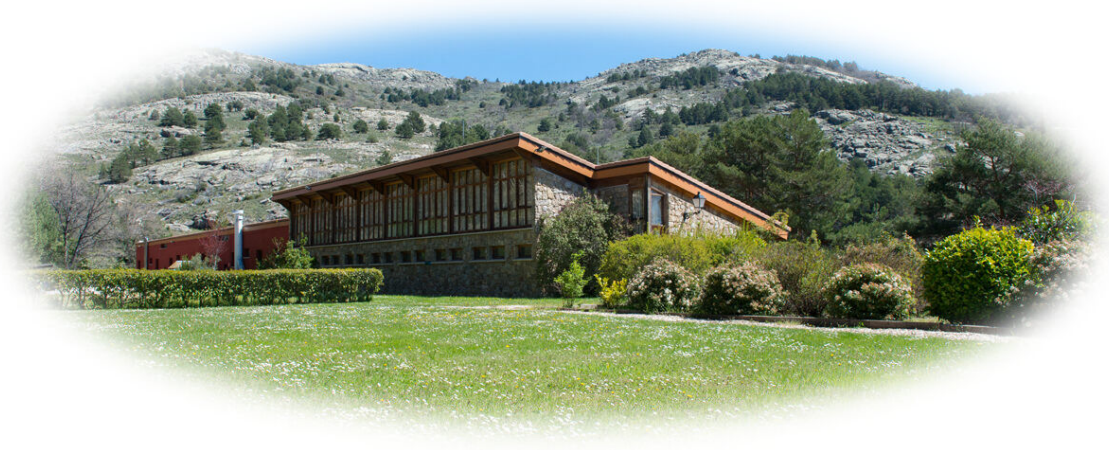
I Joint Workshop on Functional Data Analysis and Nonparametric Statistics



La Cristalera (Miraflores de la Sierra, Madrid), 6–9 June, 2023

Proceedings of I Joint Workshop on Functional Data Analysis and Nonparametric Statistics, a meeting of the Functional Data Analysis and Nonparametric Statistics Working Groups of the Spanish Society of Statistics and Operations Research (SEIO)

Seio Sociedad de Estadística e Investigación Operativa

Seio Sociedad de Estadística e Investigación Operativa

Last updated: 2023-06-09

Committees

Coordinators of SEIO working groups

Jose Ameijeiras Alonso – Nonparametric Statistics Working Group
Antonio Elías Fernández – Functional Data Analysis Working Group
Eduardo García Portugués – Nonparametric Statistics Working Group
Luis Alberto Rodríguez Ramírez – Functional Data Analysis Working Group

Scientific Committee

Antonio Cuevas – Universidad Autónoma de Madrid
Wenceslao González Manteiga – Universidade de Santiago de Compostela
María Dolores Martínez Miranda – Universidad de Granada

Organizing Committee

Javier Álvarez Liébana – Universidad Complutense de Madrid
Jose Ameijeiras Alonso – Universidade de Santiago de Compostela
Antonio Elías Fernández – Universidad de Málaga
Eduardo García Portugués – Universidad Carlos III de Madrid
Andrea Meilán Vila – Universidad Carlos III de Madrid
Luis Alberto Rodríguez Ramírez – Universidad Autónoma de Madrid
José Luis Torrecilla Noguerales – Universidad Autónoma de Madrid

Editors of book of abstracts: Organizing Committee

Workshop website: https://iwfdanp.webnode.es/

Sponsors


Seio
Sociedad
de Estadística
e Investigación
Operativa


Seio
Sociedad
de Estadística
e Investigación
Operativa


UAM
Universidad Autónoma
de Madrid


FUAM
Fundación
Universidad Autónoma
de Madrid

Contents

Foreword

We are honored to present the Proceedings of the I Joint Workshop on Functional Data Analysis and Nonparametric Statistics (JW-FDA-NP). This workshop, organized in collaboration with the Scientific and Organizing Committees, represents the first joint meeting of the Functional Data Analysis (FDA) and Nonparametric Statistics (NP) Working Groups of the Spanish Society of Statistics and Operations Research (SEIO). These working groups were created in 2009 and 2019, respectively, and aim to shepherd the research activities in these areas under the umbrella of SEIO. The present workshop is the fourth organized by the FDA Working Group, after those of Castro Urdiales (2019) and Getafe (2017, 2015), and the first organized by the NP Working Group.

The research intersections and common members between the FDA and NP communities are substantial, and hence the idea of gathering both in a joint workshop naturally appeared in the two working groups. The purpose is to bring together researchers who have contributed to each or both areas to disseminate their research, interact, and explore intersections. The workshop design dedicates specific days for each area, mixes senior and junior researchers, and tries to allocate enough time for the speakers to delve into details in the expositions of their works. As the venue for this special occasion, we have selected the Oberwolfach-esque residence of La Cristalera in Miraflores de la Sierra. We hope that this location can contribute to the event's success and create a memorable experience for the participants.

We express our gratitude to the authors of the abstracts of these proceedings for their highly-valuable scientific contributions. We gratefully acknowledge the members of the Scientific Committee for their efforts in ensuring the scientific quality of the workshop. We also thank the members of the Organizing Committee for their work in making the workshop a reality.

The workshop was possible due to the funding of the SEIO to the FDA and NP Working Groups, and due to the V PRICIT program from the Universidad Autónoma de Madrid.

We hope you enjoy your time at the workshop!

The coordinators of SEIO Working Groups on FDA and NP
Madrid, 1 June 2023

Schedule

Tuesday 6 June (FDA day)

- **16:00 – 16:30** Registration
- **16:30 – 17:00** Opening ceremony
- **17:00 – 18:00** **Plenary session** FDA *Chair: Antonio Cuevas*
  - Statistical depth in multivariate and function spaces (**Stanislav Nagy**)
- **18:00 – 18:05** Short break
- **18:05 – 18:50** **Invited session** FDA *Chair: Javier Álvarez Liébana*
  - Functional regression models with functional response: New approaches and a comparative study (**Manuel Febrero-Bande**)
- **18:50 – 19:25** **Contributed session** FDA *Chair: José Luis Torrecilla*
  - Optimal classification of Gaussian processes (**Alberto Suárez**)
- **19:25 – 21:00** Free time
- **21:00** Dinner

Wednesday 7 June (FDA day)

- **9:00 – 10:00** Plenary session FDA *Chair: José Luis Torrecilla*
  - Functional data classification and RKHS (**José R. Berrendero**)
- **10:00 – 10:05** Short break
- **10:05 – 11:20** Contributed session FDA *Chair: Luis Alberto Rodríguez*
  - 10:05 – 10:30. scikit-fda: A Python package for functional data analysis (**Carlos Ramos-Carreño**)
  - 10:30 – 10:55. Multivariate functional outlier detection using the FastMUOD indices (**Oluwasegun Ojo**)
  - 10:55 – 11:20. Random effect inclusion in a functional logistic regression model (**Cristhian Leonardo Urbano-León**)
- **11:20 – 11:50** Coffee break
- **11:50 – 13:20** Invited session FDA *Chair: Luis Alberto Rodríguez*
  - 11:50 – 12:35. Fast and fair simultaneous confidence bands for functional parameters (**Dominik Liebl**)
  - 12:35 – 13:20. Censored functional data (**Ewa Strzalkowska-Kominiak**)
- **13:20 – 13:25** Short break
- **13:25 – 14:15** Contributed session FDA *Chair: Javier Álvarez Liébana*
  - 13:25 – 13:50. Theoretical properties of isotropic functional random variables (**Marc Vidal**)
  - 13:50 – 14:15. Domain selection for Gaussian process data: An application to electrocardiogram signals (**Nicolás Hernández**)
- **14:15 – 16:00** Lunch
- **16:00 – 17:30** Invited session FDA *Chair: Antonio Elías Fernández*
  - 16:00 – 16:45. Recent advances in functional times series (**María D. Ruiz-Medina**)
  - 16:45 – 17:30. Single-index mixture cure models: An application to cardiotoxicity in breast cancer patients (**Ricardo Cao**)
- **17:30 – 18:00** Coffee break
- **18:00 – 18:25** Contributed session FDA *Chair: José Luis Torrecilla*
  - 18:00 – 18:25. Bayesian RKHS-based methods in functional regression (**Antonio Coín**)
- **18:25 – 21:00** Free time
- **21:00** Dinner

Thursday 8 June (NP day)

- **9:00 – 10:00** **Plenary session** NP *Chair: Eduardo García Portugués*
  - L_p inference for multivariate location based on data-based simplices (**Davy Paindaveine**)
- **10:00 – 10:05** Short break
- **10:05 – 11:20** **Contributed session** NP *Chair: Andrea Meilán Vila*
  - 10:05 – 10:30. Causal survival embeddings: Non-parametric counterfactual inference under censoring (**Carlos García-Meixide**)
  - 10:30 – 10:55. Smooth k-sample tests under left truncation (**Adrián Lago**)
  - 10:55 – 11:20. A bootstrap bandwidth selector for the smoothed Beran's estimator with application to length-of-stay times in hospital of COVID-19 patients (**Rebeca Peláez**)
- **11:20 – 11:50** Coffee break
- **11:50 – 13:20** **Invited session** NP *Chair: Jose Ameijeiras Alonso*
  - 11:50 – 12:35. Kernel density estimation on the polysphere and its applications (**Andrea Meilán-Vila**)
  - 12:35 – 13:20. Nonparametric estimation of the shape functions (**Juan Carlos Pardo-Fernández**)
- **13:20 – 13:25** Short break
- **13:25 – 13:50** **Contributed session** NP *Chair: Eduardo García Portugués*
  - 13:25 – 13:50. Kernel-based model predictive control for fluid flows in presence of noise (**Luigi Marra**)
- **13:50 – 14:00** Hiking briefing
- **14:00 – 15:20** Lunch
- **15:20 – 16:05** **Invited session** NP *Chair: Jose Ameijeiras Alonso*
  - Goodness-of-fit tests for regression models with a doubly truncated response (**Jacobo de Uña-Álvarez**)
- **16:05 – 16:40** **Contributed session** NP *Chair: Jose Ameijeiras Alonso*
  - Statistical analysis of non-convexity measures (**Beatriz Pateiro-López**)
- **16:40 –** Hiking
- **21:00** Dinner BBQ

Friday 9 June (NP day)

- **9:00 – 10:00** **Plenary session** NP *Chair: Lola Martínez Miranda*
  - An overview on regression methods for circular variables (**Rosa M. Crujeiras**)
- **10:05 – 11:35** **Invited session** NP *Chair: Andrea Meilán Vila*
  - 10:05 – 10:50. Addressing the population comparison problem for point processes on linear networks (**María Isabel Borrajo**)
  - 10:50 – 11:35. Finite difference methods for kernel smoothing (**José E. Chacón**)
- **11:35 – 12:05** Coffee break
- **12:05 – 13:20** **Contributed session** NP *Chair: Eduardo García Portugués*
  - 12:05 – 12:30. Application of informational measurements combined with kernel density estimator to describe epileptic seizures via EEG (**Antonio Squicciarini**)
  - 12:30 – 12:55. Estimation of distance correlation: A simulation-based comparative study (**Ricardo Cao***)
  - 12:55 – 13:20. Estimating axial symmetry directions of a distribution by means of random projections (**Manuel Hernández-Banadik**)
- **13:20 – 13:50** Closing ceremony
- **13:50 – 16:00** Lunch

Abstracts: Plenary speakers

Functional data classification and RKHS

7th June
09:00-10:00
FDA session

José R. Berrendero[1]

[1]Universidad Autónoma de Madrid - ICMAT

We will consider two different settings related to functional data classification: the Gaussian model (in which it is assumed that the regressors given the responses are trajectories of Gaussian stochastic processes) and the logistic model (which imposes conditions on the distribution of the responses given the regressors). We will analyze the role of the theory of reproducing kernel Hilbert spaces (RKHS) to study the relationship between both models, and discuss the usefulness of this theory in the analysis of variable selection methods. The talk will be based on parts of [1] and [2], and related work in process.

Keywords: Functional data analysis; RKHS; Supervised classification; Variable selection.

Acknowledgements This research has been partially supported by Grants PID2019-109387GB-I00 from the Spanish Ministry of Science and Innovation, Grant CEX2019-000904-S funded by MCIN/AEI/10.13039/501100011033

References

[1] J.R. Berrendero, A. Cuevas, J.L. Torrecilla. *On the use of reproducing kernel Hilbert spaces in functional classification.* J. Am. Stat. Assoc. **113** (2018) 1210-1218.

[2] J.R. Berrendero, B. Bueno-Larraz, A. Cuevas. *On functional logistic regression: some conceptual issues.* TEST (to appear).

9th June
09:00-10:00
NP session

An overview on regression methods for circular variables

Rosa M. Crujeiras[1]

[1] Universidade de Santiago de Compostela - CITMAga

In this talk, we will revise nonparametric regression tools for settings involving a circular covariate and/or response. In recent years, there have been important developments on kernel regression methods for fitting regression curves in circular-linear, linear-circular and circular-circular regression. The existing approaches somehow adapt to the circular nature of the covariate and/or response the ideas in Euclidean local-polynomial methods. This review will consider these previous approaches, originally devised for (classical) mean regression, adapted to these new contexts. We will also consider more complex scenarios where our data may exhibit dependence ([2]), functional covariates ([3]) or where the mean may not be sufficient for capturing the data relation ([1]). Finally, we will also introduce other tools based on local-likelihood. The different proposals will be motivated with real data examples. Our goal is to provide an overview talk that, given the expert audience, may inspire future developments in this field.

Keywords: Circular data; Local-likelihodd; Local-linear fit; Real data.

Acknowledgements R.M. Crujeiras work is supported by AEI project "Complex Dynamics and Nonparametric Inference" (PID2020-116587GB-I00). The contributions of this talk are joint work with A. Meilán-Vila, M. Alonso-Pena and M. Francisco-Fernández.

References

[1] M. Alonso-Pena, R.M. Crujeiras. *Analyzing animal escape data with circular nonparametric multimodal regression*. Ann. Appl. Stat. **19** (2023) 130-152.

[2] A. Meilán-Vila, R.M. Crujeiras, M. Francisco-Fernández. *Nonparametric estimation of circular trend surfaces with application to wave direction*. Stoch. Environ. Res. Risk Assess. **35** (2021) 923–939.

[3] A. Meilán-Vila, R.M. Crujeiras, M. Francisco-Fernández. *Nonparametric estimation for a functional-circular regression model*. Stat. Papers (2023).

Statistical depth in multivariate and function spaces

6th June
17:00-18:00
FDA session

Stanislav Nagy[1] and George Wynne[2]

[1]Charles University of Prague; [2]University of Bristol

Statistical depths extend elements of nonparametric data analysis, such as ranks, orderings, or quantiles, to multivariate [4] and other complex spaces [1, 2]. We give an overview of several classical notions of depth applicable in multivariate spaces, and discuss the challenges encountered when considering those depths in function spaces and for infinite-dimensional data.

We observe that several widely used depths [1, 4] have been investigated also outside of statistics, under different names. Leveraging those connections, we derive new theoretical results for depths, both in the finite-dimensional [3] and functional [5] setting.

Keywords: Depth; Functional data analysis; Halfspace depth; h-depth.

Acknowledgements This work was partially supported by Czech Science Foundation, project n. 23-05737S.

References

[1] A. Cuevas, M. Febrero, R. Fraiman. *On the use of the bootstrap for estimating functions with functional data*. Comput. Statist. Data Anal. **51(2)** (2006) 1063-1074.

[2] S. López-Pintado, J. Romo. *On the concept of depth for functional data*. J. Am. Stat. Assoc. **104(486)** (2009) 718-734.

[3] S. Nagy, C. Schütt, E.M. Werner. *Halfspace depth and floating body*. Stat. Surv. **13** (2019) 52-118.

[4] J.W. Tukey. *Mathematics and the picturing of data*. In Proceedings of the International Congress of Mathematicians **2** (1975) 523-531.

[5] G. Wynne, S. Nagy. *Statistical depth meets machine learning: kernel mean embeddings and depth in functional data analysis*. arXiv preprint arXiv:2105.12778 (2022).

8th June
09:00-10:00
NP session

L_p inference for multivariate location based on data-based simplices

Alexander Dürre[1] and *Davy Paindaveine*[2]
[1]Leiden University; [2]Universitè Libre de Bruxelles

The fundamental problem of estimating the location of a d-variate probability measure under an L_p loss function is considered. The naive estimator, that minimizes the usual empirical L_p risk, has a known asymptotic behaviour but suffers from several deficiencies for $p \neq 2$, the most important one being the lack of equivariance under general affine transformations. We introduce a collection of L_p location estimators that minimize the size of suitable ℓ-dimensional data-based simplices. For $\ell = 1$, these estimators reduce to the naive ones, whereas, for $\ell = d$, they are equivariant under affine transformations and generalize the famous [3] median. The proposed class also contains the spatial median. Under very mild assumptions, we derive an explicit Bahadur representation result for each estimator in the class and establish asymptotic normality. Under a centro-symmetry assumption, we also introduce companion tests for the problem of testing the null hypothesis that the location μ of the underlying probability measure coincides with a given location μ_0. We compute asymptotic powers of these tests under contiguous local alternatives, which reveals that asymptotic relative efficiencies with respect to traditional parametric Gaussian procedures for hypothesis testing coincide with those obtained for point estimation. Monte Carlo exercises confirm our asymptotic results.

Keywords: Affine equivariance/affine invariance; L_p loss functions; Oja median; Random simplices; Spatial median.

References

[1] A. Dürre, D. Paindaveine. *Affine-equivariant inference for multivariate location under L_p loss functions*. Ann. Stat. **50** (2022) 2616-2640.

[2] A. Dürre, D. Paindaveine. *Multivariate L_p location testing: Wald tests and Lagrange multiplier tests based on simplices* (submitted).

[3] H. Oja. *Descriptive statistics for multivariate distributions*. Stat. Probab. Lett. **1** (1983) 327-332.

Abstracts: Invited speakers

Addressing the population comparison problem for point processes on linear networks

9th June
10:05-10:50
NP session

María Isabel Borrajo[1]

[1]Universidade de Santiago de Compostela - CITMAga

Data sets representing the spatial location of a series of observations appear in a wide variety of scenarios, for example, trees in a forest, earthquakes in a region or traffic accidents in road networks. The latter is an example of point patterns which do not lay on a two-dimensional subregion of the plane, but which are constricted to a one-dimensional subset. These types of patterns are said to lay on a linear network. Analysing point processes on linear networks presents greater complexities than working on any Euclidean space, mainly because of the associated metric space.

A vastly studied problem in Statistics is population comparison, i.e., determine whether two (or more) samples are generated by the same stochastic process. This problem also arises when dealing with point processes, for example, the distribution of two species of flora in a forest, outbreaks of natural or caused forest fires, car-car and car-motorcycle collisions on a road network.

In the spatial point processes domain, this comparison problem has already been addressed, however this is not the case for point processes on other different domains. Inferential methods, as the ones proposed for the Euclidean plane, have not yet been developed regarding point processes on linear networks. In this work we study the two-sample problem for point processes on linear networks, proposing two specific testing methods, based on a Kolmogorov-Smirnov and a Cramer-von-Mises type statistics. A thorough simulation study is accomplished to detail the finite sample performance of our proposals. The test statistics are also applied to traffic collisions in Rio de Janeiro (Brazil).

Keywords: Linear network; Point processes; Two-sample problem.

Acknowledgements The author acknowledge the support from grant PID2020-116587GB-I00 funded by MCIN/AEI/10.13039/501100011033 and the European Union.

7th June
16:45-17:30
FDA session

Single-index mixture cure models: An application to cardiotoxicity in breast cancer patients

Ricardo Cao[1], Beatriz Piñeiro-Lamas[1] and Ana López-Cheda[1]

[1]Universidade da Coruña - CITIC

Standard survival models assume that the event of interest would always happen if there was a sufficient follow-up time. However, this is not always realistic. For instance, HER2-positive breast cancer patients usually receive trastuzumab. Although this therapy has antitumor efficacy, it can cause a problem in the heart, known as cardiotoxicity, in some patients. In this context, there will be a fraction of individuals that will never suffer the side effect, just because they are not susceptible to it. They are said to be cured, in the sense that no matter how long you observe them, they will never experience the final event. To study the time until the cardiotoxicity appears, mixture cure models are appropriate. They allow to estimate both the probability of being cured and the survival function of the uncured population, depending on some covariates. In the literature, nonparametric estimation of both functions is limited to continuous unidimensional covariates [1, 2]. We fill this important gap by considering multidimensional and functional covariates, and proposing a single-index model for dimension reduction. A dataset related to cardiotoxicity from the University Hospital of A Coruña is considered.

Keywords: Censored data; Cure models; Single-index models.

Acknowledgements This research has been supported by MINECO Grant MTM2017-82724-R, and by the Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2020-14 and Centro de Investigación del Sistema Universitario de Galicia ED431G 2019/01), all of them through the ERDF. The second author acknowledges financial support from Axudas Predoutorais da Xunta de Galicia, with reference ED481A-2020/290.

References

[1] A. López-Cheda, R. Cao, M.A. Jácome, I. Van Keilegom. *Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models.* Comput. Stat. Data Anal. **105** (2017) 144-165.

[2] A. López-Cheda, M.A. Jácome, R. Cao. *Nonparametric latency estimation for mixture cure models.* TEST **26** (2017) 353-376.

Finite difference methods for kernel smoothing

9th June
10:50-11:35
NP session

José E. Chacón[1]

[1]Universidad de Extremadura

The naive kernel density estimator can be motivated, in analytical terms, as an application of a central difference formula to approximate the derivative of the cumulative distribution function. The relationship between kernel smoothing and finite difference methods typically ends here, but in fact the latter have much more to offer for the former. It was not until relatively recently that higher-order difference approximations were derived in explicit form. In this talk we will show how these higher-order formulas find their application in kernel smoothing to obtain higher-order naive kernel density estimators, naive kernel estimators of arbitrary density derivatives and naive kernel estimators of densities with bounded support.

Keywords: Bounded support; Density estimation; Derivative estimation; Finite difference method; Kernel smoothing.

Acknowledgements This research has been partially funded by the grants PID2019-109387GB-I00 and PID2021-124051NB-I00 from the Spanish Ministerio de Ciencia e Innovación.

6th June
18:05-18:50
FDA session

Functional regression models with functional response: New approaches and a comparative study

Manuel Febrero-Bande[1], Morteza Amini[2], Mohammad Darbalaei[1,2] and Manuel Oviedo-de la Fuente[3]

[1]Universidade de Santiago de Compostela; [2]University of Tehran;
[3]Universidade da Coruña

This paper proposes three new approaches for additive functional regression models with functional responses. The first one is a reformulation of the linear regression model, and the last two are on the yet scarce case of additive nonlinear functional regression models. One of the nonlinear models is based on constructing a Spectral Additive Model (the word "Spectral" refers to the representation of the covariates in a $\mathcal{L}_2$ basis), which is restricted (by construction) to Hilbertian spaces. The other one extends the kernel estimator, and it can be applied to general metric spaces since it is only based on distances. We include our new approaches as well as real data-sets in an R package. The performances of the new proposals are compared with previous ones ([2], [3]), which are reviewed from the theoretical and practical point of view. The simulation results show the advantages of the nonlinear proposals and the small loss of efficiency when the simulation scenario is truly linear. Finally, a visualization tool is also provided for checking the linearity of the relationship between a single covariate and the response. Simulation codes, packages and real datasets are included in a GitHub repository ([1]).

Keywords: Functional regression; Functional response; Nonlinear models.

Acknowledgements

The research by Manuel Febrero-Bande and Manuel Oviedo-de la Fuente has been partially supported by the Spanish Grant PID2020-116587GB-I00 funded by MCIN/AEI/10.13039/501100011033. The research by Mohammad Darbalaei and Morteza Amini is based upon research funded by Iran National Science Foundation (INSF) under project No. 99014748.

References

[1] M. Febrero-Bande, M. Oviedo-de la Fuente. *FRMFR* GitHub repository (2023). https://github.com/moviedo5/FRMFR/

[2] J. Goldsmith, F. Scheipl, L. Huang, J. Wrobel, Ch. Di, J. Gellar, J. Harezlak, M.W. McLean, B. Swihart, L. Xiao, C.M. Crainiceanu, Ph. Reiss. *refund: regression with functional data* (2023). R package version 0.1-30. https://cran.r-project.org/web/packages/refund/index.html

[3] L. Ruiyan, X. Qi. *FRegSigCom: functional regression using signal compression approach* (2019). R package version 0.3.0. https://cran.r-project.org/web/packages/FRegSigCom/index.html

Fast and fair simultaneous confidence bands for functional parameters

7th Jun
11:50-12:35
FDA session

Dominik Liebl[1] and Matthew Reimherr[2]
[1]University of Bonn; [2]Penn State University

Quantifying uncertainty using confidence regions is a central goal of statistical inference. Despite this, methodologies for confidence bands in Functional Data Analysis are still underdeveloped compared to estimation and hypothesis testing. In this work, we present a new methodology for constructing simultaneous confidence bands for functional parameter estimates. Our bands possess a number of positive qualities: (1) they are not based on resampling and thus are fast to compute, (2) they are constructed under the fairness constraint of balanced false positive rates across partitions of the bands' domain which facilitates the typical global, but also novel local interpretations, and (3) they do not require an estimate of the full covariance function and thus can be used in the case of fragmentary functional data. Simulations show the excellent finite-sample behavior of our bands in comparison to existing alternatives. The practical use of our bands is demonstrated in two case studies on sports biomechanics and fragmentary growth curves.

Keywords: Functional data analysis; Hypothesis testing; Kac-Rice formula; Simultaneous inference; Statistical fairness guarantees.

8th June
11:50-12:35
NP session

Kernel density estimation on the polysphere and its applications

Andrea Meilán-Vila[1] and Eduardo García-Portugués[1]

[1]Universidad Carlos III de Madrid

Polyspherical data refers to observations on the product of hyperspheres, represented as $S^{d_1} \times \ldots \times S^{d_r}$, where $d_1, \ldots, d_r \geq 1$ and S^d denotes a hypersphere of dimension $d \geq 1$. The polysphere encompasses various geometric shapes, including the circle (when $r = d_1 = 1$), sphere (when $r = 1$ and $d_1 = 2$), and torus (when $d_1 = \ldots = d_r = 1$).

The objective of this research is to introduce and investigate a kernel density estimator specifically designed for the analysis of this data. The primary asymptotic characteristics of the estimator, including bias, variance, pointwise normality, and optimal bandwidth, are derived. As an application of the proposed estimator, a novel fully nonparametric dimension-reduction technique is introduced in this work, with the objective of identifying the primary source of variability in hippocampus shapes parameterized by s-reps [1, 2]. This methodology involves estimating density ridges for high-dimensional polyspherical data.

Keywords: Density estimation; Directional data; Nonparametric statistics; Skeletal representations.

Acknowledgements The authors acknowledge support from grant PID2021-124051NB-I00, funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe".

References

[1] E. García-Portugués, A. Meilán-Vila. Hippocampus shape analysis via skeletal models and kernel smoothing. In: Y. Larriba (Ed.) *Statistical Methods at the Forefront of Biomedical Advances* (to appear). Springer, Cham.

[2] S.M. Pizer, S. Jung, D. Goswami, J. Vicory, X. Zhao, R. Chaudhuri, J.N. Damon, S. Huckemann, J. Marron. Nested sphere statistics of skeletal models. In: M. Breuß, A. Bruckstein, P. Maragos (Eds.) *Innovations for Shape Analysis* (2013). Mathematics and Visualization. Springer, Berlin, Heidelberg.

Nonparametric estimation of the shape functions

8th June
12:35-13:20
NP session

Juan Carlos Pardo-Fernández[1] and María Dolores Jiménez-Gamero[2]
[1]Universidade de Vigo; [2]Universidad de Sevilla

Arriaza *et al.* (2019) [1] introduced the right and left shape functions, which enjoy interesting properties in terms of describing the global form of a distribution. In this talk, we will propose nonparametric estimators of those functions. The estimators involve nonparametric estimation of the quantile and density functions. Pointwise and uniform consistency are proved under general regularity assumptions, as well as the limit in law. Simulations are included to study the practical performance of the proposed estimators. The analysis of a real data set illustrates the methodology.

Keywords: Nonparametric estimation; Pointwise convergence; Shape functions; Weak convergence.

Acknowledgements Work supported by the Grant PID2020-118101GB-I00, Ministerio de Ciencia e Innovación (MCIN/AEI/10.13039/501100011033).

References

[1] A. Arriaza, A. Di Crescenzo, M.A. Sordo, A. Suárez-Llorens. *Shape measures based on the convex transform order*. Metrika **82** (2019) 99-124.

7th June
16:00-16:45
FDA session

Recent advances in functional times series

María D. Ruiz-Medina[1], Diana P. Ovalle-Múñoz[1] and Antoni Torres-Signes[2]
[1]Universidad de Granada; [2]Universidad de Málaga

This talk summarizes recent contributions in the field of functional time series. Specifically, the first part of this talk is focused on long-range dependence functional time series, with values in the space of square-integrable functions on a compact and connected two points homogeneous space. In particular, weak-consistent semiparametric estimation of the spectral density operator is addressed in the Gaussian case (see, e.g., [2]; [3]). Some limit theorems on spectral asymptotics, and its potential applications are introduced, covering, in particular, the field of point pattern analysis (see [1]). Secondly, under a weak-dependent scenario, Fréchet regression is analyzed in the context of curve processes (response and regressor) with values in a Riemannian manifold (see [4]). Local linear Fréchet regression is also extended to this context. Finally, some ongoing research lines in collaboration with other research teams are commented under both, short- and long- range dependence, functional time series scenarios

Keywords: Fréchet regression; Long-memory functional time series; Riemannian manifold-valued curve processes; Semiparametric functional estimation; Spectral analysis.

Acknowledgements This work has been supported in part by Andalusian Research Group FQM-147, and project IMAG. CEX2020-001105-M MCIN/AEI/10.13039/501100011033

References

[1] M.P. Frías, A. Torres-Signes, M.D. Ruiz-Medina. *Point pattern analysis and classification on compact two-point homogeneous spaces evolving time.* Stoch. Environ. Res. Risk Assess. **37** (2023) 2145-2158.

[2] D. Ovalle-Muñoz, M.D. Ruiz-Medina. *LRD spectral analysis of multifractional functional time series on manifolds.* arXiv preprint arXiv:2212.06228 (2022).

[3] M.D. Ruiz-Medina. *Spectral analysis of long range dependence functional time series.* Fract. Calc. Appl. Anal. **25** (2022) 1426-1458.

[4] A. Torres-Signes, M.P. Frías, M.D. Ruiz-Medina. *Multivariate manifold-valued curve regression in time.* arXiv preprint arXiv:2208.12585 (2022).

Censored functional data

Ewa Strzalkowska-Kominiak[1]

[1]Universidad Carlos III de Madrid

7th June
12:35-13:20
FDA session

Functional data analysis plays an increasingly important role in medical research since patients are followed over time. Thus, the measurements of a particular biomarker for each patient are often registered as curves. Hence, it is of interest to estimate the mean function under certain condition as an average of the observed functional data over a given period of time. This, however, often results difficult since, in this kind of follow-up studies, there are always some persons who drop-out of the study before it ends. For these patients, only a partial functional observation is available and, assuming that the loss of information happens completely at random, may lead to wrong conclusions. In this talk we propose an estimator for the functional mean when the functions may be censored from the right and so only partly observed. We present a simulation study and an application to a real data example on lung growth measured by changes in pulmonary function for US girls.

Keywords: Censored data; Functional data analysis; Survival analysis.

References

[1] A. Delaigle, P. Hall. *Classification using censored functional data.* J. Am. Stat. Assoc. **108** (2013) 1269-1283.

[2] E. Strzalkowska-Kominiak, J. Romo. *Censored functional data for incomplete follow-up studies.* Stat. Med. **40** (2021) 2821-2838.

[3] W. Wang. *Proportional hazard regression models with unknown link function and time-dependent covariates.* Stat. Sin. **14** (2004) 885-905.

8th June
15:20-16:05
NP session

Goodness-of-fit tests for regression models with a doubly truncated response

Jacobo de Uña-Álvarez[1]

[1]Universidade de Vigo

In Survival Analysis, Epidemiology or Reliability, among other fields, doubly truncated data may appear [3]. Double truncation means that the target variable is observed only when it falls within two random limits, which are also available in such a case. Unlike other phenomena of data incompleteness, nonparametric maximum-likelihood estimation with doubly truncated data does not have a closed form; this results in complicated asymptotics. An omnibus goodness-of-fit test for a regression model with a doubly truncated response will be introduced. In the spirit of [2], the test statistic will be based on the distance between two empirical integrated regression functions: one purely nonparametric, and the other one driven by the model to be tested. The underlying process will be a marked empirical process based on weighted residuals, where the weights remove the observational bias induced by the double truncation. The asymptotic null distribution of the test statistic will be obtained for both a fully specified and a parametric regression model. A bootstrap algorithm will be proposed in order to approximate the null distribution of the test in practice. The method will be illustrated with both simulated and real data. A projection approach to deal with multiple covariates [1] will be discussed.

Keywords: Random truncation; Survival analysis.

Acknowledgements Work supported by the grant PID2020-118101GB-I00, Ministerio de Ciencia e Innovación (MCIN/AEI/10.13039/501100011033).

References

[1] J.C. Escanciano. *A consistent diagnostic test for regression models using projections.* Econ. Theory **22** (2006) 1030-1051.

[2] W. Stute. *Nonparametric model checks for regression.* Ann. Stat. **25** (1997) 613-641.

[3] J. de Uña-Álvarez, C. Moreira, R.M. Crujeiras. *The statistical analysis of doubly truncated data with applications in R* (2021). Hoboken, NJ: John Wiley.

Abstracts: Contributed speakers

Bayesian RKHS-based methods in functional regression

7th June
18:35-19:00
FDA session

José R. Berrendero[1], *Antonio Coín*[1] and Antonio Cuevas[1]
[1]Universidad Autónoma de Madrid

We propose a novel Bayesian approach for functional linear and logistic regression models based on the theory of reproducing kernel Hilbert spaces (RKHS's). These newly introduced models build upon the RKHS associated with the covariance function of the underlying stochastic process, and can be viewed as a finite-dimensional approximation to the classical functional regression paradigm. The corresponding functional model (or the functional logistic equation in the case of binary response) is determined by a function living on a dense subspace of the RKHS of interest, which has a tractable parametric form based on linear combinations of the kernel. By imposing a suitable prior distribution on this functional space we can perform data-driven inference via standard Bayes methodology, estimating the posterior distribution through Markov chain Monte Carlo (MCMC) methods. We show that several prediction strategies derived from this approximate posterior are competitive against other usual alternatives in both simulated examples and real data sets. We also provide a theoretical consistency result for the posterior distribution based on an application of Doob's theorem to our RKHS setting.

Keywords: Bayesian inference; Functional data; Linear regression; Logistic regression; Reproducing kernel Hilbert space.

Acknowledgements Research partially supported by grant PRE2020-095147 of the Spanish Ministry of Science and Innovation (MICINN). We also wish to acknowledge the computational resources provided by the CCC-UAM.

References

[1] J. R. Berrendero, B. Bueno-Larraz, A. Cuevas. *An RKHS model for variable selection in functional linear regression.* J. Multivar. Anal. **170** (2019) 25-45.

[2] J. R. Berrendero, B. Bueno-Larraz, A. Cuevas. *On functional logistic regression: some conceptual issues.* TEST **32** (2023) 321–349.

[3] J. W. Miller. *Consistency of mixture models with a prior on the number of components.* Depend. Model. **11** (2023) 1-9.

8th June
10:05-10:30
NP session

Causal survival embeddings: Non-parametric counterfactual inference under censoring

Carlos García-Meixide[1]

[1]ETH Zürich

We propose a new nonparametric estimator for counterfactual survival functions under right-censoring using reproducing kernel Hilbert spaces (RKHS). We prove results regarding its asymptotic behavior, illustrate its practical performance through a simulation study, and apply it to the SPRINT trial; obtaining results in agreement with parallel analyses in the cardiology literature. Our method provides a novel tool for estimating counterfactual survival functions in observational studies under incomplete information, with potential applications in biomedicine.

Keywords: Survival analysis; RKHS; causality.

References

[1] V. Chernozhukov, I. Fernández-Val, B. Melly. *Inference on counterfactual distributions.* Econometrica **81** (2013) 2205-2268.

[2] T. A. Gerds, J. Beyersmann, L. Starkopf, S. Frank, M. J. van der Laan, M. Schumacher. *The Kaplan-Meier integral in the presence of covariates: A review.* From Statistics to Mathematical Finance: Festschrift in Honour of Winfried Stute (2017) 25-41.

[3] T. Martinussen. *Causality and the Cox regression model.* Ann. Rev. Stat. Appl. **9** (2022) 249-259.

[4] K. Muandet, M. Kanagawa, S. Saengkyongam, S. Marukatat (2021). *Counterfactual mean embeddings.* J. Mach. Learn. Res. **22** (2021) 1-71.

[5] A.W. Van der Vaart. *Asymptotic Statistics* (2000). Cambridge University Press.

Estimating axial symmetry directions of a distribution by means of random projections

9th June
12:55-13:20
NP session

Manuel Hernández-Banadik[1], Alejandro Cholaquidis[1], Juan Cuesta-Albertos[2] and Ricardo Fraiman[1]

[1]Universidad de la República; [2]Universidad de Cantabria

Suppose we have i.i.d. data $X_1, \ldots, X_n \in \mathbf{R}^d$ with distribution F_X, such that $\mathbf{E}(X) = 0$. We say that F_X is axially symmetric with respect to the direction $u \in \mathbf{S}^{d-1}$, if $F_X = F_{(2uu^t - I_d)X}$, i.e., X has the same distribution as $(2uu^t - I_d)X$. Several hypothesis tests [1, 2] exist in the literature to determine if the observed data, given a direction $u \in \mathbf{S}^{d-1}$, come from a distribution that is symmetric with respect to the direction generated by u.

The goal of this work is to detect the set of all symmetry directions from an i.i.d. sample. An estimator for such a set is given by

$$\hat{u} = \arg \min_{u \in \mathbf{S}^{d-1}} \sum_{j=1}^{k} \| F_{\langle X, h_j \rangle} - F_{\langle (2uu^t - I_d)X, h_j \rangle} \|_\infty,$$

where $h_1, \ldots, h_k$ are random directions with uniform distribution in $\mathbf{S}^{d-1}$, and the dimension of the problem is reduced using random projections. The objectives of this work are to prove the consistency of the estimator and derive its asymptotic distribution. Examples with simulated and real data are also presented

Keywords: Axial symmetry; Directional statistics; Random projections.

References

[1] Š. Hudecová, M. Šiman. *Testing axial symmetry by means of directional regression quantiles.* Electron. J. Stat. **15** (2021) 2690-2715.

[2] Š. Hudecová, M. Šiman. *Testing axial symmetry by means of integrated rank scores.* J. Nonparametr. Stat. (2022) 1-17.

7th June
13:50-14:15
FDA session

Domain selection for Gaussian process data: An application to electrocardiogram signals

Nicolás Hernández[1] and Gabriel Martos[2]

[1]University College of London; [2]Universidad Torcuato Di Tella

Gaussian Processes and the Kullback-Leibler divergence have been deeply studied in Statistics and Machine Learning. This paper marry these two concepts and introduce the local Kullback-Leibler divergence to learn about intervals where two Gaussian Processes differ the most. We address subtleties entailed in the estimation of local divergences and the corresponding interval of local maximum divergence as well. The estimation performance and the numerical efficiency of the proposed method are showcased via a Monte Carlo simulation study. In a medical research context, we assess the potential of the devised tools in the analysis of electrocardiogram signals.

Keywords: Domain selection; Electrocardiogram signals; Gaussian processes; Intervals of local maximum divergence; Kullback-Leibler divergence.

Smooth k-sample tests under left truncation

8th June
10:30-10:55
NP session

Adrián Lago[1], Ingrid Van Keilegom[2], Juan Carlos Pardo-Fernández[1] and Jacobo de Uña-Álvarez[1]

[1]Universidade de Vigo; [2]KU Leuven

Left truncation arises in many different applied fields due to the impossibility of observation of every individual that experiments the event of interest, frequently as a result of the way a study is designed or limitations on the measurement instruments. Truncation causes an observational bias which also induces bias in the estimators of different population quantities, such as the density function. A proper estimation of it can be derived from the Lynden-Bell estimator of the survival function for left-truncated data, proposed in [2]. Let us now consider k independent populations. It is a common applied problem to determine whether the target variables are the same in every population. To do that, a test based on an integral distance between the estimator of the density function in each population and the one of the pooled sample is proposed. Its asymptotic distribution will be studied and, due to the difficulty of its application in practice, a bootstrap resampling plan will be proposed to approximate the null distribution of the test statistic. The appropriateness of the method will be studied via Monte Carlo simulations. As the test is based on density estimators, the choice of the bandwidth plays an important role on its performance, thus it will be carefully studied. Moreover, a bandwidth to maximize the power under the alternative hypothesis will be proposed, based on a double-bootstrap algorithm as the one of [1]. Lastly, the proposed test will be compared to other tests in the literature for left-truncated data, such as the Kolmogorov-Smirnov and the log-rank, under different simulation scenarios. The performance of the test will be exemplified with real data regarding pregnancy times.

Keywords: Bootstrap; k-sample problem; Survival Analysis.

References

[1] R. Cao, I. Van Keilegom. *Empirical Likelihood Tests for two-sample problems via nonparametric density estimation*. Can. J. Stat. **34** (2006) 61-77.

[2] D. Lynden-Bell. *A method of allowing for known observational selection in small samples applied to 3CR quasars*. Monthly Notices R. Astron. Soc. **155** (1971) 95-118.

8th June
13:25-13:50
NP session

Kernel-based model predictive control for fluid flows in presence of noise

Luigi Marra[1], Andrea Meilán-Vila[2] and Stefano Discetti[1]

[1]Universidad Carlos III de Madrid (Department of Aerospace Engineering);
[2]Universidad Carlos III de Madrid (Department of Statistics)

The prediction and control of fluid flows is a major research topic due to its wide-ranging practical applications in many scientific fields. In this context, Model Predictive Control (MPC) has emerged as a powerful technique offering the ability to handle nonlinear dynamical systems with complex constraints [2]. The implementation of MPC requires a robust model to predict the behavior of dynamical systems from time series of sensor data. The presence of noise renders this task particularly challenging in applications.

In this work, we propose an automated procedure, which self-tunes the control parameters involved in the MPC process using a black-box optimization via Bayesian methods [3]. The effect of noise in the accuracy of the prediction of the system behavior is reduced by local polynomial regression [4] on time series of the control sensor output. The algorithm has been successfully applied to control the wake of the fluidic pinball [1], yielding promising results in scenarios with different noise levels.

Keywords: Fluid flow control; Local polynomial regression; Model predictive control; Noise robustness; Time series.

Acknowledgements The authors acknowledge the support from the research project PREDATOR-CM-UC3M. This project has been funded by the call "Estímulo a la Investigación de Jóvenes Doctores/as" within the frame of the Convenio Pluri-anual CM UC3M and the V PRICIT (V Regional Plan for Scientific Research and Technological Innovation).

References

[1] N. Deng, B.R. Noack, M. Morzyński, L.R. Pastur. *Low-order model for successive bifurcations of the fluidic pinball*. J. Fluid Mech. **884** (2020) 1-39.

[2] E. Kaiser, J. Kutz, S. Brunton. *Sparse identification of nonlinear dynamics for model predictive control in the low-data limit*. Proc. R. Soc. **474** (2018) 1-25.

[3] L. Hewing, K.P. Wabersich, M. Menner, M.N. Zeilinger. *Learning-based model predictive control: Toward safe learning in control*. Annu. rev. control robot. auton. syst. **3** (2020) 269-296.

[4] Q.J. Nottingham, D.F. Cook. *Local linear regression for estimating time series data*. Comput. Stat. Data Anal. **37** (2001) 209-217.

Estimation of distance correlation: A simulation-based comparative study

9th June
12:30-12:55
NP session

Blanca Monroy-Castillo[1], Amalia Jácome[1] and *Ricardo Cao*[1,*]

[1]Universidade da Coruña

Distance correlation [4] is a new class of multivariate dependence coefficients applicable to random vectors of arbitrary and not necessarily equal dimension. Even more, distance correlation, unlike the Pearson's correlation coefficient, is zero only if the random vectors are independent. Since its introduction, distance correlation has had many applications in, for example, life science [3], variable selection [7, 1] and has been extended to different contexts [5, 6]. In [6] an unbiased version of the squared sample distance covariance is proposed. Then, in [2] proved that the unbiased estimator turns out to be a U-statistic. In this work, a simulation study is developed to compare the estimations of distance correlation by means of the estimation proposed in [4] and the estimation through U-statistics [2]. The study shows the efficiency (MSE) and compares the computational time for both methods under different dependence structures.

Keywords: Distance correlation; Simulation; U-statistics.

References

[1] M. Febrero-Bande, W. Gónzalez-Mantiega, M. Oviedo de la Fuente. *Variable selection in functional additive regression models*. Comput. Stat. **34** (2019) 469-487.

[2] X. Huo, G.J. Szekely. *Fast computing for distance covariance*. Technometrics. **58** (2016) 435-447.

[3] J. Kong, B.E. Klein, R. Klein, G. Wahba. *Using distance correlation and SS-ANOVA to assess associations of familial relationships, lifestyle factors, diseases and mortality*. PNAS. **62** (2012) 20352-20357.

[4] G.J. Szekely, M.L. Rizzo, N.K. Bakirov. *Measuring and testing dependence by correlation of distances*. Ann. Stat. **35** (2007) 2769-2794.

[5] G.J. Szekely, M.L. Rizzo. *Brownian distance covariance*. Ann. Appl. Stat. **42** (2014) 2382-2412.

[6] G.J. Szekely, M.L. Rizzo. *Partial distance correlation with methods for dissimilarities*. Ann. Stat. **3** (2009) 1236-1265.

[7] C. Yeningün, M.L. Rizzo. *Variable selection in regression using maximal correlation and distance correlation*. J. Stat. Comput. Simul. **85** (2015) 1692-1705.

*Presented by the last author due to unforeseen unavailability of the first author.

7th June
10:30-10:55
FDA session

Multivariate functional outlier detection using the FastMUOD indices

Oluwasegun Ojo[1,3], Antonio Fernández Anta[1], Marc Genton[2] and Rosa E. Lillo[3]

[1]IMDEA Networks Institute; [2]King Abdullah University;
[3]Universidad Carlos III de Madrid

We present definitions and properties of the fast massive unsupervised outlier detection (FastMUOD) indices, used for outlier detection (OD) in functional data. FastMUOD detects outliers by computing, for each curve, an amplitude, magnitude, and shape index meant to target the corresponding types of outliers. Some methods adapting FastMUOD to outlier detection in multivariate functional data are then proposed. These include applying FastMUOD on the components of the multivariate data and using random projections. Moreover, these techniques are tested on various simulated and real multivariate functional datasets. Compared with the state of the art in multivariate functional OD, the use of random projections showed the most effective results with similar, and in some cases improved, OD performance. Based on the proportion of random projections that flag each multivariate function as an outlier, we propose a new graphical tool, the magnitude-shape-amplitude (MSA) plot, useful for visualizing the magnitude, shape and amplitude outlyingness of multivariate functional data.

Keywords: FastMUOD; Functional data; Multivariate data; Outlier classification.

References

[1] O.T. Ojo, A. Fernández Anta, R.E. Lillo, C. Sguera. *Detecting and classifying outliers in big functional data*. Adv. Data Anal. Classif. **16** (2022) 725–760.

[2] O. T. Ojo, A. Fernández Anta, M. G. Genton, R. E. Lillo. *Multivariate functional outlier detection using the fast massive unsupervised outlier detection indices*. Stat. **12** (2023).

Statistical analysis of non-convexity measures

8th June
16:05-16:40
NP session

Alejandro Cholaquidis[1], Ricardo Fraiman[1], Leonardo Moreno[2] and *Beatriz Pateiro-López*[3]

[1]Universidad de la República; [2]Universidad de la República - Instituto de Estadística; [3]Universidade de Santiago de Compostela - CITMAga

Convex sets and convex functions play an important role in many areas of mathematics and statistics. In some practical applications (like, for instance, in many set estimation problems, or in optimization problems where the domain is not convex) convexity can be a quite restrictive hypothesis. In this regard, to gain an insight into how much a set or a function departures from convexity can be useful to tune the parameters, or adapt (or even change) the algorithms or the estimation methods. In the literature, there are various measures of non-convexity that have been introduced for both sets and functions. These measures can be either geometric or topological in nature. In this study, we focus on the statistical analysis of some of these non-convexity measures for a set S. Specifically, we aim to estimate these measures using a sample of points from S. Additionally, we introduce a new measure of non-convexity and discuss the different notions of non-convexity. We prove the consistency of the proposed estimators and find their asymptotic distribution. We also examine the practical implementation of these estimators and provide an example of their use in analyzing real data.

Keywords: Convex hull; Convexity measure; Shape analysis.

Acknowledgements The authors gratefully acknowledge support from grant FCE-1-2019-1-156054, ANII (Uruguay), grant PID2020-116587GB-I00 funded by MCIN/AEI/10.13039/501100011033 and ED431C 2021/24 funded by Consellería de Cultura, Educación e Universidade.

References

[1] M. Fradelizi, M. Madiman, A. Marsiglietti, A. Zvavitc. *The convexification effect of Minkowski summation*. EMS Surv. Math. Sci. **5** (2018) 1-64.

[2] E. Rahtu, M. Salo, J. Heikkila. *A new convexity measure based on a probabilistic interpretation of images*. IEEE Trans. Patt. Anal. Mach. Intell. **28** (2006) 1501–1512.

8th June
10:55-11:20
NP session

A bootstrap bandwidth selector for the smoothed Beran's estimator with application to length-of-stay times in hospital of COVID-19 patients

Rebeca Peláez[1], Ricardo Cao[1] and Juan M. Vilar[1]

[1]University of A Coruña - CITIC

In the biomedical field, estimating the probability of a patient surviving beyond a specified time, T, given a covariate, X, is a significant problem. This involves estimating the conditional survival function, $S(t|x)$. In many cases, T is right randomly censored, meaning that some survival times are unknown because the study concludes before all individuals have experienced the event of interest.

The most commonly used nonparametric estimator of $S(t|x)$ under censoring was introduced by [1]. This and other usual estimators in the literature are based on covariate smoothing. In [2], the smoothed Beran's estimator of $S(t|x)$, smoothed in both the covariate and the time variable, is proposed. Simulation studies show its good performance. Asymptotic expressions of the estimator were proved, but there is no available method to choose the unknown smoothing parameters involved. In this work, a resampling technique to approximate the optimal bandwidths is proposed. Our approach combines the obvious bootstrap and the smoothed boostrap for the covariate and the time variables. The construction of bootstrap-based confidence regions for the conditional survival function is also addressed. Simulation studies show a reasonable behaviour of the proposals.

These procedures are applied to obtain nonparametric estimations and confidence regions of the conditional survival function of length-of-stay in hospital for COVID-19 patients in Galicia, Spain. The study leds to deeper insights into differences in hospitalised virus patients based on their age, sex and pre-existing conditions such as obesity or COPD.

Keywords: Bandwidth selection; Bootstrap; Censored data; Conditional survival function; Confidence regions.

Acknowledgements This research has been supported by MICINN Grant PID2020-113578RB-100, by the Xunta de Galicia (Grupo de Referencia Competitiva ED431C-2020-14 and Centro Singular de Investigación de Galicia ED431G 2019/01), all of them through the ERDF.

References

[1] R. Beran. *Nonparametric regression with randomly censored survival data* (1981). Technical report, University of California.

[2] R. Peláez, R. Cao, J.M. Vilar. *Nonparametric estimation of the conditional survival function with double smoothing.* J. Nonparametr. Stat. **34** (2022) 1063-1090.

scikit-fda: A Python package for functional data analysis

7th June
10:05-10:30
FDA session

Carlos Ramos-Carreño[1], José Luis Torrecilla[2] and Alberto Suárez[1]

[1]Universidad Autónoma de Madrid (Department of Computer Science);
[2]Universidad Autónoma de Madrid (Department of Mathematics)

We present the new developments in *scikit-fda* [2], a Python package for the statistical analysis of functional data. The package *scikit-fda* provides structures for the representation and manipulation of functional data either in discretized form or as a basis expansion. It includes methods for preprocessing (e.g., smoothing, registration, dimensionality reduction) exploratory analysis (e.g., robust statistics, visualization, outlier detection) and for machine learning (clustering, classification, and regression). The library is fully integrated in the scientific Python ecosystem. This environment is widely used by both researchers and developers, specially in the fields of machine learning and data science. In particular, *scikit-fda* conforms to the application programming interface (API) of *scikit-learn*, a powerful machine learning library in Python [1]. This design makes it possible reuse the tools available in *scikit-learn* for hyperparameter tuning and model selection. Furthermore, the functional data methods of *scikit-fda* can be combined with the multivariate utilities from *scikit-learn* and other compatible libraries in the same workflow. Finally, the package *scikit-fda* incorporates extensive documentation (https://fda.readthedocs.io), including tutorials and examples [3]. This makes it a useful didactical resource for people interested in functional data analysis.

Keywords: Functional data; Machine learning; Python; Software.

Acknowledgements We want to express our gratitude to all the contributors to the *scikit-fda* package. We acknowledge financial support from the Spanish Ministry of Education and Innovation, projects PID2019-106827GB-I00/ AEI/10.13039/501100011033 and PID2019-109387GB-I00, and from the Spanish Ministry of Science, Innovation and Universities (MICIU) with grant FPU18/00047.

References

[1] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, G. Varoquaux. *API design for machine learning software: experiences from the scikit-learn project*. ECML PKDD Workshop: Languages for Data Mining and Machine Learning (2013) 108–122.

[2] C. Ramos-Carreño, J.L. Torrecilla, M. Carbajo-Berrocal, P. Marcos, A. Suárez. *scikit-fda: A Python package for functional data analysis*. arXiv preprint arXiv:2211.02566 (2022).

[3] C. Ramos-Carreño, J.L. Torrecilla, Y. Hong, A. Suárez. *scikit-fda: Computational tools for machine learning with functional data*. IEEE 34th International Conference on Tools with Artificial Intelligence (2022) 213–218.

Application of informational measurements combined with kernel density estimator to describe epileptic seizures via EEG

9th June
12:05-12:30
NP session

Antonio Squicciarini[1], Elio Valero-Toranzo[2] and Alejandro Zarzo-Altarejos[1]

[1]Universidad Politécnica de Madrid - GI-TACA; [2]Universidad Rey Juan Carlos

Epilepsy is a common neurological disorder that can have a significant negative impact on patients' quality of life, and approximately one-third of patients are resistant to drug treatment. Manual seizure recognition in electroencephalogram (EEG) signals is a time-consuming and specialized process, which has led to the development of automatic and explainable seizure detection tools. While entropy measures have shown promise as feature extraction methods for EEG signal analysis [1][3][2], the potential of kernel density estimator as an inference method has not been explored in previous studies. In this study, we investigate the capabilities of general information measurements, including Fisher differential information, Shannon, Rényi, and Tsallis differential entropy, combined with a non-parametric kernel inference approach. We apply these measures to differentiate between ictal and non-ictal phases in EEG signals using time-dependent information plots and information planes. Our results demonstrate the effectiveness of this methodology in characterizing seizure periods and identifying the most informative EEG channel. Moreover, the use of different information measurement tools highlights their unique detection abilities. Notably, the Fisher-Shannon information plane proved to be effective in capturing distinct time-evolution trends during the epilepsy period.

Keywords: EEG; Epilepsy; Entropy; Kernel-based estimation.

Acknowledgements I would like to acknowledge "Programa Propio - Universidad Politécnica de Madrid (UPM)" for providing funding for the research.

References

[1] P. Boonyakitanont, A. Lek-uthai, K. Chomtho, J. Songsiri. *A review of feature extraction and performance evaluation in epileptic seizure detection using EEG*. Biomed. Signal Process. Control **57** (2020) 1-16.

[2] D.M. Mateos, R. Guevara Erra, R. Wennberg, J.L. Perez Velazquez. *Measures of entropy and complexity in altered states of consciousness*. Cogn. Neurodyn. **12** (2018) 73–84.

[3] O.A. Rosso, M.T. Martin, A. Figliola, K. Keller, A. Plastino. *EEG analysis using wavelet-based information tools*. J. Neurosci. Methods **153** (2006) 163–182.

Optimal classification of Gaussian processes

7th June
18:00-18:35
FDA session

Adrián Muñoz-Perera[1] and *Alberto Suárez*[1]

[1]Universidad Autónoma de Madrid

In this work, we derive optimal rules for the discrimination of Gaussian processes using two different methods. The first one takes as a starting point the Bayes classification rule for the discretized processes [4]. The second method is based on the theory of reproducing kernel Hilbert spaces (RKHS). In particular, it utilizes the properties of the covariance function of the process, which is the inner product of the associated RKHS. Separate derivations are given for the case in which the measures of the two processes are mutually absolutely continuous, and when they are mutually singular (orthogonal). In the former, the optimal rules obtained coincide with those expressed in terms of the Radon-Nikodym derivative between the two measures [1, 2]. In the latter, near-perfect classification (zero Bayes error) is obtained [3]. The optimal classification rule in this case arises from singular terms in the optimal classification rule for the discretely monitored process. Alternatively, they can be derived from the spectrum of the kernel.

Keywords: Functional data; Ornstein-Uhlenbeck processes; Optimal classification; Reproducing kernel Hilbert spaces.

Acknowledgements The authors acknowledge financial support from PID2019-106827GB-I00/AEI/10.13039/501100011033.

References

[1] A. Baíllo, A. Cuevas, J.A. Cuesta-Albertos. *Supervised classification for a family of Gaussian functional models*. Scand. J. Stat. **83** (2011) 480-498.

[2] J.R. Berrendero, A. Cuevas, J.L. Torrecilla. *On the use of reproducing kernel Hilbert spaces in functional classification*. J. Am. Stat. Assoc. **113** (2018) 1210-1218.

[3] A. Delaigle, P. Hall. *Achieving near perfect classification for functional data*. J. R. Stat. Soc. Ser. B Methodol. **74** (2012) 267-286.

[4] J.L. Torrecilla, C. Ramos-Carreño, M. Sánchez-Montañés, A. Suárez. *Optimal classification of Gaussian processes in homo- and heteroscedastic settings*. Stat. Comput. **30** (2020) 1091-1111.

7th June
10:55-11:20
FDA session

Random effect inclusion in a functional logistic regression model

Cristhian Leonardo Urbano-León[1], Manuel Escabias[1] and Ana M. Aguilera[1]

[1]University of Granada

In this work we explore the possibility of including a random effect to extend the functional logistic regression model presented in [1] for modelling a scalar binary responses variable from a functional predictor when independence between the functional curves cannot be assumed. Our proposal is motivated by a functional data set which consists of records of the position of the joints (angles they form with certain axes) with respect to the total gait cycle in different experimental conditions for the same set of individuals. This model assumes that curves of the functional predictor and the parameter function belong to the same finite-dimensional subspace $L^2(T)$ of squared integrable functions over the same closed real interval T, so the basis expansion can be used for functional data as in [2].

Keywords: Basis expansion; Gait cycle; Functional data analysis, Functional logistic regression; Random effects.

Acknowledgements The authors acknowledge the support by the research group FQM-307 of the Government of Andalusia (Spain) and by the project PID2020-113961GB-I00 of the Spanish Ministry of Science and Innovation (also supported by the FEDER programme). The authors also acknowledge the financial support of the Consejería de Conocimiento, Investigación y Universidad, Junta de Andalucía (Spain), and the FEDER programme for project A-FQM-66-UGR20. Additionally, the authors acknowledge financial support by the IMAG–María de Maeztu grant CEX2020-001105-M/AEI/10.13039/501100011033.

References

[1] M. Escabias, A.M. Aguilera, M.J. Valderrama. *Principal component estimation of functional logistic regression: discussion of two different approaches*. J. Nonparametr. Stat. **16** (2004) 365-384.

[2] C.L. Urbano-Leon, M. Escabias, D.P. Ovalle-Muñoz, J. Olaya-Ochoa. *Scalar variance and scalar correlation for functional data*. Mathematics **11** (2023) 1-20.

Theoretical properties of isotropic functional random variables

7th June
13:25-13:50
FDA session

Marc Vidal[1,2,3] and Ana M. Aguilera[2]

[1]Ghent University; [2]Universidad de Granada; [3]Max Planck Institute

Isotropic random variables have certain topological features that can enhance the application of statistical techniques among a wide variety of random processes. Here, we first delve into the definition of independence in Hilbert spaces to relate it to the notion of isotropy of a functional random variable. We define an operator that allows mapping a functional random variable into an isotropic one that is conditionally bounded [2]. Through simulation studies and real data examples, we study the optimization of this transformation by defining a subclass of operators derived from the factorization of the precision operator composed in a secondary space with a particular geometry. We further study the asymptotic behaviour of the proposed estimators.

Keywords: Correlation operator; Independence; Mahalanobis distance; Whitening operator.

Acknowledgements This research was partially supported by the Mathusalem funding from the Flemish Government, the project FQM-307 of the Government of Andalusia (Spain) and by the project PID2020-113961GB-I00 of the Spanish Ministry of Science and Innovation (also supported by the FEDER programme). We also acknowledge the financial support of the IMAG-María de Maeztu grant CEX2020-001105-M/AEI/10.13039/501100011033.

References

[1] M. Vidal, M. Rosso, A.M. Aguilera. *Bi-smoothed functional independent component analysis for EEG artifact removal*. Mathematics. **9** (2022) 1-17.

[2] M. Vidal, A.M. Aguilera. *Novel whitening approaches in functional settings*. Stat. **12** (2023) 1-10.

List of all participants

1. Álvarez Liébana, Javier (Universidad Complutense de Madrid)
2. Ameijeiras Alonso, Jose (CITMAga; Universidade de Santiago de Compostela)
3. Berrendero, José R. (Universidad Autónoma de Madrid; ICMAT)
4. Borrajo, María Isabel (Universidade de Santiago de Compostela)
5. Cao, Ricardo (Universidade da Coruña; CITIC)
6. Chacón, José E. (Universidad de Extremadura)
7. Coín, Antonio (Universidad Autónoma de Madrid)
8. Crujeiras, Rosa M. (CITMAga; Universidade de Santiago de Compostela)
9. Cuevas, Antonio (Universidad Autónoma de Madrid)
10. Diz Castro, Daniel (Universidade de Santiago de Compostela)
11. Elías Fernández, Antonio (Universidad de Málaga)
12. Febrero Bande, Manuel (Universidade de Santiago de Compostela)
13. García Meixide, Carlos (ETH Zürich)
14. García Portugués, Eduardo (Universidad Carlos III de Madrid)
15. Hernández, Nicolás (University College London)
16. Hernández Banadik, Manuel (Universidad de la República)
17. Lago Balseiro, Adrián (Universidade de Vigo)
18. Liebl, Dominik (University Bonn)
19. Marra, Luigi (Universidad Carlos III de Madrid)
20. Martínez Miranda, María Dolores (University of Granada)
21. Meilán Vila, Andrea (Universidad Carlos III de Madrid)
22. Nagy, Stanislav (Charles University)
23. Ojo, Oluwasegun (Universidad Carlos III de Madrid)
24. Ovalle Muñoz, Diana Paola (Universidad de Granada)

25. Paindaveine, Davy (Université libre de Bruxelles)
26. Pardo Fernández, Juan Carlos (Universidade de Vigo)
27. Pateiro López, Beatriz (Universidade de Santiago de Compostela)
28. Peláez, Rebeca (Universidade da Coruña)
29. Ramos Carreño, Carlos (Universidad Autónoma de Madrid)
30. Rodríguez Ramírez, Luis Alberto (Universidad Autónoma de Madrid)
31. Ruiz Medina, María Dolores (Universidad de Granada)
32. Squicciarini, Antonio (Universidad Politécnica de Madrid)
33. Strzalkowska-Kominiak, Ewa (Universidad Carlos III de Madrid)
34. Suárez, Alberto (Universidad Autónoma de Madrid)
35. Torrecilla, José Luis (Universidad Autónoma de Madrid)
36. de Uña Álvarez, Jacobo (Universidade de Vigo)
37. Urbano León, Cristhian Leonardo (Universidad de Granada)
38. Vidal, Marc (Ghent University; University of Granada; Max Planck Institute (Leipzig))
39. Vidal García, María (Universidade de Santiago de Compostela)