2017
International Workshop
on Functional
and Operatorial Statistics

IWFOS

Book of abstracts

Fourth International Workshop on Functional and Operatorial Statistics

A Coruña (Spain) – Fundación Barrié

June 15-17, 2017

Organizing committee

- Germán ANEIROS (Co-chair). Universidade da Coruña, Spain.
- Enea BONGIORNO (Co-chair). Università del Piemonte Orientale, Italy.
- Ricardo CAO. Universidade da Coruña, Spain.
- Rubén FERNÁNDEZ-CASAL. Universidade da Coruña, Spain.
- Mario FRANCISCO-FERNÁNDEZ. Universidade da Coruña, Spain.
- Salvador NAYA. Universidade da Coruña, Spain.
- Paula RAÑA. Universidade da Coruña, Spain.
- Ernesto SALINELLI. Università del Piemonte Orientale, Italy.
- Javier TARRÍO-SAAVEDRA. Universidade da Coruña, Spain.
- Sylvie VIGUIER-PLA. Institut de Mathématiques de Toulouse, France.
- José Antonio VILAR. Universidade da Coruña, Spain.
- Juan Manuel VILAR. Universidade da Coruña, Spain.

Scientific committee

- Ricardo CAO (Co-chair). Universidade da Coruña, Spain.
- Philippe VIEU (Co-chair). Institut de Mathématiques de Toulouse, France
- Ana AGUILERA. Universidad de Granada, Spain.
- Germán ANEIROS. Universide da Coruña, Spain.
- Graciela BOENTE. Universidad de Buenos Aires, Argentina.
- Pedro DELICADO. Universitat Politècnica de Catalunya, Spain.
- Aldo GOIA. Università del Piemonte Orientale, Italy.
- Wenceslao GONZÁLEZ-MANTEIGA. Universidad de Santiago de Compostela, Spain.
- Sonja GREVEN. Ludwig-Maximilians-Universität München, Germany.
- Siegfried HÖRMANN. Université Libre de Bruxelles, Belgium.
- Marie HUSKOVA. Charles University, Czech Republic.
- Steve MARRON. University of North Carolina, USA.
- Juhyun PARK. Lancaster University, United Kingdom.
- Dimitris POLITIS. University of California, USA.
- Juan ROMO. Universidad Carlos III, Spain.
- Piercesare SECCHI. Politecnico di Milano, Italy.
- Han Lin SHANG. Australian National University, Australia.
- Sara VAN DE GEER. Eldgenössische Technische Hochschule Zürich, Switzerland.

Thursday June 15th

Session 9:30 – 11:00 Chair: Philippe Vieu

Invited talk 9:30

Clustering functional data using projections

Presenter: A. Delaigle
Co-Authors: P. Hall and T. Pham

We show that, in the functional data context, by appropriately exploiting the functional nature of the data, it is possible to cluster the observations asymptotically perfectly. We demonstrate that this level of performance can often be achieved by the *k*-means algorithm as long as the data are projected on a carefully chosen finite dimensional space. In general, the notion of ideal cluster is not clearly defined. We derive our results in the setting where the data come from two populations whose distributions differ at least in terms of means, and where an ideal cluster corresponds to one of these two populations. We propose an iterative algorithm to choose the projection functions in a way that optimises clustering performance, where, to avoid peculiar solutions, we use a weighted least-squares criterion. We apply our iterative clustering procedure on simulated and real data, where we show that it works well.

Contributed talk 10:20

Classification of Gaussian data: a comparison between finite and infinite-dimensional models

Presenter: J.R. Berrendero

We review several basic results for supervised classification of Gaussian observations. Our main purpose is to point out the similarities and to highlight the differences between the classification of finite-dimensional vectors and that of infinite-dimensional functional data. The theory of reproducing kernel Hilbert spaces (RKHS) provides a convenient framework for this task: it furnishes an appropriate language to formulate the results for functional data and includes finite-dimensional vectors as particular cases. Some of the issues we will discuss are: a) expressions for optimal rules and Bayes errors, b) the reasons why the near perfect classification phenomenon appears in classification of functional data but is mostly irrelevant for finite-dimensional observations, c) classification based on Mahalanobis distance, d) relationship between Gaussian models and logistic regression. This talk is based on several joint works (partly in progress) with B. Bueno-Larraz, A. Cuevas and J.L. Torrecilla.

Contributed talk 10:40

Boosting generalized additive models for location, scale and shape for functional data

Presenter: S. Greven
Co-authors: A. Stöcker, S. Brockhaus, S. Schaffer, B. von Bronk and M. Opitz

We extend Generalized Additive Models for Location, Scale, and Shape (GAMLSS) to regression with functional response. GAMLSS [1] are a flexible model class allowing for modeling multiple distributional parameters at once, such that for each parameter an individual additive predictor can be specified. By extending this to functional regression we may for example simultaneously model the point-wise mean and variance of response curves over time. In addition, a variety of marginal probability distributions can be used for the responses, transcending the exponential family. The model is fitted combining gradient boosting based functional regression [2, 3] and boosting based GAMLSS [4, 5]. This provides inherent model selection and regularization.
We apply the functional GAMLSS to analyze bacterial interaction in *E. coli* and show how the consideration of the variance structure fruitfully extends usual growth models. Employing historical function-on-function effects, we model the growth curve of one bacterial strain as pointwise gamma distributed with mean and standard deviation depending on the complete growth history of a competitive strain up to the current time point. By extending functional response regression beyond the mean, our approach offers new insights into underlying processes and helps overcome overly restrictive model assumptions.

References

[1] Rigby, R. A. and Stasinopoulos, D. M.: Generalized Additive Models for Location, Scale and Shape. J R Stat Soc Ser C Appl Stat. **54(3)**, 507–554 (2005)

[2] Brockhaus, S., Fuest, A., Mayr, A. and Greven, S.: Signal Regression Models for Location, Scale and Shape with an Application to Stock Returns (*Submitted*)

[3] Brockhaus, S., Melcher, M., Leisch, F. and Greven, S.: Boosting flexible functional regression models with a high number of functional historical effects. Stat Comput (*To appear*)

[4] Mayr, A., Fenske, N., Hofner, B., Kneib, T. and Schmid, M. Generalized additive models for location, scale and shape for high dimensional data: a flexible approach based on boosting. J R Stat Soc Ser C Appl Stat. **61(3)**, 1467–9876 (2012)

[5] Thomas, J., Mayr, A., Bischl, B., Schmid, M., Smith, A. and Hofner, B. Stability selection for component-wise gradient boosting in multiple dimensions (*Submitted*)

Session 11:30 – 12:50 — Chair: Ana Aguilera

Invited talk 11:30

Some practical uses of Radon-Nikodym derivatives in functional data analysis

Presenter: A. Cuevas

The Radon-Nikodym derivative of a Gaussian process with respect to another one can be explicitly calculated in many cases. It turns out that these derivatives have interesting applications in different problems of Functional Data Analysis, including

(a) Binary classification.

(b) Detection of “anomalous” observations in a sequence.

(c) Definition of depth measures.

We will briefly discuss these topics, presenting some results and commenting on some work in progress. This talk is based on common research work with José R. Berrendero, B. Bueno, R. Fraiman and J.L. Torrecilla.

Contributed talk 12:10

Robust fusion methods for Big Data

Presenter: R. Fraiman
Co-Authors: C. Aaron, A. Cholaquidis and B. Ghattas

We address one of the important problems in Big Data, namely how to combine estimators from different subsamples by robust fusion procedures, when we are unable to deal with the whole sample.

Contributed talk 12:30

Estimation of the mean vector in large dimension

Presenter: T. Bodnar
Co-Authors: N. Parolya and O. Okhri

In this paper we derive the optimal linear shrinkage estimator for the large-dimensional mean vector using random matrix theory. Under weak conditions imposed on the the underlying data generating process, we find the asymptotic equivalents to the optimal shrinkage intensities, prove their asymptotic normality, and estimate them consistently. The obtained non-parametric estimator for the high-dimensional mean vector has a simple structure and is proven to minimize asymptotically with probability 1 the quadratic loss.

Session 15:00 – 16:20 Chair: Sonja Greven

Invited talk 15:00

Estimating invertible functional time series

Presenter: A. Aue
Co-Author: J. Klepsch

This contribution discusses the estimation of an invertible functional time series through fitting of functional moving average processes. The method uses a functional version of the innovations algorithm and dimension reduction onto a number of principal directions. Several methods are suggested to automate the procedures. Empirical evidence is presented in the form of simulations and an application to traffic data.

Contributed talk 15:40

Detection of periodicity in functional time series

Presenter: G. Nisol
Co-Authors: S. Hörmann and P. Kokoszka

We derive several tests for the presence of a periodic component in a time series of functions. We consider both the traditional setting in which the periodic functional signal is contaminated by functional white noise, and a more general setting of a contaminating process which is weakly dependent. Several forms of the periodic component are considered. Our tests are motivated by the likelihood principle and fall into two broad categories, which we term multivariate and fully functional. Overall, for the functional series that motivate this research, the fully functional tests exhibit a superior balance of size and power. Asymptotic null distributions of all tests are derived and their consistency is established. Their finite sample performance is examined and compared by numerical studies and application to pollution data.

Contributed talk 16:00

Grouped multivariate functional time series method: An application to mortality forecasting

Presenter: H.L. Shang
Co-Author: Y. Yang

Age-specific mortality rates are often disaggregated by different attributes, such as sex and state. Forecasting age-specific mortality rates at the sub-national levels may not add up to the forecasts at the national level. Further, the independent forecasts may not utilize correlation among sub-populations to improve forecast accuracy. Using Japanese mortality data, we extend the grouped univariate functional time series methods to grouped multivariate functional time series forecasting methods.

Session 16:50 – 18:10 Chair: Aldo Goia

Invited talk 16:50

Confidence and prediction intervals in semi-functional partial linear regression

Presenter: P. Raña
Co-Authors: G. Aneiros, P. Vieu and J. Vilar

Semi-functional partial linear regression model allows to deal with a nonparametric and a linear component within the functional regression. Naïve and wild bootstrap procedures are proposed to approximate the distribution of the estimators for each component in the model, and their asymptotic validities are obtained in the context of dependence data, under α-mixing conditions. Based on that bootstrap procedures, confidence intervals can be obtained for each component in the model, which can be also extended to deal with prediction intervals and prediction densities.

Contributed talk 17:30

Parameter regimes in partially functional linear regression for panel data

Presenter: D. Liebl
Co-Author: F. Walders

We introduce a novel semiparametric partially functional linear regression model for panel data. The parametric model part is completely time varying, whereas the functional non-parametric component is allowed to vary over a set of different (functional) parameter regimes. These parameter regimes are assumed latent and need to be estimated from the data additionally to the unknown model parameters. We develop asymptotic theory for the suggested estimators including rates of convergence as $n, T \to \infty$. Our statistical model is motivated from economic theory on asset pricing. It allows to identify different risk regimes, governing the pricing of idiosyncratic risk in stock markets. For our application we develop necessary theoretical ground and offer a vast empirical study based on high-frequency stock-level data for the S&P 500 Index.

Contributed talk 17:50

High-dimensional functional time series forecasting

Presenter: Y. Gao
Co-Authors: H.L. Shang and Y. Yang

In this paper, we address the problem of forecasting high-dimensional functional time series through a two-fold dimension reduction procedure. Dynamic functional principal component analysis is applied to reduce each infinite-dimension functional time series to a vector. We use factor model as a further dimension reduction technique so that only a small number of latent factors are preserved. Simple time series models can be used to forecast the factors and forecast of the functions can be constructed. The proposed method is easy to implement especially when the dimension of functional time series is large. We show the superiority of our approach by both simulation studies and an application to Japan mortality rates data.

Friday June 16th

Session 9:00 – 10:20 — Chair: Han Lin Shang

Invited talk 9:00

Functional data analysis of neuroimaging signals associated with cerebral activity in the brain cortex

Presenter: L.M. Sangalli
Co-Authors: E. Lila and J.A.D. Aston

We consider the problem of performing principal component analysis of functional data observed over two-dimensional manifolds. The method is illustrated via the analysis of neuroimaging signals associated with cerebral activity in the brain cortex.

Contributed talk 9:40

Functional linear regression models for scalar responses on remote sensing data: an application to Oceanography

Presenter: N. Acar-Denizli
Co-Authors: P. Delicado, G. Basarir and I. Caballero

Remote Sensing (RS) data obtained from satellites are a type of spectral data which consist of reflectance values recorded at different wavelengths. This type of data can be considered as a functional data due to the continous structure of the spectrum. The aim of this study is to propose Functional Linear Regression Models (FLRMs) to analyze the turbidity in the coastal zone of Guadalquivir estuary from satellite data. With this aim different types of FLRMs for scalar response have been used to predict the amount of Total Suspended Solids (TSS) on RS data and their results have been compared.

Contributed talk 10:00

Functional data analysis in kinematics of children going to school

Presenter: M. Escabias
Co-authors: A.M. Aguilera, J.M. Heredia-Jiménez and E. Orantes-González

Traditionally gait analysis has examined discrete measures as descriptors of gait to compare different experimental situations. Functional data analysis (FDA) uses information from the entire curves and trajectories, thus revealing the nature of the movement. The aim of our study is to develop some repeated measures FDA methodologies to analyze kinematics of children's trunks while transporting a backpack and a trolley with different loads.

Poster Session 10:20 – 11:00

See pages 13 – 23

Session 11:20 – 12:40 — Chair: Pedro Delicado

Invited talk 11:20

Phase variation, Fréchet means, and procrustes analysis in Wasserstein space

Presenter: V.M. Panaretos
Co-author: Y. Zemel

A distinguishing characteristic of functional data is that they can vary both in amplitude and phase. Amplitude variation is linear in nature, generally encoded in the covariance operator, and representins fluctuations "in the *y*-axis". Phase variation is intrinsically nonlinear, and refers to fluctuations "in the *x*-axis", in particular random domain deformations

$T : [0,1]^d \to [0,1]^d$ (warpings) that produce different (and unknown) domain scales for each realisation of the random function. A key challenge in functional data analysis is to separate these two types of variation. The same challenge can arise when one considers functional data that are probability measures or point processes, say $\{\Pi_i\}$, and the purpose of our contribution is to investigate the problem of their registration: given unobservable random homeomorphisms $T_i : [0,1]^d \to [0,1]^d$ that deform the domain of definition, we wish to recover both $\{\Pi_i\}$ and $\{T_i\}$ from warped observations $\widetilde{\Pi}_i = (T_i)\#\Pi_i$. Here, "#" represents the pushforward, $T\#\Pi(A) = \Pi(T^{-1}(A))$. In other words, if a given realisation of Π contains the points $x_1, \ldots, x_m$, then one instead observes the warped collection $T(x_1), \ldots, T(x_m)$. We demonstrate that the canonical framework to analyse such random measures is within the formalism of Wasserstein spaces and optimal transportation. Following first principles, we show that the problem of separating amplitude and phase variation is equivalent to the problem of finding a Fréchet mean in Wasserstein space and of constructing the optimal multicoupling of a collection of marginal distributions. We then show how this can be exploited in order to get consistent fully nonparametric estimators. The problem turns out to be qualitatively very different depending on whether $d = 1$ or $d > 1$, because in the former case Wasserstein space is flat, whereas in the latter it is positively curved. In the case $d = 1$, we are able to get explicit closed form estimators, and obtain parametric rates of convergence as well as central limit theorems. The multidimensional case $d > 1$ is much more challenging, as here Wasserstein space is positively curved. Still, exploiting the tangent bundle structure of Wasserstein space, we illustrate how to deduce the Fréchet mean via gradient descent. We show that this is equivalent to a Procrustes analysis for the warp/registration maps, thus only requiring successive solutions to pairwise optimal coupling problems, and establish nonparametric consistency. Our results for the case $d = 1$ are from Panaretos & Zemel [1], while our results concerning the case $d > 1$ are from Zemel & Panaretos [3, 4]. For a more extensive presentation, including an accessible introduction to optimal transportation, we refer to the forthcoming monograph by Panaretos & Zemel [2] and the PhD thesis of Zemel [5].

References

[1] Panaretos, V.M. and Zemel, Y.: Amplitude and Phase Variation of Point Processes. Annals of Statistics **44(2)**, 771–812 (2016)

[2] Panaretos, V.M. and Zemel, Y.: Foundations of Statistics in the Wasserstein Space *in preparation*

[3] Zemel, Y. and Panaretos, V.M.: Fréchet Means in Wasserstein Space: Gradient Descent and Procrustes Analysis. Technical Report #1-16, Chair of Mathematical Statistics, EPFL, http://smat.epfl.ch/reports/1-16.pdf, February 2016.

[4] Zemel, Y. and Panaretos, V.M.: Fréchet Means and Procrustes Analysis in Wasserstein Space. arXiv:1701.06876.

[5] Zemel, Y.: Fréchet Means in Wasserstein Space: Theory and Algorithms. PhD Thesis, École Polytechnique Fédérale de Lausanne, February 2017.

Contributed talk 12:00

Differential interval-wise testing for local inference in Sobolev spaces

Presenter: A. Pini
Co-Authors: L. Spreafico, S. Vantini and A. Vietti

We present a local non-parametric inferential technique - namely, the differential interval-wise testing, or D-IWT - able to test the distributional equality of two samples of functional data embedded in Sobolev spaces. D-IWT can impute differences between the two samples to specific parts of the domain and to specific orders of differentiation. The proposed technique is applied to the functional data analysis of a data set of tongue profiles.

Contributed talk 12:20

A diagonal componentwise approach for ARB(1) prediction

Presenter: J. Álvarez-Liébana
Co-Author: M.D. Ruiz-Medina

This paper extends to the Banach-valued framework previous strong-consistency results derived, in the context of diagonal componentwise estimation of the autocorrelation operator of autoregressive Hilbertian processes, and the associated plug-in prediction. The Banach space B considered here is $B = \mathscr{C}([0,1])$, the space of continuous functions on $[0,1]$ with the supremum norm.

Session 15:00 – 16:20 Chair: Marie Husková

Invited talk 15:00

Applied functional registration

Presenter: J.A.D. Aston

Registration (warping) in functional data is often seen to simply be a way of removing a nuisance process before any true analysis of the data takes place. However, this talk will concentrate on warping as an integral part of the data generating process and indeed often the most interesting element of the data analysis. Different concepts in registration will be examined, such as using it as a part of a generative data model, using it to determine the interval of a functional observation or even allowing higher dimensional statistical analysis to be framed taking into account both function and shape. The ideas and concepts will be illustrated throughout using examples from linguistics, forensics and brain imaging.

Contributed talk 15:40

Tests for separability in nonparametric covariance operators of random surfaces

Presenter: S. Tavakoli
Co-Authors: D. Pigoli and J.A.D. Aston

We consider the problem of testing for separability in nonparametric covariance operators of multidimensional functional data is considered. We cast the problem in a tensor product of Hilbert space framework, where the role of the partial trace operator is emphasized, and the tests proposed are computationally tractable. An applications to acoustic phonetic data is also presented.

Contributed talk 16:00

Variable selection in functional additive regression models

Presenter: M. Febrero-Bande
Co-Authors: W. González-Manteiga and M. Oviedo de la Fuente

This paper considers the problem of variable selection when some of the variables have a functional nature and can be mixed with other type of variables (scalar, multivariate, directional, etc). Our proposal begins with a simple null model and sequentially selects a new variable to be incorporated into the model. For the sake of simplicity, this paper only uses additive models. However, the proposed algorithm may assess the type of contribution (linear, non linear, ...) of each variable. The algorithm have showed quite promising results when applied to real data sets.

Session 16:50 – 18:10 Chair: Germán Aneiros

Invited talk 16:50

Statistical functional depth

Presenter: A. Nieto-Reyes
Co-Author: H. Battey

This presentation is a summary of the paper [1], which formalizes the definition of statistical functional depth, with some extensions on the matter.

References

[1] Nieto-Reyes, A., Battey, H.: A topologically valid definition of depth for functional data. Statist. Sci. **31**, 61–79 (2016)

Contributed talk 17:30

Depth analysis for sparse functional data

Presenter: C. Sguera
Co-Authors: J.L.Torrecilla Noguerales, E. García Portugués and S. López-Pintado

Data depth is a well-known and useful notion in functional data analysis. It provides a center-outward ranking for a sample of curves. This ordering allows the definition of descriptive statistics such as medians, trimmed means and central regions for functional data. Moreover, data depth is often used as a building block for developing outlier-detection techniques and for robustifying standard statistical methods. Functional depths have been originally proposed for sample of curves that are measured on a common and dense grid. In practice, this is not always the case, since curves are often observed at subject-dependent and/or sparse grids. The main approach in the literature for dealing with this situation is based on estimating the individual trajectories on an artificially created common and dense grid of points and using these estimated curves as observed data in the notion of depth. Up to the date, the proposals based on this *reconstruct-then-depth* approach ignored the inherent uncertainty associated to the preliminary curve estimation step. In this work, we propose two alternative approaches for functional depths that explicitly address sparsity. First, we propose versions of existing functional depths that deal directly with sparse data, without requiring reconstruction, in a *data-as-is* approach. Second, we design a general procedure that allows the reconstruct-then-depth approach to take uncertainty into account. Preliminary results illustrate that the performances of the proposed alternatives are competitive for different levels and types of sparsity, and show the benefits of bearing in mind uncertainty in both simulated and real sparse settings.

Contributed talk 17:50

Linear causality in the sense of Granger with stationary functional time series

Presenter: M. Saumard

In this paper, we investigate the causality in the sense of Granger for functional time series. The concept of causality for functional time series is defined and a statistical procedure of testing the hypothesis of non-causality is proposed. This procedure is based on a test of equality of covariance operators for dependent processes.

Saturday June 17th

Session 9:00 – 10:20 — Chair: Wenceslao González-Manteiga

Invited talk 9:00

Essentials of backward nested descriptors inference

Presenter: S.F. Huckemann
Co-Author: B. Eltzner

Principal component analysis (PCA) is a popular device for dimension reduction and their asymptotics are well known. In particular, principal components through the mean span the data with decreasing residual variance, as the dimension increases, or, equivalently maximize projected variance, as the dimensions decrease, and these spans are nested in a backward and forward fashion – all due to Pythagoras Theorem. For non-Euclidean data with no Pythagorian variance decomposition available, it is not obvious what should take the place of PCA and how asymptotic results generalize. For spaces with high symmetry, for instance for spheres, backward nested sphere analysis has been successfully introduced. For spaces with less symmetry, recently, nested barycentric subspaces have been proposed. In this short contribution we sketch how to arrive at asymptotic results for sequences of random nested subspaces.

Contributed talk 9:40

Two-sample tests for multivariate functional data

Presenter: S.G. Meintanis
Co-Authors: Q. Jiang and L. Zhu

We consider two–sample tests for functional data with observations which may be uni– or multi–dimensional. The new methods are formulated as L2–type criteria based on empirical characteristic functions and are convenient from the computational point of view.

Contributed talk 10:00

Permutation tests in the two-sample problem for functional data

Presenter: A. Cabaña
Co-authors: A.M. Estrada, J. Peña and A.J. Quiroz

We propose two kind of permutation tests for the two sample problem for functional data. One is based on nearest neighbours and the other based on functional depths.

Poster Session 10:20 – 11:00

See pages 13 – 23

Session 11:20 – 13:00 — Chair: Ricardo Cao

Invited talk 11:20

Registration for exponential family functional data

Presenter: J. Goldsmith
Co-author: J. Wrobel

We consider the problem of aligning curves from exponential family distributions. The approach is based on the combination of alignment and functional principal components analysis, and is facilitated by recent extensions of FPCA to non-Gaussian settings. Our work is motivated by the study of physical activity using accelerometers, wearable devices that provide around-the-clock monitoring of activity and produce non-Gaussian measurements. We apply the proposed methods to activity counts using a Poisson distribution, and to a binary “active” vs “inactive” indicator using a binomial distribution. After alignment, the trajectories show clear peaks of activity in the morning and afternoon with a dip in the middle of the day.

Contributed talk 12:00

Registration of functional data using local variations

Presenter: A. Chakraborty
Co-Author: V.M. Panaretos

We study the problem of registration of functional data which have been subjected to random deformation (warping) of their time scale (see, e.g., [1] for a nice survey of the existing methodologies). The separation of this phase variation (variation in the "horizontal" axis) from the amplitude variation (variation in the "vertical" axis) is crucial for properly conducting further analyses, which otherwise can be severely distorted. We provide precise conditions under which the two forms of variation are identifiable, under minimal assumptions on the form of the random warp maps. We show using counterexamples that that these conditions are sharp. We then study a non-parametric registration method based on a "local variation measure", which connects functional registration with the problem of optimal transportation (see, e.g., [2]). The proposed method consistently estimates the warp maps from discretely observed data, without requiring any tuning parameters, as the number of observations grows and the measurement grid becomes dense. A detailed theoretical investigation of the strong consistency and the weak convergence properties of the resulting functional estimators is carried out, as well as of the impact of deviating from the identifiability conditions. Simulation and real data analyses demonstrate the good finite sample performance of our method.

References

[1] Marron, J.S., Ramsay, J.O., Sangalli, L.M. and Srivastava, A.: Functional data analysis of amplitude and phase variation. Statist. Sci. **30(4)**, 468–484 (2015)

[2] Villani, C.: Topics in optimal transportation. American Mathematical Society, Providence (2003)

Contributed talk 12:20

Parameter estimation of the functional linear model with scalar response with responses missing at random

Presenter: P. Galeano
Co-Authors: M. Febrero-Bande and W. González-Manteiga

This contribution considers estimation of the parameters of the functional linear model with scalar response when some of the responses are missing at random. We consider two different estimation methods of the functional slope of the model and analyze their characteristics. Simulations and the analysis of a real data example provides some insight into the behavior of both estimation procedures.

Contributed talk 12:40

Functional data analysis of "Omics" data: how does the genomic landscape influence integration and fixation of endogenous retroviruses?

Presenter: M.A. Cremona
Co-Authors: R. Campos-Sánchez, A. Pini, S. Vantini, K.D. Makova and F. Chiaromonte

We consider thousands of endogenous retrovirus detected in the human and mouse genomes, and quantify a large number of genomic landscape features at high resolution around their integration sites and in control regions. We propose to analyze this data employing a recently developed functional inferential procedure and functional logistic regression, with the aim of gaining insights on the effects of genomic landscape features on the integration and fixation of endogenous retroviruses.

Poster contributions

Session: Friday 16th and Saturday 17th, 10:20 – 11:00

A general sparse modeling approach for regression problems involving functional data

Presenter: G. Aneiros
Co-Author: P. Vieu

This presentation aims to introduce an approach for dealing with sparse regression models when functional variables are involved in the statistical sample. The idea is not to restrict to any specific variable selection procedure, but rather to present a two-stage methodology allowing to adapt efficiently any multivariate procedure to the functional framework. These ideas can be applied to any kind of functional regression models, including linear, semi-parametric or non-parametric models.

A time-dependent PDE regularization to model functional data defined over spatio-temporal domains.

Presenter: E. Arnone
Co-Authors: L. Azzimonti, F. Nobile and L.M. Sangalli

We propose a method for the analysis of functional data defined over spatio-temporal domains when prior knowledge on the phenomenon under study is available. The model is based on regression with Partial Differential Equations (PDE) penalization. The PDE formalizes the information on the phenomenon and models the regularity of the field in space and time.

Regression models with correlated errors based on functional random design

Presenter: K. Benhenni
Co-Authors: S. Hedli-Griche and M. Rachdi

This work deals with the study of the estimation of the functional regression operator when the explanatory variable takes its values in some abstract space of functions. The main goal is to establish the exact rate of convergence of the mean squared error of the functional version of the Nadaraya-Watson kernel estimator when the errors come from a stationary process under long or short memory and based on random functional data. Moreover, these theoretical results are checked through some simulations with regular (smooth) and irregular curves and then with real data.

Some applications of RKHS in functional data analysis

Presenter: B. Bueno-Larraz
Co-Authors: J.R. Berrendero and A. Cuevas

The theory of RKHS's (Reproducing Kernel Hilbert Spaces) has found a surprisingly large number of applications in different fields, including statistics. We will consider here two examples in which the RKHS space associated with the underlying process plays a outstanding role.
First, in functional regression problems with scalar response, we show that the formulation of the problem in RKHS terms has some advantages, especially when we aim at dimension reduction methods via variable selection. The basic idea of our approach is to use so-called Loève's congruence (a basic tool in the RKHS theory) in the definition of the model. The usefulness of this approach in variable selection problems comes from the fact that the finite-dimensional linear model based on the selected points can be seen as a particular case of the RKHS-based linear functional model.
Second, some closely related ideas (relying on the use of a smoothed version of the RKHS inner product) can also be used to define a functional counterpart of the classical Mahalanobis distance. This is joint work (partially in progress) with José R. Berrendero and Antonio Cuevas.

Nonparametric mean estimation for big-but-biased data

Presenter: R. Cao
Co-Author: L. Borrajo

Crawford [1] has recently warned about the risks of the sentence *with enough data, the numbers speak for themselves*. Some of the problems coming from ignoring sampling bias in big data statistical analysis have been recently reported by Cao [2]. The problem of nonparametric statistical inference in big data under the presence of sampling bias is considered in this work. The mean estimation problem is studied in this setup, in a nonparametric framework, when the biasing weight function is known (unrealistic) as well as for unknown weight functions (realistic). Two different scenarios are considered to remedy the problem of ignoring the weight function: (i) having a small sized simple random sample of the real population and (ii) having observed a sample from a doubly biased distribution. In both cases the problem is related to nonparametric density estimation. Asymptotic expressions for the mean squared error of the estimators proposed are considered. This leads to some asymptotic formula for the optimal smoothing parameter. Some simulations illustrate the performance of the nonparametric methods proposed in this work.

Acknowledgement This research has been supported by MINECO Grant MTM2014-52876-R and Xunta de Galicia Grants ED431C-2016-015 and ED431G/01, all of them through the European Regional Development Fund (ERDF). The contract of the first author has received financial support from the Xunta de Galicia and the European Union through the European Social Fund (ESF).

References

[1] Crawford, K.: The hidden biases in big data. Harvard Business Review, April 1st. Available at https://hbr.org/2013/04/the-hidden-biases-in-big-data (2013)

[2] Cao, R.: Inferencia estadística con datos de gran volumen. La Gaceta de la RSME, **18**, 393–417 (2015)

Functional GARCH models

Presenter: C. Cerovecki

The availability of financial series at very high frequencies suggests the use of *functional time series* methodology. For example, if $p_t(u)$ denotes the price of an asset on day t at time u then we can consider the intra-day log-returns $y_t(u) = \log p_t(u) - \log p_t(u-h)$ where h is some fixed time window. Such series are usually described by *volatility models*. Although there exists a wide variety of multivariate generalizations of these models, that can be found e.g. in [2], the first attempt to generalize them to *functional time series* was made in [3] for the ARCH(1) and latter in [1] for the GARCH(1,1). We work in a similar framework, i.e. we assume that $(y_t)_{t\in\mathbb{Z}}$ is a strictly stationary process valued in $L^2[0,1]$ and which satisfies the following equations

$$\begin{aligned} y_t(u) &= \sigma_t(u)\cdot\eta_t(u), \quad \forall u\in[0,1] \\ \sigma_t^2 &= \delta + \alpha(y_{t-1}^2) + \beta(\sigma_{t-1}^2), \end{aligned}$$

where $(\eta_t)_{t\in\mathbb{Z}}$ is some i.i.d. sequence in $L^2[0,1]$, and where y_{t-1}^2 and σ_{t-1}^2 denote the pointwise squares of the functions y_{t-1} and σ_{t-1}. We wish to estimate the function $\delta\in L^2[0,1]$ and the operators $\alpha,\beta : L^2[0,1]\to L^2[0,1]$. To fix ideas, let's consider the particular case where

$$\sigma_t^2(u) = \sum_{k=1}^{\infty} d_k\psi_k(u) + a\int_0^1 y_{t-1}^2(v)dv + b\int_0^1 \sigma_{t-1}^2(v)dv, \quad d_1,d_2,\ldots,a,b\in\mathbb{R}.$$

Here we let the intercept be a fully functional object expressed in some appropriate basis $(\psi_k)_{k\geq 1}$, whereas the operators α and β are radically simplified. One can interpret the first covariate as the so-called Realized Volatility - usually defined as $\sum_{j=0}^{\lfloor 1/h\rfloor} y_{t-1}^2(jh)$ - of day $t-1$. In this simple example the profile of the volatility process $E[\sigma_t^2]$ is proportional to δ. Although this model is interesting for practical interpretations, our goal is to allow for more complex α and β. We propose several approaches to do so. The difficulty is to find a trade-off between the parametric approach inherent to GARCH modeling and our infinite dimensional setting.

We propose various improvements to [1], e.g., we provide an optimal condition for the existence of stationary solutions, including the important case when the process has no second-order moments. Moreover, we extend it to higher order models (with $p,q>1$). Finally, we propose an estimation procedure that is consistent and has the usual asymptotic properties under mild assumptions. It is inspired by the *quasi maximum likelihood estimator* rather than by the *least squares estimator*. Finally we compare these estimators on various simulated and real data sets.

References

[1] Aue, A., Horváth, L. and Pellatt., D. F.: Functional generalized autoregressive conditional heteroskedasticity. Journal of Time Series Analysis, 38(1), pp.3–21. (2016).

[2] Francq, C. and Zakoïan, J. M.: GARCH models: structure, statistical inference and financial applications. John Wiley & Sons (2011).

[3] Hörmann, S., Horváth, L. and Reeder, R.: A functional version of the ARCH model. Econometric Theory, 29(02), pp.267-288. (2013).

Testing for white noise in functional time series

Presenter: V. Characiejus
Co-Authors: P. Bagchi and H. Dette

We propose a new procedure for white noise testing of a functional time series. Our approach is based on an explicit representation of the L^2-distance between the spectral density operator and its best (L^2)-approximation by a spectral density operator corresponding to a white noise process. The estimation of this distance can be accomplished by sums of periodogram kernels and it is shown that an appropriately standardized version of the estimator is asymptotically normally distributed under the null hypothesis (of functional white noise) and under the alternative. As a consequence we obtain a simple test (using the quantiles of the normal distribution) for the hypothesis of a white noise functional process. In particular the test does neither require the estimation of a long run variance (including a fourth order cumulant) nor resampling procedures to calculate critical values. Moreover, in contrast to all other methods proposed in the literature our approach also allows to test for "relevant" deviations from white noise and to construct confidence intervals for a measure which measures the discrepancy of the underlying process from a functional white noise process. See [1] for more details.

References

[1] Bagchi, P., Characiejus, V., and Dette, H.: A simple test for white noise in functional time series. `arXiv:1612.04996` (2016)

Mixed effects modeling with warping for functional data using B-spline

Presenter: E. Devijver
Co-Authors: G. Claeskens and I. Gijbels

Functional data usually present two types of variability: a phase variability and an amplitude variability. In this work, we propose to use a non linear functional mixed model to carry both variabilities, in order to improve the modeling and the interpretation. An observation Y_i for the individual i is suppose to be the sum of a mean curve μ and an individual curve U_i, both composed with a parametric warping function denoted by $w(.;\theta_i)$ for θ_i the warping parameter for the individual i. For $i \in \{1,\dots,n\}$, on a grid time $(t_{i,j})_{1\leq j\leq T_i} \in \mathscr{T}$,

$$Y_i(t_{i,j}) = \mu(w^{-1}(t_{i,j};\theta_i)) + U_i(w^{-1}(t_{i,j};\theta_i)) + \varepsilon_{i,j}.$$

Remark that time points may be different across individuals. In our modeling, we propose to consider θ_i as a random variable, to consider individual variation in the warping function. They are represented by a linear mixed effect model, to interpret variabilities among individuals. The warping function we consider is the following: for $t \in \mathscr{T}$,

$$w^{-1}(t;\theta_i) = \frac{\int_0^t \exp(\sum_{l=1}^{m^w} \theta_{i,l} B_l(u))du}{\int_0^1 \exp(\sum_{l=1}^{m^w} \theta_l B_l(u))du};$$

where $(B_l)_{1\leq l\leq m^w}$ is a B-spline. Remark that it does not rely on landmarks, which are not necessarily easy to define. In this presentation, I will describe a procedure to estimate the parameters of this model, and theoretical results.

Prediction bands for functional data based on depth measures

Presenter: A. Elías Fernández
Co-Authors: R. Jiménez Recaredo

Suppose a sample of random functions $\{Y_1,\dots,Y_n\}$ with values on $C(I)$, $I=[a,b]$ and a function Y_{short} that it is observed in a domain such that $I_{short}\subseteq I$. In this work, we deal with the problem of extending the partially observed function Y_{short} to the unobserved domain.

We propose a non parametric methodology for solving this problem motivated by the work of Sugihara and May (1990). The novelty of our approach relies on the selection of subsamples that make the function to predict a deep datum in the range of observation. The central regions delimited by the deepest curves of the selected subsamples provide tight bands that wrap not only in the observed part but even in the unobserved domain, preserving also its shape.

The involved subsampling problem is dealt by algorithms specially designed to be used in conjunction with two different tools for computing and visualizing central regions for functional data. Following Hyndman and Sun (2010), the first algorithm is based on Tukey's depth and the first two robust principal components scores. This two dimensional feature space allows us to find neighbourhoods or subsamples that surround the curve to predict in a natural way. In contrast, the second algorithm attemps to solve the problem in the functional space by applying functional depth measures and following the functional boxplot of Sun and Genton (2012).

We present two case studies for putting into practice our methodology. First, we tackle the problem of forecasting the Spanish electricity demand during the last three months of 2016 with a data sample of daily functions from 2014. On other hand, we propose an exercise of missing interval imputation with a data set of Spanish daily temperatures measured in 53 weather stations along the country.

The performance of both algorithms is similar for samples where dimensional reduction does not lead to considerable loss of information. Furthermore, the methodology is easy to adapt to a wide range of processes and the proposed algorithms could be considered with other depth measures or distances.

References

[1] Sugihara, G. and May, R. M.: Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature, **344** (6268):734741 (1990).

[2] Sun, Y. and Genton, M. G.: Functional boxplots. Journal of Computational and Graphical Statistics, **20**(2):316334 (2011).

[3] Hyndman, R. J. and Shang, H. L.: Rainbow plots, bagplots, and boxplots for functional data. Journal of Computational and Graphical Statistics, **19**(1):2945 (2010).

Quantile regression for functional data: an approximation using PIRLS

Presenter: M. Franco-Villoria
Co-Authors: R. Ignaccolo and M. Scott

Quantile regression allows estimation of the relationship between response and explanatory variables at any percentile of the distribution of the response (conditionally on the explanatory variables). We extend quantile regression to the functional case, rewriting the quantile regression model as a generalized additive model where both the functional covariates and the functional coefficients are parametrized in terms of B-splines. Parameter estimation is done using a penalized iterative reweighed least squares (PIRLS) algorithm. We evaluate the performance of the model by means of a simulation study.

An asymptotic factorization of the Small–Ball Probability: theory and estimates

Presenter: A. Goia
Co-Authors: J.-B. Aubin and E.G. Bongiorno

This work reviews recent results on an asymptotic factorization of the Small–Ball Probability of a $\mathscr{L}^2_{[0,1]}$–valued process, as the radius of the ball tends to zero. This factorization involves a volumetric term, a pseudo–density for the probability law of the process, and a correction factor. Estimators of the latter two factors are introduced and some of their theoretical properties considered.

On the Geometric Brownian Motion assumption for financial time series

Presenter: A. Goia
Co-Authors: E.G. Bongiorno and P. Vieu

The Geometric Brownian Motion type process is commonly used to describe stock price movements and is basic for many option pricing models. In this paper a new methodology for recognizing Brownian functionals is applied to financial datasets in order to evaluate the compatibility between real financial data and the above modeling assumption. The method rests on using the volumetric term which appears in the factorization of the small–ball probability of a random curve.

Parametric, semiparametric and nonparametric modelling in survival analysis

Presenter: I. Horová

In summarizing survival data, there are two functions of central interest, namely, survival and hazard function. These functions complement each other and present different viewpoints to the data. In practice, both of them can depend of some characteristics. We focus on an estimation the conditional hazard function. If an appropriate probability distribution of survival time is known then parametric methods are suitable (see [2]). However, this distribution is often unknown. Semiparametric methods do not require this strong assumption. The most frequently used semiparametric model was proposed by D.R. Cox [1]. This model is based on the assumption of proportionality of the hazards. In many cases this assumption is very restrictive. An alternative approach is to use nonparametric methods. We focus on kernel smoothing techniques (see [3]). Kernel methods belong to effective nonparametric methods. They also possess good statistical properties of estimations.
We conduct an extensive simulation study in which parametric, semiparametric and nonparametric models are compared. The attention is also paid to application to medical data. It is shown that the kernel method is able to capture changes in the conditional hazard function in both time and covariate directions.

References

[1] Cox, D. R.: Regression Models and Life-Tables. Journal of the Royal Statistical Society. Series B **34 (2)**, 187–220 (1972)

[2] Klein, J. P., Moeschberger, M. L. : Survival analysis: techniques for censored and truncated data. Springer Science & Business Media, (2005)

[3] Spierdijk, L.: Nonparametric conditional hazard rate estimation: a local linear approach. Computational Statistics & Data Analysis **52 (5)**, 2419–2434 (2008)

Change point detection for multivariate observations based on characteristic functions

Presenter: M. Husková
Co-Authors: Z. Hlávka and S. Meintanis

There have been proposed and studied a number of statistical procedures based on empirical characteristics functions for various setups, e.g., are goodness-of-fit tests, model specification tests and tests for detection of changes. The overview paper on such procedures was published by S. Meintanis (2016).
The aim of the talk is to present an extension of procedures based on empirical characteristic functions to detection of a change in distribution function of multivariate observations. The proposed statistics may be written in convenient closed-form expressions. This kind of simplicity is important when dealing with multivariate data. The talk will include results on large sample behavior, discussion on computational aspects and the implementation of the procedures on the basis of suitable resampling techniques. Also results of a Monte Carlo study for the finite-sample properties of the methods along with some empirical applications will be presented.

References

[1] Meintanis S.G.: A review of testing procedures based on the empirical characteristic function. South African Statist. J. 50, 1-14 (2016)

Modeling mandibular shape variation using functional ANOVA models

Presenter: L. Ippoliti
Co-Authors: L. Fontanella and P. Valentini

The shape changes of human mandible observed in a longitudinal study is addressed. A classical evaluation of growth and shape changes in the human mandible is performed by means of conventional cephalometric analyses of lateral radiographs of the craniofacial complex. In general, conventional metrical approaches have proved to be insufficient for the analysis of size and shape changes of complex anatomical forms such as the human mandible. For example, lines and angles measured by traditional methods are not able to provide information about where the growth and shape change has occurred. In this work, it will be shown that functional data analysis [1], represents a natural approach for representing the shape variations of a mandible and that functional linear models provide an elegant statistical framework for functional variability decomposition. First, the data and the problem of curve registration are described. Then, a Bayesian functional ANOVA model [2] is introduced to describe the different sources of variability as well as to summarize the main modes of variation of the mandibles. Finally, some computational issues are addressed and methods for adjusting inference in the presence of data-location error are discussed.

References

[1] Ferraty F, Vieu P: Nonparametric functional data analysis. Springer, New York (2006)

[2] Kaufman, C.G., Sain, S.R.: Bayesian functional ANOVA modeling using Gaussian process prior distributions. Bayesian Analysis. **1**, 123–149 (2010)

Uniform in the smoothing parameter consistency results in functional regression

Presenter: L. Kara-Zaïtri
Co-Authors: A. Laksaci, M. Rachdi and P. Vieu

This paper focuses on uniform in bandwidth and uniform in nearest neighbors consistencies of both kernel and kNN type estimators involving functional data. We established in previous works results in this topic for a selection of non-parametric conditional operators. Our interest here is to adapt that approach for studying the generalized nonparametric regression function.

Local bandwidth selectors for functional kernel regression

Presenter: J. Koláček
Co-Authors: D. Kuruczová

Nonparametric regression methods based on kernel smoothing rely on selection of suitable bandwidth parameter in order to minimize the mean squared error. We focus on the functional regression, i.e. the case when the predictor is a functional random variable. In finite-dimensional setting, global selection of bandwidth (same bandwidth is used for all data points) is often sufficient. Due to sparsity of infinite dimensional space, local bandwidth selection (different bandwidth is chosen for each data point) seems to be more suitable approach. In our study, we focus on local bandwidth selection methods. Furthermore, we propose local bandwidth selection method based on penalizing functions. Using simulation studies, we compare local bandwidth selection methods with each other and with their global counterparts.

Acknowledgement This work was supported by the grant GA15-06991S of the Czech Science Foundation.

Functional quantile regression: Local linear modelisation

Presenter: A. Laksaci
Co-Authors: Z. Kaid

A nonparametric local linear estimator of the conditional quantiles of a scalar response variable Y given a random variable X taking values in a semi-metric space. We establish the almost complete consistency and the asymptotic normality of this estimate. We prove that the asymptotic proprieties of this estimate are closely related to some topological characteristics of the data. Finally, a Monte Carlo study is carried out to evaluate the performance of this estimate.

Predicting the physiological limits of sport stress tests with functional data

Presenter: M. Matabuena
Co-Authors: M. Francisco-Fernández and R. Cao

This work aims at illustrating the enormous potential of continuous monitoring of the athlete, jointly with the application of statistical techniques for functional data in the analysis and control of sports performance. It is shown that using low intensity exercise, one can predict the performance of a group of athletes without forcing them to fatigue. This is the first indirect methodology proposed in the scientific literature that allows to estimate in a precise way physical fitness without producing fatigue. The areas of application of this procedure are not only limited to sport science. They are diverse and include, among others, medicine and education.

Shape classification through functional data reparametrization and distribution-based comparison

Presenter: P. Montero-Manso
Co-Authors: J.A. Vilar

Shape classification can be performed by identifying each shape with a probability distribution function, usually by sampling inside the shape or along its surface. These distributions can be treated as functional data and compared for classification purposes. In this work, for the task of classification, we take advantage of the capabilities of recently proposed distances between multivariate distributions coming from the homogeneity testing literature. By lifting the restriction of comparing one dimensional distributions, these distances allow us to propose alternative mappings from shape to distribution that are able to improve discriminatory power. These mappings involve reparametrizing the shapes as multivariate functional data. The methods are applied to a real dataset of otolith shapes.

Acknowledgement This work has received financial support from the Spanish Ministerio de Economía y Competitividad (grant MTM2014-52876-R), the Xunta de Galicia (Centro Singular de Investigación de Galicia accreditation ED431G/01 2016-2019 and Grupos de Referencia Competitiva ED431C2016-015) and the European Union (European Regional Development Fund - ERDF)

An overview of consistency results for depth functionals

Presenter: S. Nagy

Data depth is a nonparametric tool which may serve as an extension of quantiles to general data. Any viable depth must posses the uniform strong consistency property of its sample version. In this overview, a concise summary of the available uniform consistency results for most of the depths for functional data is given. Extensions of this theory towards random surfaces, imperfectly observed, and discontinuous functional data are studied.

Dynamical modelling of functional data using warped solutions of ODEs

Presenter: N. Olsen
Co-Authors: A. Tolver

For many samples of functional data (e.g. data from biomechanical systems) it is unreasonable to believe that timing may be changed without altering the amplitude of the signal. We present a new model that jointly models the phase and amplitude variation in a functional sample by regarding functional variables as warped solutions to a common inhomogenous differential equation. One such example is the following where the estimated trajectory x_i is solution to the second-order differential equation

$$x_i''(t) = \theta''(v_i(t)) + a(v_i(t))[x_i'(t) - \theta'(v_i(t))] + b(v_i(t))[x_i(t) - \theta(v_i(t))]$$

where the flexibility comes from the variation induced by the warping functions $v_i(t)$ and differences in initial conditions.

For estimation, we use an iterative procedure that alternates between updating warps and updating parameters of the differential equation, including starting values. For this purpose we have developed specialized software that includes an RKF45 solver for finding solutions to ODEs.

We demonstrate the potential of the procedure on a data set consisting of multivariate functional samples of arm movements. We apply the proposed method and show that it is able to model the variation very well, and that the estimated trajectories are very close to the original signals.

Family-wise error rate on domain subsets: a unified framework for local inference in functional data analysis

Presenter: A. Pini
Co-Author: K. Abramowicz, L. Schelin, S. Sjöstedt de Luna and S. Vantini

We consider a functional test of the hypotheses H_0 against H_1 (e.g., a test on parameters of a functional-on-scalar linear model), where we aim at selecting the parts of the domain where H_0 is violated, while controlling the probability of false discoveries. It is straightforward to define an unadjusted p-value function, associating a p-value to each point of the domain. Such function only point-wise controls the probability of type I error, so it can not be used for domain-selection purposes since it would not provide any global control of the probability of type-I error. This is why the definition of an adjusted p-value function provided with a stronger control is often required. We here require the control of the probability of falsely rejecting the null hypothesis on subsets of the domain (control of the family-wise error rate, FWER on subsets). We compare different methods to define the adjusted p-value function. The methods that we discuss belong to a general set of methods based on the following steps: a family $\mathscr{S}$ of subsets of the domain is defined; the restriction of the null hypothesis is tested on every element of the family; the adjusted p-value of each point is computed as the maximum p-value of the tests of every element containing that point. We consider several methods where the choice of $\mathscr{S}$ is either fixed or data-driven. We show theoretically that it is not possible to strongly control the FWER and at the same time perform domain selection with a fixed choice of $\mathscr{S}$. Conversely, we show how it is possible – with a data-driven choice of $\mathscr{S}$ – to control asymptotically the FWER and perform domain selection. Finite-sample properties of all the methods are further investigated in a simulation study.

Functional data analysis in neural processing

Presenter: O. Pokora
Co-Authors: J. Koláček and T.W. Chiu

Evoked potentials (EPs) reflect neural processing. They are widely used to study sensory perception. However, methods of analyzing EP have been limited mostly to the conventional ensemble averaging of EP response trials to a repeated stimulus, and less so to single-trials analysis. We applied the functional data analysis (FDA) approach to study auditory EP in the rat model of tinnitus, in which overdoses of salicylate (SS) are known to alter sound perception characteristically, as the same way as in humans. Single-trial auditory EPs were analyzed, after being collected on a daily basis from an awake rat. Single-trial EP integrals were generated with sound stimuli presented systematically over an intensity range. The results were approximated using the cubic spline to give sets of smoothed response-level functions. Comparisons between daily intensity-series for each of three types of sound were done using cross-distance measures in the form of metric and first-derivative based semimetric. From the results of FDA, the first-derivative form was found to provide a clearer separation, when EP data were compared between SS and the control groups. This is also true when the daily data were compared within the more variable SS-group itself. In addition, at the high intensity region, where SS-action is presumably strong, we also observed characteristic changes in two statistical parameters, mean and skewness, of the cross-distance representations. Results suggested that FDA is a sensitive approach for EP studies and it can become a powerful tool for the research in neural science, particularly neuropharmacology.

Acknowledgement This work was supported by the grant GA15-06991S of the Czech Science Foundation.

Schilling test for the two sample problem with image data

Presenter: M. Sabaté
Co-Authors: A. Cabaña and A.J. Quiroz

Let $\{X_N(t)\}$ and $\{Y_M(t)\}$ be collections of functional samples with distributions P and Q. The considered test is a generalization to functional data of the k-nearest neighbors multivariate two-sample test of Schilling [2] as appears in Cabaña, Estrada, Peña and J Quiroz [1]. Under $H_0 : P = Q$ vs $H_A : P \neq Q$, let Z_i be the combined sample obtained by concatenating X_i and Y_i. Define the indicator function $I_i(r)$ as 1 if Z_i and the r-nearest neighbors are from the same family sample and 0 otherwise. The nearest neighbors are defined using the $L^2([0,1])$. We define the following statistic: $T_{(N+M),k} = \frac{1}{(N+M)k} \sum_{i=1}^{N+M} \sum_{r=1}^{k} I_i(r)$. We present results of an application to the two sample problem in image selection. Consider a sample of images as arrays of correlated points that can be indexed as $\{X_i(t)\}$ where i stands for the image label and t for the pixel position. We treat the images as functions of the pixel position and perform the previous stated test. Note that the $L^2([0,1])$ distance is used, and therefore we don't have restrictions in the sample space dimension. We use this property to consider RGB images in a tridimensional space.

References

[1] A. Cabaña, A.M. Estrada, J.I Peña, A. J Quiroz.: Permutation tests in the two-sample problem for functional data. (unpublished)

[2] Schilling, Mark F.: Multivariate Two-sample Tests Based on Nearest Neighbor. Journal of the American Statistical Association 81 (395). [American Statistical Association, Taylor and Francis, Ltd.]: 799–806. (1986) doi:10.2307/2289012

Speed prediction of the sport horse from accelerometric and gyroscopic data

Presenter: A. Schmutz
Co-Authors: J. Jacques, L. Chèze and P. Martin

We present a predictive method for horse speed with functional predictors.The dataset characteristic is that the number of measurements per individual is variable.The method used is an unsupervised clustering method for multivariate functional data based on a multivariate principal component analysis followed by the multivariate nonparametric regression via the functional kernel estimator proposed by Ferraty and Vieu.

Hotelling in Wonderland

Presenter: A. Stamm
Co-Authors: A. Pini and S. Vantini

While Hotelling's T^2 statistic is traditionally defined as the Mahalanobis distance between the sample mean and the true mean induced by the inverse of the sample covariance matrix, we hereby propose an alternative definition which allows a unifying and coherent definition of Hotelling's T^2 statistic in any Hilbert space independently from its dimensionality and sample size. In details, we introduce the definition of random variables in Hilbert spaces, the concept of mean and covariance in such spaces and the relevant operators for formulating a proper definition of Hotelling's T^2 statistic relying on the concept of Bochner integral.

Multi-resolution clustering of time dependent functional data with applications to climate reconstruction

Presenter: J. Strandberg
Co-Authors: K. Abramowicz and S. Sjöstedt de Luna

A multi-resolution approach used to cluster dependent functional data is presented. Given a lattice of (time) points, a function is observed at each grid point. We assume that there are latent (unobservable) groups that vary slowly over time. We consider the case when at different time scales (resolutions) different groupings arise, with groups being characterised by distinct frequencies of the observed functions. We propose and discuss a non-parametric double clustering based method, which identifies latent groups at different scales. We present an application of the introduced methodology to varved lake sediment data, aiming at reconstructing winter climatic regimes in northern Sweden at different resolutions during the last six thousand years.

Functional data analysis approach of Mandel's h and k statistics in Interlaboratory Studies

Presenter: J. Tarrío-Saavedra
Co-Authors: M. Flores, S. Naya and R. Fernández-Casal

In this work, functional versions of Mandels's *h* and *k* statistics for outlier laboratory detection in interlaboratory studies (ILS) are presented. The critical values of *h* and *k* outlier test are approximated using bootstrap resampling, and their characteristic graphics are obtained. Thermogravimetric data are simulated to study the performance of the proposed d^H and d^K functional test statistics.

Contribution of functional approach to the classification and the identification of acoustic emission source mechanisms

Presenter: O.I. Traore
Co-Authors: P. Cristini, N. Favretto-Cristini, L. Pantera, P. Vieu and S. Viguier-Pla

In a context of nuclear Reactivity Initiated Accident, we describe acoustic emission signals, for which a problem of classification is open. As classical approaches with a reduced number of variables do not give satisfactory discrimination, we propose to use the envelopes of the received signals. We perform a *k*-means clustering and discuss the first results of this approach.

A statistical analysis of the cardiod growth model

Presenter: P. Valentini
Co-Authors: J.T. Kent, K.V. Mardia and L. Ippoliti

This work discusses the "revised cardioid growth model" [1] applied to objects represented by a continuous closed line in the real plane. The model is concerned with a functional regression that relates two star-shaped outlines Y and X observed at two different temporal instants [2]. The purpose of this work is to carry out a statistical assessment on the strengths and weaknesses of the cardiod growth model. To facilitate the statistical analysis, we emphasize a modified version of the model, which differs from the original model in three ways: a) growth is modelled on a log scale rather than a linear scale; b) there are two parameters to model growth (essentially an intercept and slope parameter) instead of a single slope parameter in the original model; c) explicit assumptions are introduced to model the statistical error. An application of the model for a longitudinal analysis of a set of outlines representing craniofacial data is discussed.

References

[1] Ramanathan N., Chellappa R., Biswas S.: Computational methods for modeling facial aging: A survey. Visual Languages and Computing. **3**, 131–144 (2009)

[2] Mardia K.V., Kent J.T., Goodall C.R., Little J.A.: Kriging and splines with derivative information. Biometrika. **83**, 207–221 (1996)

Testing for stationarity of functional time series in the frequency domain

Presenter: A. van Delft
Co-Author: A. Aue

Interest in functional time series has spiked in the recent past with papers covering both methodology and applications being published at a much increased pace. We contribute to the research in this area by proposing stationarity tests for functional time series based on frequency domain methods. Setting up the tests requires a delicate understanding of periodogram- and spectral density operators that are the functional counterparts of periodogram- and spectral density matrices in the multivariate world. Properties of the proposed statistics are derived both under the null hypothesis of stationary functional time series and under the smooth alternative of locally stationary functional time series. The methodology is theoretically justified through asymptotic results. Evidence from simulation studies and an application to annual temperature curves suggests that the tests work well in finite samples.

Locally stationary functional time series

Presenter: A. van Delft
Co-Authors: M. Eichler

Inference methods for functional data have received a lot of attention the last few years. So far, the literature on time series of functional data has focused on processes of which the probabilistic law is either constant over time or constant up to its second-order structure. Especially for long stretches of data it is desirable to be able to weaken this assumption. We introduce a framework that allows for meaningful statistical inference of functional data of which the dynamics change over time. That is, we put forward the concept of local stationarity in the functional setting and establish a class of processes that have a functional time-varying spectral representation. Time-varying functional ARMA processes are investigated and shown to be functional locally stationary according to the proposed definition. Important in our context is the notion of a time-varying spectral density operator of which the properties are studied and uniqueness is derived. The framework is used to construct an estimator of the spectral density operator based on a segmented periodogram tensor. This estimator is consistent and shown to be asymptotically Gaussian. The latter is established using a weaker tightness criterion than what is commonly thought to be required.

On asymptotic properties of functional conditional mode estimation with both stationary ergodic and responses MAR

Presenter: P. Vieu
Co-Authors: N. Ling and Y. Liu

This contribution deals with functional conditional mode estimation given a functional explanatory variable with both stationary ergodic and responses missing at random (MAR). More precisely, we propose the estimators for functional conditional density and conditional mode respectively in this case. The main results of the work are the establishment of the asymptotic properties of such estimators.

Commutator of projectors and of unitary operators

Presenter: S. Viguier-Pla
Co-Author: A. Boudou

We define and study the concept of commutator for two projectors, for a projector and a unitary operator, and for two unitary operators. Then we state several properties of these commutators. We recall that projectors and unitary operators are linked with the spectral elements of stationary processes. We establish relations between these commutators and some other tools related to the proximity between processes.

Random functional variable and Fourier series

Presenter: J. Zelinka

This paper presents how a functional random variable can be expressed in the form of Fourier series. This expansion can be used for the definition of components of the functional random variable and for the approximation of the random curves as the partial sum of the Fourier series. Thus we can estimate the distribution of the components, simulate the functional random variable with given components and compute some characteristics of the distribution of its norm.

Supporters


Fundación Barrié
50
AÑOS


Vicerreitoría de Política Científica,
Investigación e Transferencia
UNIVERSIDADE DA CORUÑA


gmodes
citic


#SGAPEIO
Sociedade Galega para a
Promoción da Estatística
e da Investigación de
Operacións


S E
I O


Rede de Tecnoloxías
Cloud e Big Data
para
hpc
XUNTA
DE GALICIA
Expte. R2016/045


INSTITUT
de MATHEMATIQUES
de TOULOUSE


UNIVERSITÉ
FRANCO
ITALIENNE
UNIVERSITÀ
ITALO
FRANCESE


ISNPS


Springer


Ayuntamiento de A Coruña
Concello da Coruña