Sankhyā : The Indian Journal of Statistics
2014, Volume 76-A, Part 2, pp. 328-355

On the Local Linear Modelization of the Conditional Distribution for Functional Data

Jacques Demongeot
Université J. Fourier, Grenoble, France
Ali Laksaci
Université Djillali Liabès, Sidi Bel-Abbès, Algeria
Mustapha Rachdi
Université de Grenoble-Alpes, Grenoble Cedex 09, France
Saâdia Rahmani
Université de Saida, Saida, Alegria

Abstract

In this paper,we investigate the problem of the local linear estimation of the cumulative distribution function of a real random variable Y conditioned by a functional variable X (valued in an infinite dimensional space). The almost-complete and the mean square consistencies, with rates, of the constructed estimator are stated. We precise that the exact expression involved in the leading terms of the mean squared error is given. We point out, also, that the accuracy of our asymptotic results leads to interesting perspectives from a practical point of view. Thus, we discuss the features of our functional local modeling and the applicability of our asymptotic result on some statistical problems such as the choice of the smoothing parameters and the determination of confidence intervals. Moreover, a simulation study has been conducted in order to highlight, on a finite sample, the superiority of our method to the standard kernel method, in the functional framework.

AMS (2000) *subject classification.* Primary 62G05; Secondary 62G07, 62G08, 62G35, 62G20.
Keywords and phrases. Functional data, local linear estimator, conditional distribution, nonparametric model, small balls probability.

1 Introduction

Consider n pairs of independent random variables (X_i, Y_i) for $i = 1, \ldots, n$ that we assume drawn from the pair (X, Y) which is valued in $\mathfrak{F} \times \mathbb{R}$, where $\mathfrak{F}$ is a semi-metric space equipped with a semi-metric d. For $x \in \mathfrak{F}$, we assume that there exists a regular version of the conditional probability of

Y given $X = x$, which is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$. Our purpose is to study the local linear estimation of the conditional cumulative distribution function of Y given $X = x$, which we will denote by $F^x(\cdot)$.

Recall that, as indicated in Fan and Gijbels (1996), the function $F^x(\cdot)$ can be viewed as a regression model with the response variable $H(h_H^{-1}(\cdot - Y_i))$, where H is some cumulative distribution function and $(h_H = h_{H,n})$ is a sequence of positive real numbers. This consideration is motivated by the fact that:

$$\mathbb{E}[H(h_H^{-1}(y - Y_i))|X_i = x] \to F^x(y) \text{ as } h_H \to 0$$

In our functional context, we combine this idea with studies presented in Barrientos-Marin *et al.* (2010), Demongeot *et al.* (2011), Demongeot *et al.* (2013), which are dedicated to the nonparametric functional regression estimation, to construct a functional local linear estimator of the conditional distribution function, based on the minimization, with respect to the pair (a, b), of the following criterion:

$$\sum_{i=1}^{n} \left(H(h_H^{-1}(y - Y_i)) - a - b\beta(X_i, x)\right)^2 K(h_K^{-1}\delta(x, X_i)) \tag{1}$$

where $\beta(.,.)$ and $\delta(.,.)$ are known bi-functional operators defined from $\mathfrak{F}^2$ into $\mathbb{R}$, K is a kernel and $h_K = h_{K,n}$ (resp. $h_H = h_{H,n}$) is chosen as a sequence of positive real numbers. More precisely, the functional local linear estimator $\widehat{F}^x(y)$ of $F^x(y)$ is then $\widehat{a}$ which is the first component of the pair (a, b) solution of the minimization problem (1). However, if the bi-functional operator β is such that, $\forall z \in \mathfrak{F}$, $\beta(z, z) = 0$, then the quantity $\widehat{F}^x(y)$ is explicitly defined by:

$$\widehat{F}^x(y) = \frac{\sum_{i,j=1}^{n} W_{ij} H(h_H^{-1}(y - Y_j))}{\sum_{i,j=1}^{n} W_{ij}} \tag{2}$$

where $W_{ij} = \beta(X_i, x)\left(\beta(X_i, x) - \beta(X_j, x)\right) K(h_K^{-1}\delta(x, X_i))K(h_K^{-1}\delta(x, X_j))$, with the convention $0/0 = 0$.

Notice that, questions on the statistical modelization of the functional data have known a growing interest among theoretical and applied statistics (cf. Bosq, 2000; Ramsay and Silverman, 2002a, 2002b for parametric models and Ferraty and Vieu (2006) for the nonparametric case). Recall also that, in functional statistics, the estimation of the conditional cumulative distribution function has great importance. In fact, it is involved in many applications, such as reliability, survival analysis, ... Moreover, there are