Journal of Multivariate Analysis 146 (2016) 105–118

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Adaptive estimation in the functional nonparametric regression model

Gaëlle Chagny [a], Angelina Roche [b,*]

[a] LMRS, UMR CNRS 6085, Université de Rouen, France
[b] MAP5 UMR CNRS 8145, Université Paris Descartes, France

ARTICLE INFO

Article history:
Received 29 December 2014
Available online 22 July 2015

AMS 2010 subject classifications:
62G08
62H12

Keywords:
Functional data analysis
Regression estimation
Nonparametric kernel estimators
Bandwidth selection

ABSTRACT

In this paper, we consider nonparametric regression estimation when the predictor is a functional random variable (typically a curve) and the response is scalar. Starting from a classical collection of kernel estimates, the bias–variance decomposition of a pointwise risk is investigated to understand what can be expected at best from adaptive estimation. We propose a fully data-driven local bandwidth selection rule in the spirit of the Goldenshluger and Lepski method. The main result is a nonasymptotic risk bound which shows the optimality of our tuned estimator from the oracle point of view. Convergence rates are also derived for regression functions belonging to Hölder spaces and under various assumptions on the rate of decay of the small ball probability of the explanatory variable. A simulation study also illustrates the good practical performances of our estimator.

1. Introduction

In many fields, more and more collected data can nowadays be considered as curves, thanks to advanced recording. Developing statistics for Functional Data Analysis (F.D.A.) is thus a great challenge as evidenced by the recent monography of Horváth and Kokoszka [32] and the increasing number of contributed books on this subject (see e.g. [13]). Among all the methods explored to deal with such complex observations, nonparametric statistics have been widely investigated (see [21] for a clear account). The subject of this work is one of the most studied problem, nonparametric regression estimation with functional covariate. Let

$$Y = m(X) + \varepsilon, \tag{1}$$

where Y is a real random variable, X a random variable which takes values in a separable infinite-dimensional Hilbert space $(\mathbb{H}, \langle\cdot,\cdot\rangle, \|\cdot\|)$ (it can be $\mathbb{L}^2(I)$, the set of squared-integrable functions on a subset I of $\mathbb{R}$, or a Sobolev space), and $m : \mathbb{H} \to \mathbb{R}$ the target function to recover. The random variable ε stands for a noise term. We suppose that ε and X are independent and that ε is centered with $\mathbb{E}[\varepsilon^2]^{1/2} = \sigma < \infty$. The specificity thus stands in the dimension which is infinite in two aspects: first, the framework is a functional one (the covariate X is a stochastic process which lives in an infinite-dimensional space), and then, no assumption is made on the shape of the function m to estimate. We assume that we observe a data sample $\{(X_i, Y_i), i = 1, \ldots, n\}$ distributed like the couple (X, Y), that is $Y_i = m(X_i) + \varepsilon_i$, with ε_i's independent identically distributed (*i.i.d.* in the sequel) like ε, and independent from the X_i's.

* Corresponding author.
E-mail addresses: gaelle.chagny@gmail.com (G. Chagny), angelina.roche@parisdescartes.fr (A. Roche).

http://dx.doi.org/10.1016/j.jmva.2015.07.001
0047-259X/

The aim of this contribution is to provide an adaptive optimal strategy to estimate the regression function m. We explore the pointwise risk of a collection of kernel estimates to define a data-driven bandwidth selection method. The resulting estimator is shown to be optimal in a nonasymptotic way.

The starting point of our work is the following collection of Nadaraya–Watson-type estimators. Let $K : \mathbb{R} \to \mathbb{R}$ be a kernel function (that is $\int_{\mathbb{R}} K(u)du = 1$), $K_h : u \mapsto K(u/h)/h$, for $h > 0$, and

$$\widehat{m}_h(x) := \sum_{i=1}^{n} W_h^{(i)}(x) Y_i \quad \text{where } W_h^{(i)}(x) := \frac{K_h(\|X_i - x\|)}{\sum_{j=1}^{n} K_h(\|X_j - x\|)}, \tag{2}$$

for any $(x, y) \in \mathbb{H} \times \mathbb{R}$. These estimators have first been introduced in the functional context by Ferraty and Vieu [19] who have then widely studied their asymptotic properties: almost complete convergence, asymptotic normality, asymptotic expansions for the mean integrated squared error (see [24] and references therein). A clear account about convergence rates (upper and lower bounds for the pointwise risk of (2) under concentration assumptions for the process X, as well as minimax rates) is provided by Mas [38]. This kind of estimates also gives rise to other kernel-based strategies to recover m in the same functional context. Local linear methods in the spirit of local polynomial estimators have been explored by Baíllo and Grané [7], Barrientos-Marin et al. [8], Boj et al. [12] and Berlinet et al. [10], k-nearest neighbors methods are examined by Kudraszow and Vieu [33], a recursive kernel approach is suggested by Amiri et al. [2] and Reproducing Kernel Hilbert Space-based methods are reported by Avery et al. [6].

All kernel-based strategies have one point in common: they heavily depend on the choice of the smoothing parameter h (see (2)), the so-called bandwidth. Heuristically, a large value for h leads to an estimator with large bias (but small variance), while a too small value leads to high variability. The most commonly used method to select the bandwidth of a functional regression kernel estimate is the leave-one-out cross validation. The first algorithm is proposed by Ferraty and Vieu [20], and shown to be asymptotically optimal by Rachdi and Vieu [39]. It is a global method that do not depend on the point (curve) of estimation. A local version, also optimal in an asymptotic way, is proposed by Benhenni et al. [9]. More recently, Bayesian strategies have been studied [41,42], but only for simulation purposes.

The question which motivate our study is the following: for functional regression estimation in model (1), how to define a bandwidth selection criterion which is proved to automatically balance the bias–variance trade-off in a nonasymptotic way? The bandwidth has to be chosen for application purpose, but the criterion should also be theoretically justified, with nonasymptotic risk bounds.

While the same question is the subject of a wide literature in classical nonparametric estimation, only two papers, to our knowledge, focus on this problem in the framework of F.D.A., and none of them are related to functional regression. First, a method based on empirical prediction-risk minimization is explored by Antoniadis et al. [4]: oracle inequalities are proved, but the criterion is specific to functional time series prediction. Then, inspired both by the advances about the Lepski methods [29] and model selection, Chagny and Roche [17] investigate a global bandwidth selection method to estimate a cumulative distribution function (c.d.f. in the sequel) conditionally to a functional covariate, shown to be optimal both in the oracle and minimax sense, for an integrated error.

The contribution of this work is the following. We build a bandwidth selection device, for the estimators (2) in functional regression in such a way that the resulting estimator is optimal in a nonasymptotic point of view. Since Benhenni et al. [9] observe that a local choice can improve significantly the precision of estimation in a functional context, the criterion is defined in a local way, at a fixed curve. To be more precise, our study is based on the pointwise risk: for x_0 a fixed point in $\mathbb{H}$, the risk of an estimator $\widehat{m}(x_0)$ computed at the curve x_0, is in the sequel,

$$\mathbb{E}\left[(\widehat{m}(x_0) - m(x_0))^2\right].$$

We obtain an exact bias–variance decomposition (Proposition 1) which permits to understand what we can expect at best from adaptive estimation, which is the subject of Section 3. The bandwidth selection, at a fixed curve, is automatically performed in the spirit of Goldenshluger and Lepski [29]. The resulting estimator achieves the same performance as the one which would have been selected if the regularity index of the target function had been known, up to a constant and to a logarithm factor, as it is proved in our main result, Theorem 2. The result holds whatever the sample size. Convergence rates are also deduced in Section 3.4 for functions m belonging to Hölder spaces, and under various concentration assumptions on the process X. These assumptions are on the rate of decay of the small ball probability

$$\varphi^{x_0} : h > 0 \mapsto \mathbb{P}(\|X - x_0\| \leq h), \quad x_0 \in \mathbb{H},$$

which influences the rate, as usual. The faster the small ball probability decreases, the smaller the rate of convergence of the estimator is. The rates (Proposition 3) are quite slow when X really take values in an infinite-dimensional space. They nevertheless match with the lower bounds of Mas [38]. This is the so-called "curse of dimensionality". Practical issues are discussed in Section 4 and the performances of our bandwidth selection criterion are compared with cross-validated criteria and k-nearest neighbors method. Finally, the main steps of the proofs are gathered in Section 6. A supplementary material (see Appendix A) is available with further simulation results, as well as detailed proofs for the main results.

One can notice that similar questions are behind the previous study of Chagny and Roche [17]. The same kind of methodology has been applied, leading to similar nonasymptotic upper-bounds to estimate a conditional c.d.f. However, the