МАТЕМАТИЧКИ ВЕСНИК
64, 4 (2012), 275–285
December 2012

оригинални научни рад
research paper

ASYMPTOTIC DISTRIBUTION OF ROBUST k-NEAREST NEIGHBOUR ESTIMATOR FOR FUNCTIONAL NONPARAMETRIC MODELS

Mohammed Kadi Attouch and Tawfik Benchikh

Abstract. We propose a family of robust nonparametric estimators for robust regression function based on k-nearest neighbour (k-NN) method. We establish the asymptotic normality of the estimator under the concentration properties on small balls of the probability measure of the functional explanatory variables.

1. Introduction

The main goal of this paper is to study the robust nonparametric estimator of regression function with k-nearest neighbour method when the covariates have functional nature. The study of this model is motivated by the fact that the robust estimator is insensible to the presence of outliers.

In many practical situations, one is faced with functional-type phenomena. It is now possible to take into account their functional nature thanks to technological improvements permitted to collect data discretized on thinner grids. The statistical problems involved in the modelization of functional random variables have received an increasing interest in recent literature, we only refer to the good overviews in parametric models given by Bosq [5], Ramsay & Silverman [23,24] and the monograph of Ferraty & Vieu [16] for the prediction problem in functional nonparametric statistics via the regression function, the conditional mode and the conditional quantile estimation by the kernel method. The first asymptotic results of robust estimator in the functional nonparametric context is given by Cadre [7] who studied the estimator of the median, Azzedine et al. [3] obtained the rate of almost complete convergence of this class of estimator and Crambes et al. [8] present results dealing with L_p error for independent and dependent functional data. The asymptotic normality of robust non parametric regression function has been

2010 AMS Subject Classification: 62G05, 62G08, 62G20, 62G35.

Keywords and phrases: Asymptotic distribution; functional data; k-nearest neighbour; robust estimation; small balls probability.

This work was supported by “Agence Nationale de Recherche Universitaire (A.N.D.R.U.)” in P.N.R., No.46/15. and No.46/1, Algeria.

established by Attouch et al. [1,2] in the both cases independent and dependent functional data.

In cases where data is sparse, the k-NN kernel estimate has a significant advantage over the classical kernel estimate. The k-NN kernel estimate is also automatically able to take into account the local structure of the data. This advantage, however, may turn into a disadvantage. If there is an outlier in the data set, the local prediction may be bad. To avoid this, a robust non-parametric regression combined to k-NN method can give better results.

The literature of the k-NN method for estimation of the regression function date bakes to Royall [21] and Stone [22] and has received, since, continuous developments (Mack [19] derived the rates of convergence for the bias and variance as well as asymptotic normality in multivariate case, Collomb [9] studied different types of convergence (probability, a.s, a.co) of the estimator. Devroye [12] obtained the strong consistency and the uniform convergence). For the functional data studies, the k-NN kernel estimate was first introduced in the monograph of Ferraty & Vieu [16], Burba et al. [6] obtained the rate of almost complete convergence of the regression function using the k-NN method for independent data.

The principal aim of this paper is to apply the k-NN method to establish the asymptotic normality of the estimator for independent and identically distributed observations. Nonparametric k-NN method jointly with the robust method can have the property of automatically removing irrelevant variables in a regression model, this method permits the construction of adapted neighborhood with a consideration of the local structure of the data. To highlight the k-NN studies comparatively to the classical regression we give a real data application.

The paper is organized as follows: the following section is dedicated to presenting our model of study. Then, we give hypotheses and state our main result in Section 3. In Section 4 we illustrate the effectiveness of the robust k-NN method in presence of outlier data. All proofs are given in the appendix.

2. Models and estimators

Let $(X_1, Y_1), \dots (X_n, Y_n)$ be n independent pairs, identically distributed as (X, Y) which is a random pair valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space, $d(\cdot, \cdot)$ denoting the semi-metric. For any x in $\mathcal{F}$, we consider ψ a real-valued Borel function satisfying some regularity conditions to be stated below. The nonparametric model studied in this paper, denoted by θ_x, is implicitly defined as a zero with respect to (w.r.t.) t of the following equation

$$\Psi(x, t) = \mathbb{E}\left[\psi(Y, t) \mid X = x\right] = 0 \tag{1}$$

We suppose that, for all $x \in \mathcal{F}$, θ_x exists and is the unique zero w.r.t. t of (1) (see, for instance Koul and Stute [18]) for the existence and uniqueness of θ_x).

The k nearest neighbour (k-NN) estimate of $\Psi(x, t)$ is defined by

$$\widehat{\Psi}_{kNN}(x, t) := \frac{\sum_{i=1}^{n} K(H_{n,k}^{-1} d(x, X_i)) \psi(Y_i, t)}{\sum_{i=1}^{n} K(H_{n,k}^{-1} d(x, X_i))}, \quad \forall t \in \mathbb{R}$$

where K is a kernel function and $H_{n,k}(\cdot)$ is defined as follows:

$$H_{n,k}(x) = \min\Big\{h \in \mathbb{R}^+ \;\Big|\; \sum_{i=1}^{n} \mathbf{1}_{B(x,h)}(X_i) = k\Big\}.$$

In the case of the non-random bandwidth $h := h_n$ (sequence of positive real numbers which goes to zero as n goes to infinity), the robust kernel estimate of $\Psi(x,t)$ (introduced in Azzedine et al. [3]) is defined by

$$\widehat{\Psi}(x,t) := \frac{\sum_{i=1}^{n} K(h^{-1}d(x,X_i))\psi(Y_i,t)}{\sum_{i=1}^{n} K(h^{-1}d(x,X_i))}, \quad \forall t \in \mathbb{R}.$$

A natural estimator of θ_x denoted by $\widehat{\theta_x}$, is a zero w.r.t. t of the

$$\widehat{\Psi}_{kNN}(x,t) = 0. \tag{2}$$

The robust method used here belongs to the class of M-estimates introduced by Huber [17].

3. Hypotheses and results

From now on, x stands for a fixed point in $\mathcal{F}$, N_x denotes a fixed neighborhood of x and we set $\lambda_\gamma(u,t) = \mathbb{E}[(\psi(Y,t))^\gamma \mid X = u]$ and $\Gamma_\gamma(u,t) = \mathbb{E}[(\psi'(Y,t))^\gamma \mid X = u]$, for $\gamma \in \{1,2\}$. We need the following hypotheses gathered together for easy references.

(H1) There exists a nonnegative differentiable ϕ-strictly increasing and a nonnegative function g such that

$$\mathbb{P}\left(X \in B(x,r)\right) = \phi(r) \cdot g(x) \text{ where } B(x,r) = \{x' \in \mathcal{F} \mid d(x,x') < r\}.$$

(H2) The function ψ is continuous differentiable, strictly monotone and bounded w.r.t. the second component and its derivative $\partial\psi(y,t)/\partial t$ is bounded and continuous at θ_x uniformly in y.

(H3) The function $\lambda_\gamma(\cdot,\cdot)$ satisfies the Lipschitz's condition w.r.t. the first one, that is: there exists a strictly positive constant b_γ such that:

$$\exists C_1 > 0,\ \forall (u_1,u_2) \in N_x \times N_x,\ \forall t \in \mathbb{R},\ |\lambda_\gamma(u_1,t) - \lambda_\gamma(u_2,t)| \le C_1 d^{b_\gamma}(u_1,u_2).$$

(H4) The function $\Gamma_\gamma(\cdot,\cdot)$ satisfies the Lipschitz's condition w.r.t. the first one, that is: there exists a strictly positive constant d_γ such that

$$\exists C_2 > 0,\ \forall (u_1,u_2) \in N_x \times N_x,\ \forall t \in \mathbb{R},\ |\Gamma_\gamma(u_1,t) - \Gamma_\gamma(u_2,t)| \le C_2 d^{d_\gamma}(u_1,u_2).$$

(H5) For each sequence $u_n \downarrow 0$ as $n \longrightarrow \infty$ of positive real numbers, there exists a function $\beta(\cdot)$ such that:

$$\forall t \in [0,1],\ \lim_{u_n \to 0} \frac{\phi(tu_n)}{\phi(u_n)} = \beta(t) \text{ and } \frac{\log(n)}{n u_n^2 \phi^2(u_n)} \longrightarrow 0 \text{ as } n \longrightarrow \infty$$

(H6) The sequence of positive real numbers $k_n = k$ satisfies:

$$\frac{k}{n} \longrightarrow 0 \text{ and } \log(n)/k \to 0 \text{ as } n \to \infty.$$

(H7) The kernel K is a positive function supported on $[0,1]$. Its derivative K' exists and is such that there exist two constants C_3 and C_4 with $-\infty < C_3 < K'(t) < C_4 < 0$ for $0 \le t \le 1$.

(H8) The derivative of the real function $\varphi_x(s) = \mathbb{E}[\psi(Y,\theta_x) \mid d(X,x) = s$, at 0 exists.

Comments on the hypotheses

1. The assumption (H1) is classical for the explanatory variable X. This assumption replaces the condition of a strict positivity of the density of explanatory variable usually assumed in finite-dimensional case. The decomposition of the concentration measure as product of two independent functions has been adopted in Masry [20] and used after by many authors (see Attouch et al. [1] for comment and some explicit examples).
2. Condition (H2) controls the robustness properties of our model. This assumption keeps the same conditions on the function ψ given by Boente and Rodriguez [4] in the multivariate case.
3. Hypotheses (H3)–(H4) are regularity conditions which characterize the functional space of our model and are needed to evaluate the bias term in our asymptotic properties.
4. The function $\beta(\cdot)$ defined in (H5) plays a crucial role in our asymptotic approach. It permits to give the variance term explicitly. We quote the following examples (which can be found in Ferraty et al. [14]:

   1) $\beta(u) = u^\gamma$, when $\phi(h) = h^\tau$ for some $\tau > 0$,

   2) $\beta(u) = \delta_1(u)$ where $\delta_1(u)$ is Dirac's function, when $\phi(h) = h^\tau \exp\left\{-\frac{C}{h^p}\right\}$ for some $\tau > 0$ and $p > 0$,

   3) $\beta(u) = \mathbf{1}_{]0,1]}(u)$, when $\phi(h) = \frac{1}{|\ln h|} + o\left(\frac{1}{|\ln h|}\right)$.
5. The assumptions (H6)–(H8) are technical conditions imposed for the brevity of proofs.

Now we are in position to give our main result.

Theorem 1. *Assume that (H1)–(H8) hold, then $\widehat{\theta_x}$ exists and is unique with probability tending to* 1*, and for any $x \in \mathcal{A}$, we have*

$$\left(\frac{k}{\sigma^2(x,\theta_x)}\right)^{1/2}\left(\widehat{\theta_x} - \theta_x - B_n(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \text{ as } n \to \infty \tag{3}$$

where

$$B_n(x) = \frac{n\phi^{-1}(k/n)g(x)}{\sqrt{k\alpha_2\lambda_2(x,\theta_x)}}\int_0^1 K(t)\varphi_x(t\phi^{-1}(k/n))\phi'(t\phi^{-1}(k/n))dt + o(\phi^{-1}(k/n)) \tag{4}$$

and

$$\sigma^2(x,\theta_x) = \frac{\alpha_2\lambda_2(x,\theta_x)}{\alpha_1^2 g(x)(\Gamma_1(x,\theta_x))^2} \ \ (\textit{with}\ \alpha_j = -\int_0^1 (K^j)'(s)\beta(s)ds,\ \textit{for},\ j=1,\,2), \tag{5}$$

$$\mathcal{A} = \{x \in \mathcal{F},\ g(x)\lambda_2(x,\theta_x)\Gamma_1(x,\theta_x) \neq 0\};$$

$\xrightarrow{\mathcal{D}}$ *means the convergence in distribution.*

It is easy to see that, if one imposes some regularity assumptions on the real function $\varphi_x(\cdot)$, we can give explicitly the asymptotic behavior of the term $B_n(x)$, by using Taylor's expansion of the function φ_x (see Delsol [10]). However, to remove the bias term $B_n(x)$ from Theorem 1, we need an additional condition on the k-NN parameter k.

Corollary 1. *Under the hypotheses of Theorem 1 and if the k-NN parameter k satisfies $k(\phi^{-1}(k/n))^{2b_1} \to 0$ as $n \to \infty$, we have*

$$\left(\frac{k}{\sigma^2(x,\theta_x)}\right)^{1/2}\left(\widehat{\theta_x} - \theta_x\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \ \textit{as}\ n \to \infty. \tag{6}$$

The detailed proof of Theorem 1 is postponed to appendix.

4. Applications

4.1 Conditional confidence interval

Our main application of the above Theorem is to build confidence interval for the true value of θ given curve $X = x$. A plug-in estimate for the asymptotic standard deviation $\sigma(x,\theta_x)$ can be obtained using the estimators $\widehat{\lambda_2}(x,\widehat{\theta_x})$ and $\widehat{\Gamma_1}(x,\widehat{\theta_x})$ of $\lambda_2(x,\theta_x)$, $\Gamma_1(x,\theta_x)$ respectively. We get $\widehat{\sigma}(x,\widehat{\theta_x}) := \left(\dfrac{\widehat{\alpha_2}\widehat{\lambda_2}(x,\widehat{\theta_x})}{(\widehat{\alpha_1})^2 g(x)\widehat{\Gamma_1}(x,\widehat{\theta_x})}\right)^{1/2}$.
Then $\hat{\sigma}(x,\widehat{\theta_x})$ can be used to get the following approximate $(1-\beta)$ confidence interval for θ_x

$$\widehat{\theta_x} \pm t_{1-\zeta/2} \times \left(\frac{\widehat{\sigma}_n^2(x,\widehat{\theta_x})}{k}\right)^{1/2}$$

where $t_{1-\zeta/2}$ denotes the $1-\zeta/2$ quantile of the standard normal distribution.

Here we point out that the estimators $\widehat{\lambda_2}(x,\widehat{\theta_x})$ and $\widehat{\Gamma_1}(x,\widehat{\theta_x})$ will be calculated, for $x \in \mathcal{A}$, in the same way as in (2). We estimate empirically α_1 and α_2 by

$$\widehat{\alpha_1} = \frac{1}{kg(x)}\sum_{i=1}^n K_i \quad \text{and} \quad \widehat{\alpha_2} = \frac{1}{kg(x)}\sum_{i=1}^n K_i^2,$$

where $K_i = K\left(\dfrac{d(x, X_i)}{\phi^{-1}(k/n)}\right)$. This last estimation is justified by the fact that, under (H1), (H5) and (H6), we have, (see Ferraty & Vieu [16, p. 44])

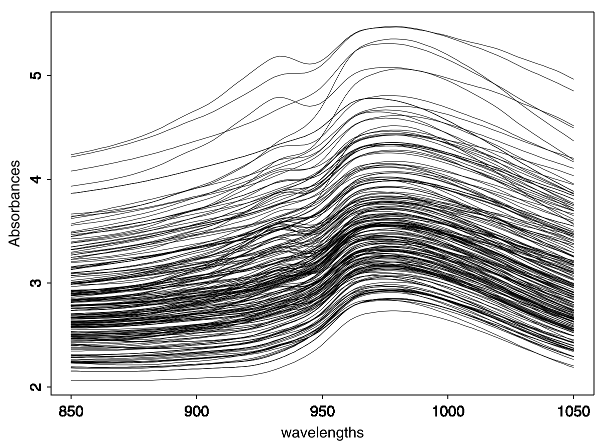
$$\frac{1}{kg(x)}\mathbb{E}[K_1^j] \to \alpha_j, \quad j = 1, 2.$$

Finally, the approximate $(1-\zeta/2)$ confidence band, for any $x \in \mathcal{A}$, is

$$[a_-(x), a_+(x)] \text{ where } a_\pm(x) = \widehat{\theta_x} \pm t_{1-\zeta/2} \times \left(\frac{\sum_{i=1}^n K_i^2 \widehat{\lambda_2}(x, \widehat{\theta_x})}{\left(\sum_{i=1}^n K_i\right)^2 \widehat{\Gamma_1}(x, \widehat{\theta_x})}\right)^{1/2}.$$

4.2. A real data application

Now we apply the described method to some chemiometrical real data. This data come from a quality control problem in the food industry and it concerns a sample of finely chopped meat. The data are available on the web site `http://lib.stat.cmu.edu/datasets/tecator`.



The 215 spectrometric curves, $\{X_i(t),\ t \in [850, 1050],\ i = 1, \cdots, 215.\}$

This figure plots absorbance versus wavelength (850–1050) for 215 selected pieces of meat. Note that, the main goal of spectrometric analysis is to allow for the discovery of the proportion of some specific chemical content (see Ferraty & Vieu [16] for further details related to spectrometric data). At this stage one would like to use the spectrometric curve X to predict Y the proportion of protein content in the piece of meat.

Thus, in order to show the superiority of our prediction method we compare the three different methods, k-NN regression given in Burba et al. [6], Robust regression in Attouch et al. [1] and k-NN robust regression presented in this paper.

In order to introduce the outliers in this sample, we multiply by 100 the response variable of a number of observations. In practice, we consider 215 observations split into two samples: learning sample (170 observations) and test sample (45 observations).

For both methods, the kernel K is the quadratic function, and we use a standard L^2 semi-metric; this choice is motivated by the fact that the shape of these spectrometric curves is very smooth (see Ferraty & Vieu [16] for more motivations of this choice). We proceed by the following algorithm:

The optimal bandwidths h and k optimal nearest neighbor obtaining from the training sample by the cross-validation method (see Burba et al. [6]). We put

$$\widehat{Y_i^j} = \widehat{\theta}(X_{i^*}) \qquad \text{and} \qquad \widetilde{Y_i^j} = \widehat{r}(X_{i^*}) \quad i = 171, \dots, 215.$$

where X_{i^*} is the nearest curve to X_i in the training sample and $\widehat{r}$ is the kernel estimate of the classical regression function defined by

$$\widehat{r}(x) = \frac{\sum_{i=171}^{215} K(h_n^{-1} d(x, X_i)) Y_i}{\sum_{i=171}^{215} K(h_n^{-1} d(x, X_i))}.$$

Whereas the optimal number of neighbors k_{opt} is defined by $k_{opt} = \text{argmin}_k \, CV(k)$, where

$$CV(k) = \sum_{i=171}^{215} (Y_i - \widehat{r}^i_{kNN}(X_i))^2$$

with

$$\widehat{r}^i_{kNN}(x) = \frac{\sum_{j=171, j\neq i}^{215} \psi(Y_j) K(d(X_j, x)/h_k(x))}{\sum_{j=171, j\neq i}^{215} K(d(X_j, x)/h_k(x))}.$$

The error used to evaluate this comparison is the mean of absolute error (MAE) expressed by

$$\frac{1}{45} \sum_{i=171}^{215} |Y_i - \widehat{T}(X_i)|$$

where $\widehat{T}$ designate the estimator used: robust regression, and k-NN in classical or robust regression.

In our simulations, we worked with several functions ($L_1 - L_2$, Androws, Tuckey, Cauchy ...), but we found that the best results are obtained when the $L_1 - L_2$ function $\left(\psi(t) = \dfrac{t}{\sqrt{1 + t^2/2}}\right)$ is used.

The results are given in Table 1. We observe that in the presence of outliers, the k-NN robust regression gives better results than the k-NN regression and robust regression, in sense that, even if the MAE value of the both methods increases substantially relatively to the number of the perturbed points, it remains very low for the k-NN robust one.

Number of perturbation	0 value	5 values	15 values
MAE Robust reg.	0.0450763	0.0789672	0.1598794
MAE K-NN reg.	0.0385238	0.039065	0.03928697
MAE K-NN Robust reg.	0.03742967	0.03756078	0.03776294

Table 1. Comparison between both methods in the presence of outliers

5. Conclusion and perspectives

As already mentioned in the introduction, the k-NN estimate is prone to outliers. This disadvantage is treated clearly in this paper by robust k-NN kernel regression estimation.

This work can be generalized to the dependent data (see Attouch et al. [2]. In nonparametric statistics, uniform convergence is considered as a preliminary step to obtain sharper results, in this context, it would be very interesting to extend the results of Ferraty et al. [15] to robust k-NN estimate.

Obtaining explicit expressions of the dominant terms of centered moments can be envisaged when we obtain the asymptotic normal result (see, Delsol [10]). This approach can be investigate in future work.

6. Appendix

For $i = 1, \dots, n$, we consider quantities $K_i(x,u) = K\left(u^{-1}d(x,X_i)\right)$, $\psi_i(t) = \psi(Y_i,t)$, and let

$$\begin{aligned} \widehat{\Psi}_N(x,t,u) &= \frac{1}{n\mathbb{E}\left[K_1(x,u)\right]}\sum_{i=1}^n K_i(x,u)\psi_i(t), \\ \widehat{\Psi}_D(x,u) &= \frac{1}{n\mathbb{E}[K_1(x,u)]}\sum_{i=1}^n K_i(x,u), \end{aligned} \tag{7}$$

so that $\widehat{\Psi}_{kNN}(x,t) = \frac{\widehat{\Psi}_N(x,t,H_{n,k})}{\widehat{\Psi}_D(x,H_{n,k})}$.

Denote by $c_n(t,u) = \widehat{\Psi}_N(x,t,u)$, $c = \Psi(x,t)$ and let $\zeta \in]0,1[$. We choose D_n^- and D_n^+ as

$$\phi(D_n^-) = \sqrt{\zeta}\frac{k}{n},\ \phi(D_n^+) = \frac{1}{\sqrt{\zeta}}\frac{k}{n}.$$

Using the Taylor's expansion of order one around θ_x, we get

$$c_n(\widehat{\theta_x},u) = c_n(\theta_x,u) + (\widehat{\theta_x}-\theta_x)c_n'(\xi_n,u).$$

with $\xi_n \in (\widehat{\theta_x},\theta_x)$. Because of the definition of $\widehat{\theta_x}$, we have

$$\widehat{\theta_x}-\theta_x = \frac{-c_n(\theta_x,H_{n,k})}{c_n'(\xi_n,H_{n,k})}. \tag{8}$$

Finally, we have the following decomposition:

$$\left(\frac{k}{\sigma^2(x,\theta_x)}\right)^{1/2}\widehat{\theta_x}-\theta_x = \left(\frac{k}{\sigma^2(x,\theta_x)}\right)^{1/2}\frac{-c_n(\theta_x,H_{n,k})}{c_n'(\xi_n,H_{n,k})}$$

$$= \left(\frac{k}{\sigma^2(x,\theta_x)}\right)^{1/2} \frac{-c_n(\theta_x, H_{n,k}) + c_n(\theta_x, D_n^+)}{c_n'(\xi_n, H_{n,k})} - \left(\frac{k}{\sigma^2(x,\theta_x)}\right)^{1/2} \frac{c_n(\theta_x, D_n^+)}{c_n'(\xi_n, H_{n,k})}$$

$$= -\left(\frac{k}{\sigma^2(x,\theta_x)}\right)^{1/2} \frac{c_n(\theta_x, D_n^+) - \mathbb{E}\left[c_n(\theta_x, D_n^+)\right]}{c_n'(\xi_n, H_{n,k})}$$
$$+ \left(\frac{k}{\sigma^2(x,\theta_x)}\right)^{1/2} \frac{-c_n(\theta_x, H_{n,k}) + c_n(\theta_x, D_n^+)}{c_n'(\xi_n, H_{n,k})} - \left(\frac{k}{\sigma^2(x,\theta_x)}\right)^{1/2} \frac{\mathbb{E}\left[c_n(\theta_x, D_n^+)\right]}{c_n'(\xi_n, H_{n,k})} \tag{9}$$

Then, to state asymptotic normality, we show that the numerator of the first term of the right hand side of (9) suitably normalized is asymptotically normally distributed, the numerator of the second term is equal to $B_n(x)$ and the denominator converges in probability to $\Gamma_1(x, \theta_x)$.

On the one hand, the asymptotic normality of

$$\left(\frac{k}{\sigma^2(x,\theta_x)}\right)^{1/2} \left(c_n(\theta_x, D_n^+) - \mathbb{E}\left[c_n(\theta_x, D_n^+)\right]\right)$$

was proved in Lemma 2.1 in Attouch et al. [1] by choosing the bandwidth parameter as $h := h_n = D_n^+$.

Note that, Attouch et al. [1] in Lemma 2.2 prove that

$$\mathbb{E}\left[c_n(\theta_x, h_n)\right] = \frac{h}{\phi(h)\alpha_1} \int_0^1 K(t)\varphi(th)\phi'(th)\,dt.$$

Then, we deduced in the case $h := h_n = D_n^+$, that

$$\left(\frac{k}{\sigma^2(x,\theta_x)}\right)^{1/2} \mathbb{E}\left[c_n(\theta_x, D_n^+)\right]$$
$$= \frac{n\phi^{-1}(k/n)g(x)}{\sqrt{k\alpha_2\lambda_2(x,\theta_x)}} \int_0^1 K(t)\varphi(t\phi^{-1}(k/n))\phi'(t\phi^{-1}(k/n))\,dt \tag{10}$$

On the other hand, by hypothesis (H7), and the fact that $\mathbf{1}_{\{D_n^- \le H_{n,k} \le D_n^+\}} \xrightarrow{a.co.} 1$ when $\frac{k}{n} \longrightarrow 0$ (see Burba et al. [6]), we have

$$c_n(\theta_x, D_n^+) \le c_n(\theta_x, H_{n,k}) \le c_n(\theta_x, D_n^-).$$

Using the fact that

$$\left|c_n(\theta_x, H_{k,n}) - c_n(\theta_x, D_n^+)\right| \le \left|c_n(\theta_x, D_n^-) - c_n(\theta_x, D_n^+)\right|$$
$$\le \left|c_n(\theta_x, D_n^+) - \mathbb{E}\left[c_n(\theta_x, D_n^+)\right]\right| + \left|\mathbb{E}\left[\theta_x, c_n(D_n^+)\right] - \mathbb{E}\left[c_n(\theta_x, D_n^-)\right]\right|$$
$$+ \left|\mathbb{E}\left[c_n(\theta_x, D_n^-) - c_n(\theta_x, D_n^-)\right]\right|,$$

hypothesis (H5) and Lemma 3.4 of Azzedine et al. [3] give that

$$\left(\frac{k}{\sigma^2(x,\theta_x)}\right)^{1/2} \left|c_n(\theta_x, D_n^+) - \mathbb{E}\left[c_n(\theta_x, D^+)\right]\right| = O_{a.co.}\left(\sqrt{\frac{\log(n)}{n(D_n^+)^2(\phi(D_n^+))^2}}\right)$$

and

$$\left(\frac{k}{\sigma^2(x,\theta_x)}\right)^{1/2}\left|c_n(\theta_x,D_n^-)-\mathbb{E}\left[c_n(\theta_x,D^-)\right]\right|=O_{a.co.}\left(\sqrt{\frac{\log(n)}{n(D_n^-)^2(\phi(D_n^-))^2}}\right).$$

The equiprobability of the couples (X_i,Y_i), hypotheses (H1) and (H6) allow to write

$$\begin{aligned}
&\left|\mathbb{E}\left[c_n(\theta_x,D^+)\right]-\mathbb{E}\left[c_n(\theta_x,D^-)\right]\right|\\
&\quad\le\left|\mathbb{E}\left[c_n(\theta_x,D^+)\right]-\lambda_1(x,t)\right|+\left|\mathbb{E}\left[c_n(\theta_x,D^-)\right]-\lambda_1(x,t)\right|\\
&\quad\le\left|\mathbb{E}\left[\frac{1}{\mathbb{E}\left[K_1\right]}K_1\mathbf{1}_{B(x,D_n^+)}(X_1)\Big(\lambda_1(X_1,t)-\lambda_1(x,t)\Big)\right]\right|\\
&\qquad+\left|\mathbb{E}\left[\frac{1}{\mathbb{E}\left[K_1\right]}K_1\mathbf{1}_{B(x,D_n^-)}(X_1)\Big(\lambda_1(X_1,t)-\lambda_1(x,t)\Big)\right]\right|\\
&\quad\le C\ \left((D_n^+)^{b_1}+(D_n^-)^{b_1}\right)
\end{aligned}$$

As $b_1>1$, the definition of D_n^+ and D_n^-, hypothesis (H5) and (10) permit to obtain (4).

On the other hand, to establish the convergence in probability of denominator in (8), note that

$$\left|c_n'(\xi_n,D_n^+)-\Gamma_1(x,\theta_x)\right|\le\left|c_n'(\xi_n,D_n^+)-c_n'(\theta_x,D_n^+)\right|+\left|c_n'(\theta_x,D_n^+)-\Gamma_1(\theta_x,x)\right|. \tag{11}$$

Concerning the first term, observe that

$$\left|c_n'(\xi_n,D_n^+)-c_n'(x,\theta_x,D_n^+)\right|\le\sup_{y\in\mathbb{R}}\left|\frac{\partial\psi(y,\xi_n-\theta_x)}{\partial t}\right|,$$

using the fact that $\frac{\partial\psi(y,t)}{\partial t}$ is continuous at θ_x uniformly in y, and the convergence in probability of $\widehat{\theta_x}$ to θ_x (see Corollary 2.4 in Attouch et al. [1], we deduce that the first term of (11) converges in probability to 0.

However, the limit of second term is a consequence of the following inequality and the Lemma 2.4 of Attouch et al. [1]

$$\left|c_n'(\theta_x,D_n^+)-\Gamma_1(\theta_x,x)\right|\le\left|c_n'(\theta_x,D_n^+)-\Gamma_1(\theta_x,x)\right|.$$

Acknowledgement. The authors would like to thank the editor and the anonymous referees, whose remarks permit us to improve substantially the quality of the paper and clarify some points.

REFERENCES

[1] M. Attouch, A. Laksaci, E. Ould-Saïd, *Asymptotic distribution of robust estimator for functional nonparametric models*, Comm. Statist. Theory and Methods **38** (2009), 1317–1335.

[2] M. Attouch, A. Laksaci, E. Ould-Saïd, *Asymptotic normality of a robust estimator of the regression function for functional time series*, J. Korean Statist. Soc. **39** (2010), 489–500.

[3] N. Azzedine, A. Laksaci, E. Ould Saïd, *On the robust nonparametric regression estimation for functional regressor*, Statist. Probab. Lett. **78** (2008), 3216–3221.

[**4**] G. Boente, D. Rodriguez, *Robust estimators of high order derivatives of regression function*, Statist. Probab. Lett. **7** (2006), 1335–1344.

[**5**] D. Bosq, *Linear processes in function spaces. Theory and Application*, Lectures Notes in Statistics. Springer Verlag, New-York, Vol. **149**, 2000.

[**6**] F. Burba, F. Ferraty, P. Vieu, *k-nearest neighbour method in functional nonparametric regression*, J. Nonparametr. Stat. **21** (2009), 453–469.

[**7**] B. Cadre, *Convergent estimators for the L_1-median of a Banach valued random variable*, Statistics **35** (2001), 509–521.

[**8**] C. Crambes, L. Delsol, A. Laksaci, *Robust nonparametric estimation for functional data*, J. Nonparametr. Stat. **20** (2008), 573–598.

[**9**] G. Collomb, *Estimation de la régression par la méthode des k points les plus proches avec noyau: quelques propriétés de convergence ponctuelle*, in: *Nonparametric Asymptotic Statistics*, Lecture Notes in Mathematics, Springer-Verlag, **821** (1980), 159–175.

[**10**] L. Delsol, *Advances on asymptotic normality in nonparametric functional time series analysis*, Statistics **43** (2009), 13–33.

[**11**] L. Delsol, *Régression non paramtrique-Fonctionnelle: les expressions asymptotiques des moments*, Ann. I.S.U.P. **LI** (2007), 43–67.

[**12**] L.P. Devroye, *The uniform convergence of nearest neighbour regression function estimators and their application in optimization*, IEEE Trans. Inform. Theory, **24** (1978), 142–151.

[**13**] F. Ferraty, A. Laksaci, A. Tadj, P. Vieu, *Rate of uniform consistency for nonparametric estimates with functional variables*, J. Statist. Plann. Inference **140** (2010), 335–352.

[**14**] F. Ferraty, A. Mas, P. Vieu, *Nonparametric regression on functional data: Inference and practical aspects*, Austral. New Zealand J. Statist. **49** (2007), 267–286.

[**15**] F. Ferraty, I. Van Keilegom, P. Vieu, *On the Validity of the Bootstrap in Non-Parametric Functional Regression*, Scand. J. Statist. **37** (2010), 286–306.

[**16**] F. Ferraty, P. Vieu, *Nonparametric functional data analysis. Theory and Practice*, Springer-Verlag, 2006.

[**17**] R.J. Huber, (1964). *Robust estimation of a location parameter*, Ann. Math. Statist. **35** (1964), 73–101.

[**18**] H.L. Koul, W. Stute, *Regression model fitting with long memory errors*, J. Statist. Plann. Inference **71** (1998), 35–56.

[**19**] Y.P. Mack, *Local properties of kNN regression estimates*, SIAM J. Algebra Discrete Methods **2** (1981), 311–323.

[**20**] E. Masry, *Nonparametric regression estimation for dependent functional data: asymptotic normality*, Stoch. Proc. Appl. **115** (2005), 155–177.

[**21**] R.M. Royall, *A class of nonparametric estimates of a smooth regression function*, Ph.D. Diss., Stanford University, 1966.

[**22**] C.J. Stone, *Consistent nonparametric regression*, Ann. Statist **5** (1977), 595–645.

[**23**] J.Q. Ramsay, B.W. Silverman, *Functional Data Analysis*, Springer Series in Statistics, 1997.

[**24**] J.Q. Ramsay, B.W. Silverman, *Applied Functional Data Analysis. Methods and Case Studies*, Second edition, Springer-Verlag, 2005.

(received 06.02.2011; in revised form 11.01.2012; available online 15.03.2012)

Lab. de Mathématiques, Univ. Djillali Liabès, Sidi Bel Abbès, BP 89, Sidi Bel Abbès 22000, Algeria

E-mail: `attou_kadi@yahoo.fr, benchikh@yahoo.fr`