Statistics and Probability Letters 113 (2016) 7–15

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Estimation of the noise covariance operator in functional linear regression with functional outputs

Christophe Crambes [a,*], Nadine Hilgert [b], Tito Manrique [a,b]

[a] UMR 5149 IMAG, Université de Montpellier, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France
[b] UMR 0729 MISTEA, INRA, 2 Place Pierre Viala, 34060 Montpellier Cedex 2, France

ARTICLE INFO

Article history:
Received 30 June 2015
Received in revised form 2 February 2016
Accepted 11 February 2016
Available online 20 February 2016

Keywords:
Functional linear regression
Functional response
Noise covariance operator

ABSTRACT

This work deals with the estimation of the noise in functional linear regression when both the response and the covariate are functional. Namely, we propose two estimators of the covariance operator of the noise. We give some asymptotic properties of these estimators, and we study their behavior on simulations.

1. Introduction

We consider the following functional linear regression model where the functional output $Y(\cdot)$ is related to a random function $X(\cdot)$ through

$$Y(t) = \int_0^1 \mathcal{S}(t,s)X(s)ds + \varepsilon(t). \quad (1)$$

Here $\mathcal{S}(\cdot,\cdot)$ is an unknown integrable kernel: $\int_0^1\int_0^1 |\mathcal{S}(t,s)|dtds < \infty$, to be estimated. ε is a noise random variable, independent of X. The functional variables X, Y and ε are random functions taking values on the interval $I = [0, 1]$ of $\mathbb{R}$. Considering this particular interval is equivalent to considering any other interval $[a, b]$ in what follows. For the sake of clarity, we assume moreover that the random functions X and ε are centered. The case of non centered X and Y functions can be equivalently studied by adding an additive non random intercept function in model (1).

In all the sequel we consider a sample $(X_i, Y_i)_{i=1,\ldots,n}$ of independent and identically distributed observations, following (1) and taking values in the same Hilbert space $H = \mathbb{L}^2([0, 1])$, the space of all real valued square integrable functions defined on $[0, 1]$. The objective of this paper is to estimate the unknown noise covariance operator Γ_ε of ε and its trace $\sigma_\varepsilon^2 := tr(\Gamma_\varepsilon)$ from these data sets. The estimation of the noise covariance operator Γ_ε is well known in the context of multivariate multiple regression models, see for example Johnson and Wichern (2007, section 7.7). The question is a little more tricky in the context of functional data. Answering it will then make possible the construction of hypothesis testing in connection with model (1).

Functional data analysis has given rise to many theoretical results applied in various domains: economics, biology, finance, etc. The monograph by Ramsay and Silverman (2005) is a major reference that gives an overview on the subject

* Corresponding author.
E-mail address: christophe.crambes@univ-montp2.fr (C. Crambes).

http://dx.doi.org/10.1016/j.spl.2016.02.006

and highlights the drawbacks of considering a multivariate point of view. Novel asymptotic developments and illustrations on simulated and real data sets are also provided in Horváth and Kokoszka (2012). We follow here the approach of Crambes and Mas (2013) that studied the prediction in the model (1) revisited as:

$$Y = SX + \varepsilon, \tag{2}$$

where $S : H \to H$ is a general linear integral operator defined by $S(f)(t) = \int_0^1 \mathcal{S}(t,s)f(s)ds$ for any function f in H. The authors showed that the trace σ_ε^2 is an important constant involved in the square prediction error bound that participate to determine the convergence rate. The estimation of σ_ε^2 will thus provide details on the prediction quality in model (1).

In this context of functional linear regression, it is well known that the covariance operator of X cannot be inverted directly (see Cardot et al., 1999), thus a regularization is needed. In Crambes and Mas (2013), it is based on the Karhunen–Loève expansion and the functional principal component analysis of the (X_i). This approach is also often used in functional linear models with scalar output, see for example Cardot et al. (1999).

The construction of the estimator $\hat{S}$ is introduced in Section 2. Section 3 is devoted to the estimation of Γ_ε and its trace. Two types of estimators are given. Convergence properties are established and discussed. The proofs are postponed in Section 5. The results are illustrated on simulation trials in Section 4.

2. Estimation of S

2.1. Preliminaries

We denote respectively $\langle \cdot, \cdot \rangle_H$ and $\|.\|_H$ the inner product and the corresponding norm in the Hilbert space H. We shall recall that $\langle f, g\rangle_H = \int_0^1 f(t)g(t)dt$, for all functions f and g in $\mathbb{L}^2([0,1])$. In contrast, $\langle \cdot, \cdot \rangle_n$ and $\|.\|_n$ stand for the inner product and the Euclidean norm in $\mathbb{R}^n$. The tensor product is denoted $\otimes$ and defined by $f \otimes g = \langle g, . \rangle_H f$ for any functions $f, g \in H$.

We assume that X and ε have a second moment, that is: $\mathbb{E}[\|X\|_H^2] < \infty$ and $\mathbb{E}[\|\varepsilon\|_H^2] < \infty$. The covariance operator of X is the linear operator defined on H as follows: $\Gamma := \mathbb{E}[X \otimes X]$. The cross covariance operator of X and Y is defined as $\Delta := \mathbb{E}[Y \otimes X]$. The empirical counterparts of these operators are: $\hat{\Gamma}_n := \frac{1}{n}\sum_{i=1}^n X_i \otimes X_i$ and $\hat{\Delta}_n := \frac{1}{n}\sum_{i=1}^n Y_i \otimes X_i$.

An objective of the paper is to study the trace σ_ε^2. We thus introduce the nuclear norm defined by $\|A\|_{\mathcal{N}} = \sum_{j=1}^{+\infty} |\mu_j|$, for any operator A such that $\sum_{j=1}^{+\infty} |\mu_j| < +\infty$ where $(\mu_j)_{j\geq 1}$ is the sequence of the eigenvalues of A. We denote $\|.\|_\infty$ the operator norm defined by $\|A\|_\infty = \sup_{\|u\|=1} \|Au\|$.

2.2. Spectral decomposition of Γ

It is well known that Γ is a symmetric, positive trace-class operator, and thus diagonalizable in an orthonormal basis (see for instance Hsing and Eubank, 2015). Let $(\lambda_j)_{j\geq 1}$ be its non-increasing sequence of eigenvalues, and $(v_j)_{j\geq 1}$ the corresponding eigenfunctions in H. Then Γ decomposes as follows:

$$\Gamma = \sum_{j=1}^{\infty} \lambda_j v_j \otimes v_j,$$

For any integer k, we define $\Pi_k := \sum_{j=1}^k v_j \otimes v_j$ the projection operator on the sub-space $\langle v_1, \dots, v_k \rangle$. By projecting Γ on this sub-space, we get :

$$\Gamma|_{\langle v_1,\dots,v_k\rangle} := \Gamma\Pi_k = \sum_{j=1}^{k} \lambda_j v_j \otimes v_j.$$

2.3. Construction of the estimator of S

We start from the moment equation

$$\Delta = S\Gamma. \tag{3}$$

On the sub-space $\langle v_1, \dots, v_k\rangle$, the operator Γ is invertible, more precisely $(\Gamma\Pi_k)^{-1} = \sum_{j=1}^k \lambda_j^{-1} v_j \otimes v_j$. As a consequence, with Eq. (3) and the fact that $\Pi_k\Gamma\Pi_k = \Gamma\Pi_k$ we get, on the sub-space $\langle v_1, \dots, v_k\rangle$, $\Delta\Pi_k = (S\Pi_k)(\Gamma\Pi_k)$. We deduce that $S\Pi_k = \Delta\Pi_k(\Gamma\Pi_k)^{-1}$.

Now, taking $k = k_n$, denoting $\hat{\Pi}_{k_n} := \sum_{j=1}^{k_n} \hat{v}_j \otimes \hat{v}_j$ and the generalized inverse $\hat{\Gamma}_{k_n}^{+} := (\hat{\Gamma}_n\hat{\Pi}_{k_n})^{-1}$, we are able to define the estimator of S. We have

$$\hat{\Gamma}_n = \sum_{j=1}^{\infty} \hat{\lambda}_j \hat{v}_j \otimes \hat{v}_j = \sum_{j=1}^{n} \hat{\lambda}_j \hat{v}_j \otimes \hat{v}_j,$$