PLS regression for multivariate functional data

Cristian Preda[a*] and Gilbert Saporta [b]

[a]Université Lille 1, Villeneuve d'Ascq, France
[b]CNAM, Paris, France

Keywords: multivariate functional data, PLS, regression

Introduction

Functional data or data represented by curves, is generally considered as sample paths of a real-valued stochastic process with continuous time, $X = \{X_t\}_{t\in[0,T]}$. Most of the approaches dealing with functional data consider the univariate case, i.e. $X(t) \in \mathbb{R}$, $\forall t \in [0,T]$, a path of X being represented by a single curve. Despite its evident interest, the multivariate case,

$$\mathbf{X}(t) = (X_1(t), \ldots, X_p(t)) \in \mathbb{R}^p, \quad p \geq 2$$

is, curiously, rarely considered in literature. In this case a path of $\mathbf{X}$ is represented by a set of p curves. The dependency between the p measures provides the structure of $\mathbf{X}$. One finds in [1] a brief example of bi-dimensional functional data, $\mathbf{X}(t) = (X_1(t), X_2(t)) \in \mathbb{R}^2$, as a model for gait data (knee and hip measures) used in the context of functional principal analysis as an extension of the univariate case. For a more theoretical framework, we must go back to the pioneer works of [2] on random variables with values into a general Hilbert space. In [3]) the author provides a complete analysis of multivariate functional data from the point of view of factorial methods (principal components and canonical analysis). Recently, [4] considered model-based clustering for multivariate functional data and [5] introduced linear tools, similar to principal component analysis, for analysing such data.

In this paper we consider the linear regression model with multivariate functional random variable predictor and vectorial response,

$$\mathbb{E}(\mathbf{Y}|\mathbf{X} = \mathbf{x}) = \int_0^T \sum_{i=1}^{p} \beta_i(t) x_i(t) dt, \quad \mathbf{Y} \in \mathbb{R}^q, \beta_i \in (L_2([0,T]))^q, \forall i = 1, \ldots p. \tag{1}$$

As an extension of the PLS approach for the functional linear regression model proposed in [6] for univariate functional data, we develop the PLS estimation in the case of multivariate functional predictor. The Tucker criterion provides the PLS components as eigen-vectors of the product of the Escoufier's operators associated to the response and the predictor.

We present the PLS estimation when the predictor is approximated in a finite dimensional space of functions. A simulation study illustrates our methodology.

References

[1] J. O. Ramsay and B. W. Silverman, *Functional data analysis*, Springer, New York, second ed., 2005.

[2] P. Besse, "tude descriptive d'un processus," *Thèse de doctorat* $3^{ème}$ *cycle* **Université Paul Sabatier, Toulouse.**, 1979.

*Corresponding author. E-mail: cristian.preda@polytech-lille.fr, Université de Lille 1, Villeneuve d'Ascq, France.

[3] G. Saporta, “Mthodes exploratoires d’analyse de donnes temporelles,” *Cahiers du Buro* **37–38**, 1981.

[4] J. Jacques and C. Preda, “Model-based clustering of multivariate functional data,” *Computational Statistics and Data Analysis* **71**, 2014.

[5] J.-M. Chiou and H.-G. Muller, “Linear manifold modelling of multivariate functional data,” *Journal of the Royal Statistical Society: Series B* **DOI: 10.1111/rssb.12038**, 2013.

[6] C. Preda and G. Saporta, “PLS regression on a stochastic process,” *Computational Statistics and Data Analysis* **49**, 2005.