Journal of Biometrics & Biostatistics
ISSN: 2155-6180

Escabias et al., J Biom Biostat 2012, 3:8
DOI: 10.4172/2155-6180.1000e120

Editorial Open Access

Functional Data Analysis in Biometrics and Biostatistics

Manuel Escabias, Mariano J Valderrama* and M Carmen Aguilera-Morillo

Unit of Biometrics and Statistics, Faculty of Pharmacy, University of Granada, 18071-Granada, Spain

Introduction

Functional data analysis is one of the areas of statistics that has generated most interest in recent years, from both theoretical and applied standpoints. This interest is reflected in the growing number of articles on this question in recent years, from the two that appeared in 1997 to the 83 published in 2011, according to the ISI Web of Knowledge database. This growth became particularly evident following the publication of the first specialist book in the field, "Functional data analysis" by J.O. Ramsay and B. Silverman in 1997.

From the practical point of view, one of the fields where this topic has aroused special interest is that of health sciences, the environment and biology, where in recent years 222 articles have been published, with more being added every year, according to the ISI Web of Knowledge.

What is meant by functional data?

Functional data are defined as discrete observations of a phenomenon that can be represented by smooth curves which reflect the dependence structure between neighbouring points, so that the phenomenon can be evaluated for any point of time. Some classic examples from the literature on functional data analysis are temperature data, rainfall data and growth data [1].

By treating this type of data from the standpoint of functional data analysis, it is possible to avoid the problems encountered with the classical multivariate approach, which considers such data as observations of different variables, which by their very nature will be strongly correlated, especially between neighbouring observations (variables).

The need to be able to evaluate a functional datum for any point in time leads us to define the representation of functional data as smooth curves. One of the methods most commonly adopted to do so is to perform this representation by means of function bases. On the one hand, this approach reduces the computational dimension of the problem, while on the other, it enables matrix algebra to be used in handling the models without imposing excessive practical limitations on the analysis of functional data [1].

Applications of FDA in Biometrics and Biostatistics

As indicated above, different methods of functional data analysis have been increasingly used in biometrics and biostatistics in recent years. Thus, Ratcliffe et al. [2] predicted human foetal heart rate responses to curves of repeated vibroacoustic stimulation. Escabias et al. [3] established the relationship between the risk of drought and curves of temperatures. Aguilera et al. [4] modelled the probability of lupus flare from curves that measured the time evolution of stress levels in patients with systemic erythematous lupus. Valderrama et al. [5] and Escabias et al. [6] used different curves of meteorological and climatic variables to model and forecast airborne cypress pollen concentration and olive pollen peaks. James [7] used functional data methods from a randomized placebo controlled trial of the drug D-penicillamine on patients with primary biliary cirrhosis of the liver. Finally, Wu and Muller [8] used functional data analysis methods to study the dependence of trajectories of viral load on those of CD4 cell counts, which are important markers for evaluating antiviral therapies in treating AIDS.

Main methods used in FDA

Functional principal component analysis (FPCA)

The main objective of FPCA is to extract from a set of curves the aspects which characterise them, to reveal the complexity of the data, to observe the different types of curves to be found, and to understand the structure of variability, covariances and correlations within the curves, as measured by variability, covariance and correlation surfaces. Furthermore, by means of functional PCA we can create a linear representation of a set of curves in which the random component is represented by vectors and the systematic component by functions or curves. This method also reduces the dimensions of the problem by representing curves in terms of a finite number of functions. A successful application of FPCA with Fourier basis expansions has been developed by Valderrama et al. [5].

Functional partial least squares (PLS)

An alternative approach to functional PCA is that of the functional PLS methodology, which shares the same objectives but where in extracting the components, rather than taking into account the variability between the curves, their relationship with a response variable is considered. Aguilera et al. [9] proposed its formulation in terms of basis expansion.

Functional discriminant and functional cluster analysis

The goal of functional discriminant analysis is to classify individuals according to the common features of a functional variable. In other words, the curves are classified into groups such that the curves of each group are as similar as possible regarding certain characteristics while the curves of different groups differ as much as possible in this respect.

Kayano et al. [10] used functional cluster analysis to model the three-dimensional (3D) protein structural data that determines the 3D arrangement of amino acids in individual proteins. Matsui et al. [11] used functional discriminant analysis to classify handwritten characters written in the air with one finger. Linear discriminant analysis has also been generalised for functional data classification [12]. Preda et al. [13]

***Corresponding author:** Mariano J. Valderrama, Professor of Statistics, Unit of Biometrics and Statistics, Faculty of Pharmacy, University of Granada, 18071-Granada, Spain, E-mail: valderra@ugr.es

Received November 20, 2012; **Accepted** November 22, 2012; **Published** November 29, 2012

Citation: Escabias M, Valderrama MJ, Aguilera-Morillo MC (2012) Functional Data Analysis in Biometrics and Biostatistics. J Biom Biostat 3:e120. doi:10.4172/2155-6180.1000e120

used linear discriminant analysis to classify the quality of biscuits in terms of the resistance of the dough used.

Functional ANOVA and regression analysis

In general, functional regression models are used for modelling relationships between functional and non-functional variables. Thus, classical variables can be modelled from curves, or curves can be modelled from curves.

When the explanatory variable is categorical and the response variable is functional, we wish to determine whether there are differences in the functional variable among the different categories of the explanatory variable. This problem is known as functional analysis of the variance. When the explanatory variable is functional and the response is scalar, our aim is to predict the response from the explanatory variable. This functional regression model is the one of the most commonly used. When both the response and the explanatory variables are functional, we wish to know how they are related and, more exactly, how the explanatory variable influences the response variable.

In spectroscopy, where the data are curves measured as functions of wavelengths, functional linear regression and ANOVA models with B-spline expansions were used by Saeys et al. [14]. A hidden process regression model for functional data was used by Chamroukhi et al. [15], based on an experimental study for curve discrimination.

The functional logistic regression model is used for modelling a dichotomous response from a functional explanatory variable. Although this model has the same goals as functional discriminant analysis and even as the functional ANOVA model, because it can also classify curves, the relationship between the response and the functional variables can be interpreted numerically, and this feature is of great interest in fields such as medicine, epidemiology and environmental studies. When there are multiple responses, ordinal and nominal regression models are used.

A functional nominal logistic model has been considered for predicting land use with the temporal evolution of coarse-resolution remote sensing data [16]. These authors proposed a quadrature method to approximate the linear predictor of the model from discrete data, and functional PCA to reduce the dimension of the problem, expressing the functional parameters in terms of spline interpolated eigen functions.

Multicollinearity in regression models affects the estimation of model parameters and therefore the interpretation of the curves and the relationships among the variables. Multicollinearity in functional regression models is apparent when the curves and the parameters are expressed in terms of basis functions, in the linear model, on one hand, and in the logistics model, on the other. Various alternatives based on different types of functional PCA have been proposed to resolve problems of multicollinearity [17].

CARMA models

One of the most recent areas of research is the development of time-continuous CARMA models given as solutions to stochastic differential equations, as described by Bosq and Blanke [18]. In medicine, this approach is especially suitable for forecasting the development of a patient's physiological records, such as ECGs, over a period of time.

Acknowledgements

This research was supported by Project MTM2010-20502 from Dirección General de Investigación del Ministerio de Educación y Cultura, Spain.

References

1. Ramsay JO, Silverman BW (2002) Applied Functional Data Analysis. Springer-Verlag, New York.
2. Ratcliffe SJ, Heller GZ, Leader LR (2002) Functional data analysis with application to periodically stimulated foetal heart rate data. II: functional logistic regression. Stat Med 21: 1115-1127.
3. Escabias M, Aguilera AM, Valderrama MJ (2005) Modelling environmental data by functional principal component logistic regression. Environmetrics 16: 95-107.
4. Aguilera AM, Escabias M, Valderrama MJ (2008) Discussion of different logistic models with functional data. Application to Systemic Lupus Erythematosus. Comput Stat Data Anal 53: 151-163.
5. Valderrama MJ, Ocaña FA, Aguilera AM, Ocaña-Peinado FM (2010) Forecasting pollen concentration by a two-step functional model. Biometrics 66: 578-585.
6. Escabias M, Valderrama MJ, Aguilera AM, Santofimia ME, Aguilera-Morillo MC (2012) Stepwise selection of functional covariates in forecasting peak levels of olive pollen. Stoch Environ Res Risk Assess. Accepted for publication (DOI: 10.1007/s00477-012-0655-0).
7. James GM (2002) Generalized linear models with functional predictors. J R Stat Soc Series B Stat Methodol 64: 411-432.
8. Wu S, Müller HG (2011) Response-adaptive regression for longitudinal data. Biometrics 67: 852-860.
9. Aguilera AM, Escabias M, Preda C, Saporta G (2010) Using basis expansions for estimating functional PLS regression: Applications with chemometric data. Chemometrics and Intelligent Laboratory Systems 104: 289-305.
10. Kayano M, Dozono K, Konishi S (2010) Functional cluster analysis via orthonormalized Gaussian basis expansions and its application. Journal of Classification 27: 211-230.
11. Matsui H, Araki T, Konishi S (2011) Multiclass functional discriminant analysis and its application to gesture recognition. Journal of Classification 28: 227-243.
12. James GM, Hastie TJ (2001) Functional linear discriminant analysis for irregularly sampled curves. J R Stat Soc Series B Stat Methodol 63: 533-550.
13. Preda C, Saporta G, Lévéder C (2007) PLS classification of functional data. Comput Stat 22: 223-235.
14. Saeys W, De Ketelaere B and Dairus P (2008) Potential applications of functional data analysis in chemometrics. J Chemom 22: 335-344.
15. Chamroukhi F, Samé A, Govaert G, Aknin P (2010) A hidden process regression model for functional data description. Application to curve discrimination. Neurocomputing 73: 1210-1221.
16. Cardot H, Faivre R, Goulard M (2003) Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data. J Appl Stat 30: 1185-1199.
17. Escabias M, Aguilera AM, Valderrama MJ (2004) Principal component estimation of functional logistic regression: discussion of two different approaches. J Nonparametr Stat 16: 365-384.
18. Bosq D, Blanke D (2007) Prediction and Inference in Large Dimensions. Wiley series in Probability and Statistics, Wiley-Dunod.