



HAL
open science

On the robustification of the kernel estimator of the functional modal regression

Azzi Amel, Laksaci Ali, Ould Saïd Elias

► To cite this version:

Azzi Amel, Laksaci Ali, Ould Saïd Elias. On the robustification of the kernel estimator of the functional modal regression. *Statistics and Probability Letters*, 2022, 181, pp.109256. 10.1016/j.spl.2021.109256 . hal-04412536

HAL Id: hal-04412536

<https://hal.science/hal-04412536>

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

On the robustification of the kernel estimator of the functional modal regression

Azzi Amel^a, Laksaci Ali^b and Ould Saïd Elias^{c,*}

^aCenter Universitaire Salhi Ahmed de Naama. Algérie

^bDepartment of Mathematics, College of Science. King Khalid University, P.O. Box 9004, Abha 62529, Saudi Arabia

^cUniv. du Littoral Côte d'Opale. Laboratoire de Math. Pures et Appliquées. IUT de Calais. 19, rue Louis David. Calais, 62228, France

ARTICLE INFO

Keywords:

Conditional mode estimate
Functional statistics
Quantile regression
Robust estimator.

ABSTRACT

A new nonparametric estimator of the conditional mode when the regressors are functionals is proposed. The main aim of this paper is to establish the almost complete convergence (with rate) of the constructed estimator is estimate under general assumptions in nonparametric functional statistics. A simulation study is carried out to examine illustrate, the finite samples behavior of the constructed estimator. Finally, a discussion highlighting the impact of this new estimator in nonparametric functional data analysis is also given.

1. Introduction

The modal regression is one of the most common predictor models in nonparametric statistics. It constitutes a valuable alternative predictor to the classical regression expressed by the conditional expectation. It is well known that the conditional mode is more informative than the classical regression in some particular cases. We cite for instance, when the conditional density is un-symmetric, has multi-modal or if the white noise has heavy-tail distribution (see Collomb *et al.* 1987).

In this work, we focus in the case when the regressors belong to a functional space. Noting that absence of graphical tool to define the structure of the relationship between the functional explanatory variables and the scalar response provides motivation for seeking unstructure models that allow to exploit this link without preliminary assumption. Thus, the nonparametric models are more flexible than the parametric approach when the observations are functional. The functional nonparametric statistics has received extensive attention during the last two decades. It was popularized by the monograph of Ferraty and Vieu (2006). Since the publication of this monograph, numerous studies have been performed to evaluate the co-variation between functional input variable and output scalar one. The reader interested by the recent developments in this emerging field of nonparametric functional data analysis may turn to the survey papers of Geenens (2015), Ling and Vieu (2018) and Chowdhury and Chaudhuri (2019). More generally, we refer to some journal special issues, Goia *et al.* 2016, Aneiros *et al.* (2019a, 2019b) for recent advances in FDA. In this context, the nonparametric estimation of the conditional mode have been investigated by many authors. For instance, we cite Ezzahrioui and Ould Saïd (2008) for the asymptotic normality of the kernel estimator of the conditional mode, Dabo-Niang *et al.* (2015) for the L_p consistency and the asymptotic normality of the same estimator in the spatial case. Ling *et al.* (2016) studied the nonparametric estimation of the conditional mode in the ergodic functional time series case. They established the almost sure consistency of the constructed estimator of the modal regression with missing at random (MAR) responses. More recent advances in the functional conditional mode estimation can be found in Bouanani *et al.* (2020) and the references therein.

While all these authors estimate the conditional mode by maximizing the kernel estimator of the conditional density function, in this paper we estimate this model using the robust estimator of the quantile regression. This estimation strategy allows to increases the robustness of the conditional mode estimation and broaden the scope of its application to the supervised classification problems. It should be noted that the robustness property is fundamental for the prediction issues. It permits to give a stable predictor that keeps its nice properties even if the conditions of the ideal situation are not verified. It is also worth noting that, our main asymptotic result is the establishment of the almost complete of the constructed under standards conditions. Because of the the new estimator is constructed from one

*Principal corresponding author

 amel.azzi@outlook.fr (A. Amel); alilak@yahoo.fr (L.A.); elias.ould-said@univ-littoral.fr (O.S.E.)
ORCID(s):

kernel, the convergence rate is obtained under less restrictive assumptions than the double kernels estimator. The easy implementation of the new estimator as well as its robustness are emphasized in this paper.

The paper is organized as follows. We present our model in Section 2. In Section 3, we give some notations, hypotheses and the main results. Section 4 is devoted to some discussions on the nice features of our approach. In the last section we conduct a simulation study to examine the feasibility of the constructed estimator.

2. The model

Let (X_i, Y_i) for $i = 1, \dots, n$ be n pairs of independent random vectors that we assume drawn from the pair (X, Y) which is valued in $\mathcal{F} \times \mathbb{R}$, where \mathcal{F} is a semi-metric space equipped with a semi-metric d . In what follows, we fix a point x in \mathcal{F} , consider a given neighborhood \mathcal{N}_x of x and assume that the regular version of the conditional distribution function of Y a given $X = x$, denoted by $F(\cdot|x)$, exists in the neighborhood \mathcal{N}_x of x . Moreover, we suppose that $F(\cdot|x)$ is strictly increasing and has a continuous density $f(y|x)$ with respect to (w.r.t.) Lebesgue's measure over \mathbb{R} . The conditional mode is defined as maximizer of the conditional density over given compact S

$$\theta(x) = \arg \max_{y \in S} f(y|x),$$

which is usually estimated nonparametrically by

$$\tilde{\theta}(x) \arg \max_{y \in S} \hat{f}(y|x),$$

where \hat{f} is the double kernel estimator of the conditional density function (see, Ferraty *et al.* (2006)). Alternatively, we focus, in this paper, another estimator based on the derivative of the conditional quantile which is more robust. Indeed, let $\alpha \in F^{-1}(S|x) = [\tau_x, \rho_x] \subset (0, 1)$, and $t(\alpha|x)$ the conditional quantile of order α given $X = x$. Using a similar ideas of Ota *et al.* (2019), in linear case, we assume that $t(\alpha|x)$ is of class C^1 , such that $f(t(\alpha|x)|x) > 0$, and we write that

$$t'(\alpha|x) = \frac{\partial t(\alpha|x)}{\partial \alpha} = \frac{1}{f(t(\alpha|x)|x)}.$$

Thus, the conditional mode can be defined as

$$\theta(x) = t(\alpha_\theta|x) \tag{1}$$

with

$$\alpha_\theta = \arg \min_{\alpha \in [\tau_x, \rho_x]} t'(\alpha|x).$$

Therefore, the natural estimator of the conditional mode is

$$\hat{\theta}(x) = \hat{t}(\hat{\alpha}_\theta|x) \tag{2}$$

where

$$\hat{\alpha}_\theta = \arg \min_{\alpha \in [\tau_x, \rho_x]} \hat{t}'(\alpha|x).$$

with \hat{t} and \hat{t}' are, respectively, estimators of the conditional quantile and its derivative. For sake of simplification, we estimate the derivative $t'(\alpha|x)$, $\alpha \in (0, 1)$ by

$$\hat{t}'(\alpha|x) = \frac{\hat{t}(\alpha + b_n|x) - \hat{t}(\alpha - b_n|x)}{2b_n}.$$

where b_n is a sequence of positive real numbers has the same role of the bandwidth parameter in the kernel smoothing. So, all it remains to construct is a robust estimator of the conditional quantile. To do that, we use the fact that the α^{th} conditional quantile $t(\alpha|x)$ can be defined as the unique solution with respect to (w.r.t) ζ of the optimization problem

$$\min_{\zeta \in \mathbb{R}} \mathbb{E} [\rho_\alpha(Y, \zeta) | X = x] \tag{3}$$

where $\rho_\alpha(Y, \zeta) = (2\alpha - 1)(Y - \zeta) + |Y - \zeta|$. Thus, the robust conditional mode estimation is obtained by taking in (2)

$$\hat{t}(\alpha|x) = \arg \min_{\zeta \in \mathbb{R}} \hat{\Psi}(\zeta, \alpha|x) \quad (4)$$

where

$$\hat{\Psi}(\zeta, \alpha|x) = \frac{\sum_{i=1}^n K(a_n^{-1}d(x, X_i))\rho_\alpha(Y_i - \zeta)}{\sum_{i=1}^n K(a_n^{-1}d(x, X_i))}, \quad \forall \zeta \in \mathbb{R}$$

with K is a kernel function and a_n is a sequence of positive real numbers which goes to zero as n goes to infinity. Of course, the existence and the unicity of the conditional mode $\theta(x)$ is insured by the continuity and the strict increasing of the conditional distribution function $F(\cdot|x)$. Whereas the estimator $\hat{\theta}(x)$ is not necessarily unique, so, the remainder of the paper concerns any value $\hat{\theta}(x)$ by the minimizers (4).

The main goal of this paper is to establish the asymptotic property of the estimator $\hat{\theta}(x)$ of $\theta(x)$ when the explanatory variable X is valued in the semi-metric space \mathcal{F} . As far as we know the present contribution is the first one that use this strategy to construct an estimator of the conditional mode even in the multivariate case. Of course, the finite dimensional case ($\mathcal{F} = \mathbb{R}^p$) can be viewed as particular case of this study. Therefore, the importance of this kind of functional data analysis is mainly due to the fact that it covers both functional and non functional cases. It worth to noting that the functional case is more interesting, not only for the fundamental problems they formulate, but also for many applications they may allow. More motivations on this topic can be found in Bosq (2000), Ramsay and Silverman (2005) and Ferraty and Vieu (2006), Aneiros *et al.*. (2019a, 2019b) and the references therein.

3. Main results

All along the paper, when no confusion will be possible, we will denote by C or C' some strictly positive generic constants and we set $B(x, h) = \{x' \in \mathcal{F} : d(x', x) < h\}$. Now, we list some required conditions that are necessary in deriving the almost complete convergence of $\hat{\theta}(x)$ of $\theta(x)$.

(H1) $\mathbb{P}(X \in B(x, r)) = \phi_x(r) > 0$. Besides, $\phi_x(r) \rightarrow 0$ as $r \rightarrow 0$.

(H2) The functions $t(\cdot|x)$ is of class $C^3([\tau_x, \rho_x])$ and $F(\cdot|x)$ such that the following Lipschitz's condition for all $(x_1, x_2) \in \mathcal{N}_x \times \mathcal{N}_x$, $|F(\zeta|x_1) - F(\zeta|x_2)| \leq Cd^\nu(x_1, x_2)$ for some $\nu > 0$, where \mathcal{N}_x denotes a neighborhood of x .

(H3) K is a function with support $(0, 1)$ such that $0 < C' < K(u) < C < \infty$.

(H4) The smoothing parameters a_n and b_n satisfy $\lim_{n \rightarrow \infty} \frac{\log n}{nb_n\phi_x(a_n)} = 0$

The conditions (H1)-(H4) are less standard in the nonparametric functional data analysis (NFDA). Condition (H1) is indispensable in NFDA and the function $\phi_x(\cdot)$ can be explicitly given for several continuous processes (see Ferraty and Vieu (2006)). While the regularity conditions is necessary to explore the functional space of the nonparametric path of the studied. Such assumption has a great impact on the bias term in the convergence rate of $\hat{\theta}(x)$. Finally conditions (H3)-(H4) concern the kernel K and the bandwidths a_n and b_n which cover the technical aspect of the estimator $\hat{\theta}(x)$. They are less restrictive than the usual technical conditions in the conditional mode estimation because the present approach requires one kernel instead to the conditional density approach that requires two kernels.

The following Theorem gives the almost-complete (a.co.) convergence of $\hat{\theta}(x)$.¹

Theorem 1. Under assumptions (H1)-(H4) and if $\inf_{\alpha \in (0,1)} \frac{\partial^3 t(\alpha|x)}{\partial \alpha} > 0$ we have:

$$\hat{\theta}(x) - \theta(x) = O(a_n^{b/2}) + O(b_n^{1/2}) + O\left(\left(\frac{\log n}{nb_n\phi_x(a_n)}\right)^{\frac{1}{4}}\right) \quad a.co.$$

¹Let $(z_n)_{n \in \mathbb{N}}$ be a sequence of real r.v.'s; we say that z_n converges almost completely (a.co.) to zero if, and only if, $\forall \epsilon > 0, \sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon) < \infty$. Moreover, let $(u_n)_{n \in \mathbb{N}^*}$ be a sequence of positive real numbers; we say that $z_n = O_{a.co.}(u_n)$ if, and only if, $\exists \epsilon > 0, \sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon u_n) < \infty$: This kind of convergence implies both almost sure convergence and convergence in probability (cf. Sarda and Vieu (2000) for details).

Remark 1. *Let's point out that, as all smoothing analysis, the bandwidth parameter a_n has a great influence on the quality of the estimator. It is well known that in the curve estimation by the kernel method, a parameter that is too small causes the appearance of artificial details in the graph of the estimator, however for a large enough value of the bandwidth a_n , the majority of the features is on the contrary erased. Therefore, the choice of the bandwidth a_n is a central question in nonparametric estimation. In the literature, we know that there are mainly three methods, i) "the rule of thumb", ii) "the plug-in" and iii) "the cross-validation". Each method has its merits and drawbacks. The very popular one is the latter and its main idea is to minimize the following criterion $CV_{a_n} = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \hat{\theta}_{-i, a_n}(X_i))^2$*

where $\hat{\theta}_{-i, a_n}(X_i)$ is the estimator of $\theta(\cdot)$ obtained by raising the observation (X_i, Y_i) in the sense of practical point of view. Even if the latter has the drawbacks that it is very variable and can give an underestimation of $(a_{n_{opt}})$, it remains the most common used method in practical point of view. Then in our simulation study, we have adopted the cross-validation method.

4. Sketch of the proof

The proof of Theorem 1 is based on some standard analytical arguments. Indeed, we write

$$\begin{aligned} \hat{\theta}(x) - \theta(x) &\leq \hat{t}(\hat{\alpha}_\theta|x) - t(\alpha_\theta|x) \\ &= \hat{t}(\hat{\alpha}_\theta|x) - t(\hat{\alpha}_\theta|x) + t(\hat{\alpha}_\theta|x) - t(\alpha_\theta|x) \\ &\leq \sup_{\alpha \in [\tau_x, \rho_x]} |\hat{t}(\alpha|x) - t(\alpha|x)| + t(\hat{\alpha}_\theta|x) - t(\alpha_\theta|x). \end{aligned} \quad (5)$$

Concerning the second term of the right quantity we use the Taylor expansion to prove that

$$t(\hat{\alpha}_\theta|x) - t(\alpha_\theta|x) = (\hat{\alpha}_\theta - \alpha_\theta)t'(\alpha_\theta^*|x), \quad \alpha_\theta^* \text{ being in } (\hat{\alpha}_\theta, \alpha_\theta). \quad (6)$$

On the other hand, because of α_θ is minimizer of $t'(\cdot|x)$, then

$$t'(\hat{\alpha}_\theta|x) - t'(\alpha_\theta|x) = (\hat{\alpha}_\theta - \alpha_\theta)^2 t'''(\alpha_\theta^{**}|x), \quad \text{where } \alpha_\theta^{**} \in (\hat{\alpha}_\theta, \alpha_\theta). \quad (7)$$

Moreover,

$$\begin{aligned} t'(\hat{\alpha}_\theta|x) - t'(\alpha_\theta|x) &= t'(\hat{\alpha}_\theta|x) - \hat{t}'(\hat{\alpha}_\theta|x) + \hat{t}'(\hat{\alpha}_\theta|x) - t'(\alpha_\theta|x) \\ &\leq |t'(\hat{\alpha}_\theta|x) - \hat{t}'(\hat{\alpha}_\theta|x)| + |\min_\alpha \hat{t}'(\alpha|x) - \min_\alpha t'(\alpha|x)| \\ &\leq 2 \sup_{\alpha \in [\tau_x, \rho_x]} |\hat{t}'(\alpha|x) - t'(\alpha|x)| \end{aligned} \quad (8)$$

Combining Equations (5)-(8) to write that

$$|\hat{\theta}(x) - \theta(x)| \leq C \left(\sup_{\alpha \in [\tau_x, \rho_x]} |\hat{t}(\alpha|x) - t(\alpha|x)| + \sqrt{\sup_{\alpha \in [\tau_x, \rho_x]} |\hat{t}'(\alpha|x) - t'(\alpha|x)|} \right).$$

Now, using the definition $\hat{t}'(\alpha|x)$ to get, for n large enough

$$\begin{aligned} \hat{t}'(\alpha|x) - t'(\alpha|x) &= \frac{\hat{t}(\alpha + b_n|x) - t(\alpha + b_n|x) + t(\alpha - b_n|x) - \hat{t}(\alpha - b_n|x)}{2b_n} \\ &\quad + \frac{t(\alpha + b_n|x) - t(\alpha|x) + t(\alpha|x) - t(\alpha - b_n|x) - 2b_n t'(\alpha|x)}{2b_n} \\ &\leq C b_n^{-1} \sup_{\alpha \in (\tau_x - b_n, \rho_x + b_n)} |\hat{t}(\alpha|x) - t(\alpha|x)| + O(b_n) \end{aligned}$$

So, all what is left to be evaluated that, for n large enough, we have $\sup_{\alpha \in (0,1)} |\hat{t}(\alpha|x) - t(\alpha|x)|$. Thus, Theorem 1 is consequence of the following Proposition

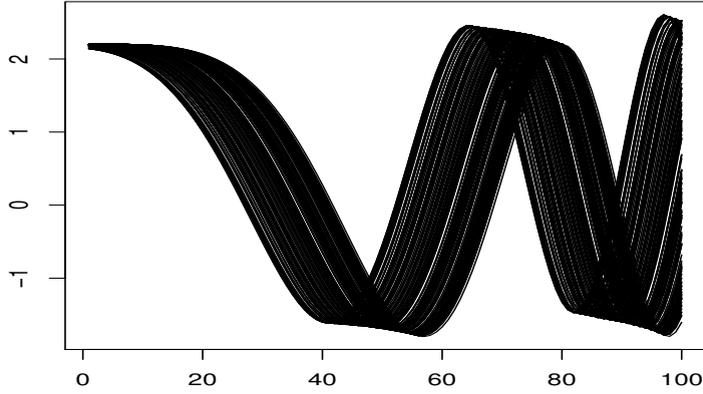


Figure 1: Functional explanatory variables

Proposition 1. Under assumptions of Theorem 1, we have for n large enough

$$\sup_{\alpha \in (0,1)} |\hat{t}(\alpha|x) - t(\alpha|x)| = O(a_n^b) + O_{a.co.} \left(\left(\frac{\log n}{n \phi_x(a_n)} \right)^{\frac{1}{2}} \right).$$

The proof of this proposition is based on the Bahadur representation of the conditional quantile. The latter is consequence of the following lemmas.

Lemma 1. Let F_n be a sequence of decreasing real random functions and A_n be a random real sequence such that $A_n = o_{a.co.}(1)$ and $\sup_{|\delta| \leq M} |F_n(\delta) + \lambda \delta - A_n| = o_{a.co.}(1)$ for certain constants $\lambda, M > 0$.

Then, for any real **random** sequence δ_n such that $F_n(\delta_n) = o_{a.co.}(1)$, we have: $\sum_{n=1}^{\infty} \mathbb{P} \{ |\delta_n| \geq M \} < \infty$.

The detailed proof can be obtained from the authors.

5. Numerical analysis

The main aim of this section is to show the easy implementation of the proposed estimator in the prediction analysis and to examine its insensitivity to the outliers. To do that, we generate a sample of functional random observations $(X_i)_{i=1, \dots, n}$ using the following formula

$$X_i(t) = 0.2(\cos(W_1 t))^2 + 2\cos((W_1 + t)^2) + 0.2(W_1 t) * 2$$

where W_1 is random variable distributed as uniform on $(0, 0.5)$. All the curves X_i 's are discretized on the same grid generated from 100 equispaced measurements in $(0, \pi)$. The generated curves are displayed in Fig. 1.

Next, we draw a scalar response observations using the following nonparametric regression model

$$Y_i = R(X_i) + \epsilon_i, \quad i = 1, \dots, n.$$

where the operator R is defined by

$$R(x) = 2 \int_0^{\pi} x(t) dt.$$

Observe that the conditional distribution of Y given X is closely linked to the distribution of the white noise ϵ_i . So, to highlight the robustness of $\hat{\theta}$ compared to the classical estimator

$$\tilde{\theta}(x) = \arg \max \frac{\sum_{i=1}^n K(a_n^{-1}d(x, X_i))K(b_n^{-1}(y - Y_i))}{\sum_{i=1}^n b_n K(a_n^{-1}d(x, X_i))}$$

we consider two types of white noise distributions: heavy tailed and light tailed distributions. Specifically, we simulate, for the first case, ϵ_i from Levy-distribution $m = 0, s = 2$ which is heavy tailed distribution. While for the second situation, we use the Gumbel distribution $\mu = 0.5, \beta = 2$ (as light tailed distribution) to generate ϵ_i . Recall that the heavy tailed distributions tend to have many outliers with very high values.

Undoubtedly, the easy implementation of $\hat{\theta}$ depends to the flexibility of the choice of parameters involved in its definition. Precisely, the practical use of the estimator $\hat{\theta}$ is based on the selection of the kernel K and the bandwidth parameters (a_n, b_n) and the semi-metric d . Concerning the prediction by the estimator $\tilde{\theta}$, we use the routine *funopare.mode.lcv* developed by Ferraty and Vieu (2006) which is available on the website². Specifically, we applied this routine with a quadratic kernel on the interval $(0, 1)$ and the metric of L_2 between the B-splines smoothed curves. The optimal bandwidths in $\tilde{\theta}$ were chosen by the cross-validation method over a discrete set of bandwidth values defined by the nearest neighbors. Therefore, in order to conduct a fair comparison between the two predictors, we use the same kernel, the same metric and the same rule to select the smoothing parameters in $\hat{\theta}(x)$. Finally, the finite sample behaviour of the estimators $\hat{\theta}$ and $\tilde{\theta}$ is examined in the prediction context by splitting randomly for several times into two subsets: the learning sample (100 observations) and the testing sample (50 observations). The learning sample is used to compute the estimators $\hat{\theta}$ and $\tilde{\theta}$ as predictors of the response values of the testing sample. In order to compare the stability of both estimators against different frequencies of outliers in the data, we repeat this prediction process 50 times. For each time, we calculate the Mean Absolute Error (MAE) values

$$\text{MAE}(\tilde{\theta}) = \frac{1}{50} \sum_{i=1}^{50} |Y_i - \tilde{\theta}(X_i)|,$$

where the observations (X_i, Y_i) represent the testing sample and $\tilde{\theta}$ means either $\hat{\theta}$ or $\tilde{\theta}$. The prediction error for both situations (heavy tailed distribution and light-tailed distribution) are plotted in the following figures 2, 3.

It is clear that, even if both estimators has a similar precision in certain situation (see the light-tailed distribution case, figure 3), the prediction error of the estimator $\hat{\theta}$ is more stable than the predictor $\tilde{\theta}$. In sense that the variability of $\text{MAE}(\hat{\theta})$ is lower than $\text{MAE}(\tilde{\theta})$. This statement is confirmed by the interquartile range of the MAE (IQR-MAE). In particular, the IQR-MAE for the heavy-tailed distribution case is equal to 5.11 for the predictor $\tilde{\theta}$ versus 2.86 for the predictor $\hat{\theta}$. However there is small difference for the light-tailed distribution, the IQR-MAE of predictors ($\tilde{\theta}$ and $\hat{\theta}$) are respectively 0.86, 0.69. In conclusion, we can say that the estimator $\hat{\theta}$ enhance considerably the robustness of the modal regression, as predictor model, by decreasing its sensitivity to d the outliers.

6. Discussion and Comments

6.0.1. On the features of the estimation method

Compared to the classical estimator based on the conditional density estimation, our new estimator has a several advantages. Indeed, the use of the L_1 conditional quantile allows to increases the robustness of the conditional mode estimation and reduces the parameters involved in its estimator which is very beneficial of the computational time of the predictor. In addition, the new estimator offers the possibility to increase the scope of application of the modale regression using the principe of the binary quantile regression. In order to highlight the main features of the new estimator compared to the classical one, we discuss in the rest of this section the principal motivations of the new estimator.

²<http://www.lsp.ups-tlse.fr/staph/npfda>

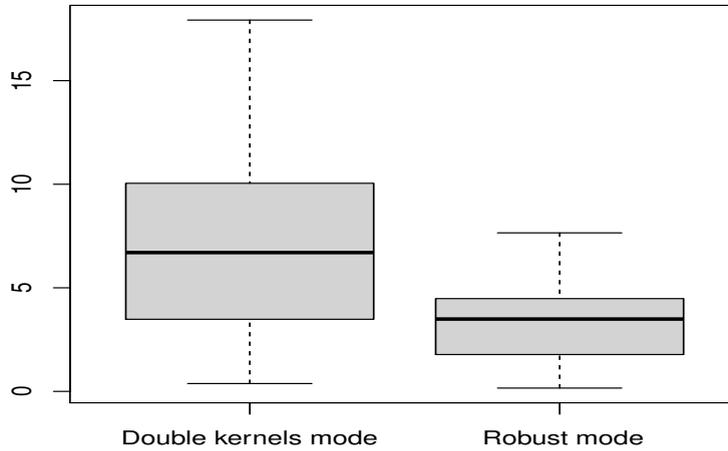


Figure 2: The MAE error in the heavy distribution case

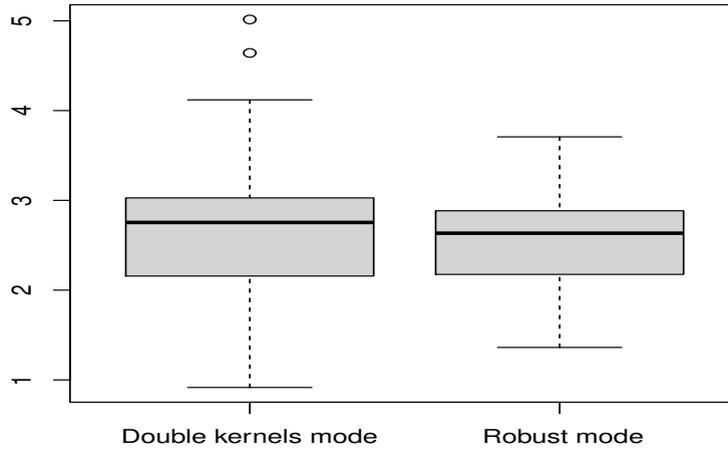


Figure 3: The MAE error in the light distribution case

Feature 1: The robustness and the accuracy

It is well known that the convergence rate or the accuracy for any estimator is closely linked to the considered assumptions. In sense that good properties are usually obtained under strong hypotheses that are very difficult to check in practice. For this reason, the robustness of the estimation method is very important for the practical issues. It permits to keep good properties even when the hypotheses of the ideal situation are not fully verified. Thus, the main nice feature of our approach is the possibility to construct a new estimator with high robustness property without impacting the accuracy of the conditional mode function. Indeed, it is clear that the obtained convergence rate has the same structure as all previous studies in nonparametric analysis by kernel smoothing in functional statistics. But, the use of

the quantile regression improve the robustness of the conditional mode function. On the other hand, the asymptotic property of the new estimator is also depends on the estimation of the sparsity function $t'(\alpha|x)$. In this work, we have opted for a simple, fast and robust estimator of $t'(\alpha|x)$ which can be implemented easily in practice using the L_1 quantile regression. However, in the multivariate statistics literature, there exit some alternative estimators of $t'(\alpha|x)$ constructed from the density estimation of the residue (see, Jones (1992) or Belloni et al. (2019) and the references therein) . So it will be interesting in the future to study if these alternative estimators of $t'(\alpha|x)$ can improve the convergence rate of the conditional mode estimation without loss in robustness.

Feature 2: Supervised classification by conditional mode function

The second main feature of the present approach is the possibility to employ the conditional mode as classifier in the supervised curves classification. It concerns the case where the output variable Y is a binary random variable which is expressed by $Y = \mathbb{1}_{\{C\}}$ where C is a given class. This question has been studied by Ferraty and Vieu (2006). Their algorithm is based on the estimation $\mathbb{P}(C|X) = \mathbb{E}[Y|X]$ to decide on the affectation of X to the class C . Specifically, the curve x is affected to the class C if $\mathbb{P}(C|X = x)$ is greater than $\mathbb{P}(\bar{C}|X = x)$, where \bar{C} is the complementary set of C . The estimation of these conditional probabilities is performed by the functional version of the Nadaraya-Watson's estimator of the $\mathbb{E}[Y|X]$. Now, using the binary quantile regression ideas developed in Manski (1985) to define an alternative classifier based on the modal regression. Indeed, considered the following binary regression

$$Y = r(X) + \varepsilon \quad \text{and the class } C \text{ is characterised by } C = \{Y > 0\}.$$

Therefore, by some simple analytical arguments we prove that

$$\mathbb{P}(C|X) \leq 1 - \alpha \Leftrightarrow t(\alpha|x) \leq 0.$$

Thus, our new classification rule affect the new curve x to the class C according the sign of $\hat{\theta}(x)$.

6.1. Some future tracks

The estimation of the conditional mode using the quantile regression offers some important tracks for the future. In fact the quantile regression behaves in various statistical procedures. It can be treated as linear model (see, Cardot *et al.* 2005), semiparametric model (see Crambes *et al.* 2016) as well as nonparametric model. At this stage, the present estimation method can be combined to the different versions of the quantile regression provide a linear estimator of the conditional mode or semiparametric estimator of the modal regression. Therefore, the first naturel prospect of this work is to study the linear (or semiparametric) estimation of the regression model. Noting that the parametric technique is very fast and does not require a large sample size and does not affected by the dimensionality curse. On the other hand, the local linear estimation of the conditional quantile studied by Al-Awadhi *et al.* (2019), allows to construct an alternative nonparametric estimator of the conditional mode. The latter is also important to reduce the bias term of the classical kernel estimator. Thus the study of the asymptotic property of the local linear estimator of the regression model using the quantile function is also an important open question of the present work. A challenging task would be to consider an extension of our results to the censored data and dependent observations, which requires nontrivial mathematics, that goes well beyond the scope of the present paper.

Acknowledgement

The authors are indebted to the Editor-in-Chief, the Associate Editor and the three referees for their very valuable comments and suggestions which led to a considerable improvement of the manuscript.

References

- Al-Awadhi, F., Kaid, Z., Laksaci, A., Ouassou, I. and Rachdi, M. (2019). Functional data analysis: local linear estimation of the L_1 -conditional quantiles. *Stat. Methods Appl.*, 28, 217–240.
- Aneiros-Perez, G., Cao, R., Fraiman, R. and Vieu, P. (2019a). Editorial for the special issue on functional data analysis and related topics. *J. Multivariate Anal.*, 170, 1–3.
- Aneiros PÁrez, G., Cao, R. and Vieu, P. (2019b). Editorial on the special issue on functional data analysis and related topics. *Comput. Statist.*, 34, 447–450.
- Belloni, A., Chernozhukov, and Kato. K. , (2019) Valid post-selection inference in high-dimensional approximately sparse quantile regression models. *Journal of the American Statistical Association*, 114, 749–758.
- Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and applications*. Lecture Notes in Statistics, 149, Springer. New York.

- Bouanani, O., Rahmani, S., Laksaci, A. and Rachdi, M. (2020). Asymptotic normality of conditional mode estimation for functional dependent data. *Indian J. Pure Appl. Math.*, **51**, 465–481.
- Chowdhury, J. and Chaudhuri, P. (2019). Nonparametric depth and quantile regression for functional data. *Bernoulli*, **25**(1), 395–423
- Cardot, H., Crambes, C. and Sarda, P. (2005). Quantile regression when the covariates are functions. *J. Nonparametr. Statist.* **17**, 841–856.
- Collomb, G., Hardle, W. and Hassani, S. (1987). A note on prediction via conditional mode estimation. *J. Statist. Plann. and Inf.*, **15**, 227-236.
- Crambes, C., Hilgert, N. and Manrique, T. (2016). Estimation of the noise covariance operator in functional linear regression with functional outputs. *Statist. Probab. Lett.*, **113**, 7–13.
- Dabo-Niang, S., Kaid, Z. and Laksaci A. (2015). Asymptotic properties of the kernel estimate of spatial conditional mode when the regressor is functional. *AStA Adv. Stat. Anal.* **99**, 131–160.
- Ezzahrioui, M. and Ould Saïd, E. (2008). Asymptotic normality of nonparametric estimators of the conditional mode function for functional data *J. Nonparametric Statist.* , **20**, 3-18.
- Ferraty, F., Vieu, P. (2006). *Nonparametric functional data analysis*. Springer. Series in Statistics, Springer New York.
- Geenens, G. (2015). Moments, errors, asymptotic normality and large deviation principle in nonparametric functional regression. *Statist. Probab. Lett.* **107**, 369–377.
- Goia, A. and Vieu, P. (2016). An introduction to recent advances in high/infinite dimensional statistics. *J. Multivariate Anal.*, **46**, 1–6.
- Jones, M.C. (1992) Estimating densities, quantiles, quantile densities and densities quantiles, *Ann. Inst. Statist. Math.* **44**, 721-727.
- Ling, N., Liu, Y. and Vieu, P. (2016). Conditional mode estimation for functional stationary ergodic data with responses missing at random. *Statistics*, **50** , 991–1013.
- Ling, N. and Vieu, P. (2018). Nonparametric modeling for functional data: selected survey and tracks for future. *Statistics*, **52**, 934–949.
- Manski, C. Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics*, **27**, no 3, 313-333, (1985).
- Ota, H., Kato, K. and Hara, S. (2019). Quantile regression approach to conditional mode estimation. *Electron. J. Statist.* **13**, 3120–3160.
- Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis, Second Edition*. Springer, New York.