leqslant Ci^{-a}$, assumptions (H1) and (H2) may be replaced, thanks to [4], by the following ones:

$$\exists \nu > 0, \quad \Theta_2(h_n) = \mathrm{O}\big(F(h_n)^{1+\nu}\big), \quad \text{with } a > \frac{(1+\nu)p-2}{\nu(p-2)}, \tag{9}$$

$$\exists \gamma > 0, \quad nF^{1+\gamma}(h_n) \to +\infty \quad \text{and} \quad a > \frac{2}{\gamma}+1. \tag{10}$$

4. Pointwise asymptotic confidence bands

To get asymptotic confidence bands one needs to estimate the constants appearing in Theorem 3.1. Whereas M_1 and M_2 seem to be easily estimated, the bias term is more difficult to study. To avoid this problem one makes an additional assumption on h_n which will make the bias negligible with respect to $\sqrt{nF(h_n)}^{-1}$. Taking

$$\hat{M}_2(x) := \frac{1}{n\hat{F}(h_n)}\sum_{i=1}^{n} K^2\left(\frac{d(X_i,x)}{h_n}\right), \qquad \hat{M}_1(x) := \frac{1}{n\hat{F}(h_n)}\sum_{i=1}^{n} K\left(\frac{d(X_i,x)}{h_n}\right),$$

where $\hat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{[d(X_i,x),+\infty[}(t)$, we get the following corollary:

Corollary 4.1. *Under assumptions of Theorem* 3.1*, if* $h_n\sqrt{nF(h_n)} \to 0$*, then*:

$$\frac{\hat{M}_1}{\sqrt{\hat{M}_2\hat{\sigma_\epsilon^2}}}\sqrt{n\hat{F}(h_n)}\big(\hat{r}(x) - r(x)\big) \to N(0,1),$$

where $\hat{\sigma_\epsilon^2}$ *is an estimator converging in probability to* σ_ϵ^2*.*

5. Asymptotic expressions of $\mathbb{L}^q$ errors

We get directly from the asymptotic law of Z_n and the uniform integrability of $|Z_n|^q$ the asymptotic expressions of the moments of order q and the $\mathbb{L}^q$ errors of the kernel estimator. To get the uniform integrability of $|Z_n|^q$ we need additional assumptions:

$$\exists t,\ 2 \leqslant t < p,\ \forall k \leqslant t,\ \exists \nu_k > 0, \quad \Theta_k(s) = \mathrm{O}\big(F(s)^{1+\nu_k}\big)$$

$$\text{with } \nu_k \leqslant \nu_{k-1}+1 \text{ and } a > \max_{2\leqslant k\leqslant t}(k-1)\frac{(1+\nu_k)p-t}{\nu_k(p-t)}, \tag{11}$$

$$\exists u,\ 2 \leqslant u \leqslant p,\ \exists M, \quad \max_{i,n}\mathbb{E}\big[|\epsilon_i|^u \mid X_1,\ldots,X_n\big] \leqslant M \text{ a.s.} \tag{12}$$

From a statistical point of view, to choose the smoothing parameter, one often focuses on the $\mathbb{L}^q$ errors of the kernel estimator. That is the aim of the following theorem in which we explicit their asymptotic dominant terms whether q is even or not. In particular the case q odd, leads us to introduce the two following sequences of polynomials: $P_{2m+1}(u) = \sum_{l=0}^{m} a_{m,l}u^{2l+1}$ and $Q_{2m+1}(u) = \sum_{l=0}^{m} b_{m,l}u^{2l}$, where

$$a_{m,l} = \frac{(2m+1)!}{(2l+1)!2^{(m-l)}(m-l)!},$$

$$b_{m,l} = \sum_{j=m-l+1}^{m}\left[C_{2m+1}^{2j+1}\frac{2^j j!}{2^{j+l-m}(j+l-m)!} - C_{2m+1}^{2j}\frac{(2j)!2^{j+l-m}(j+l-m)!}{2^j j!(2(j+l-m))!}\right] + C_{2m+1}^{2(m-l)+1}2^{m-l}(m-l)!,$$

and the function $\psi_m(u) = (2G(u)-1)P_{2m+1}(u) + 2g(u)Q_{2m+1}(u)$ (with the convention that a sum on an empty set of indices equals zero).

Theorem 5.1. *Under assumptions* (1)–(8) *and taking* $\ell = p$ *in the independent case or* (1)–(12) *and taking* $\ell = 2[\frac{\min(t,u)}{2}]$ *in the arithmetic* α*-mixing case, one gets*:

(i) $\forall 0 \leqslant q < \ell$,

$$\mathbb{E}\big[|\hat{r}(x) - r(x)|^q\big] = \mathbb{E}\left[\left|h_n\phi'(0)\frac{M_0}{M_1} + W\sqrt{\frac{M_2\sigma_\epsilon^2}{nF(h_n)M_1^2}}\right|^q\right] + \mathrm{o}\left(\frac{1}{(nF(h_n))^{\frac{q}{2}}}\right),$$

(ii) $\forall 0 \leqslant 2m < \ell$,

$$\mathbb{E}\big[|\hat{r}(x) - r(x)|^{2m}\big] = \sum_{k=0}^{m} \frac{(\frac{M_2\sigma_\epsilon^2}{M_1^2})^k(\frac{M_0}{M_1}\phi'(0))^{2(m-k)}(2m)!}{(2(m-k))!k!2^k}\frac{h_n^{2(m-k)}}{(nF(h_n))^k} + \mathrm{o}\left(\frac{1}{(nF(h_n))^m}\right),$$

(iii) $\forall 0 \leqslant 2m+1 < \ell$,

$$\mathbb{E}\big[|\hat{r}(x) - r(x)|^{2m+1}\big] = \left(\frac{M_2\sigma_\epsilon^2}{M_1^2 nF(h_n)}\right)^{m+\frac{1}{2}}\psi_m\left(\frac{h_n\phi'(0)M_0\sqrt{nF(h_n)}}{\sqrt{M_2\sigma_\epsilon^2}}\right) + \mathrm{o}\left(\frac{1}{(nF(h_n))^{m+\frac{1}{2}}}\right).$$

6. Conclusions and perspectives

The above results complete [2] and [5] and give explicit constants in the asymptotic law and $\mathbb{L}^q$ errors instead of giving only upper bounds. Consequently, they may be used to provide confidence bands or choose the smoothing parameter that is asymptotically optimal with regard to the $\mathbb{L}^q$ error. Besides they generalise former works on the $\mathbb{L}^1$ (see [8]) and $\mathbb{L}^2$ errors with multivariate variables to the functional case and for other orders.

For the future, it would be of interest to get an 'integrated version' of asymptotic normality of $\hat{r}(x)$ notably to construct some specification testing procedures. It would also be interesting to get similar results for the integrated errors. Finally, we might attempt to extend our results to a more general case of α-mixing variables introduced in [1].

Acknowledgements

I would like to thank the referee for his constructive comments and suggestions to improve this Note. I am also grateful to the group STAPH with special thanks to F. Ferraty and P. Vieu.

References

[1] J. Dedecker, C. Prieur, New dependence coefficients. Examples and applications to statistics, Probab. Theory Related Fields 132 (2005) 203–236.
[2] F. Ferraty, A. Mas, P. Vieu, Advances on nonparametric regression for functional data, ANZ Journal of Statistics (2007), in press.
[3] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis, Springer-Verlag, New York, 2006.
[4] E. Liebscher, Central limit theorem for α-mixing triangular arrays with applications to nonparametric statistics, Mathematical Methods of Statistics 10 (2) (2001) 194–214.
[5] E. Masry, Nonparametric regression estimation for dependent functional data: asymptotic normality, Stochastic Process. Appl. 115 (1) (2005) 155–177.
[6] E. Rio, About the Lindeberg method for strongly mixing sequences, ESAIM 1 (1995) 35–61.
[7] E. Rio, Théorie asymptotique des processus aléatoires faiblement dépendants, Mathématiques et applications, vol. 31, Springer-Verlag, Berlin, 2000.
[8] M.P. Wand, On exact $\mathbb{L}^1$ rates in nonparametric kernel regression, Scand. J. Statist. 17 (3) (1990) 251–256.