

SESIONES ESPECIALES

Congreso RSME 2013



S17

Estadística

Mar 22, 11:00 - 11:55, Aula 7 — Wenceslao González Manteiga:
Goodness-of-fit tests for regression models: the functional data case

Mar 22, 12:00 - 12:25, Aula 7 — Juan Carlos Pardo Fernández:
On the use of the characteristic function of the residuals to test for the equality of regression curves

Mar 22, 12:30 - 12:55, Aula 7 — Carla Moreira:
Goodness-of-fit tests for a semiparametric model under random double truncation

Mar 22, 13:00 - 13:25, Aula 7 — Pablo Martínez Camblor:
A general bootstrap algorithm for hypothesis testing

Mar 22, 17:00 - 17:55, Aula 7 — Ricardo Cao:
Maximum likelihood estimation for conditional distribution single-index models under censoring

Mar 22, 18:00 - 18:25, Aula 7 — María del Carmen Iglesias Pérez:
CLT of nonparametric regression estimator with truncated, censored and dependent data

Mar 22, 18:30 - 18:55, Aula 7 — Javier Roca Pardiñas:
Métodos estadísticos de selección de variables en modelos de regresión multivariantes. Aplicaciones en medioambiente y salud

Mar 22, 19:00 - 19:25, Aula 7 — Mar Rodríguez Gironde:
Model building in non proportional hazard regression

Goodness-of-fit tests for regression models: the functional data case

Wenceslao González-Manteiga¹, Juan Cuesta-Albertos², Manuel
Febrero-Bande¹, Eduardo García-Portugués¹

In this talk the topic of the goodness-of-fit for regression models with functional covariates is considered. Although several papers have been published in the last two decades for the checking of regression models, the case where the covariates are functional is quite recent and has become of interest in the last years. We will review the very recent advances in this area and we will propose a new goodness-of-fit test for the null hypothesis of a functional linear model with scalar response. Our test is based on a generalization to the functional framework of a previous one, designed for the goodness-of-fit of regression models with multivariate covariates using random projections. The test statistic is easy to compute using geometrical and matrix arguments, and simple to calibrate in its distribution by a wild bootstrap on the residuals. Some theoretical aspects are derived and the finite sample properties of the test are illustrated by a simulation study. Finally, the test is applied to real data for checking the assumption of the functional linear model and a graphical tool is introduced.

Keywords: Functional Data, Goodness-of-fit, Bootstrap/resampling, Statistical computing

MSC 2010: 62G05, 62G09, 62G10

References

- [1] E. GARCÍA-PORTUGUÉS, W. GONZÁLEZ-MANTEIGA AND M. FEBRERO-BANDE, A goodness-of-fit for the functional linear model. *arXiv:1205.6167* (second revision in *Journal of Computational and Graphical Statistics*).
- [2] L. HORVATH AND P. KOKOSZKA, *Inference for Functional Data With Applications*. Springer, 2012. New York.
- [3] J. O. RAMSAY AND B. W. SILVERMAN, *Functional data analysis*. Springer, 2005. New York.

¹Department of Statistics and Operations Research
University of Santiago de Compostela
Faculty of Mathematics

²Department of Mathematics
University of Cantabria
Faculty of Sciences

On the use of the characteristic function of the residuals to test for the equality of regression curves

Juan Carlos Pardo-Fernández¹

In this talk we will present a new procedure to test for the equality of k regression curves in a fully nonparametric context. The test is based on the comparison of two empirical estimators of the characteristic functions of the regression residuals in each population. The asymptotic behaviour of the test statistic is studied in detail. It is shown that, under the null hypothesis, the distribution of the test statistic converges to a combination of independent χ_1^2 random variables. Under certain restrictions on the populations, the asymptotic null distribution of the test statistic is χ_{k-1}^2 . The practical performance of the test based on the asymptotic null distribution is investigated by means of simulations.

This is joint work with M. Dolores Jiménez-Gamero (Universidad de Sevilla) and Anouar El Ghouch (Université catholique de Louvain).

Keywords: Comparison of regression curves, Empirical characteristic function, Regression residuals.

MSC 2010: 62G08, 62G10

¹Departamento de Estatística e Investigación Operativa
Universidade de Vigo
Facultade de Ciencias Económicas e Empresariais
Campus Universitario As Lagoas-Marcosende
36310 Vigo (Pontevedra)
juancp@uvigo.es

Goodness-of-fit tests for a semiparametric model under random double truncation

Carla Moreira^{1,2}, Jacobo de Uña-Álvarez¹, Ingrid Van Keilegom³

Doubly truncated data are commonly encountered in areas like medicine, astronomy, economy, among others. A semiparametric estimator of a doubly truncated random variable has been proposed by [1]. Their estimator is based on a parametric specification of the distribution function of the truncation times. This semiparametric estimator outperforms the nonparametric maximum likelihood estimator when the parametric information is correct, but might behave badly when the assumed parametric model is far off. We introduce several goodness-of-fit tests for the parametric model. The proposed tests are investigated through simulations. For illustration purposes, the tests are also applied to data on the induction time to AIDS for blood transfusion patients.

Keywords: Double truncation, goodness-of-fit tests

MSC 2010: 62N86, 62G10

References

- [1] MOREIRA, C. AND DE UÑA-ÁLVAREZ, J., A semiparametric estimator of survival for doubly truncated data. *Statistics in Medicine* **29**, 3147–3159 (2010).

¹Department of Statistics and OR
University of Vigo
Lagoas - Marcosende, 36 310, Vigo - Spain
carla@uvigo.es, jacobode@uvigo.es

²Centro de Matemática da Universidade do Minho
Campus de Gualtar, 4710 - 057 Braga - Portugal

³Institute of Statistics, Biostatistics and Actuarial Sciences
Université catholique de Louvain
Voie du Roman Pays 20, B 1348 Louvain-la-Neuve - Belgium
ingrid.vankeilegom@uclouvain.be

A general bootstrap algorithm for hypothesis testing

Pablo Martínez-Cambor^{1,2}, Norberto Corral²

The bootstrap, introduced and explored in detail by Efron [1] [2], is a intensive computer-based method originally devoted to estimate the standard deviations, confidence intervals and bias of the studied statistic. This technique is useful in a wide variety of statistical procedures, however, its use for hypothesis testing, when the data structure is complex, is not straightforward and each case must be particularly treated. In this work, a general bootstrap algorithm (gBA) for hypothesis testing is studied. The considered method preserves the data structure of each group independently and, in spite of most of the authors match that the resampling under the null is critical to the proper construction of bootstrap test (see, for example, Hall and Wilson [3] or Westfall and Young [4]), in the introduced procedure, the null hypothesis is only used in order to compute the bootstrap statistic values (not at the resampling, as usual). The asymptotic distribution is developed and several case studies are discussed.

Keywords: Gini index, Survival model, Competing risk, Cumulative incidence function

MSC 2010: 62F03, 62F40

References

- [1] B. EFRON, (1979) Bootstrap methods: another look at the jackknife, *Annals of Statistics*, **7**(1), 1-26.
- [2] B. EFRON, (1982) The jakknife, the bootstrap and other resampling plans, *Regional Conference Series in Applied Mathematics*, CBMS-NSF.
- [3] P. HALL, S.R. WILSON, (1991) Two guidelines for bootstrap hypothesis testing, *Biometrics*, **47**, 757-762.
- [4] P.H. WESTFALL, S.S. YOUNG, (1993) *Resampling-based Multiple Testing: Examples and Methods for p-value Adjustment*, Wiley, New York.

¹Oficina de Investigación Biosanitaria de Asturias
pablomc@ficyt.es

²Departamento de Estadística e IO y DM
Universidad de Oviedo
norbert@uniovi.es

Maximum likelihood estimation for conditional distribution single-index models under censoring^{*}

Ricardo Cao¹, Ewa Strzalkowska-Kominiak²

A new likelihood approach is proposed for the problem of semiparametric estimation of a conditional distribution or density under censoring. Consistency and asymptotic normality for two versions of the maximum likelihood estimator of the parameter vector in the single index model are proved. The single-index model considered can be seen as a useful tool for credit scoring and estimation of the default probability in credit risk. A data-driven bandwidth selection procedure is proposed. It allows to choose the smoothing parameter involved in our approach. The finite sample performance of the estimators has been studied by simulations, where the new method has been compared with the method proposed by Bouaziz and Lopez (2010) [1]. To the best of our knowledge this is the only existing competitor in this context. The simulation study shows the good behavior of the proposed method.

Keywords: conditional density function, credit risk, kernel estimation, survival analysis

MSC 2010: 62G07, 62G20, 62N01, 62N02

References

- [1] O. BOUAZIZ, O. LOPEZ, Conditional density estimation in a censored single-index model. *Bernoulli* **16**, 514–542 (2010).

¹Departamento de Matemáticas
Universidade da Coruña
Facultade de Informática, Campus de Elviña, 15071 A Coruña, Spain
rcao@udc.es

²Departamento de Estadística
Universidad Carlos III de Madrid
C/ Madrid, 126, 28903 Getafe (Madrid) Spain
ewa.strzalkowska@uc3m.es

^{*}MINECO Grants MTM2008-00166 and MTM2011-22392 (EU ERDF support included)

CLT of nonparametric regression estimator with truncated, censored and dependent data

**Liang, Han-Ying¹; de Uña-Álvarez, Jacobo²; Iglesias-Pérez, María del
Carmen³**

We propose a method to estimate the regression function under a left truncation and right censored model. It is assumed that the observations form a stationary α -mixing sequence. The new estimators are based on the idea of the Nadaraya-Watson and the local lineal smoothers applied to a certain transformation of the data. Asymptotic normality of the estimators is established. Their finite sample behavior is also investigated in a simulation study.

Keywords: Asymptotic normality, Nadaraya-Watson and local linear smoothing, regression function, truncated and censored data, α -mixing.

MSC 2010: 62N01, 62E20.

¹Department of Mathematics
Tongji University
Shanghai 200092. P.R. China
hyliang83@yahoo.com

²Departamento de Estadística e Investigación Operativa
Universidad de Vigo
Facultad de Ciencias Económicas y Empresariales.
Campus Lagoas-Marcosende. 36310 Vigo. España
jacobou@uvigo.es

³Departamento de Estadística e Investigación Operativa
Universidad de Vigo
Escuela de Ingeniería Forestal.
Campus A Xunqueira. 36005 Pontevedra. España
mcigles@uvigo.es

Métodos estadísticos de selección de variables en modelos de regresión multivariantes. Aplicaciones en salud

Javier Roca Pardiñas¹

En muchas aplicaciones biomédicas, el interés reside en ser capaz de distinguir entre dos posibles estados de una variable de respuesta dada, dependiendo de los valores de determinados predictores. Si el número de predictores es elevado, o si existe una redundancia entre ellos, entonces será muy importante decidir sobre el mejor subconjunto de predictores que dará lugar el modelo con mejor capacidad de discriminación. Con este objetivo, se han considerado los modelos aditivos generalizados de respuesta binaria, y se ha utilizado la “receiver operating characteristic (ROC)” con el fin de determinar y comparar la capacidad discriminatoria de tales modelos. En este estudio se han desarrollado pruebas basadas en el remuestreo bootstrap que permiten: a) obtener el número óptimo de predictores; y b) el modelo o modelos con ese número de predictores, que presentan el mayor AUC (área bajo la curva ROC). El método propuesto se aplicó a un sistema de diagnóstico asistido por ordenador dedicado a la detección precoz del cáncer de mama.

¹Departamento de Estadística e I.O
Universidade de Vigo
E.U. Ingeniería Industrial
C/Torrecedeira 86, 36208
Vigo (Pontevedra)
roca@uvigo.es

Model building in non proportional hazard regression

Mar Rodríguez-Girondo¹, Thomas Kneib², Carmen Cadarso-Suárez³, Emad Abu-Assi⁴

Recent developments of statistical methods allow for a very flexible modeling of covariates affecting survival times via the hazard rate, including also the inspection of possible time-dependent associations. Despite their immediate appeal in terms of flexibility, these models typically introduce additional difficulties when a subset of covariates and the corresponding modeling alternatives have to be chosen, i.e. for building the most suitable model for given data. This is particularly true when potentially time-varying associations are given. We propose to conduct a piecewise exponential representation of the original survival data to link hazard regression with estimation schemes based on the Poisson likelihood to make recent advances for model building in exponential family regression accessible also in the non proportional hazard regression context. A two-stage stepwise selection approach, an approach based on doubly penalized likelihood and a componentwise functional gradient descent approach will be adapted to the piecewise exponential regression problem and compared via an intensive simulation study. An application to prognosis after discharge for patients who suffered a myocardial infarction supplements the simulation to demonstrate the pros and cons of the approaches in real data analyses.

Keywords: Survival analysis, Variable selection, Model Choice, Boosting, Generalized additive models

MSC 2010: 62P10, 62N01, 62O7

¹siDOR Research Group
University of Vigo
Facultade de Ciencias Económicas e Empresariais
Campus Lagoas-Marcosende
Vigo, SPAIN
margirondo@uvigo.es

²Center for Statistics and Econometrics
Georg August University
Platz der Göttinger Sieben 5
Göttingen, GERMANY
tkneib@uni-goettingen.de

³Unit of Biostatistics, Department of Statistics
University of Santiago de Compostela
Facultade de Medicina e Odontoloxía
Rúa San Francisco s/n
Santiago de Compostela, SPAIN
carmen.cadarso@usc.es

⁴Department of Cardiology
Hospital Clínico de Santiago de Compostela
Facultade de Medicina e Odontoloxía
A Choupana s/n
Santiago de Compostela, SPAIN
eabuassi@yahoo.es