

Workshop on
Advances in Statistical Inference
July 19–20, 2018

Facultade de Ciencias Económicas e Empresariais
Aula Seminario 8

Thursday, July 19th

Testing problems

11:30-12:00

gBA para contrastar envejecimiento DPRL

Alba Franco Pereira (Universidad Complutense de Madrid)

Abstract: En este trabajo se propone un método para llevar a cabo un contraste de hipótesis que permita concluir cuando una variable verifica la noción de envejecimiento DPRL (función percentil de la vida residual decreciente). La motivación de este trabajo es un trabajo de Franco-Pereira and de Uña-Álvarez (2013) en el que se estudiaba un nuevo estimador de la función percentil de la vida residual con restricción de monotonía. Este estimador monótono es recomendable cuando se cumple la restricción de monotonía por tener menor error cuadrático medio que el empírico.

El estadístico del contraste propuesto está basado en una distancia entre el estimador monótono y el estimador empírico de la función percentil de la vida residual. Obtenemos la distribución asintótica del estadístico del contraste bajo la hipótesis de decrecimiento de la función percentil de la vida residual y proponemos un algoritmo bootstrap para utilizar en la práctica. Además, se muestran los resultados de un estudio de simulación y una aplicación a datos reales.

Este es un trabajo conjunto con Pablo Martínez Cambor y Jacobo de Uña Álvarez.

12:00-12:30

A model specification test for the variance function in non-parametric regression

Juan Carlos Pardo Fernández (Universidade de Vigo)

Abstract: The problem of testing for the parametric form of the conditional variance is considered in a fully non-parametric regression model. A test statistic based on a weighted L_2 -distance between the empirical characteristic functions of residuals constructed under the null hypothesis and under the alternative is proposed and studied theoretically. The null asymptotic distribution of test statistic is obtained and employed to approximate the critical values. Finite sample properties of the proposed test are numerically investigated in several Monte Carlo experiments. The developed results assume independent data. Their extension to dependent observations is also discussed.

This is joint work with M. Dolores Jiménez Gamero and is based on the paper:

J.C. Pardo-Fernández and M.D. Jiménez-Gamero: A model specification test for the variance function in nonparametric regression. To appear in *AStA - Advances in Statistical Analysis*.

12:30-13:00

A characterization of multinormality and corresponding tests of fit for i.i.d. and GARCH data

M.D. Jiménez Gamero (Universidad de Sevilla)

Abstract: A novel characterization of multivariate normality is provided. It incorporates both the characteristic function and the moment generating function. It is employed to construct a class of affine invariant, consistent and easy-to-use goodness-of-fit tests for normality. The test statistics are suitably weighted L^2 -statistics. It is studied their asymptotic behavior both for i.i.d. observations as well as in the context of testing that the innovation distribution of a multivariate GARCH model is Gaussian. It is also studied the finite-sample behavior of the new tests and compared with alternative existing tests.

The talk is based on the paper:

N. Henze, M.D. Jiménez-Gamero, S.G. Meintanis: A novel characterization of multinormality and corresponding tests of fit, including for GARCH models. To appear in *Econometric Theory*, DOI 10.1017/S0266466618000154

Problems in biostatistics

15:30-16:00

Nueva medida de acuerdo nominal entre observadores: coeficiente Alfa

María Álvarez Hernández (Centro Universitario de la Defensa - Escuela Naval Militar)

Abstract: En muchos campos de la ciencia es frecuente la necesidad de evaluar el grado de acuerdo o concordancia entre R observadores que clasifican independientemente a n sujetos dentro de K categorías no ordenadas. Las medidas más populares son los coeficientes kappa de Cohen ($R = 2$), kappa de Fleiss ($R > 2$) y kappa de Hubert ($R \geq 2$), los cuales dependen fuertemente de las distribuciones marginales. En 2004 se definió el coeficiente Delta para el caso $R = 2$, que no tenía los defectos del coeficiente kappa de Cohen. El coeficiente Delta es válido en cualquier circunstancia, esté fijada o no la distribución marginal de uno de los dos observadores, pero presenta el inconveniente de estar definido bajo el supuesto de que uno de los dos observadores es un estándar. En este trabajo se define un nuevo coeficiente Alpha que es válido para $R \geq 2$, está definido para el caso de que no exista un estándar, es una extensión del coeficiente Delta y no se deja afectar por las distribuciones marginales. Una ventaja adicional de este modelo Alpha es que no solo permite medir el grado de acuerdo global, sino también el grado de acuerdo en cada categoría.

16:00-16:30

The three-state progressive model under various types of censoring and truncation

Carla Moreira (Instituto de Saúde Pública da Universidade do Porto)

Abstract: The statistical analysis of consecutive gap times is an issue of much importance in a number of fields, including engineering, economy, epidemiology, and survival analysis. Particular difficulties appear, for example, when information on a cohort is obtained through intermittent visits or successive cross-sections; in such cases, special combinations of left-truncated, right-censored and interval censored data will appear. We describe one of such complicated settings in a three-state progressive model, and we introduce an inverse-probability-weighted type estimator for the joint distribution of two

gap times which takes the aforementioned censoring and truncation issues into account. The performance of the proposed estimator is investigated through simulations, considering scenarios with independent and dependent gap times. For illustration purposes, the estimator is applied to data from the EPIPorto adults' cohort study.

This is joint work with Jacobo de Uña Álvarez and Ana Cristina Santos.

16:30-17:00

Studying causal effects in Cox models from observational designs

Pablo Martínez Camblor (Dartmouth Institute for Health Policy and Clinical Practice)

Abstract: Instrumental variables (IV) methods are widely used for estimating average treatments effects in the presence of unmeasured confounders. However, the capability of existing IV procedures, and most notably the two-stage residual inclusion (2SRI) algorithm recommended for use in non-linear contexts, to account for unmeasured confounders in the Cox proportional hazard model is unclear. After a brief introduction about the causality problem, we show that instrumenting an endogenous treatment induces an individual frailty and prove that it is consistent under certain conditions. The finite sample size behavior is studied across a broad set of conditions via Monte Carlo simulations. Finally, the proposed methodology is used to estimate the average effect of carotid endarterectomy versus carotid stenting on the mortality of patients suffering from carotid artery disease. Results suggest that the 2SRI-frailty estimator generally reduces the bias of both point and interval estimators compared to traditional 2SRI.

Friday, July 20th

Estimation and testing in non-standard settings

10:00-10:30

Comparing several multiple testing methods for discrete uniform p -values

Marta Cousido Rocha (Universidade de Vigo)

Abstract: Several multiple testing procedures for discrete homogeneous p -values are compared through simulations. Specifically, we consider the q -value approach based on several estimators for the proportion of true null hypotheses π_0 : Storey (2002) (continuous, possibly heterogeneous p -values), Liang (2016) (discrete uniform, homogeneous p -values) and Chen et al. (2014) (in the particular case of discrete homogeneous p -values). As a benchmark, we consider the q -value approach based on the true π_0 . The scenario is that of the two-sample problem with low sample size, along a large number of locations or genes. The test statistics are the standard Student's t test, a permutation test based on the absolute deviation between the sample means, the Kolmogorov-Smirnov two-sample test, and a permutation test based on the L_2 distance between the empirical characteristic functions pertaining to the two samples. The simulation study includes differences in both location and scale/shape, and various degrees of departure between the samples. The main conclusion is that the specific estimator for π_0 influences the power a lot, and that the approaches for discrete p -values may or may not improve the original q -value procedure in Storey (2002) for continuous p -values depending on the specific situation.

References:

- Chen, X., Doerge, R.W., and Heyse, J.F. (2014). Methodology Multiple testing with discrete data: proportion of true null hypotheses and two adaptive FDR procedures. arXiv:1410.4274v2
- Liang, K. (2016). False Discovery Rate Estimation for Large-Scale Homogeneous Discrete p -Values. *Biometrics*, 72, 639-648.
- Storey, J. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 64, 479-498.

10:30-11:00

Tests de bondad de ajuste para la detección de trastornos en experimentos de NGS y estimación de parámetros

Norman Jiménez Otero (Universidade de Vigo)

Abstract: Los experimentos de secuenciación de nueva generación (NGS) son comunes en investigaciones biomédicas y conllevan desafíos metodológicos relacionados con la alta dimensión y naturaleza compleja de los datos. En este trabajo revisamos algunos de los problemas que surgen en la detección de alteraciones genéticas, y un procedimiento de bondad de ajuste para la detección de trastornos. Además proponemos un estimador cuando los parámetros de la secuenciación no son accesibles.

11:00-11:30

Smoothing parameter selection in set estimation

Paula Saavedra Nieves (Instituto Tecnológico de Matemática Industrial)

Abstract: Set estimation theory deals with the problem of reconstructing a set in the Euclidean space from a random finite sample of points whose distribution is closely related to it. Our interest is focused on support and density level set estimation under r -convexity assumption. This is a much more flexible and general shape condition than convexity. In addition, the smoothing parameter r represents a geometric characteristic of the set to be reconstructed and, therefore, it is unknown. In this talk, we will select its optimal value from the data in order to derive fully automatic estimators for the support and density level sets.

Advances in regression

12:30-13:00

Regresión restringida para datos altamente dependientes con restricciones de igualdad e desigualdade nos coeficientes. Aplicación a un caso real

Tomás Cotos Yáñez (Universidade de Vigo)

Abstract: Neste traballo estúdanse os modelos lineais de regresión con co-variables altamente correlacionadas ou cando p é moi próximo a n , coa complexidade de que hai restriccións nos coeficientes. Concretamente centrarémonos na regresión Ridge. Motivado pola aplicación a un caso real,

determinación do desprazamento químico de ^{13}C NMR (Nuclear Magnetic Resonance), propóñense diferentes alternativas de estimación deste tipo de modelos lineais de forma iterativa e en dúas etapas.

Este é un traballo conxunto con Ana Pérez González e Luis Muñoz.

13:00-13:30

Modelos GAMLSS. Aplicación al cálculo de percentiles de rendimiento en remo

M. Carmen Iglesias Pérez (Universidade de Vigo)

Abstract: Los modelos aditivos generalizados de posición, escala y forma (GAMLSS) son un tipo de modelos de regresión semiparamétricos en los cuales se asume una distribución paramétrica para la variable respuesta, pero los parámetros de dicha distribución pueden ser modelados como funciones suaves de variables explicativas. Los GAMLSS se vienen usando en los últimos años en diferentes áreas de investigación para la construcción y cálculo de percentiles. En este trabajo se presenta una introducción de estos modelos y una aplicación al cálculo de percentiles de rendimiento en una prueba máxima de 2000 metros sobre ergómetro de remo.

13:30-14:00

Predicting pollution incidents through semiparametric quantile regression models

Javier Roca Pardiñas (Universidade de Vigo)

Abstract: We present a method to forecast pollution episodes using measurements of the pollutant concentration along time. Specifically, we use a backfitting algorithm with local polynomial kernel smoothers to estimate a semiparametric additive quantile regression model. We also propose a statistical hypothesis test to determine critical values, i.e., the values of the concentration that are significant to forecast the pollution episodes. This test is based on a wild bootstrap approach modified to suit asymmetric loss functions, as occurs in quantile regression. The validity of the method was checked with both simulated and real data, the latter from SO₂ emissions from a coal-fired power station located in Northern Spain.