CD01

Complex Data Analysis
Análisis de Datos Complejos

Organizers   **Organizadores**   **Antolatzaileak**

Belén Pulido Bravo   **Jose Ameijeiras Alonso**

(IBiDat - Universidad Carlos III de Madrid)   (CITMAga, Univ. Santiago de Compostela)

Diego Bolón Rodríguez

(Université Libre de Bruxelles)

Description   **Descripción**   **Deskribapena**

This session aims to explore various advanced statistical techniques for dealing with the complexity of contemporary data, as well as some of their applications. In particular, three different areas of study will be considered: functional data analysis, where the data can be considered as a function that depends on some continuous variable, non-parametric statistics, where no assumptions are made about the underlying distribution of the data, and survival analysis, where some data cannot be directly observed due to the presence of censoring.

Esta sesión pretende explorar diversas técnicas estadísticas avanzadas para abordar la complejidad de los datos contemporáneos, así como algunas de sus aplicaciones. En particular, se considerarán tres áreas de estudio diferentes: el análisis de datos funcional, donde los datos pueden considerarse como una función que depende de alguna variable continua, la estadística no paramétrica, donde no se hacen suposiciones sobre la distribución subyacente de los datos, y el análisis de supervivencia, donde algunos datos no pueden ser observados directamente por la presencia de censura.

MSC Codes	Códigos MSC	MSC Kodeak

62-XX
(primary)

62R10; 62G05; 62N02
(secondary)

Slots	Bloques	Blokeak

1.B (Aula 0.16); 1.C (Aula 0.16); 2.A (Aula 0.16)

QR Code	Código QR	QR Kodea

Session Schedule	Horario de la Sesión	Saioaren Ordutegia

M14 | 15:00-15:20 | 0.16
Improved Seasonal Adjustment for Long Time Series
Cheyenne Amoroso (CITIC - Universidade da Coruña)

M14 | 15:30-15:50 | 0.16
An independent component approach for scalar-on-function linear regression.
Helena Ortiz Alcalá (University of Granada)

M14 | 16:00-16:20 | 0.16
New versions of the epigraph and hypograph indices based on areas between curves
Belén Pulido Bravo (IBiDat - Universidad Carlos III de Madrid)

M14 | 17:30-17:50 | 0.16
Cure models in cardio-oncology: dealing with functional and image covariates
Beatriz Piñeiro-Lamas (Universidade da Coruña)

M14 | 18:00-18:20 | 0.16

Regularization techniques for Cox regression in high-dimensional survival analysis

Pilar González-Barquero (Universidad Carlos III de Madrid)

M14 | 18:30-18:50 | 0.16

Random slope mixed models for Small Area Estimation of complex data

Naomi Diz-Rosales (CITIC - Universidade da Coruña)

M14 | 19:00-19:20 | 0.16

ROC curves with time-dependent variables

Arís Fanjul Hevia (Universidad de Oviedo)

J16 | 11:00-11:20 | 0.16

Bootstrap prediction intervals for quantile autoregression

Silvia Novo (Universidad Carlos III de Madrid)

J16 | 11:30-11:50 | 0.16

Different applications and extensions of the distance covariance

Laura Freijeiro González (Universidad de Oviedo)

J16 | 12:00-12:20 | 0.16

Quick quantile segmentation

Luis Alberto Rodríguez Ramírez (Georg August-Universität Göttingen)

J16 | 12:30-12:50 | 0.16

High density region estimation on manifolds

Diego Bolón (Université Libre de Bruxelles)

Tuesday 14	**Martes 14**	**Asteartea 14**
15:00-15:20	**15:00-15:20**	**15:00-15:20**
[Room 0.16]	**[Aula 0.16]**	**[Gela 0.16]**

Improved Seasonal Adjustment for Long Time Series
Cheyenne Amoroso
(CITIC - Universidade da Coruña)

A common task in economics is seasonal adjusting time series, often using Tramo-Seats methodology. However, long time series complicate identifying a single model, especially with structural changes in the data. New methodologies are proposed for cases with two identified ARIMA models and a transition period, modeled as a weighted average of the two events through a time-dependent weighting function. These approaches are evaluated through simulations to assess improvements and robustness.

Joint work with Carolina García-Martos, Germán Aneiros-Pérez, José Antonio Vilar, Manuel Oviedo de la Fuente and Mario Francisco Fernandez.

Tuesday 14	**Martes 14**	**Asteartea 14**
15:30-15:50	**15:30-15:50**	**15:30-15:50**
[Room 0.16]	**[Aula 0.16]**	**[Gela 0.16]**

An independent component approach for scalar-on-function linear regression.
Helena Ortiz Alcalá
(University of Granada)

The estimation of scalar-on-function linear regression model presents challenges due to the infinite dimension of functional data. In this context, we propose a new estimation method based on Independent Component Analysis, which is obtained by maximising kurtosis rather than variability or covariance with the response, classical solutions found in the literature. The performance and advantages of this new approach in comparison with other methods will be tested through a simulation study.

Joint work with Marc Vidal, Christian Acal and Ana M. Aguilera.

Tuesday 14	**Martes 14**	**Asteartea 14**
16:00-16:20	**16:00-16:20**	**16:00-16:20**
[Room 0.16]	**[Aula 0.16]**	**[Gela 0.16]**

New versions of the epigraph and hypograph indices based on areas between curves
Belén Pulido Bravo
(IBiDat - Universidad Carlos III de Madrid)

Ordering functions is a well-known problem in FDA. The statistical depth provides a criterion to order curves from center to outwards, while the epigraph and hypograph indices give an ordination from top to bottom or vice versa. This work proposes new definitions of these indices based on areas between curves, enhancing their ability to isolate outliers. These indices can be considered in several data analysis problems, demonstrating good performance on both synthetic and real datasets.

Joint work with Rosa E. Lillo and Alba M. Franco-Pereira.

Tuesday 14	**Martes 14**	**Asteartea 14**
17:30-17:50	**17:30-17:50**	**17:30-17:50**
[Room 0.16]	**[Aula 0.16]**	**[Gela 0.16]**

Cure models in cardio-oncology: dealing with functional and image covariates
Beatriz Piñeiro-Lamas
(Universidade da Coruña)

In medical practice, mixture cure models can be used to estimate the probability of experiencing a side effect and the distribution of the time until its appearance. Single-index mixture cure models, which avoid the curse of dimensionality in vector covariates, are extended to functional and image covariates, commonly used in medicine. The effect of Tissue Doppler Imaging in breast cancer patients receiving cardiotoxic therapies is analyzed.

Joint work with Ana López-Cheda and Ricardo Cao.

Tuesday 14　　　　　**Martes 14**　　　　　**Asteartea 14**
18:00-18:20　　　　　**18:00-18:20**　　　　　**18:00-18:20**
[Room 0.16]　　　　　**[Aula 0.16]**　　　　　**[Gela 0.16]**

Regularization techniques for Cox regression in high-dimensional survival analysis
Pilar González-Barquero
(Universidad Carlos III de Madrid)

Variable selection methods are crucial in high-dimensional settings, as the presence of numerous covariates complicates decision-making processes. In this context, Cox regression models are unfeasible since they present an infinite number of possible solutions for the regression coefficients. This study focuses on the proposal and evaluation of weight calculation methods for adaptive Lasso and their comparison with other regularization techniques for Cox regression in high-dimensional scenarios.

Joint work with Rosa E. Lillo and Álvaro Méndez-Civieta.

arXiv:2406.19213

Tuesday 14　　　　　**Martes 14**　　　　　**Asteartea 14**
18:30-18:50　　　　　**18:30-18:50**　　　　　**18:30-18:50**
[Room 0.16]　　　　　**[Aula 0.16]**　　　　　**[Gela 0.16]**

Random slope mixed models for Small Area Estimation of complex data
Naomi Diz-Rosales
(CITIC - Universidade da Coruña)

The increasing complexity of data in societal challenges, whether economic or health, highlights the need for flexible modelling tools. This study presents an innovative approach using random slope mixed models for small area estimation. We compute maximum likelihood estimates and random effects predictors, and evaluate the effectiveness of the methodology through simulation studies, demonstrating its relevance in addressing various social challenges.

Joint work with María José Lombardía and Domingo Morales.

arXiv:2404.18493

ROC curves with time-dependent variables
Arís Fanjul Hevia
(Universidad de Oviedo)

The Receiver Operating Characteristic (ROC) curve is a statistical tool that combines the concepts of sensitivity and specificity to evaluate a diagnostic marker's discriminatory capability. This work aims to explore existing alternatives in the literature for incorporating time-dependent variables into ROC curve analysis, particularly when comparing different diagnostic markers.

Joint work with Wenceslao González Manteiga and Juan Carlos Pardo Fernández.

Bootstrap prediction intervals for quantile autoregression
Silvia Novo
(Universidad Carlos III de Madrid)

New methods are proposed to compute prediction intervals in quantile autoregression models, both under homoscedasticity and in general quantile autoregression models. The proposed methods are based on quantile estimation, bootstrap multipliers to mimic the variability in coefficient estimation, and bootstrap replicates of future values. The consistency of the proposed methods is proven. Simulations and a real data analysis are provided to evaluate their finite-sample performance.

Joint work with César Sánchez-Sellero.

Thursday 16	**Jueves 16**	**Osteguna 16**
11:30-11:50	**11:30-11:50**	**11:30-11:50**
[Room 0.16]	**[Aula 0.16]**	**[Gela 0.16]**

Different applications and extensions of the distance covariance
Laura Freijeiro González
(Universidad de Oviedo)

In recent years, different extensions of the well-known distance covariance coefficient of Székely et al. (2007) have been proposed. These are devoted to quantifying distinct types of dependence, providing a deeper understanding of the underlying structure of the data. Additionally, these can be employed to characterize not only dependence but other structures as well. A brief review of these methodologies will be provided, jointly with some practical applications and extensions.

Joint work with Wenceslao González Manteiga and Manuel Febrero Bande.

Thursday 16	**Jueves 16**	**Osteguna 16**
12:00-12:20	**12:00-12:20**	**12:00-12:20**
[Room 0.16]	**[Aula 0.16]**	**[Gela 0.16]**

Quick quantile segmentation
Luis Alberto Rodríguez Ramírez
(Georg August-Universität Göttingen)

Change point estimation is formulated as finding the maximum of a gain function that improves data segmentation. Searching through all candidates requires O(n) evaluations, which can be computationally expensive. We propose optimistic search methods with O(log n) evaluations, exploiting the gain function's structure. This talk presents asymptotic consistency results for robust gain functions using empirical process theory. Further, efficiency bounds for optimistic search methods are also given.

Joint work with Housen Li and Axel Munk.

High density region estimation on manifolds
Diego Bolón

(Université Libre de Bruxelles)

Highest density regions (HDRs) are the sets where the density function of the data exceeds a given (and usually high) threshold. We introduce a new HDR estimator for manifold data that combines an underlying density estimator with some prior geometric information. The consistency of the new estimator is proven, and its consistency rate is derived. Finally, the performance in practice of the new HDR estimator is illustrated with a real data example.

Joint work with Rosa M. Crujeiras and Alberto Rodriguez-Casal.