Journal of Statistical Planning and Inference 147 (2014) 1–23

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

A partial overview of the theory of statistics with functional data

Antonio Cuevas *

Departamento de Matemáticas, Universidad Autónoma de Madrid, Spain

ARTICLE INFO

Article history:
Received 30 June 2012
Received in revised form
23 January 2013
Accepted 7 April 2013
Available online 16 April 2013

Keywords:
Functional data
Functional classification
Functional regression
Bootstrap in FDA

ABSTRACT

The theory and practice of statistical methods in situations where the available data are functions (instead of real numbers or vectors) is often referred to as *Functional Data Analysis* (FDA). This subject has become increasingly popular from the end of the 1990s and is now a major research field in statistics.

The aim of this expository paper is to offer a short tutorial as well as a partial survey of the state of the art in FDA theory. Both the selection of topics and the references list are far from exhaustive. Many interesting ideas and references have been left out for the sake of brevity and readability.

In summary, this paper provides:

(a) A discussion on the nature and treatment of the functional data.
(b) A review of some probabilistic tools especially suited for FDA.
(c) A discussion about how the usual centrality parameters, mean, median and mode, can be defined and estimated in the functional setting.
(d) Short accounts of the main ideas and current literature on regression, classification, dimension reduction and bootstrap methods in FDA.
(e) Some final comments regarding software for FDA.

1. Introduction

1.1. Some historical perspective

The problems of statistical inference can be primarily classified according to the nature of the sample space $\mathcal{X}$ (where the available data live) and that of the parameter space Θ, where the target "parameter" is supposed to belong. To a certain extent, the progress of the mathematical statistics can be described in terms of the conquest of new broader more sophisticated structures for $\mathcal{X}$ and Θ, in particular those corresponding to infinite-dimensional spaces. The denomination *Abstract Inference* was used by Grenander (1981) to provide a particularly insightful view of this progress towards generality in the statistical theory.

From this perspective, the theory of statistics with functional data, often denoted Functional Data Analysis (FDA), corresponds to a last-generation statistics where $\mathcal{X}$ (and, in many cases, also Θ) is an infinite-dimensional function space.

* Tel.: +34 914973810; fax: +34 914974889.
E-mail address: antonio.cuevas@uam.es

0378-3758/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2013.04.002

So, according to the above mentioned classification, FDA could be placed in the general development of statistical theory as indicated in the following informal sketch (where n denotes the sample size):

Statistical theory	$\mathcal{X}$	Θ	**Dating back to**
Classical parametric inf.	$\mathbb{R}$	$\Theta \subset \mathbb{R}$	1920s
Multivariate analysis	$\mathbb{R}^d$ $(n >> d)$	$\Theta \subset \mathbb{R}^k$ $(n >> k)$	1940s
Nonparametrics	$\mathbb{R}^d$ $(n >> d)$	A function space	1960s
High dimensional problems	$\mathbb{R}^d$ $(n < d)$	$\Theta \subset \mathbb{R}^k$	2000s
Functional Data Analysis	A function space	$\mathbb{R}^k$, or a function space	1990s

Hence, in simple words, we might say FDA refers usually to those statistical problems where the available data consist on a sample of n functions $x_1 = x_1(t), \ldots, x_n = x_n(t)$ defined on a compact interval of the real line, say $[0,1]$. Additional (real or functional) variables are often incorporated, for example in the regression models.

Other more sophisticated models are possible, where [0, 1] is replaced by a d-dimensional interval or the sample functions are vector-valued. Also, FDA bears some affinity with those statistical problems, often referred to as "inference in stochastic processes" where the sample information is given by a partial trajectory $x(t)$, $t \in [0,T]$ of a stochastic process $\{X(t),\ t \geq 0\}$. In this case, the length T of the observation interval plays the role of the sample size n.

Of course the FDA theory has incorporated many standard tools of the classical parametric or multivariate statistics. For example, the dimension reduction tools. However, the infinite-dimensional nature of the sample space poses especial problems which allow us to classify FDA as a genuinely new branch of the statistical theory.

1.2. Some monographs and general references

The book by Ramsay and Silverman (2005), whose first edition was published in 1997, must be cited as a major landmark in the history of FDA. This book has a practical orientation, targeted to a wide scientific audience. It has developed a crucial role in the popularization of FDA. The associated software, freely provided by the authors, became soon an effective toolbox for an increasing number of researchers, flooded by a new abundance of experimental data coming from on-line monitoring of different experiments. Another book by the same authors, Ramsay and Silverman (2002), is focussed on the illustration of the main FDA techniques through the study of specific case studies with real data.

The book by Ferraty and Vieu (2006) represented a second-generation view of the subject. It incorporates further mathematical insights, including a more detailed treatment on the non-trivial asymptotic issues involved in FDA, together with some discussion of several relevant issues as the use of semi-metrics and the so-called "small ball probabilities" phenomenon, which is in the basis of many theoretical difficulties in FDA. However, again, the practical aspects played a major role among the aims of this book.

The FDA French school include many other references deserving mention: some researchers are grouped in the STAPH team (www.math.univtoulouse.fr/staph/) whose earliest contribution to the topic is perhaps the paper by Dauxois et al. (1982), a pioneering contribution to the study of principal components for functional data.

The paper by Bosq (1991) is another path-breaking reference in the topic (not considered here) of functional-valued time series. The corresponding general theory of auto-regressive functional processes is given in Bosq (2000). The monograph by Bosq and Blanke (2007) deals mainly with the use of nonparametric approaches in statistical functional problems. Whereas the orientation of these two books is mostly theoretical, they are both, in a way, extremely practical as they jointly provide a fascinating account of the main mathematical tools involved in FDA.

The book by Horváth and Kokoszka (2012) is a fresh addition to the current general literature on FDA. It offers a well-balanced mixture of theoretical aspects (e.g., the useful Chapter 2 on Hilbert space theory) and applications (in particular, detailed discussions of real data examples and up-to-date information on software). About 40% of the book length (from Chapters 13 to 18) is devoted to the analysis of dependent functional data, including functional time series, change point detection and spatial statistics with functional data.

Special issues devoted to FDA topics have been published by different journals, including *Statistica Sinica*, issue 14, 3 (2004). *Computational Statistics*, 22, 3 (2007), *Computational Statistics & Data Analysis*, 51, 10 (2007), *Journal of Multivariate Analysis*, 101, 2 (2010).

Among the survey and overview papers, let us mention, e.g., Rice (2004), Müller (2005), González-Manteiga and Vieu (2011) (which includes an extensive bibliography), Delsol et al. (2011) (especially oriented to practical issues and real-data applications), etc.

The recent collective book Ferraty and Romain (2011) consist of 16 chapters, from different authors, with up-to-date surveys of the main topics of FDA.

1.3. The purpose and contains of this paper

The aim is to provide a personal perspective of the current theory and practice of FDA. Such a view is necessarily limited by several obvious constrains, including the space limitations and the author's awareness of the different subjects.