C. R. Acad. Sci. Paris, Ser. I 355 (2017) 824–829

Contents lists available at ScienceDirect

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

www.sciencedirect.com

Statistique

Estimation locale linéaire de la régression non paramétrique fonctionnelle par la méthode des k plus proches voisins

Local linear estimate of the regression operator by the kNN method

Mohammed Attouch, Ali Laksaci, Fatima Rafaa

Laboratoire de Statistique et Processus Stochastiques, Université Djillali Liabes, Sidi Bel Abbes, Algérie

INFO ARTICLE

Historique de l'article :
Reçu le 3 février 2017
Accepté après révision le 24 mai 2017
Disponible sur Internet le 13 juin 2017

Présenté par le comité de rédaction

RÉSUMÉ

Nous étudions l'estimation locale linéaire de l'opérateur de régression lorsque la variable explicative prend ses valeurs dans un espace semi-métrique. Nous construisons un estimateur par la méthode des k plus proches voisins. Deux propriétés asymptotiques de cet estimateur seront établies. Dans la première partie, nous prouvons la convergence presque complète ponctuelle, tandis que, dans la deuxième, nous montrons la convergence presque complète uniforme sur le nombre de voisins.

ABSTRACT

We consider the problem of the local linear estimation of the regression operator when the regressor is functional. We construct an estimator by the kNN method and we study its asymptotic properties. Precisely, we establish the almost complete consistency of this estimator with rate both pointwise and uniform on the number of neighbor cases.

Abridged English version

Let us introduce n pairs of random variables (X_i, Y_i) for $i = 1, \dots, n$ that we assume drawn from the pair (X, Y), which is valued in $\mathcal{F} \times \mathcal{R}$, where $\mathcal{F}$ is a semi-metric space equipped with a semi-metric d. Furthermore, we assume that the link between X and Y is modeled by the following relation

$$Y = r(X) + \epsilon,$$

where r is an operator from $\mathcal{F}$ to $\mathcal{R}$ and ϵ is a random error variable such that $\mathbb{E}[\epsilon|X] = 0$.

In this work, we combine the ideas of the local linear approach to those of the k-NN method to estimate the regression operator r. Noting that, in functional statistics, there are several ways for extending the local linear approach. Here we adopt the fast version proposed by Barrientos-Marin et al. [2], for whom the regression function $r(.)$ is locally approximated by

Adresses e-mail : attou_kadi@yahoo.fr (M. Attouch), alilak@yahoo.fr (A. Laksaci), rafaa_fatima@yahoo.com (F. Rafaa).

http://dx.doi.org/10.1016/j.crma.2017.05.007
1631-073X/

$$r(x_0) = a + b\beta(x_0, x) + o(\beta(x_0, x)), \forall x_0 \text{ in a neighborhood of } x \text{ in } \mathcal{F} \tag{1}$$

where $\beta(.,.)$ is a known function from $\mathcal{F}^2$ into $\mathcal{R}$ such that, $\forall \xi \in \mathcal{F}$, $\beta(\xi, \xi) = 0$. The coefficients a and b are estimated by

$$\min_{a,b\in\mathcal{R}} \sum_{i=1}^{n} (Y_i - a - b\beta(X_i, x))^2 K(H_{k,x}^{-1}\delta(x, X_i))$$

where $\delta(.,.)$ is a function of $\mathcal{F} \times \mathcal{F}$ such that $d(.,.) = |\delta(.,.)|$, K is a kernel and

$$H_{k,x} = \min\left\{h \in \mathcal{R}^+ \text{ such that } \sum_{i=1}^{n} \mathbb{1}_{B(x,h)}(X_i) = k\right\}.$$

These estimates are explicitly expressed by

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n} \beta_i(\beta_i - \beta_j)K_iK_j} \begin{pmatrix} \sum_{i=1}^{n}\sum_{j=1}^{n} \beta_i(\beta_i - \beta_j)K_iK_jY_j \\ \sum_{i=1}^{n}\sum_{j=1}^{n} (\beta_j - \beta_i)K_iK_jY_j \end{pmatrix}$$

where $\beta_i = \beta(X_i, x)$ et $K_i = K(H_{k,x}^{-1}\delta(x, X_i))$.

By using (1) and the fact that $\beta(x, x) = 0$, we define the kNN local linear estimate of $r(x)$ by

$$\widehat{r}(x) = \hat{a} = \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n} \beta_i(\beta_i - \beta_j)K_iK_j} \sum_{i=1}^{n}\sum_{j=1}^{n} \beta_i(\beta_i - \beta_j)K_iK_jY_j.$$

The main goal of this Note is to study the almost complete consistency of $\widehat{r}(x)$. Precisely, our main results are given in the following theorems (see the French version for the notations and assumptions)

Theorem 0.1. *Under the hypotheses (H1)–(H7) we have,*

$$\widehat{r}(x) - r(x) = O\left(\left(\phi_x^{-1}\left(\frac{k}{n}\right)\right)^{\beta} + \left(\sqrt{\frac{\log n}{k}}\right)\right) \qquad \text{a.co.}$$

Theorem 0.2. *Under the hypotheses (U1)–(U7) we have,*

$$\sup_{k_{1,n}\le k\le k_{2,n}} |\widehat{r}(x) - r(x)| = O\left(\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\beta}\right) + O_{\text{a.co.}}\left(\sqrt{\frac{\log n}{k_{1,n}}}\right).$$

1. Introduction

Le cadre général de cette note est la modélisation locale linéaire de la fonction de régression. Plus précisément, nous construisons, par la méthode des k plus proches voisins (kNN), un estimateur pour ce modèle et nous établissons ses propriétés asymptotiques dans le cas où la variable explicative est fonctionnelle.

Notons que la modélisation locale linéaire est une approche alternative à l'estimation de Nadaraya–Watson (NW), qui a plus d'avantages sur cette dernière (voir [8]). En particulier, le plus grand avantage de la méthode des polynômes locaux sur la méthode du noyau est la réduction considérable du biais de l'estimateur. Motivée par cette supériorité sur l'estimation de NW, l'estimation locale linaire de la régression fonctionnelle a fait l'objet de plusieurs travaux en statistique fonctionnelle. Nous citons, par exemple, [2,1,5]. Dans la plupart de ces travaux, les auteurs utilisent une pondération issue de la méthode du noyau. Dans cette note, nous construisons notre estimateur à l'aide de la méthode des k plus proches voisins. Rappelons que l'importance de l'estimation par la méthode de kNN provient essentiellement du fait que le paramètre de lissage est bien adapté aux données. La souplesse et l'efficacité, de cette méthode, ont motivé plusieurs auteurs à introduire cette approche en statistique fonctionnelle. Parmi les travaux précurseurs dans cette thématique, nous citons [3,9–11].

Dans cette note, nous allons combiner les idées de la régression locale linaire à ceux de la méthode du kNN pour construire un estimateur localement adaptatif avec un biais réduit. De plus, l'estimateur construit est plus pratique que celui de la méthode à noyau, car le nombre optimal des voisins est à sélectionner sur un ensemble fini (voir [3], pour plus de discussions). Comme résultat asymptotique, nous établissons la convergence presque complète de l'estimateur obtenu. Nous traitons, dans un premier temps, le cas où le nombre des voisins est fixé. Ensuite, nous généralisons ce résultat au cas uniform sur ce nombre. Nous mentionnons également le fait que la convergence uniforme sur le nombre des voisins a un grand impact en pratique. En effet, dans la plupart des cas, l'estimateur optimal est associé à un nombre des voisins optimal, qui est fonction des observations. Dans cette situation, la convergence presque complète pour un nombre des voisins fixé est insuffisante pour utiliser cet estimateur en pratique. La version uniforme est nécessaire pour analyser le comportement asymptotique de cet estimateur. Notons que ces questions de la convergence uniforme sur le paramètre de lissage ont été largement étudiées dans l'estimation par la méthode à noyau (voir [7,4]). Cependant, à l'exception de l'article [9], qui traite la régression classique, ce problème n' a pas été abordé dans la méthode de kNN.

Nous présentons notre modèle et son estimateur dans le paragraphe suivant. Dans le troisième paragraphe, nous étudions la convergence presque complète de notre estimateur lorsque le nombre des voisins est fixé. La version uniforme de cette convergence est donnée au quatrième paragraphe. L'impact pratique de nos résultats est discuté dans le dernier paragraphe.

2. Le modèle et son estimateur

Soit (X, Y) un couple de variables aléatoires à valeurs dans $\mathcal{F} \times \mathcal{R}$, où $\mathcal{F}$ est un espace semi-métrique. On note d la semi-métrique sur $\mathcal{F}$.

Généralement, l'estimation de la fonction de régression r par la méthode des polynômes locaux est fondée sur une condition de régularité permettant d'approximer localement r par un polynôme. En analyse statistique des données fonctionnelles, plusieurs approches par polynômes locaux ont été proposées. Dans cette note, nous procédons par l'approximation suivante :

$$r(x_0) = a + b\beta(x_0, x) + o(\beta(x_0, x)), \forall x_0 \text{ dans un voisinage de } x \text{ de } \mathcal{F} \tag{2}$$

où $\beta(\cdot,\cdot)$ est une fonction de $\mathcal{F} \times \mathcal{F}$ à valeurs dans $\mathcal{R}$ vérifie $\beta(\xi, \xi) = 0$, pour tout $\xi \in \mathcal{F}$.

Nous mentionnons que cette approximation locale a été introduite par [2], dont l'objectif de construire une version rapide de l'estimateur local linéaire fonctionnel. Il s'agit d'une condition qui contrôle le taux d'accroissement de l'opérateur de régression r relativement à un bi-fonctionnel β. Cette considération nous permet de remplacer le développement de Taylor dans le cas vectoriel. Il est à noter que cette condition n'est pas très restrictive. En effet, elle est vérifiée, par exemple, dans un espace semi métrique, $\mathcal{F}$, muni de la semi-métrique d'indice fonctionnel simple $(d(x, x') = |<\theta, x - x'>|;, \theta \in \mathcal{F})$, pour un opérateur de $r(x) = f(<\theta_0, x>)$; $(\theta_0 \in \mathcal{F})$ et un bi-fonctionnel $\beta(x, x') =<\theta_0, x - x'>$, où f est une fonction à taux d'accroissement fini.

Sous la décomposition (2), les deux paramètres a et b sont à estimer en cherchant la solution du problème de minimisation :

$$\min_{a,b\in\mathcal{R}} \sum_{i=1}^{n} (Y_i - a - b\beta(X_i, x))^2 K(H_{k,x}^{-1}\delta(x, X_i)) \tag{3}$$

où $\delta(\cdot,\cdot)$ est une fonction de $\mathcal{F} \times \mathcal{F}$ à valeurs dans $\mathcal{R}$ vérifie $|\delta(\xi, \xi')| = d((\xi, \xi'))$, K est un noyau et $H_{k,x} = \min\left\{h \in \mathcal{R}^+ \text{ tel que } \sum_{i=1}^{n} \mathbb{1}_{B(x,h)}(X_i) = k\right\}$. La solution de ce problème est explicitement donnée par

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n} \beta_i(\beta_i - \beta_j)K_iK_j} \begin{pmatrix} \sum_{i=1}^{n}\sum_{j=1}^{n} \beta_i(\beta_i - \beta_j)K_iK_jY_j \\ \sum_{i=1}^{n}\sum_{j=1}^{n} (\beta_j - \beta_i)K_iK_jY_j \end{pmatrix}$$

où $\beta_i = \beta(X_i, x)$ et $K_i = K(H_{k,x}^{-1}\delta(x, X_i))$.

En utilisant (2) et le fait que $\beta(x, x) = 0$, on estime $r(x)$ par

$$\widehat{r}(x) = \hat{a} = \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{n} \beta_i(\beta_i - \beta_j)K_iK_j} \sum_{i=1}^{n}\sum_{j=1}^{n} \beta_i(\beta_i - \beta_j)K_iK_jY_j.$$

Nous rappelons que la littérature sur l'estimation par la méthode de kNN est très abondante dans le cas de la statistique vectorielle (voir [6] pour une liste de références).

3. Convergence presque complète ponctuelle

Par la suite, on fixe un point x de $\mathcal{F}$, on note par N_x un voisinage de x et par C, C' des constantes strictement positives. On pose $\phi_x(r_1, r_2) = \mathbb{P}(r_2 \le \delta(X, x) \le r_1)$ et on considère les hypothèses suivantes :

(H1) La fonction $\phi_x(r) = \phi_x(-r, r)$, $\quad r > 0$ est continue au voisinage de 0 et vérifie $\phi_x(r) > 0$, $\forall\ r > 0$.

(H2) $\forall x_1, x_2 \in N_x \ \ |r(x_1) - r(x_2)| \le C_1 d^{\beta_0}(x_1, x_2); \quad \beta_0 > 0.$
(H3) La fonction $\beta(.,.)$ vérifie $\forall x' \in \mathcal{F},\ C\,|\delta(x,x')| \le |\beta(x,x')| \le C'\,|\delta(x,x')|$.
(H4) K est un noyau continu, dérivable et à support compact $[-1,1]$.
(H5) Le nombre de voisins k est tel que $k/n \to 0$ et $\lim_{n\to\infty} \log n/k = 0$
(H6) Pour toute suite $h_n \to 0$ on a :

$$\text{Il existe un } n_0,\ \text{tel que } \forall n > n_0,\ \frac{1}{\phi_x(h_n)} \int_{-1}^{1} \phi_x(zh_n, h_n) \frac{\mathrm{d}}{\mathrm{d}z}\left(z^2 K(z)\right) \mathrm{d}z > C > 0,$$

$$\text{et} \quad h_n \int_{B(x,h_n)} \beta(u,x)\,\mathrm{d}P(u) = o\left(\int_{B(x,h_n)} \beta^2(u,x)\,\mathrm{d}P(u)\right)$$

où $B(x,r) = \left\{x' \in \mathcal{F}/|\delta(x',x)| \le r\right\}$
(H7) Les moments conditionnels de Y vérifient, $\forall m \ge 2$, que $E[Y^m|X=\cdot] := \sigma_m(\cdot)$ est une fonction continue telle que $\sigma_m(X) \le Cm!$

Remarque 1. Nos hypothèses sont standard, les conditions (H1)–(H2), (H4), (H5) et (H7) ont été utilisées pour l'estimation kNN de la régression classique (voir [3]), tandis que les conditions (H3) et (H6) sont similaires à celles utilisées par [2] pour la méthode locale linéaire.

On obtient le résultat suivant.

Théorème 3.1. *Sous les hypothèses (H1)–(H7), on a*

$$\widehat{r}(x) - r(x) = O\left(\left(\phi_x^{-1}\left(\frac{k}{n}\right)\right)^{\beta_0} + \left(\sqrt{\frac{\log n}{k}}\right)\right) \qquad \text{p.co.} \tag{4}$$

Schéma de la démonstration. Il suffit de combiner les techniques de [3] a ceux utilisées par [2]. La démonstration détaillée peut être obtenue sur simple demande.

4. Convergence presque complète uniforme sur *k*

Dans cette section, nous établissons la version uniforme de la vitesse de convergence précédente. À cet objectif, on suppose que $k \in (k_{1,n}, k_{2,n})$, et on introduit les conditions suivantes :

(U1) la condition (H1) est vérifiée pour un $\phi_x()$ telle que pour tout $s \in (0,1)$,

$$\lim_{r\to 0} \frac{\phi_x(sr)}{\phi_x(r)} = \tau_x(s)\ ;$$

(U2) la condition (H2) est vérifiée ;
(U3) la condition (H3) est vérifiée ;
(U4) K est un noyau continu, dérivable de dérivée continue et à support compact $(-1/2, 1/2)$. De plus, la classe de fonctions

$$\mathcal{K}_k = \{\cdot \mapsto \gamma^{-k} K(\gamma^{-1}\delta(x,\cdot))\beta^k(x,\cdot),\ \gamma > 0\} \text{ est mesurable ponctuellement}$$

vérifiant :

$$\sup_Q \int_0^1 \sqrt{1 + \log \mathcal{N}(\epsilon \|F\|_{Q,2}, \mathcal{K}_k, d_Q)}\,\mathrm{d}\epsilon < \infty$$

où le supremum est pris sur toutes les mesures de probabilité Q de l'espace $\mathcal{F}$ avec $|F\,|_{Q,2} < \infty$ et F est la fonction enveloppe de la classe $\mathcal{K}_k$. Ici, d_Q est la métrique $L_2(Q)$ et $\mathcal{N}(\epsilon, \mathcal{K}_k, d_Q)$ est le nombre minimal de boules ouvertes (par rapport à la métrique $L_2(Q)$) de rayon ϵ qui sont nécessaires pour couvrir la classe de fonction $\mathcal{K}_k$;
(U5) les deux suites $k_{1,n}, k_{2,n}$ sont telles que

$$\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right) \to 0 \quad \text{and} \quad \frac{\log n}{\min(n\phi_x^{-1}\left(\frac{k_{1,n}}{n}\right), k_{1,n})} \to 0\ ;$$

(U6) la deuxième partie de (H6) est vérifiée et la première est remplacée par

$$K(1/2)\tau_x(1/2) - \int_{-1/2}^{1/2} K'(s)\tau_x(s)\,\mathrm{d}s > 0$$

$$\text{et} \qquad (1/4)K(1/2)\tau_x(1/2) - \int_{-1/2}^{1/2} (s^2 K'(s))\tau_x(s)\,\mathrm{d}s > 0;$$

(U7) la condition (H7) est vérifiée.

Remarque 2. De même que le premier résultat, les conditions utilisées sont classiques. En effet, toutes les conditions supplémentaires sont similaires à celles utilisées par [9].

On obtient le résultat suivant :

Théorème 4.1. *Sous les hypothèses (U1)–(U7), on a*

$$\sup_{k_{1,n}\le k\le k_{2,n}} |\widehat{r}(x) - r(x)| = O\left(\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\beta_0}\right) + O_{\text{p.co.}}\left(\sqrt{\frac{\log n}{k_{1,n}}}\right). \tag{5}$$

Schéma de la démonstration. La preuve est basée sur des idées similaire à ceux utilisées par [9]. La démonstration détaillée peut être obtenue sur simple demande.

5. L'impact des résultats asymptotiques

Comme c'est indiqué dans l'introduction, l'une des conséquences les plus importantes de notre étude asymptotique est la précision de la vitesse de convergence de l'estimateur optimal. En effet, étant donné un échantillon $(X_i, Y_i)_{i=1,\ldots,n}$ de (X, Y), l'estimateur optimal est associé à un nombre de voisins optimal k_{opt} tel que

$$k_{\text{opt}} = \arg\min_k \text{CV}(k),$$

où

$$\text{CV}(k) = \sum_{l=1}^n (Y_l - \widehat{r}^{-l}(X_l))^2$$

avec

$$\widehat{r}^{-l}(X_l) = \frac{\sum_{i=1,\neq l}^n \sum_{j=1,\neq l}^n \beta(X_i, X_l)(\beta(X_i, X_l) - \beta(X_j, X_l))K(H_{k,X_l}^{-1}\delta(X_l, X_i))K(H_{k,X_l}^{-1}\delta(X_l, X_j))Y_j}{\sum_{i=1,\neq l}^n \sum_{j=1,\neq l}^n \beta(X_i, X_l)(\beta(X_i, X_l) - \beta(X_j, X_l))K(H_{k,X_l}^{-1}\delta(X_l, X_i))K(H_{k,X_l}^{-1}\delta(X_l, X_j))}. \tag{6}$$

Il est clair que le Théorème 3.1 n'assure pas la convergence de cet estimateur optimal, car le paramètre k_{opt} est une fonction aléatoire. Dans cette situation, on fait appel au Théorème 4.1, qui permet de déduire la vitesse de convergence suivante.

Corollaire 5.1. *Si $k_{\text{opt}} \in (k_{1,n}, k_{2,n})$ et si les conditions du Théorème 4.1 sont vérifiées, alors*

$$|\widetilde{r}_{\text{opt}}(x) - r(x)| = O\left(\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\beta_0}\right) + O_{\text{p.co.}}\left(\sqrt{\frac{\log n}{k_{1,n}}}\right).$$

où $\widetilde{r}_{\text{opt}}$ est l'estimateur optimal associé au nombre des voisins optimal k_{opt}.

Remerciements

Le rapporteur anonyme est vivement remercié pour ses commentaires détaillés et constructifs.

Références

[1] A. Baìllo, A. Grané, Local linear regression for functional predictor and scalar response, J. Multivar. Anal. 100 (2009) 102–111.
[2] J. Barrientos-Marin, F. Ferraty, P. Vieu, Locally Modelled Regression and Functional Data, J. Nonparametr. Stat. 22 (2010) 617–632.
[3] F. Burba, F. Ferraty, P. Vieu, *k*-Nearest neighbor method in functional nonparametric regression, J. Nonparametr. Stat. 21 (2009) 453–469.
[4] P. Deheuvels, S. Ouadah, Uniform-in-bandwidth functional limit laws, J. Theor. Probab. 26 (2013) 697–721.
[5] J. Demongeot, A. Naceri, A. Laksaci, M. Rachdi, Local linear regression modelization when all variables are curves, Stat. Probab. Lett. 121 (2017) 37–44.
[6] L. Devroye, L. Gyorfi, A. Krzyzak, G. Lugosi, On the strong universal consistency of nearest neighbor regression function estimates, Ann. Stat. 22 (1994) 1371–1385.
[7] U. Einmahl, D. Mason, Uniform in bandwidth consistency of kernel-type function estimators, Ann. Stat. 33 (2005) 1380–1403.
[8] J. Fan, I. Gijbels, Local Polynomial Modelling and Its Applications, Chapman & Hall, London, 1996.
[9] L-Z. Karaa, A. Laksaci, M. Rachdi, P. Vieu, Data-driven *k*NN estimation in nonparametric functional data analysis, J. Multivar. Anal. 153 (2017) 176–188.
[10] N. Kudraszow, P. Vieu, Uniform consistency of *k*NN regressors for functional variables, Stat. Probab. Lett. 83 (2013) 1863–1870.
[11] H. Lian, Convergence of functional *k*-nearest neighbor regression estimate with functional responses, Electron. J. Stat. 5 (2011) 31–40.