IJAMS Peer-reviewed journal
International Journal of Applied Mathematics and Simulation

On the kernel conditional density estimator with functional explanatory variable

Nour El Hayet LADAOURI[1]. Mouloud CHERFAOUI;[2]

DOI: https://doi.org/10.69717/ijams.v2.i1.117

abstract

This article focuses on the relationship between a scalar-explained random variable Y and a functional explanatory random variable X. In fact, through this work, the aim is to estimate the conditional probability density $f(y/x)$ when the explanatory variable X is functional using the kernel method. The performance of the estimator is also analyzed with respect to sample size, assumptions on the smoothing parameters (i.e., independence between the smoothing parameters in the X and Y directions, and equality between them), and the choice of norm used in the estimation process.

keywords

Conditional density, Functional explanatory variables, Norms, Errors, Simulation.

2020 Mathematics Subject Classification

Primary 62G05; 62G07 · Secondary 62R10. ·

1. Introduction

The first work on the nonparametric kernel conditional density estimation, when the explanatory variable is functional, was introduced by Ferraty et al. A. Ferraty and Vieu, 2006; F. Ferraty, 2006. With the same approach followed by Rosenblatt Rosenblatt, 1969, for the real explanatory variable, the authors have constructed and analyzed the kernel conditional density estimator in the functional explanatory variable. Since these two works, the literature has developed on the kernel estimation of the conditional density in the framework of functional explanatory variables, its derivatives, and its applications in other fields, we can cite for example the works of: Ezzahrioui and Ould Saïd M. Ezzahrioui, 2010; M. Ezzahrioui and E, 2005, Ezzahrioui E. [N. Ezzahrioui, 2008, Laksaci A. Laksaci and Mechab, 2010, Laksaci and Mechab A. N. Laksaci, 2007, Ferraty et al. A. Ferraty and Vieu, 2008; F. Ferraty et al., 2010 and Dabo-Niang Dabo-Niang, 2007.

Let (X, Y) be a couple of random variables in $\mathbb{F}\times\mathbb{R}$, where $(\mathbb{F}, \|.\|)$ is a functions space equipped with a norm $\|.\|$, i.e X is a (time-dependent) functional random variable depending in time ($X \equiv X(t)$). Let $(X_i, Y_i)_{1\leq i\leq n}$ be n independent pairs, identically distributed as the couple (X, Y) and (x, y) be a fixed element of $\mathbb{F}\times\mathbb{R}$. The kernel estimator of $f(y/x)$ in this context is defined by:

$$\hat{f}_{ab}(y/x) \quad = \quad \frac{\sum_{j=1}^{n} K_1\left(\frac{\|x-X_j\|_p}{a}\right) K_2\left(\frac{y-Y_j}{b}\right)}{b\sum_{j=1}^{n} K_1\left(\frac{\|x-X_j\|_p}{a}\right)}, \tag{1}$$

[1]Mathematics Department, University of Biskra, Biskra, Algeria, nourelhayet.ladaouri@univ-biskra.dz
[2]Research unit LaMOS (System Modeling and optimization), University of Béjaïa, Béjaïa, Algeria, m.cherfaoui@univ-biskra.dz

Communicated Editor: A. Necir
Manuscript received *Jan 22, 2025;* ***revised*** *April 15, 2025;* ***accepted*** *May 25, 2025;* ***published on line*** *Jun 02, 2025..*

where K_1 is a real asymmetric kernel function on $\mathbb{R}_+$, K_2 is a real symmetric kernel function on $\mathbb{R}$, $a > 0$ and $b > 0$ are the smoothing parameters in the X and Y directions respectively, and $\|.\|_p$ is a p-distance defined on $\mathbb{F}$ (for more details on the p-distance see Section 2).

In order to establish the mean square convergence of the estimator (1), Laksaci A. N. Laksaci, 2007 defined the formulas for bias and variance and the asymptotic square error of the estimator. Also, Ferraty et al. F. Ferraty, 2006, showed the both pointwisely and uniformly almost complete convergence of the estimator in question. Ferraty et al. A. Ferraty and Vieu, 2006 estimated the J^{th} order of derivative of the estimator 1.

To simplify the formula (1), we propose another version, similar to that used by Youndjé Youndjé, 1996 in the case of one explanatory variable, under the hypothesis that the two smoothing parameters a and b are equal ($h = a = b$). So, the expression (1) can be rewritten as follows:

$$\hat{f}_h(y/x) = \frac{\sum_{j=1}^{n} K_1\left(\frac{\|x-X_j\|_p}{h}\right) K_2\left(\frac{y-Y_j}{h}\right)}{h\sum_{j=1}^{n} K_1\left(\frac{\|x-X_j\|_p}{h}\right)}, \tag{2}$$

2. Concept of the norm

To study data, we often need to have a notion of distance between them. In mathematics, a distance (or metric) $d(.,.)$ is an application that formalizes the intuitive idea of distance, i.e. it represents the length that separates two points.

One of the most popular idea in mathematics to calculate a distance between two points is to use a $p-$norm that we denote by $\|.\|_p$. Thus, to study the estimators introduced in the previous section, it is interesting to recall some common norms used in such estimators.

Let $\mathbb{F}$ be a set of functions and $x(t)$ and $y(t)$ two functions defined in $\mathbb{F}$. The most commonly used norms in practice to measure the distance between these two functional points (these two functions) are presented in the following table:

Name	Parameter	Expression
Manhattan norm	1-norm	$\int_{\mathbb{F}} \lvert x(t) - y(t)\rvert\, dt$
Euclidian norm	2-norm	$\left(\int_{\mathbb{F}}(x(t) - y(t))^2 dt\right)^{\frac{1}{2}}$
Minkowski norm	p-norm	$\left(\int_{\mathbb{F}} \lvert x(t) - y(t)\rvert^p\, dt\right)^{\frac{1}{p}}$
Tchebychev norm	∞-norm	$\lim_{p\to\infty} \left(\int_{\mathbb{F}} \lvert x(t) - y(t)\rvert^p\, dt\right)^{\frac{1}{p}} = \sup_t \lvert x(t) - y(t) \rvert$

Table 1: Some norms which are generally used as distance in the functional framework.

3. Choice of kernel and smoothing parameter

From the two expressions (1) and (2), it is clear that the implementation of these estimators relies on the prior fixing of the kernel, the smoothing parameter, and the norm $\|.\|_p$. In our work, we focus particularly on the problem of choosing the smoothing parameter, because the choice of the kernel function remains the same as the univariate density. Furthermore, because $\|u\|_p$ is always a positive quantity, the real kernel K_1 should have positive support, consequently, we must use asymmetric density functions for the kernel K_1 (see Chen, 1999, 2000). While K_2 we must use a symmetric density functions because $\frac{y-Y_j}{h} \in \mathbb{R}$ (see Silverman, 2018). However, the problem of choosing the smoothing parameter has received serious attention because the numerical and graphical characteristics of the designed kernel estimator are very sensitive to the variation of the smoothing parameter, where small values of this parameter (compared to the optimal smoothing parameter) generate the phenomenon of under-smoothing, while large values of this parameter generate the phenomenon of over-smoothing.

Similarly to the cross-validation approach followed by Youndjé Youndjé, 1996 when the explanatory variable is scalar, Rachdi and Vieu Rachdi and Vieu, 2007 and Benhenni et al. Benhenni et al., 2007 proposed, respectively, a global leave-out-one-curve and a local adaptive leave-out-one-curve cross-validation procedure for the regression operator estimation in functional data. Laksaci et al. A. Laksaci et al., 2013 constructed the global and local leave-out-one-curve cross-validation procedures in the context of conditional density when the explanatory variable is functional.

Global and local bandwidth selection rules

The idea of this approach is based on minimizing the integrated squared error, which is weighted by the probability measure, $dP_X(x)$, of the functional variable X and some non-negative weighting functions W_1 and W_2 associated to the variables x and y respectively. That is to say, they considered the integrated squared error defined by the following expression:

$$ISE(\hat{f}, f) = \int\int \left(\hat{f}(y/x) - f(y/x)\right)^2 W_1(x)W_2(y)dP_X(x)dy. \tag{3}$$

So, the mean integrated squared error will be given as follows:

$$MISE(\hat{f}, f) = \int\int \mathbb{E}\left(\hat{f}(y/x) - f(y/x)\right)^2 W_1(x)W_2(y)dP_X(x)dy. \tag{4}$$

Discretizing the expression (3) allows us to obtain an approximation of the mean square error given by:

$$ISE(\hat{f}, f) \approx \frac{1}{n}\sum_{i=1}^{n}\left(\hat{f}(Y_i/X_i) - f(Y_i/X_i)\right)^2 \frac{W_1(X_i)W_2(Y_i)}{f(X_i, Y_i)}. \tag{5}$$

Concerning the weighting functions W_1 and W_2, we recall that these functions were introduced to reduce bounds effects thanks to their support. But, in practice, Härdle and Marron Härdle and Marron, 1985 have emphasized that the role of their expressions is not very determining and that they are functions arbitrarily chosen by the user. Laksaci et al. A. Laksaci et al., 2013 took the expressions of W_1 and W_2 in their simulation as:

$$W_2(z) = \begin{cases} 1 & if\ z \in [\min Y_i, i = 1...n \times 0.9 \max Y_i, i = 1...n \times 1.1] \\ 0 & otherwise, \end{cases}$$

and

$$W_1(t) = \begin{cases} 1 & if\ \min d(t, X_i) < a_0 \\ 0 & otherwise, \end{cases}$$

where $a_0 = \min(a_q)$ and a_q is the quantile of order q of the vector of all distances between the curves.

We note that the ISE and $MISE$ functions depend on the unknown conditional density f, so, in practice, the smoothing parameters that minimize these errors are not computable. By following the same ideas as in Youndjé Youndjé, 1996 for the real case, Laksaci et al. A. Laksaci et al., 2013 proposed another function that is asymptotically equivalent to the quadratic distance given in (3). Where they suggest to replace (3) by the integrated squared error expressed as follows:

$$ISE(\hat{f}, f) = A + B - 2C,$$

where

$$A = \int\int \hat{f}(y/x)^2 W_1(x)W_2(y)dP_X(x)dy,$$
$$B = \int\int f(y/x)^2 W_1(x)W_2(y)dP_X(x)dy,$$

and

$$C = \int\int \hat{f}(y/x)f(y/x)W_1(x)W_2(y)dP_X(x)dy.$$

Since the second term B is independent of the smoothing parameter (a, b), the problem of minimizing the ISE is equivalent to that of minimizing the function $A - 2C$. Thus, to select the bandwidth (a, b) that minimizes the approximate ISE ($AISE = A - 2C$), we must first estimate the two quantities A and C whose form are as follows;

$$
\begin{aligned}
C &= \int\int \hat{f}(y/x) f(y/x) W_1(x) W_2(y) dP_X(x) dy, \\
&= \int\int \hat{f}(y/x) W_1(x) W_2(y) dP_{Y/X=x}(y) dP_X(x), \\
&= \int\int \hat{f}(y/x) W_1(x) W_2(y) dP_{(X,Y)}(x,y), \\
&= \mathbb{E}_{(X,Y)} \left(\hat{f}(Y/X) W_1(X) W_2(Y) \right),
\end{aligned}
$$

and,

$$
A = \mathbb{E}_X \left(\int \hat{f}(y/X) W_1(X) W_2(y) dy \right),
$$

where $\mathbb{E}_X$ denotes the mean associated with the distribution of the random variable X.

For the aim to minimize the function $A - 2C$, the authors have followed the idea of Rudemo Rudemo, 1982 and Rachdi and Vieu Rachdi and Vieu, 2007 where they adopted the cross-validation technique with leave-out-one-curve principle. More precisely, they constructed the following criteria, for the global smoothing parameter:

$$
\begin{aligned}
GCV(a,b) &= \frac{1}{n} \sum_{i=1}^{n} W_1(X_i) \int \left(\hat{f}_{-i}\left(y/X_i\right) \right)^2 W_2(y) dy \\
&\quad -\frac{2}{n} \sum_{i=1}^{n} \hat{f}_{-i}\left(Y_i/X_i\right) W_1(X_i) W_2(Y_i),
\end{aligned} \tag{6}
$$

and the following local criteria, for a fixed $y \in \mathbb{R}$ and $x \in \mathbb{F}$;

$$
\begin{aligned}
LCV(a,b) &= \frac{1}{n} \sum_{i=1}^{n} W_{1,x}(X_i) \int \left(\hat{f}^2_{-i}\left(z/X_i\right) \right) W_{2,y}(z) dz \\
&\quad -\frac{2}{n} \sum_{i=1}^{n} \hat{f}_{-i}\left(Y_i/X_i\right) W_{1,x}(X_i) W_{2,y}(Y_i),
\end{aligned} \tag{7}
$$

where, $W_{2,x}$ (respectively $W_{2,y}$) is some positive local weight function around x (respectively y), Laksaci et al. A. Laksaci et al., 2013 used the following local weight functions:

$$
W_{1,x} = \begin{cases} 1 \; if \; \min d(t;x) < a(x); \\ 0 \; otherwise \end{cases} \quad \text{and} \quad W_{2,y}(z) = \begin{cases} 1 \; if \; |z-y| < b(y); \\ 0 \; otherwise \end{cases}
$$

where $a(x)$ (respectively, for $b(x)$) the ball centered at x (respectively the interval centered at y) with radius $a(x)$ (respectively with radius $b(y)$) contains exactly k neighbors of x (respectively of y).

And $\hat{f}_{-i}(y/x)$ represent the kernel conditional density estimator, using the cross-validation technique computed from the set of points except the point (x_i, y_i), its formula is given by:

$$
\hat{f}_{-i}(y_i/x_i) = \frac{\sum_{j=1, j\neq i}^{n} K_1\left(\frac{\|x_i - X_j\|_p}{a}\right) K_2\left(\frac{y_i - Y_j}{b}\right)}{b \sum_{j=1, j\neq i}^{n} K_1\left(\frac{\|x_i - X_j\|_p}{a}\right)},
$$

Hence, the global (respectively, local) cross-validation procedure consists of choosing the smoothing parameters (a,b) which minimize the criteria GCV (respectively LCV).

4. Numerical application

This section aims to illustrate, via numerical examples and using the simulation approach, how to implement the conditional kernel density estimator when the explanatory variable is functional, and to verify the impact of the substitution of the kernel K_1, initially asymmetric, and the kernel K_2 by a same symmetric kernel K. Also, we focus in the choice of the norm used to calculate the distance between the points $x(t)$ and $x_i(t)$ on the quality of the designed estimator.

4.1. Presentation of the application and its parameters

In order to study the effect of using a symmetric kernel in the performance of kernel conditional density estimation when the explanatory variable is functional, let consider that $K_1 = K_2 = K$, with K is a real symmetric kernel function on $\mathbb{R}$. that is, we rewrite (1) and (2) respectively as follows:

$$\hat{f}_{ab}(y/x) = \frac{\sum_{j=1}^{n} K\left(\frac{\|x-X_j\|_p}{a}\right) K\left(\frac{y-Y_j}{b}\right)}{b\sum_{j=1}^{n} K\left(\frac{\|x-X_j\|_p}{a}\right)}, \tag{8}$$

and

$$\hat{f}_{h}(y/x) = \frac{\sum_{j=1}^{n} K\left(\frac{\|x-X_j\|_p}{h}\right) K\left(\frac{y-Y_j}{h}\right)}{h\sum_{j=1}^{n} K\left(\frac{\|x-X_j\|_p}{h}\right)}, \tag{9}$$

In order to meet our objective, we have implemented a simulator in a *Matlab* environment, whose main steps are:

1. Generate m samples $(X_i^{(l)}, Y_i^{(l)})$ of size n of a target distribution, where $l = 1, ..., m$ and $i = 1, ..., n$.
2. Compute $(\hat{a}, \hat{b})$ and $\hat{h}$ that minimize the average of the ISE associated with each estimator.
3. Calculate the both estimators (8) and (9) and compare their performance.

To realize these steps, and for calculation reasons, we proposed to discretize (to approximate) the average ISE. More precisely, we use the discretized expression of the average ISE that is given as follows: (see Bashtannyk and Hyndman, 2001):

$$\overline{ISE} = \frac{\Delta}{nN}\sum_{l=1}^{m}\sum_{j=1}^{J}\sum_{i=1}^{n}\left[\hat{f}(y_j^{'}/x_i^{(l)}) - f(y_j^{'}/x_i^{(l)})\right]^2, \tag{10}$$

where $(x_i, y_i), i = 1, ..., n$ an independent and identically distributed observations from the joint density of (X, Y), $y^{'} = \left(y_1^{'}, y_2^{'}, ..., y_J^{'}\right)$ is a vector of equidistant points in the space of Y and $\Delta = y_{j+1}^{'} - y_j^{'}$, $\forall j \in \{1, 2, ..., J-1\}$.

Consequently, the estimators of the optimal smoothing parameters, in the sense of the average of the ISE, correspond to the quantities which minimize the expression (10).

For the simulation example, we considered the following model, which represents a conditional density Y knowing $X = x$ follows a normal law with mean $\|x\|_2$ and variance 1, given by:

$$f(y/x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(y-\|x\|_2\right)^2}, \tag{11}$$

and suppose that the explanatory variables X_i from a stochastic process similar to that proposed by Delsol Delsol, 2008, and it defined by:

$$X_i = X_i(t) = a_i\cos(2\pi t) + b_i\sin(3\pi t) + c_i(t-0.45)(t-0.75)e^{(-d_i t)},$$

with $t \in [0, 1]$, $a_i \rightsquigarrow N(-1, 1)$, $b_i \rightsquigarrow N(-1, 1)$, $c_i \rightsquigarrow U[1, 5]$ and $d_i \rightsquigarrow U[1, 5]$, where N and U respectively designate a normal distribution and a uniform distribution.

An illustration example about the variation of our model which defined in (11), depending on $\|x\|_2$ is presented in figure 1. The Figure 2 represents an example of a sample of size $n = 10$ from the variable X, where the red curve in the figure 2 represents the theoretical $E(X(t))$, that is to say when $E(a) = E(b) = -1$ and $E(c) = E(d) = 3$.

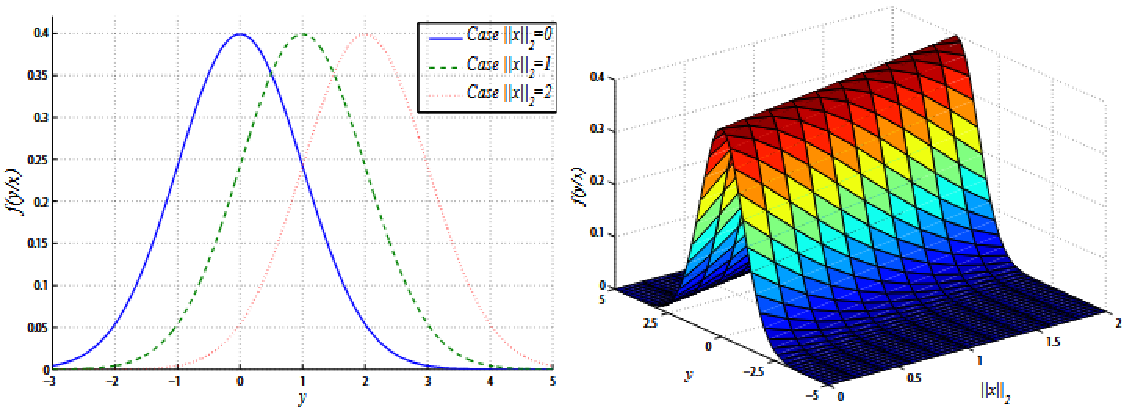

Figure 1: Illustration curves about the variation of the density $f(y/x)$ in depending on $\|x\|_2$.

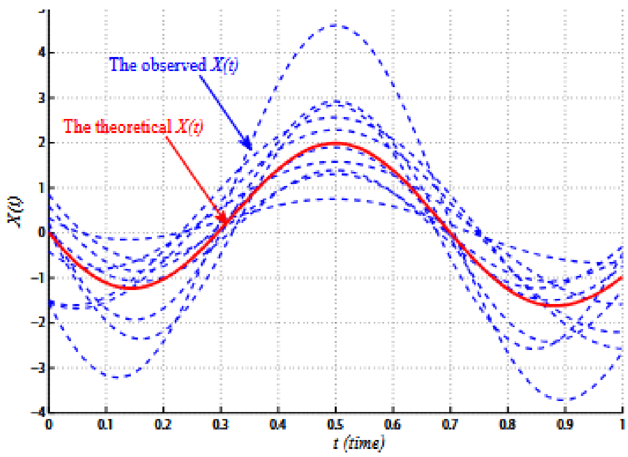

Figure 2: Ten examples of a sample xi(t) generated from the variable X.

4.2. Numerical and graphic results

To obtained the results of the numerical application, we consider to use the Gaussian kernel and the norm $p \in \{1,\ 2,\ \infty\}$ in the construction of the two versions of the conditional density estimator in question. The application was carried out on 100 samples ($m = 100$) of size $n \in \{50\ ,\ 100\ ,\ 200\ ,\ 500\ ,\ 1000\ ,\ 2000\}$ at the point $x = E(X)$ (see figure 2).

The numerical results obtained in our application are arranged in Table 2 and are presented in Figures 3–4 .

Norm	n	$a = b$		$a \neq b$	
		$\hat{h}$	$\overline{ISE}_h$	$(\hat{a}, \hat{b})$	$\overline{ISE}_{(a;b)}$
$\|.\|_1$	50	0.8739	0.0187	(23.7655 ; 0.5509)	0.0064
	100	0.8707	0.0176	(32.6362 ; 0.4953)	0.0043
	200	0.8689	0.0166	(23.3421 ; 0.419)	0.0023
	500	0.8681	0.0162	(31.3126 ; 0.3695)	0.0014
	1000	0.8672	0.0162	(28.8502 ; 0.3176)	0.001
	2000	0.8665	0.0161	(35.6687 ; 0.2659)	0.0006
$\|.\|_2$	50	0.8771	0.0176	(23.6759 ; 0.5784)	0.0062
	100	0.8703	0.0173	(21.494 ; 0.5151)	0.0039
	200	0.867	0.0173	(17.8311 ; 0.4541)	0.003
	500	0.8668	0.0165	(30.8016 ; 0.348)	0.0014
	1000	0.8667	0.016	(33.0062 ; 0.3195)	0.0008
	2000	0.8662	0.0156	(39.1335 ; 0.2431)	0.0005
$\|.\|_\infty$	50	0.8817	0.0209	(28.8429 ; 0.5458)	0.0067
	100	0.8747	0.0177	(23.7255 ; 0.4911)	0.0041
	200	0.8674	0.0176	(26.9714 ; 0.4362)	0.0024
	500	0.8676	0.0165	(29.9083 ; 0.3765)	0.0016
	1000	0.8663	0.0164	(37.9307 ; 0.2872)	0.0009
	2000	0.8655	0.0164	(41.8221 ; 0.2384)	0.0006

Table 2: Variation of $\overline{ISE}$ according to the sample size n, the norm $\|.\|_p$ and the hypothesis imposed on the smoothing parameters.

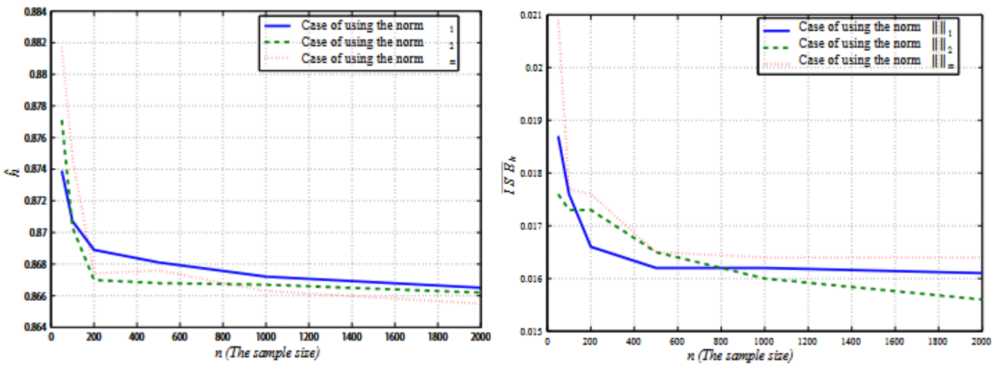

Figure 3: Variation of $\hat{h}$ and $\overline{ISE}_{\hat{h}}$ according to the sample size.

4.3. Discussion of results

Taking into account the numerical and graphical results obtained in the previous section, we note that:

- In all situations considered, the optimal smoothing parameters decrease where the sample size increases, which coincides with the following fundamental property (condition) of the smoothing parameter: $\lim_{n\to\infty} h(n) = 0$.
- Independently of the norm used, the estimators $\hat{f}_{ab}(y/E(X))$ and $\hat{f}_h(y/E(X))$ converge to $f(y/E(X))$ in L_2 (average ISE) and this can be justified by the decrease of the average ISE (convergence to zero), associated with the estimators in question, as the sample size n increases..
- The estimator $\hat{f}_{ab}(y/E(X))$ is more efficient, in the sense of the average ISE, than the estimator $\hat{f}_h(y/E(X))$ and this independently of the sample size and the norm used for the construction of these two estimators.
- The three norms used practically provide us with estimators of the same performance (average ISE). But in general, we see that there is a slight preference:
  - For the $\|.\|_2$ norm when the sample size is very small.
  - For the $\|.\|_1$ norm when the sample size is medium.
  - For the $\|.\|_2$ norm when the sample size is large.
- Because of the positivity of the distance between x and x_i ($\|x - x_i\|_p \geq 0$), the kernel must be defined on a support positive when the explanatory variable is functional. Our results show that even for a symmetric kernel we can have reasonable results.

5. Conclusion

By a small research in the literature, we can note that the conditional density when the explanatory variable is functional is a rich problem in statistics and it is in high demand in many fields of application, see Bosq, 2000, Bosq, 2000, F. Ferraty, 2006, Ramsay and Silverman, 2005. In this work, our objective was to present an illustrative numerical example of the implementation of the conditional density estimator, when the explanatory variable is functional, and we mainly focus on its performance as a function of the sample size, the hypothesis imposed on the smoothing parameters, and the norm used in its construction, and to highlight the impact of using a symmetric kernel on its quality.

Acknowledgment

The author would like to thank to express their gratitude to the anonymous reviewers for their constructive feedback and the editor for their thorough evaluation of this manuscript.

Declarations

The author declares that he has no conflicts of interest.

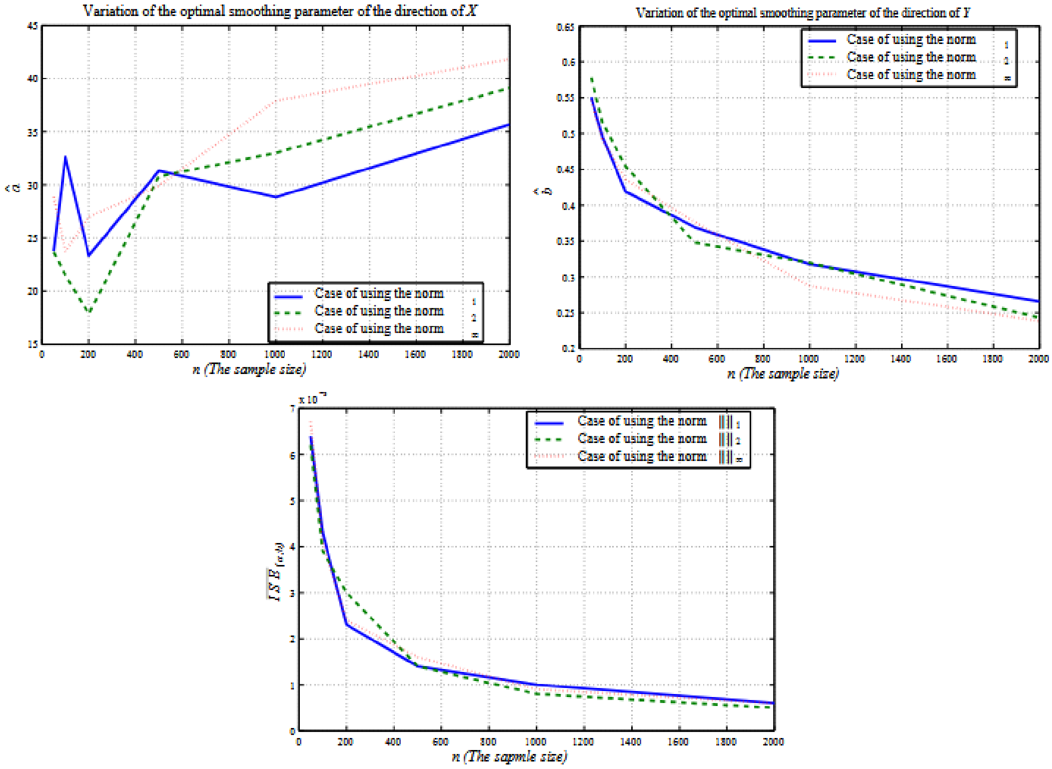

Figure 4: Variation of $(\hat{a}, \hat{b})$ and $\overline{ISE}_{(\hat{a},\hat{b})}$ according to the sample size.

References

Bashtannyk, D. M., & Hyndman, R. J. (2001). Bandwidth selection for kernel conditional density estimation. *Computational Statistics Data Analysis*, *36*(03), 279–298. https://doi.org/10.1016/S0167-9473(00)00046-3

Benhenni, K., Ferraty, F., Rachdi, M., & Vieu, P. (2007). Local smoothing regression with functional data. *Computational Statistics*, *22*(03), 353–369. https://doi.org/10.1016/S0167-9473(00)00046-3

Bosq, D. (2000). *Linear processes in function spaces: Theory and applications (vol. 149)*. Springer Science Business Media.

Chen, S. X. [N. (1999). Beta kernel estimators for density functions. *Computational Statistics Data Analysis*, *31*(2), 131–145. https://doi.org/10.1016/S0167-9473(99)00010-9

Chen, S. X. [N. (2000). Probability density function estimation using gamma kernels. *Annals of the institute of statistical mathematics*, *52*(1), 471–480. https://doi.org/10.1023/A:1004165218295

Dabo-Niang, A., S.and Laksaci. (2007). Propriétés asymptotiques d'un estimateur à noyau du mode conditionnel pour variable explicative fonctionnelle. *In Annales de l'ISUP*, *51*(3), 27–42.

Delsol, L. (2008). *Régression sur variable fonctionnelle: Estimation, tests de structure et applications (doctoral dissertation.* [Doctoral dissertation, Université Paul Sabatier-Toulouse III.].

Ezzahrioui, E. [N., M. H.and Ould-Saïd. (2008). Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. *Journal of Nonparametric Statistics.*, *20*(1), 3–18. https://doi.org/10.1080/10485250701541454

Ezzahrioui, M. (2010). On the asymptotic properties of a nonparametric estimator of the conditional mode for functional dependent data. *Statistica Neerlandika, 64*(1), 171–201.

Ezzahrioui, M., & E, O. S. (2005). *Asymptotic normality of nonparametric estimators of the conditional mode for functional data (no. 249)* (tech. rep.). Technical report.

Ferraty, A., F.and Laksaci, & Vieu, P. [N. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statistical Inference for Stochastic Processes*, *09*(1), 47–76. https://doi.org/10.1007/s11203-004-3561-3

Ferraty, A., F.and Rabhi, & Vieu, P. [N. (2008). Non-parametric estimation of the hazard function with functional explanatory variable. *REVUE ROUMAINE DE MATHEMATIQUES PURES ET APPLIQUEES*, *53*(1), 1–18. https://doi.org/10.1186/1687-1812-2012-189

Ferraty, F. (2006). *Nonparametric functional data analysis*. Springer.

Ferraty, F., Laksaci, A., Tadj, A., & Vieu, P. [N. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *Journal of Statistical planning and inference*, *140*(2), 335–352. https://doi.org/10.1016/j.jspi.2009.07.019

Härdle, W., & Marron, J. S. N. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *The Annals of Statistics*, *1985*, 1465–1481. https://doi.org/https://www.jstor.org/stable/2241365

Laksaci, A., Madani, F., & Rachdi, M. N. (2013). Kernel conditional density estimation when the regressor is valued in a semi-metric space. *Communications in Statistics-Theory and Methods*, *42*(19), 3544–3570. https://doi.org/10.1080/03610926.2011.633733

Laksaci, A., & Mechab, B. N. (2010). Estimation non paramétrique de la fonction de hasard avec variable explicative fonctionnelle: Cas des données spatiales. *Rev. Roumaine Math. Pures Appl*, *55*(1), 35–51.

Laksaci, A. N. (2007). Convergence en moyenne quadratique de l'estimateur à noyau de la densité conditionnelle avec variable explicative fonctionnelle. *In Annales de l'ISUP*, *51*(1–2), 69–80. https://doi.org/https://hal.science/hal-03635451v1

Rachdi, M., & Vieu, P. N. (2007). Nonparametric regression for functional data: Automatic smoothing parameter selection. *Journal of statistical planning and inference*, *137*(9), 2784–2801. https://doi.org/10.1016/j.jspi.2006.10.001

Ramsay, J. O., & Silverman, B. W. N. (2005). Principal components analysis for functional data. *Functional data analysis*, *2005*, 147–172. https://doi.org/10.1007/0-387-22751-2_8

Rosenblatt, M. [N. (1969). Conditional probability density and regression estimators. *Multivariate analysis II*, *25*, 31.

Rudemo, M. N. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, *1982*, 65–78. https://doi.org/10.1007/0-387-22751-2_8

Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.

Youndjé, E. N. (1996). Propriètés de convergence de l'estimateur à noyau de la densité conditionnelle. *Rev. Roumaine Math. Pures Appl*, *36*(7–8), 535–566.