# Innpar2D Workshop

## 10$^{th}$ December 2019

## Facultade de Matemáticas | Salón de Graos

**9:45 - 10:00**

**Workshop openning**

**10:00 - 10:30**

**Goodness-of-fit tests in proportional hazards models with random effects**

**María Dolores Martínez-Miranda (University of Granada)**

**Abstract.** In survival analysis the Cox hazard model is without any doubt the most used and well known semiparametric model to study the relationship between a survival time and a set of covariates. The popularity of this model can, among others, be explained by its easy interpretation and the fact that the nonparametric baseline hazard function cancels out in the likelihood, making the estimation of the parametric components as simple as in a purely parametric model. An appealing extension of the Cox model consists of adding random effects. This provides a powerful tool in a wide variety of applications, where the data have a natural clustered structure.

The assumption of linear covariate effects in the Cox model with (and without) random effects is quite strong in practice. Nevertheless, linearity is often assumed without any formal verification. This paper addresses the problem of testing the functional form of the covariate effects in a Cox model with random effects. The estimation of the model under the null is performed using (full) maximum likelihood. Under the alternative we consider orthogonal expansions and use again maximum likelihood. The test statistic is the likelihood ratio statistic and its distribution is approximated using bootstrap. The performance of the testing procedure is evaluated through simulations. The method is also applied on real survival data.

This is joint work with Wenceslao González-Manteiga and Ingrid Van Keilegom.

**10:30 - 10:50**

**Extent of occurrence estimation for invasive plants**

Paula Saavedra Nieves (Universidade de Santiago de Compostela)

**Abstract.** Support estimation deals with the problem of recontructing the compact support $S$ of an absolutely continuous random vector $X$ from an independent and identically distributed sample $\mathcal{X}_n$ of observations taken in it.

Several proposals for estimating $S$ have been considered in the literature. For instance, if we know that $S$ is convex then the convex hull of the sample is a natural support estimator. But convexity assumption may be too restrictive. In this case, it is necessary to consider more flexible shape restrictions such as $r-$convexity for $r > 0$. If we assume that $S$ is $r-$convex then the $r-$convex hull of $\mathcal{X}_n$ provides a natural reconstruction of the support. In Rodríguez-Casal (2007), it is proved that if the smoothing parameter $r$ is correctly chosen, the $r-$convex hull of the sample achieves the same convergence rates (in Hausdorff and distance in measure) as the convex hull. But, in practice, $S$ is unknown and, consequently, the real value of the geometric index $r$ too. Rodríguez-Casal and Saavedra-Nieves (2016) proposed a stochastic algorithm for selecting it from the data under the hypothesis that the sample is uniformly generated. The resulting reconstruction of $S$ is able to achieve the same convergence rates as the convex hull for estimating convex sets.

In this work, we will extend the previous result. We propose a consistent estimator of the optimal value of $r$ such that $S$ is $r-$convex under the assumption that the sample is generated from a Lipschitz continuous density bounded from below on $S$. The resulting estimator of $S$ also achieves the same convergence rates as the convex hull for estimating convex sets, but under a much more flexible scenario.

Our support estimator will be considered for reconstructing the *extent of occurrence* (EOO) of invasive plants. Roughly speaking, the International Union for Conservation of Nature defines the EOO of a specie as the convex hull of its geographical locations. We will show that this definition presents some limitations that our estimator is able to overcome.

This is a joint work with Alberto Rodríguez-Casal.

**10:50 - 11:10**

**Estimators for ROC curves with missing values using additional information of covariates**

Ana Pérez González (Universidade de Vigo)

**Abstract.** The Receiver Operating Characteristic (ROC) are a very well-known technique in medical studies where a continuous variable or marker (biomarker) is used to diagnose a disease or to evaluate the progression of a disease.

It is common to assume that all the biomarker responses are available. However, in practical situations the missing data occur frequently such as medical research, opinion polls, etc.

We are especially interested in the situation when missing responses arise and the practitioner has some covariates with predictive capability on the missingness. Under this setting, we propose two estimators of the ROC Curve using the inverse probability weighting scheme. A simulation study shows the behavior of both of them.

**11:10 - 11:40**

**On a semiparametric estimation method for AFT mixture cure models**

**Ingrid Van Keilegom (Université catholique de Louvain)**

**Abstract.** When studying survival data in the presence of right censoring, it often happens that a certain proportion of the individuals under study do not experience the event of interest and are considered as cured. The mixture cure model is one of the common models that take this feature into account. It depends on a model for the conditional probability of being cured (called the incidence) and a model for the conditional survival function of the uncured individuals (called the latency).

This work considers a logistic model for the incidence and a semiparametric accelerated failure time model for the latency part. The estimation of this model is obtained via the maximization of the semiparametric likelihood, in which the unknown error density is replaced by a kernel estimator based on the Kaplan-Meier estimator of the error distribution. Asymptotic theory for consistency and asymptotic normality of the parameter estimators is provided. Moreover, the proposed estimation method is compared with a method proposed by Lu (2010), which uses a kernel approach based on the EM algorithm to estimate the model parameters. Finally, the new method is applied to data coming from a cancer clinical trial.

**11:40 - 12:15**

Coffee break

**12:15 - 13:15**

**Project meeting**

**13:15 - 16:00**

Lunch break

**16:00 - 16:30**

**Sparse logistic regression: a robust approach**

**Ana M. Bianco (Universidad de Buenos Aires and CONICET)**

**Abstract.** Sparse statistical models correspond to situations where there are only a small number of non–zero parameters and for that reason, they are much easier to interpret than dense ones. In this talk, we focus on the logistic regression model and our aim is to address robust and penalized estimation for the regression parameter. We introduce a family of penalized weighted $M-$type estimators for the logistic regression parameter that are stable against atypical data. We explore different penalizations functions and we introduce the so–called Sign penalization. This new penalty has the advantage that it depends only on one penalty parameter, avoiding arbitrary tuning constants.

When a model has a sparse representation, discovering relevant predictive variables is a fundamen-

tal goal. We discuss the variable selection ability of the given proposals as well as their asymptotic behaviour. Through a numerical study, we compare the finite sample performance of the proposal corresponding to different penalized estimators either robust or classical, under different scenarios. A robust cross–validation criterion is also presented. The analysis of two real data sets enables to investigate the stability of the penalized estimators to the presence of outliers.

This is a joint work with Graciela Boente and Gonzalo Chebi.

**16:30 - 16:50**

**Nonparametric regression for circular variables with different groups of observations**

**María Alonso Pena (Universidade de Santiago de Compostela)**

**Abstract.** Circular data are observations which can be expressed as angles, and can be usually found as directions or periodic measurements. The behavior of circular variables can be analyzed jointly with other variables, either circular or real-valued, from a regression perspective. More specifically, within kernel-type regression models, ANCOVA models can be formulated where, in addition to a continuous predictor, a cathegorigal predictor determining different groups of observations is also considered. On the other hand, it might also be possible to identify the presence of different groups by simply looking at the data, but in a context where the variable determining the different groups is not provided. A useful tool in this setting could be modal regression, which instead of estimating the conditional mean of the response variable, estimates the conditional mode or modes. In this talk, nonparametric ANCOVA models for circular variables will be addressed and an introduction to modal regression models for angular observations will also be presented.

This is a joint work with Rosa Mª Crujeiras and Jose Ameijeiras-Alonso.

**16:50 - 17:10**

**Nonparametric tests to compare the first-order structure of inhomogeneous spatial point processes**

**Isabel Fuentes Santos (Instituto de Investigaciones Marinas, CSIC)**

**Abstract.** A common question in the analysis of multitype spatial point processes is whether two types of events have the same spatial structure. This question can arise in different scientific areas, such as ecology, environmental risk assessment, epidemiology, or criminology.

A natural way to test whether two spatial point patterns have the same spatial structure is through comparison of their first-order intensities, which describe the spatial distribution of events in the observation domain. In this line, we have develop a Cramer-von-Mises type test and a regression test based on the log-relative risk function. The former, measures the square discrepancy between their densities of event locations, we use kernel smoothing to compute the test and a smooth bootstrap procedure for calibration. The latter tests whether the ratio between the first-order intensities under comparison is spatially invariant, we first compute the kernel log-relative risk function ant test its dependence on the spatial location of events through a nonparamtric regression test. Smooth bootstrap and a permutation test have been used for calibration.

The performance of these tests has been analyzed in as simulation study, which compares our proposals with a Kolmogorov-Smirnov test with asymptotic calibration, and through application to the

analysis of wildfire patterns in Galicia and gunfire in Rio de Janeiro.

This is a joint work with Wenceslao González Manteiga and Jorge Mateu.

**17:10 - 17:40**

**Robust B-splines estimators in generalized partly linear regression under monotone constraints**

**Graciela Boente (Universidad de Buenos Aires and CONICET)**

**Abstract.** In this talk, we will consider the situation in which the observations follow an isotonic generalized partly linear model. Under this model, the mean of the responses is modelled, through a link function, linearly on some covariates and nonparametrically on an univariate regressor in such a way that the nonparametric component is assumed to be a monotone function.

A class of robust estimates for the monotone nonparametric component and for the regression parameter will be defined. The robust estimators are based on a spline approach combined with a score function which bounds large values of the deviance. Results regarding the asymptotic behaviour of the proposal will be presented. A numerical study allows to illustrate the advantages of the proposal.

This is a joint work with Daniela Rodriguez and Pablo Vena.