CrossMark

Comments on: Probability enhanced effective dimension reduction for classifying sparse functional data

Germán Aneiros[1] · Philippe Vieu[2]

1 A short historical discussion on sparsity

Sparsity is a phenomenon which has been highly discussed in the statistical literature along the last three decades when studying multivariate data. It turns out that a sample of n observations coming from a variable lying into some p-dimensional space is more and more sparsely distributed as the dimension p increases. At least two fields of Multivariate Statistics have born to face this problem. The first situation arises when p is much higher than the sample size n itself. In that setting, sparsity appears through the structure of the statistical model, and variable selection procedures play a main role (see Bühlmann and Geer 2011 for general discussion). The second situation is related to Nonparametric Statistics, where the sparseness effect leads to exponential deterioration of rates of convergence of any estimates (see Stone 1982). This fact motivated along the 1990s the development of semi-parametric statistics (see Härdle et al. 2004 for general discussion).

Going with that, in the last two decades, there has been an increasing interest for data being of continuous nature. In fact, because of both the wide scope of applied problems related to that kind of data and the various methodological challenges that the mathematical community has to face, functional data analysis (FDA) became

✉ Germán Aneiros
ganeiros@udc.es

Philippe Vieu
philippe.vieu@math.univ-toulouse.fr

[1] Universidad de A Coruña, A Coruña, Spain

[2] Institut de Mathématiques de Toulouse, Toulouse, France

Springer

a very active field of modern Statistics. In their paper, Professors Yao, Wu, and Zou mention various general monographs which have contributed to popularizing FDA (see Ramsay and Silverman 2002, 2005; Ferraty and Vieu 2006; Horváth and Kokoszka 2012). We would like to complete that by emphasizing on the various Special Issues that have been devoted in the last few years to FDA, as well on applied journals like *Computational Statistics and Data Analysis* (see González-Manteiga and Vieu 2007) or *Computational Statistics* (see Valderrama 2007) as in more theoretical ones like *Statistica Sinica* (see Davidian et al. 2004) or *Journal of Multivariate Analysis* (see Goia and Vieu 2015 for the latest one).

Because FDA is aiming to develop statistical methods for infinite-dimensional variables and because sparsity is a phenomenon linked with high-dimensional data, both topics (FDA and sparsity) are naturally crossing quite often. As discussed before for multivariate data, sparsity may appear both on the concentration of the data into the functional space and on the model. Moreover, because each functional datum is in reality observed on a finite grid, sparsity may also occur in a new way in FDA through the number of observed discretized points. The literature on this kind of sparsity phenomenon is rather developed (see for instance James et al. 2000; Yao et al. 2005; Müller and Yang 2010 for general discussions) and the authors Professors Yao, Wu, and Zou bring a new and interesting contribution.

In the following, we will make some comments on these three ways where sparsity phenomenon is crossing FDA (see Sect. 2) and will discuss some possible extensions of the new probability-based methodology for dimensionality reduction proposed by the authors (see Sect. 3) including, but not only, extensions towards the other notions of sparsity discussed in Sect. 2.

2 Sparseness and FDA

Usually, a functional dataset can be seen as a finite sample

$$\chi_1, \ldots, \chi_n,$$

coming from a random variable χ valued in some infinite-dimensional space E. In practice, each functional object χ_i is only observed in a finite number of points $\{t_i^j, \ j = 1, \ldots, n_i\}$. Thus, the observed data are

$$\chi_i^j = \chi_i(t_i^j), \quad i = 1, \ldots, n, \ j = 1, \ldots, n_i.$$

In the remainder of this section, we present some comments, and give some related bibliography, on the three types of sparsity that can appear when one deals with functional data.

(i) Sparsity in the functional space If E were finite dimensional (for instance, $E = R^p$ with euclidean topological structure), the sparsity phenomenon is usually highlighted by the fact that the small ball probability function decreases dramatically with the dimension p:

$$P(\chi \in \mathcal{B}(\chi_0, \epsilon)) \sim c\epsilon^p, \ \text{as } \epsilon \to 0,$$

where χ_0 is a fixed element in E and $\mathcal{B}(\chi_0, \epsilon)$ denotes the ball with center χ_0 and radius $\epsilon > 0$. As a consequence, nonparametric estimates have low rates of convergence of the form $n^{-\delta/(2\delta+p)}$ (see for instance Stone 1982 for discussion and examples of values of δ). In the functional setting, this problem is even more crucial since (for functional space E endorsed with usual Hilbert or Banach topology) the small ball probability is exponentially decreasing:

$$P(\chi \in \mathcal{B}(\chi_0, \epsilon)) \sim c e^{-C\epsilon^{-\gamma}}, \text{ as } \epsilon \to 0,$$

(see Li and Shao 2001 for discussion and examples) leading to nonparametric estimates having very poor rates of convergence of the form $(\log n)^{-\delta}$ (see Ferraty and Vieu 2006 for examples). The statistical community has developed various techniques to overcome this difficulty, the most popular being nonparametric statistics with other topological structures than Banach one (see Ferraty and Vieu 2006), functional adaptation of semi-parametric ideas like single index as in Ferraty et al. 2003 or Chen et al. 2011, projection pursuit as in Ferraty et al. 2013 or Chen et al. 2011, partial linear modeling as in Aneiros and Vieu 2006 or Lian 2011 (see Goia and Vieu 2014 for a short survey on semi-parametric functional statistics), or more generally, functional additive modeling as in Ferraty and Vieu 2009 or Müller and Yao 2008. The paper by Professors Yao, Wu, and Zou contains also a nicely commented bibliography on the topic.

(ii) Sparsity in the model An alternative way for controlling the dimensionality of the data consists in introducing sparsity into the statistical model. Roughly speaking, it consists in building statistical models using only a few number among the discretized variables $\chi_i(t_i^j)$ rather than the whole matrix $(\chi_i(t_i^j))_{i=1,\ldots,n, j=1,\ldots,n_i}$ or the continuous dataset $\{\chi_i, \ i = 1, \ldots n\}$. This approach is particularly adapted to situations where the number of discretized points is large and only a few of them are expected to be impact points. This question has been studied recently in the regression setting, mainly under linear assumptions as for instance in McKeague and Sen (2010), Kneip and Sarda (2011) or Aneiros and Vieu (2014) (see Aneiros and Vieu 2015a, b, respectively, for semi- and nonparametric extensions).

(iii) Sparsity in the grids While two previous notions of sparsity are strongly related with phenomenons arising also in multivariate statistics, the third one is specific to the functional setting. It concerns situations which are almost completely opposed to the one described in point ii) before in the sense that it has to front with datasets for which the number n_i of discretized observations of each functional element χ_i is small (and sometimes, in addition, the data are irregularly distributed). One can find a wide bibliographical discussion on this sparsely sampled curves situation through the paper by Professors Yao, Wu, and Zou, which presents also an innovative contribution to this field in the specific case when one has to deal with a curves classification problem. More specifically, these authors propose to construct an effective dimension reduction (EDR) space adapted to classification problems related to sparse functional data (we refer to sparsity in the grids). For that, to avoid the homogeneity in the binary response, they suggest to construct the EDR space by slicing the data based on conditional class

probabilities of observations rather than on the response itself. This idea is supported from a theoretical point of view.

3 Discussion and tracks

We begin this section by showing some concerns on the procedure proposed by Professors Yao, Wu, and Zou to estimate the EDR space. To estimate such space, estimates for both the unconditional mean $m(\cdot, \cdot)$ and the covariance operator $\sum$ (see Sect. 3) are needed, and the authors propose to apply local linear smoothing on the pooled data; that is, the sample to use consists of the discretized data from the whole set of curves. Specifically, the proposed estimators assign the same weight to each observation, in such a way that the weight received by a subject i increases as the number of observations n_i does; see Yao et al. (2005) for a theoretical study of those estimators when applied on sparse data. Actually, in the statistical literature there exist other proposals to estimate $m(\cdot, \cdot)$ and $\sum$, which present a nice asymptotic behavior (including optimal uniform convergence rates); we refer to the estimators studied in Li and Hsing (2010), where the same weight is assigned to each subject. It is worth to be noted that, apart from their optimality, another advantage of the estimators suggested in Li and Hsing (2010) is the fact that their use is justified under both sparse and dense data (actually, in practice, not always clear what kind of data are handled). We would like to know the feelings of the authors about to consider in their proposal for estimating the EDR space the estimates given in Li and Hsing (2010) instead of those studied in Yao et al. (2005). Our second concern is related to the bandwidth selector used in the nonparametric estimates. Specifically, the authors suggest to select the bandwidth parameters ignoring the within-curve correlation, this recommendation being based on Theorem 2 in Lin and Carroll (2000). Actually, it is not clear that such result is applicable in the setting of the paper by Professors Yao, Wu, and Zou. Thus, we consider that some additional discussion on bandwidth selection would be welcome.

One good measuring for the scientific impact of a paper is not only to look at the specific problem which is solved but also to the new research lines that it is opening. Clearly this paper belongs to this category, and overpassing the new method proposed by Professors Yao, Wu, and Zou for dealing with sparsely sampled curves discrimination, our feeling is that the main ideas of this paper will be used in the future in many other situations. We would like to point a few of them.

The first natural question emerging from this paper is about the possible links with other sparsity situations as described before in Sect. 2. Of course, although it might seem surprising because of linguistic similarities (since both are using some sparsity ideas), situations ii) and iii) are concerned with two different kinds of data and their crossing is certainly unrealistic. In counterpart, crossing situations i) and iii) seems much more interesting. Why would not it be possible to use your curves classification algorithm in a fully nonparametric model, that is, without any kind of linearity assumptions?

An other natural idea, which is sparsely discussed along the bibliographical parts of the paper, concerns other kinds of problems than classification. For instance, one could obviously be interested in using your probability-enhanced ideas in situations where

the response is more complicated than a simple binary one, as in regression problems with scalar (or even functional) response. Our feeling is that such an extension (at least for scalar response) would not be too difficult. One could also have in mind unsupervised classification situations, but adaptation to this case requires certainly much more developments.

References

Aneiros G, Vieu P (2006) Semi-functional partial linear regression. Stat Probab Lett 76:1102–1110

Aneiros G, Vieu P (2014) Variable selection in infinite-dimensional problems. Stat Probab Lett 94:12–20

Aneiros G, Vieu P (2015a) Partial linear modelling with multi-functional covariates. Stat Comput. doi:10.1007/s00180-015-0568-8

Aneiros G, Vieu P (2015b) Sparse nonparametric model for regression with functional covariate (preprint)

Bühlmann P, van de Geer S (2011) Statistics for high-dimensional data. Methods, theory and applications. Springer Series in Statistics. Springer, Heidelberg

Chen D, Hall P, Müller H-G (2011) Single and multiple index functional regression models with nonparametric link. Ann Stat 39:1720–1747

Davidian M, Lin X, Wang J-L (2004) Introduction to emerging issues in longitudinal and functional data analysis. Stat Sinica 14:613–614

Ferraty F, Peuch A, Vieu P (2003) Modèle à indice fonctionnel simple. Comptes Rendus Math Académie Sciences Paris 336:1025–1028

Ferraty F, Goia A, Salinelli E, Vieu P (2013) Functional projection pursuit regression. Test 22:293–320

Ferraty F, Vieu P (2006) Nonparametric functional data analysis. Springer Series in Statistics. Springer, New York

Ferraty F, Vieu P (2009) Additive prediction and boosting for functional data. Comput Stat Data Anal 53:1400–1413

Goia A, Vieu P (2015) An introduction to recent advances in high/infinite dimensional. Stat J Multivar Anal (forthcoming)

Goia A, Vieu P (2014) Some advances in semiparametric functional data modelling. Contributions in infinite-dimensional statistics and related topics. Esculapio, Bologna, pp 135–141

González-Manteiga W, Vieu P (2007) Statistics for functional data. Comput Stat Data Anal 51:4788–4792

Härdle W, Müller M, Sperlich S, Werwatz A (2004) Nonparametric and semiparametric models. Springer Series in Statistics. Springer, New York

Horváth L, Kokoszka P (2012) Inference for functional data with applications. Springer Series in Statistics. Springer, New York

James G, Hastie T, Sugar C (2000) Principal component models for sparse functional data. Biometrika 7:587–602

Kneip A, Sarda P (2011) Factor models and variable selection in high-dimensional regression analysis. Ann Stat 39:2410–2447

Li W, Shao Q-M (2001) Gaussian processes: inequalities, small ball probabilities and applications. In: Stochastic processes: theory and methods, Handbook of Statist, vol 19. North-Holland, Amsterdam, pp 533–597

Li Y, Hsing T (2010) Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. Ann Stat 38:3321–3351

Lian H (2011) Functional partial linear model. J Nonparametric Stat 23:115–128

Lin X, Carroll RJ (2000) Nonparametric function estimation for clustered data when the predictor is measured without/with error. J Am Stat Assoc 95:520–534

McKeague I, Sen B (2010) Fractals with point impact in functional linear regression. Ann Stat 38:2559–2586

Müller H-G, Yang W (2010) Dynamic relations for sparsely observed Gaussian processes. Test 19:43–45

Müller H-G, Yao F (2008) Functional additive models. J Am Stat Assoc 103:153–1544

Ramsay J, Silverman B (2002) Applied functional data analysis. Springer Series in Statistics. Springer, New York

Ramsay J, Silverman B (2005) Functional data analysis. Springer Series in Statistics, 2nd edn. Springer, New York

Stone CJ (1982) Optimal global rates of convergence for nonparametric regression. Ann Stat 10:1040–1053

Valderrama M (2007) An overview to modelling functional data. Comput Stat 22:331–334

Yao F, Müller H-G, Wang J-L (2005) Functional data analysis for sparse longitudinal data. J Am Stat Assoc 100:577–590