Computational Statistics and Data Analysis 73 (2014) 53–68

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Theoretical and practical aspects of the quadratic error in the local linear estimation of the conditional density for functional data

Mustapha Rachdi [a,*], Ali Laksaci [b], Jacques Demongeot [c], Abdel Abdali [d], Fethi Madani [e]

[a] Univ. Grenoble-Alpes, Laboratoire AGIM FRE 3405 CNRS, Equipe GEM, Université P. Mendès France (Grenoble 2), UFR SHS, BP. 47, 38040 Grenoble Cedex 09, France

[b] Laboratoire de Statistique et Processus Stochastiques, Université Djillali Liabès, BP. 89, Sidi Bel-Abbès 22000, Algeria

[c] Univ. Grenoble-Alpes, Laboratoire AGIM FRE 3405 CNRS, Equipe GEM, Faculté de Médecine de Grenoble, Université J. Fourier, 38700 La Tronche, France

[d] Univ. Grenoble-Alpes, UFR SHS, BP. 47, Université P. Mendès France (Grenoble 2), 38040 Grenoble Cedex 09, France

[e] Université Docteur Moulay Taher, 20000 Saïda, Algeria

ARTICLE INFO

Article history:
Received 6 October 2012
Received in revised form 29 October 2013
Accepted 18 November 2013
Available online 4 December 2013

Keywords:
Functional data
Local linear estimator
Conditional density
Conditional mode
Nonparametric model
Mean squared error

ABSTRACT

The problem of the nonparametric local linear estimation of the conditional density of a scalar response variable given a random variable taking values in a semi-metric space is considered. Some theoretical and practical asymptotic properties of this estimator are established. The usefulness of the estimator is highlighted through the exact expression involved in the leading terms of the quadratic error, and by conducting a computational investigation to show the superiority of this estimation method for the conditional density and then for the conditional mode. Moreover, in order to verify the pertinence of the technique, from a practical point of view, it is applied to a real dataset.
© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The observation of the functional variables has become usual due, for instance, to the development of the measuring instruments that allow one to observe them with increasing resolution (cf. Ramsay and Silverman, 2005). Then, as technology progresses, we are able to handle larger and larger datasets. At the same time, monitoring devices such as electronic equipment and sensors (for registering images, signals, etc.) have become more and more sophisticated. This high-tech revolution offers the opportunity to observe phenomena in an increasingly accurate way, but this accuracy obviously generates a large amount of data, and indeed, due to the production of statistical units sampled over a finer grid, data can be considered as observations varying over a continuum. Such continuous data (or functional data, curves, surfaces, etc.) may occur in biomechanics (e.g., human movements), in chemometrics (e.g., spectrometric curves), in econometrics (e.g., the

* Corresponding author. Tel.: +33 476825853; fax: +33 476825665.
E-mail addresses: Mustapha.Rachdi@upmf-grenoble.fr, Mustapha.Rachdi@agim.eu (M. Rachdi), alilak@yahoo.fr (A. Laksaci), Jacques.Demongeot@agim.eu (J. Demongeot), Abdel.Abdali@upmf-grenoble.fr (A. Abdali), Fethi.Madani@agim.eu (F. Madani).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.11.011

stock market index), in geophysics (e.g., spatio-temporal events such as El Niño–Southern Oscillations or the time series of satellite images), or in medicine (e.g., electro-cardiograms/electro-encephalograms). Thus, it seems natural to assume that these data are actually observations of a random variable taking its values in a functional space.

On the other hand, it is well known that standard multivariate statistical techniques show problems in the functional setting due to the infinite-dimensional nature of the data. However, the great potential for applications, in this kind of data, has encouraged new methodologies permitting to extract relevant information from functional datasets. Now, the main statistical topics (e.g., classification, inference, factor-based analysis, regression modelling, re-sampling methods, time series, random processes, etc.) are covered when the data are functional. Notice as well that, the twin challenges about this topic are the theoretical techniques requiring both elaboration of mathematical foundations and development of statistical models, as well as the practical issues of implementing these new methodologies.

In this paper, we focus on the local polynomial modelling of the conditional density function when the explanatory variable is valued in an infinite-dimensional space. Such study is motivated by the fact that the local polynomial smoothing technique has various advantages over the kernel method, such as better bias properties (cf. for example, Chu and Marron, 1991 and Fan, 1992 for an extensive discussion on the comparison between both these methods). Moreover, as noticed by Fan and Yao (2003), the conditional density provides a very informative summary on response variables, which allows us to examine the overall shape of the conditional distribution (cf. Fan et al., 1996, Van Keilegom et al., 2001 and Youndjé, 1993 and references therein).

Recall that, in nonparametric functional statistics, the first results about the conditional distribution have been obtained in Ferraty et al. (2006). They established the almost complete convergence of the kernel estimator of the conditional density and its derivatives. Hereafter, the quadratic error of this estimator has been studied by Laksaci (2007). This in turn, gave the asymptotic expansion of the exact expression involved in the leading terms of the quadratic error of the considered estimator. Recently, Ferraty et al. (2010) established the uniform almost complete convergence of the kernel estimator for some nonparametric conditional parameters, in particular, for the conditional density function.

Since the open question (cf. Ferraty and Vieu, 2006) "*How can the local polynomial ideas be adapted to infinite-dimensional settings*?", the local linear smoothing in the functional data setting has been considered by many authors. For instance, Barrientos-Marin et al. (2010), Baìllo and Grané (2009), El Methni and Rachdi (2011) develop this smoothing local linear estimation of the regression operator for i.i.d. (independent and identically distributed) functional data. The first contributions on the local polynomial modelling of the conditional density function, in the case where the explanatory variable is functional, were considered by Demongeot et al. (2010, 2011). They established the almost-complete consistency (in both point-wise and uniform sense) of a fast functional local linear estimator of the conditional density when the explanatory variable is functional and the observations are i.i.d. This last study was extended to the dependent functional data case by Demongeot et al. (2013). In what follows, we give the convergence rate in mean square of the fast functional local linear estimator considered by Demongeot et al. (2010, 2011, 2013). The expression of this convergence rate shows the superiority of this method over the kernel method, namely in the bias terms (cf. Section 4). It should be noticed that the accuracy of our asymptotic results leads to interesting perspectives from a practical point of view, in particular, minimizing mean squared errors can lead to automatically governing bandwidth selection procedures. A considerable part of this paper therefore aims at studying the adaptability of our theoretical methodology to practice using both simulated and real data.

The organization of the rest of the paper is as follows. We begin by a succinct presentation of our model in Section 2. In Section 3, we give some notations, hypotheses and we state the main theoretical results of our study. Section 4 is devoted to some discussions and comments on our results. Then, some simulation studies on (1) the conditional density, (2) the conditional mode and (3) an application on dataset, are exposed in Section 5. In Section 6, a global conclusion and some prospects are given. Finally, summarized proofs of our results are given in the Appendix.

2. Model

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 $\mathcal{F} \times \mathbb{R}$ where $(\mathcal{F}, d)$ is a semi-metric space.

Furthermore, we assume that there exists a regular version of the conditional distribution of Y given X, which is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$, and has a twice continuously differentiable probability density function denoted by $f^{Y/X}$.

We recall that the local polynomial smoothing technique is based on the following assumption: the functional parameter to be estimated must be smooth enough to be locally well approximated by a polynomial. On the other hand, in functional statistics, there are several ways for extending the local linear ideas (cf. for instance Barrientos-Marin et al., 2010, Baìllo and Grané, 2009, El Methni and Rachdi, 2011, Demongeot et al., 2010, 2011, 2013 and the references therein). Here, we adopt the fast functional local modelling for which the regression operator $m(x) = \mathbb{E}[Y|X = x]$ is such that, in a small neighborhood of a point $x \in \mathcal{F}$, $m(\cdot) \approx a + b\beta(\cdot, x)$. Following a similar reasoning as in Fan and Gijbels (1996), the conditional density function $f^{Y/X}(x, .)$ can be viewed as a regression model with the response variable $h_H^{-1}H(h_H^{-1}(\cdot - Y))$, where H is a kernel function and $h_H = h_{H,n}$ is a sequence of positive real numbers. This consideration is motivated by the fact that, under some classical conditions:

$$\mathbb{E}[h_H^{-1}H(h_H^{-1}(y - Y))|X = x] \to f^{Y/X}(x, y) \quad \text{as } h_H \to 0.$$

Download English Version:

https://daneshyari.com/en/article/416641

Download Persian Version:

https://daneshyari.com/article/416641

Daneshyari.com