# ISNI 2008
# Scientific Programme and Proceedings

Vigo, 5–7 November, 2008

Edited by Juan Carlos Pardo-Fernández and Jacobo de Uña-Álvarez

# Preface

The *International Seminar on Nonparametric Inference* – ISNI2008 is held in the Faculty of Economics and Business of the University of Vigo (Galicia, Spain) on 5–7 November, 2008. The meeting was promoted by the three Galician research groups in nonparametric statistics (Vigo, Santiago de Compostela, and A Coruña), as well as by the close collaborators in the Scientific Committee. The main goal of ISNI2008 is to facilitate the exchange of research ideas in the field of nonparametric statitics, and the establishment of new fruitful scientific collaborations.

More than one hundred researchers participate in the seminar. With the program based on seventeen plenary Invited Talks given by leading researchers in their respective areas, ISNI2008 also allows for the presentation of contributed papers. More than thirty contributions were submitted, and six of them were selected by the Scientific Commitee according to their relevance for oral presentation in plenary sessions. Important topics as Functional Data Analysis, Statistical Learning, Survival Analysis, Finance, Econometrics, Extremes, Semiparametric Modelling, Goodness-of-fit Testing, Time Series, Spatial Statistics or Resampling are covered. The *Journal of Nonparametric Statistics* will dedicate a special issue to papers inspired by the conference.

The research group in Statistical Inference, Decision and Operations Research of the University of Vigo is responsible for the local organization. Main funding came from the IAP Attraction Pole, the Spanish Ministry of Science and Innovation, the Branch of Education and Universities of the Xunta de Galicia, the University of Vigo itself, and the local Administrations. Other institutions, as the Institute of Mathematical Statistics, the Section on Nonparametric Statistics of the ASA, the Bernoulli Society, or the Spanish and Galician Societies for Statistics and Operations Research, support the meeting scientifically and help with its dissemination.

# Scientific Committee


- Ricardo Cao
  University of A Coruña, Spain
- Holger Dette
  Ruhr University of Bochum, Germany
- Wenceslao González-Manteiga
  University of Santiago de Compostela, Spain
- Gábor Lugosi
  Pompeu Fabra University of Barcelona, Spain
- Geert Molenberghs
  Hasselt University, Belgium
- Jean Opsomer
  Colorado State University, USA
- Stefan Sperlich
  University of Gottingen, Germany
- Winfried Stute
  University of Giessen, Germany
- Jacobo de Uña-Álvarez
  University of Vigo, Spain
- Ingrid Van Keilegom
  Catholic University of Louvain, Louvain-la-Neuve, Belgium
- Noël Veraverbeke
  Hasselt University, Belgium
- Philippe Vieu
  Paul Sabatier University of Tolouse, France

# Local Organizing Committee

The following members of SiDOR (Statistical Inference, Decision and Operations Research) group, University of Vigo, are responsible for the organization of local issues related to ISNI2008:

- Tomás Cotos Yáñez
- Gloria Fiestras Janeiro
- M. Carmen Iglesias Pérez
- Han-Ying Liang
- Manuel Alfredo Mosquera Rodríguez
- Carla Moreira
- José Vicente Novegil Souto
- Juan Carlos Pardo Fernández
- Ana Pérez González
- Javier Roca Pardiñas
- Celia Rodríguez Campos
- Ángelez Saavedra
- Estela Sánchez Rodríguez
- Jacobo de Uña Álvarez (main responsible)

# Contents

# I

# Scientific Programme

## Wednesday, November 5

---

**09:00** Registration

**09:30** Opening ceremony

---

**10:00** Invited talk 1 – **P. Vieu**

*Functional Data Analysis: nonparametric estimation and structural regression tests*

**10:50** Invited talk 2 – **H.-G. Müller**

*Functional Additive Modeling*

---

**11:40** Coffe-break

---

**12:10** Invited talk 3 – **J.C. Pardo-Fernández**

*New perspectives about error distribution based tests in nonparametric regression*

**13:00** Invited talk 4 – **H. Dette**

*Nonparametric analysis of covariance using quantile curves*

---

**13:50** Lunch break

---

**15:20** Poster session for contributed papers[1]

---

**16:00** Invited talk 5 – **N. Neumeyer**

*Testing for monotonicity of nonparametric regression functions*

---

**16:50** Coffee-break

---

**17:20** Invited talk 6 – **S. Sperlich**

*Inference problems with resampling*

---

[1] posters will be available along the 3 days

## Thursday, November 6

---

**09:30** Invited talk 7 – **G. Molenberghs**

*Every missing not at random model for incomplete data has got a missing at random counterpart with equal fit*

**10:20** Invited talk 8 – **W. Stute**

*Multivariate Kaplan-Meier Estimator*

---

**11:10** Coffee-break

---

**11:40** Invited talk 9 – **G. Claeskens**

*Goodness-of-fit tests in mixed models*

**12:30** Invited talk 10 – **P. Hall**

*Nonlinear methods for variable selection*

---

**13:20** Lunch break

---

**14:50** Oral session for selected contributed papers 1

14:50 J. Mora, **A. Pérez-Alonso** (presenting author): *Specification tests for the distribution of errors in nonparametric regression: a martingale approach*

15:10 **J.C. Escanciano** (presenting author), D. Jacho-Chávez: *$\sqrt{n}$-uniformly consistent density estimation in nonparametric regression models*

15:30 **R. Mukherjee** (presenting author), M. Kosorok, J. Fine: *Efficient estimation for the Accelerated Failure Time Model for forward recurrence times*

**15:50** Invited talk 11 – **A. Davison**

*Geostatistics of extremes*

---

## Friday, November 7

---

**09:30** Invited talk 12 – **G. Lugosi**

*Randomized sequential prediction: performance and algorithms*

**10:20** Invited talk 13 – **A. Antoniadis**

*Wavelet methods in statistics: some recent developments and their applications*

---

**11:10** Coffee-break

---

**11:40** Invited talk 14 – **I. Van Keilegom**

*Empirical likelihood for non-smooth criterion functions*

**12:30** Invited talk 15 – **J. Swanepoel**

*A general uniform in bandwidth consistency result with applications*

---

**13:20** Lunch break

---

**14:50** Oral session for selected contributed papers 2

14:50 **J. Dony** (presenting author): *Uniform in bandwidth consistency of the kernel–based Hill estimator*

15:10 **G. Boente** (presenting author), R. Cao, W. González-Manteiga, D. Rodriguez: *Robust inference in generalized partially linear models*

15:30 **J.A. Vilar Fernández** (presenting author), A.M. Alonso, J.M. Vilar Fernández: *Time series clustering based on nonparametric forecast*

**15:50** Invited talk 16 – **J. Opsomer**

*Shape-restricted regression in the presence of correlated errors*

---

**16:40** Coffee-break

---

**17:10** Invited talk 17 – **L. Duembgen**

*Least squares and shrinkage estimation under bimonotonicity constraints*

**18:00** Closing ceremony

---

# II

# Invited talks

# Invited talk 1

# Functional Data Analysis: nonparametric estimation and structural regression tests

**Philippe Vieu**
*Institut de Mathématiques, Equipe de Statistique et Probabilités*
*Univ. Paul Sabatier, Toulouse, France*

**Laurent Delsol and Frédéric Ferraty**
*Institut de Mathématiques, Toulouse, France*

**Abstract.** This contribution proposes to discuss some recent advances existing in the area of functional data analyis when nonparametric models and methods are used. It will be centered around regression problems, and both estimation and testing questions will be discussed.

Statistics for Functional Data is a recent field of researches that was popularized by the monographies [5] and [6]. Various statistical questions have been studied with functional data, but the previous literature (see references in [1], [5] and [6]) was concentrated around *parametric* models and methods. Since a few years *nonparametric* models have been developed for analyzing functional variables, and the monography [3] presents a wide scope of the literature in this field (including theoretical and applied issues).

In a first attempt, this talk will quickly recalls the basic ideas (extracted of [3]) making possible and efficient the nonparametric modelling of functional data analysis. A special attempt will be paid to the topological considerations used to control the effects of the infinite dimension.

In a second attempt the talk will present some recent results (extracted from [2]) linked with the using of nonparametric functional estimates for the construction of structural

testing procedures in functional regression.

This talk will be illustrated by means of examples extracted of [4].

## References

[1] Bosq, D. (2000) *Linear processes in functions spaces. Theory and Applications.* Lecture Notes in Statistics, **149**, Springer-Verlag, New York.

[2] Delsol, L. (2008). Régression sur variables fonctionnelles: Estimation, tests de structure et applications. PhD thesis, Toulouse, June 17, 2008.

[3] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis.* Springer Series in Statistics, New York.

[4] Ferraty, F. and Vieu, P. (2006). NPFDA in practice. Free access on line at *http://www.lsp.ups-tlse.fr/staph/npfda/*

[5] Ramsay, J. and Silverman, B. (1997). *Functional data analysis.* Springer Series in Statistics, New York.

[6] Ramsay, J. and Silverman, B. (2005). *Functional data analysis* (Second edition). Springer Series in Statistics, New York.

# Invited talk 2

# Functional Additive Modeling

**Hans-Georg Müller**
*University of California at Davis, USA*

**Abstract.** The functional linear regression model imposes structural constraints which are sometimes too restrictive. A nonlinear functional regression model is proposed which retains many of the nice asymptotic features of functional linear regression, but at the same time is much more flexible. This extension being analogous to the extension of multivariate regression to additive regression, the proposed model is referred to as functional additive regression. The additive model is easy to implement, and we demonstrate its usefulness in applications.

This is joint work with Fan Yao.

# Invited talk 3

# New perspectives about error distribution based tests in nonparametric regression

**Juan Carlos Pardo-Fernández**[1]
*Departamento de Estatística e Investigación Operativa. Universidade de Vigo, Spain*

**Abstract.** Statistical testing procedures in regression models based on the estimation of the error distribution have been recently proposed in the literature. In the first part of this talk we will revise the general idea of those testing procedures and give some examples. In the second part, we will discuss some new perspectives of the method.

## 1 Introduction

Let us consider a general nonparametric regression model

$$Y = m(X) + \sigma(X)\varepsilon,$$

where $Y$ is the response variable related to the covariate $X$ through the regression function $m$ and the variance function $\sigma$, and $\varepsilon$ is the regression error with distribution $F_\varepsilon(y) = P(\varepsilon \leq y)$.

Given a sample $(X_i, Y_i)$, $i = 1, \ldots, n$, from the pair $(X, Y)$, Akritas and Van Keilegom (2001) proposed estimating the error distribution, $F_\varepsilon$, by the empirical distribution of the residuals estimated nonparametrically, that is,

$$\hat{F}_\varepsilon(y) = \frac{1}{n}\sum_{i=1}^{n} I\left(\frac{Y_i - \hat{m}(X_i)}{\hat{\sigma}(X_i)} \leq y\right),$$

[1]Research supported by Ministerio de Educación y Ciencia, Xunta de Galicia and Universidade de Vigo.

where $\hat{m}$ and $\hat{\sigma}$ are appropriate nonparametric estimators of the regression function and variance function, respectively.

The estimator of the error distribution can be used to test hypothesis concerning the elements of the regression model (the regression function, the variance function and the error distribution itself). Let $H_0$ denote a hypothesis about any element of the regression model. The basic idea of the testing procedure is to compare two estimates of the error distribution: the nonparametric one introduced above, $\hat{F}_\varepsilon$, and a new one which incorporates information from the null hypothesis, $\hat{F}_{\varepsilon 0}$. Then, Kolmogorov-Smirnov and Cramér-von Mises type statistics are considered and the critical values of the test are approximated by a smooth bootstrap procedure.

## 2 Summary of the talk

In the first part of this talk, we will review the tests based on the estimation of the error distribution that have recently appeared in the literature: tests for the parametric form of the regression function (see Van Keilegom, González-Manteiga and Sánchez-Sellero, 2007), comparison of regression curves (see Pardo-Fernández, Van Keilegom and González-Manteiga, 2007), tests about the equality of error distributions (see Pardo-Fernández, 2007), and tests for the parametric form of the variance function (see Dette, Neumeyer and Van Keilegom, 2007).

In the second part of the talk we show some new perspectives of this kind of tests. Firstly, we will briefly present the estimator of the error distribution with multiple covariates studied by Neumeyer and Van Keilegom (2008), and show how these authors employ it to test for additivity. Finally, we will discuss the test for multiplicative models proposed by Dette, Pardo-Fernández and Van Keilegom (2008) in a dependent data setup.

## References

Akritas, M. G. and Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics*, **28**, 549-567.

Dette, H., Neumeyer, N. and Van Keilegom, I. (2007). A new test for the parametric form of the variance function in nonparametric regression. *Journal of the Royal Statistical Society – Series B*, **69**, 903-917.

Dette, H., Pardo-Fernández, J.C. and Van Keilegom, I. (2008). Goodness-of-fit tests for multiplicative models with dependent data. *Manuscript.*

Neumeyer, N. and Van Keilegom, I. (2008). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Manuscript.*

Pardo-Fernández, J. C. (2007). Comparison of error distributions in nonparametric regression. *Statistics and Probability Letters*, **77**, 350-356.

Pardo-Fernández, J. C., Van Keilegom, I. and González-Manteiga, W. (2007). Testing for the equality of $k$ regression curves. *Statistica Sinica*, **17**, 1115-1137.

Van Keilegom, I., González-Manteiga, W. and Sánchez-Sellero, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *TEST*, **17**, 401-415.

# Invited talk 4

# Nonparametric analysis of covariance using quantile curves

**Holger Dette**
*Ruhr-Universität Bochum, Germany*

**Abstract.** In this paper a new nonparametric estimate of conditional quantiles is proposed, that avoids the crossing problem. The method uses an initial estimate of the conditional distribution function in a first step and solves the problem of inversion and monotonization with respect to $p \in (0, 1)$ simultaneously. The asymptotic properties of the new estimate are investigated and its performance is illustrated by means of a simulation study. The results are used to construct a test for the hypothesis that $k$ quantile curves coincide.

# Invited talk 5

# Testing for monotonicity of nonparametric regression functions

**Natalie Neumeyer**
*University of Hamburg, Germany*

**Abstract.** We explain how increasing regression functions can be estimated by increasing rearrangements of unconstrained nonparametric estimators. Possibilities of applying such constrained estimators to test for monotonicity of regression functions and problems arising from the asymptotic expansion of the estimators are discussed. We propose several new tests for monotonicity based on different empirical processes of residuals. The asymptotic distributions and small sample performances are presented.

This is joint work with Melanie Birke.

# Invited talk 6

# Inference problems with resampling

**Stefan Sperlich**
*Institute for Statistics and Econometrics. Georg-August-Universität Göttingen, Germany*

**Abstract.** In statistics and econometrics, bootstrap methods are a commonly used tool for constructing confidence bands (uniform or point-wise) or to approximate critical values for specification tests. As already mentioned and explained in Härdle and Marron (1990, 1991), in most cases oversmoothing is necessary for the pre-estimation of the underlying (null hypothesis) model from which bootstrap samples will be drawn. This is what tells us theory, and also to what people usually refer to when proposing a new test or estimator.

In practice, however, basically nothing is known about how to choose that smoothing parameter (for kernel methods the bandwidth) to guarantee the claimed coverage probability or to hold the nominal size of a test. Dette et al. (2005) studied additivity tests but could not identify one that performed properly. Also Rodríguez-Poó et al. (2007) proposed an adaptive [i.e. adaptive for the bandwidth of the test, respectively the alternative, see Spokoiny (2001) and Guerre and Lavergne (2005)] omnibus test for generalized structured models but had to switch to subsampling methods. In parallel, Davidson and MacKinnon (1999), Davidson and Flachaire (2001) worked in different papers on the size distortion of bootstrap tests.

Another point is that if the bandwidth is estimated to be data adaptive, then the bootstrap samples, and consequently the whole bootstrap inference are conditioned on this bandwidth. In other words, the confidence bands for example will not hold the level for estimated or random bandwidths.

We discuss possible methods, rule of thumb and more sophisticated ones, to find a smoothing parameter that gives a test fulfilling both: holding the level and having non trivial power. Among other results, it turns out that none of the known double bootstrap methods helps. However, adapting the

method of an automatic choice of subsample size in subsam- pling tests turns out to produce promising results. In order to do such a study properly, it is necessary to discuss jointly the choice of other parameters which may significantly affect the size or coverage probability.

This talk is based on joint work with Dette and v. Lieres und Wilkau, Rodríguez-Poó and Vieu, and Barrientos-Marín.

## References

Davidson, R. and Flachaire, E. (2001). The Wild Bootstrap, Tamed at Last. *Working Papers 1000, Queen's University, Department of Economics.*

Davidson, R. and MacKinnon, J.G. (1999). The size distortion of bootstrap tests. *Econometric Theory*, **15**, 361-376.

Dette, H., von Lieres und Wilkau, C., and Sperlich, S. (2005). A Comparison of Different Nonparametric Method for Inference on Additive Models. *J. Nonparametric Statistics*, **17**, 57-81.

Guerre, E. and Lavergne, P. (2005). Data-driven rate-optimal specification testing in regression models. *Annals of Statistics*, **33**(2), 840-870.

Härdle, W. and J.S. Marron (1990). Semiparametric Comparison of Regression Curves. *Annals of Statistics*, **18**, 63-89.

Härdle, W. and J.S. Marron (1991). Bootstrap Simultaneous Bars For Nonparametric Regression. *Annals of Statistics*, **19**, 778-796.

Rodríguez-Poó, J.M., Sperlich, S., and Vieu, P. (2007). An Adaptive Specification Test for Semiparametric Models. *Preprint Centre for Statistics, zfs-2007-01.*

Spokoiny, V. (2001). Data-driven testing the fit of linear models. *Mathematical methods of statistics*, **10**, 465-497.

# Invited talk 7

# Every missing not at random model for incomplete data has got a missing at random counterpart with equal fit

**Geert Molenberghs**
*Universiteit Hasselt and Katholieke Universiteit Leuven, Belgium*

**Abstract.** Over the last decade, a variety of models to analyze incomplete multivariate and longitudinal data have been proposed, many of which allowing for the missingness to be not at random (MNAR), in the sense that the unobserved measurements influence the process governing missingness, in addition to influences coming from observed measurements and/or covariates. The fundamental problems implied by such models, to which we refer as sensitivity to unverifiable modeling assumptions, has, in turn, sparked off various strands of research in what is now termed sensitivity analysis. The nature of sensitivity originates from the fact that an MNAR model is not fully verifiable from the data, rendering the empirical distinction between MNAR and random missingness (MAR), where only covariates and observed outcomes influence missingness, hard or even impossible, unless one is prepared to accept the posited MNAR model in an unquestioning way. We show that the empirical distinction between MAR and MNAR is not possible, in the sense that each MNAR model fit to a set of observed data can be reproduced exactly by an MAR counterpart. Of course, such a pair of models will produce different predictions of the unobserved outcomes, given the observed ones. This is true for any model, whether formulated in a selection model (SeM), pattern-mixture model (PMM), or shared-parameter model (SPM) format. Specific attention will also be given to the SPM case, since we are able to provide a formal definition of MAR in this case.

Theoretical considerations are supplemented with illustrations based on a

clinical trial in onychomycosis and on the Slovenian Public Opinion survey. The implications for sensitivity analysis are discussed.

Missing data can be seen as latent variables. Such a view allows extension of our results to other forms of coarsening, such as grouping and censoring. In addition, the technology applies to random effects models, where a parametric form for the random effects can be replaced by certain other parametric (and non-parametric) form, without distorting the model's fit, latent classes, latent variables, etc.

This is joint work with M.G. Kenward, G. Verbeke, C. Beunckens and C. Sotto.

## References

Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G., and Kenward, M.G. (2008). Shared-parameter models and missingness at random. Submitted for publication.

Molenberghs, G., Beunckens, C., Sotto, C., and Kenward, M.G. (2008) Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society, Series B*, **70**, 371-388.

Verbeke, G. and Molenberghs, G. (2008). Arbitrariness of models for augmented and coarse data, with emphasis on incomplete-data and random-effects models. Submitted for publication.

# Invited talk 8

# The Multivariate Kaplan-Meier Estimator

**Winfried Stute**
*University of Giessen, Germany*

**Abstract.** Over the last two decades there have been many efforts to extend the Kaplan-Meier estimator for right-censored data to the multivariate case. In our talk we present and discuss a new estimator which comes up as a solution of an eigenvalue problem. It constitutes a bona fide estimator and has the classical Kaplan-Meier estimator as its marginals. We also study the efficiency and asymptotic properties of the new estimator.

This is joint work with Arusharka Sen (Montreal).

# Invited talk 9

# Goodness-of-fit tests in mixed models

**Gerda Claeskens**
*Katholieke Universiteit Leuven, Belgium*

**Abstract.** Availability of large sets of data, some with many variables but only few replicates, others with many repeated observations per subject, asks for advanced models. Often, one uses a mixture of random and fixed effects for describing these data. For example, in microarray experiments one typically has information on thousands of genes, with only a few replicates. This is a situation where a "classical" model with only fixed effects would fail, since the number of variables (genes) largely exceeds the number of observations (replicates). Random effects location-scale models may be applied for such purposes where the effect of the genes is assumed to follow some distribution. Often, normality is assumed. In this work we address ways in which we formally can test this hypothesis of normality in a mixed effects model.

# Invited talk 10

# Nonlinear methods for variable selection

**Peter Hall**
*University of Melbourne, Australia*

**Abstract.** The conventional approach to variable selection, based on a linear model, can perform very effectively provided the response to relevant components is approximately monotone and its gradient changes only slowly. In other circumstances, nonlinearity of response can result in significant vector components being overlooked. Even if good results are obtained by linear model fitting, they can sometimes be bettered by using a nonlinear approach. These circumstances can arise in practice, with real data, and they motivate alternative methodologies. We suggest an approach based on ranking generalised empirical correlations between the response variable and components of the explanatory vector. This technique is not prediction-based, and can identify variables that are influential but not explicitly part of a predictive model.

# Invited talk 11

# Geostatistics of extremes

**Anthony Davison**
*Ecole Politechnique Fédérale de Lausanne, Switzerland*

**Abstract.** Climatic change is forecast to change the frequency and sizes of extreme events such as major storms, heatwaves and the like, and the effects on human mortality, health and infrastructure are starting to become of major concern to public health authorities, engineers, and other planners. Predicting the possible impacts of such events necessarily entails extrapolation outside the range of the available data, and the usual basis for this is the statistics of extremes and its underlying probability models. Analysis of extreme events for single series of data is now well-established and used in a variety of disciplines, from hydrology through metallurgy to finance and insurance, but the corresponding methods for modelling events in space are underdeveloped. This talk will outline an approach to extending models for statistics of extremes to the spatial context, illustrated by examples.

# Invited talk 12

# Randomized sequential prediction: performance and algorithms

**Gabor Lugosi**
*Universitat Pompeu Fabra, Spain*

**Abstract.** We consider a family of prediction problems in which a forecaster sequentially predicts an unknown sequence. The goal of the forecaster is to predict almost as well as the best in a class of reference strategies. It was shown by Hannan and Blackwell in the 1950's that there exist randomized forecasting strategies that achieve this goal, regardless of what the unknown sequence is. Many variations of the problem has been studied, including situations in which the forecaster has limited access to the past elements of the sequence. The multi-armed bandit problem is a classical example. When the class of reference strategies is large, efficient computation of the randomized strategies becomes a nontrivial challenge. In this talk we discuss some fast algorithms that have been developed recently for a variety of problems, including the minimum-weight path problem, the "expert tracking" problem, and the minimum-weight spanning tree problem. Parts of this talk are based on joint work with Nicolo Cesa-Bianchi, Andras Gyorgy, Tamas Linder, Gyorgy Ottucsak, and Gilles Stoltz.

# Invited talk 13

# Wavelet methods in statistics: some recent developments and their applications

**Anestis Antoniadis**
*Université Joseph Fourier at Grenoble, France*

**Abstract.** The development of wavelet theory has in recent years spawned applications in signal processing, in fast algorithms for integral transforms, and in image and function representation methods. This talk attempts to synthesize some recent work on "nonlinear" wavelet methods in nonparametric curve estimation and their role on a variety of applications. We discuss in detail several wavelet shrinkage and wavelet thresholding estimators, scattered in the literature and developed, under more or less standard settings, to denoise data modeled as observations of a signal with additive noise. Most of these methods are fitted into the general concept of regularization with appropriately chosen penalty functions. The usefulness of all these methods are illustrated by means of simulations and practical examples.

# Invited talk 14

# Empirical likelihood for non-smooth criterion functions

**Ingrid Van Keilegom**
*Université catholique de Louvain, Belgium*

**Abstract.** Suppose that $X_1, \ldots, X_n$ is a sequence of independent random vectors, identically distributed as a $d$-dimensional random vector $X$. Let $\mu \in R^p$ be a parameter of interest and $\nu \in R^q$ be some nuisance parameter. The unknown, true parameters $(\mu_0, \nu_0)$ are uniquely determined by the system of equations $E\{g(X, \mu_0, \nu_0)\} = 0$, where $g = (g_1, \ldots, g_{p+q})$ is a vector of $p+q$ functions. In this paper we develop an empirical likelihood method to do inference for the parameter $\mu_0$. The results in this paper are valid under very mild conditions on the vector of criterion functions $g$. In particular, we do not require that $g_1, \ldots, g_{p+q}$ are smooth in $\mu$ or $\nu$. This offers the advantage that the criterion function may involve indicators, which are encountered when considering e.g. differences of quantiles, copulas, ROC curves, to mention just a few examples. We prove the asymptotic limit of the empirical log-likelihood ratio, and carry out a small simulation study to test the performance of the proposed empirical likelihood method for small samples. This is joint work with Noël Veraverbeke and Elisa Molanes-López.

# Invited talk 15

# A general uniform in bandwidth consistency result with applications

**Jan Swanepoel**
*North-West University, Potchefstroom Campus, South Africa*

**Abstract.** A kernel-type function estimator is sensitive to the choice of the bandwidth $h$ used. An appropriate choice of $h$ is therefore needed to produce an estimator which has, for example, a relatively small mean squared error. Such an optimal $h$ will depend on the underlying distribution of the data. Hence, a data-based choice of $h$ (usually referred to as a selector) is required. This, however, entails that the behaviour of the resulting kernel-type function estimator cannot be investigated analytically by the standard methods for estimators based on deterministic bandwidth sequences.

Einmahl and Mason (2005) introduced a method to prove uniform in bandwidth ( UIB($h$)) and uniform in location ($x$) strong consistency of the kernel density estimator, the Nadaraya-Watson regression function estimator and the conditional empirical process. They showed that their results are immediately applicable to establish uniform strong consistency of these kernel-type estimators when the bandwidth $h$ is a function of the data and/or $x$ . Dony, Einmahl and Mason (2006) derived a similar result for local polynomial regression function estimators.

In this presentation I shall discuss a new general UIB($h$) consistency theorem. The results of the above-mentioned authors can be readily derived from this theorem. It also yields UIB($h$) extensions of some results by Boos (1986). Moreover, an UIB($h$) Finkelstein-type functional Law of the Iterated Logarithm ( LIL ) can be obtained for a sequence of kernel distribution function estimator ( KDFE ) random functions. In particular, this implies an UIB($h$) Chung LIL . An UIB($h$) functional CLT for centered KDFE processes is discussed, and some oscillation results for the classical empirical process and the KDFE process are given.

It is shown that similar results also hold for the KDFE based on a nonparametric transformation of the data, introduced by Swanepoel and Van Graan (2005), as well as for the smoothed nearest neighbour type regression function estimator, proposed by Stute (1984). Finally, a practical example will be discussed regarding a nonparametric kernel-type method for transforming data to any given continuous distribution.

## References

Boos, D.D. (1986). Rates of convergence for the distance between distribution function estimators. *Metrika*, **33**, 197-202.

Dony, J., Einmahl, U. and Mason, D.M. (2006). Uniform in bandwidth consistency of local polynomial regression function estimators. *Austrian Journal of Statistics*, **35**, 105-120.

Einmahl, U. and Mason, D.M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics*, **33**, 1380-1403.

Stute, W. (1984). Asymptotic normality of nearest neighbor regression function estimates. *The Annals of Statistics*, **12**, 917-926.

Swanepoel. J.W.H. and Van Graan, F.C. (2005). A new kernel distribution function estimator based on a non-parametric transformation of the data. *Scandinavian Journal of Statistics*, **32**, 551-562.

# Invited talk 16

# Shape-restricted regression in the presence of correlated errors

**Jean Opsomer**
*Colorado State University, USA*

**Abstract.** In many data analysis situations, the shape of a regression function cannot be specified to belong to a particular parametric family, so that nonparametric estimation methods are most appropriate. However, it is also common in practice that, in addition to the often-assumed smoothness underlying most nonparametric methods, other qualitative characteristics of the function are known, such as that the function is increasing or convex. In such cases, estimating the function nonparametrically but subject to shape constraints has a number of important advantages such as improved efficiency and robustness to the choice of tuning parameters. We consider here nonparametric estimation of a regression function under shape constraints and in the presence of correlated errors. We propose methods to estimate both the mean and the correlation functions, and show that the good practical and theoretical properties of shape-restricted smoothing continue to hold in this case.

This is joint with Mary Meyer.

# Invited talk 17

# Least squares and shrinkage estimation under bimonotonicity constraints

**Lutz Duembgen**
*University of Bern, Switzerland*

**Abstract.** We describe an active set algorithm for minimization of a smooth function on the set of $(r \times c)$ matrices which are bimonotone. This algorithm can be used, for instance, to estimate a bimonotone regression function via least squares or least absolute deviations. Another application is shrinkage estimation in image denoising or, more generally, regression problems with two ordinal factors after representing the data in a suitable basis which is indexed by pairs $(i, j)$ in $(1, ..., r) \times (1, ..., c)$. Various numerical examples illustrate our methods.

This is joint work with Rudolph Beran, University of California at Davis.

# III

# Selected contributed papers

# Selected contributed paper 1

# Specification tests for the distribution of errors in nonparametric regression: a martingale approach

**Juan Mora**
*Universidad de Alicante, Spain*

**Alicia Pérez-Alonso**
*European University Institute, Italy*

**Abstract.** We discuss how to test whether the distribution of regression errors belongs to a parametric family of continuous distribution functions, making no parametric assumption about the conditional mean or the conditional variance in the regression model. More specifically, let $(X, Y)$ be a bivariate continuous random vector such that $E(Y^2)$ is finite, denote $m(x) \equiv E(Y|X = x)$, $\sigma^2(x) \equiv \mathrm{Var}(Y|X = x)$ and consider the error term $\varepsilon \equiv \{Y - m(X)\}/\sigma(X)$, which is, by definition, a zero-mean unit-variance random variable. If $F_\varepsilon(\cdot)$ denotes the c.d.f. of $\varepsilon$ and $\mathcal{F} \equiv \{F(\cdot, \theta),\, \theta \in \Theta \subset \mathbb{R}^m\}$ denotes a parametric family of zero-mean unit-variance continuous c.d.f.'s, each of them known except for the parameter vector $\theta$, we propose a testing procedure to face the hypotheses

$$\begin{aligned} \mathrm{H}_0 &: \exists\, \theta_0 \in \Theta \text{ such that } F_\varepsilon(\cdot) = F(\cdot, \theta_0), \qquad \text{vs.} \\ \mathrm{H}_1 &: F_\varepsilon(\cdot) \notin \mathcal{F}, \end{aligned}$$

when independent and identically distributed observations $\{(X_i, Y_i),\, i = 1, \ldots, n\}$, with the same distribution as $(X, Y)$, are available. In principle, one could think of using a Kolmogorov-Smirnov or a Cramér-von Mises statistic, constructed replacing errors by residuals and parameters by estimates. However, using the results derived in Akritas and Van Keilegom (2001, *Scandinavian Journal of Statistics*) the asymptotic distribution of these residual-based

statistics can be derived, and it proves to be not asymptotically distribution-free, a property that is already well-known in the literature. Then, we follow the methodology introduced in Khmaladze (1993, *Annals of Statistics*) to derive asymptotically distribution-free martingale-transformed test statistics. Finally, we derive the asymptotic distribution and the consistency of these martingale-transformed statistics under appropriate conditions. Two Monte Carlo experiments show that the transformed statistics work reasonably well in terms of size and power, and that their behaviour is not very sensitive to the choice of the smoothing value.

# Selected contributed paper 2

# $\sqrt{n}$-uniformly consistent density estimation in nonparametric regression models

**Juan Carlos Escanciano**
*Department of Economics. Indiana University*

**David Jacho-Chávez**
*Department of Economics. Indiana University*

**Abstract.** The paper introduces a $\sqrt{n}$ -consistent estimator of the probability density function of the response variable in a nonparametric regression model. The proposed estimator is shown to have a (uniform) asymptotic normal distribution, and it is computationally very simple to compute. A Monte Carlo experiment confirms our theoretical results, and an empirical application demonstrates its usefulness. The results derived in the paper adapts general U-processes theory to the inclusion of infinite dimensional nuisance parameters.

# Selected contributed paper 3

# Efficient estimation for the Accelerated Failure Time Model for forward recurrence times

**Rajat Mukherjee**
*Biostatistics. Public Health Foundation of India*

**Michael Kosorok**
*Biostatistics. University of North Carolina - Chapel Hill*

**Jason Fine**
*Statistics. University of Wisconsin - Madison*

**Abstract.** We study semiparametric efficient estimation of the regression parameter for Accelerated Failure Time models fitted to length-biased prevalent cohort data giving rise to forward recurrence times. We show that an efficient estimator in the core incident cases will still be efficient here. We derive an efficient estimator for the general setup that is also applicable for forward recurrence times.

## 1 Introduction

In prevalent cohort studies, diseased subject are enrolled at a cross-section and followed prospectively. In such applications of survival studies often the disease onset times are unknown and we only observe the time from sampling to the event of interest. This is the forward recurrence time (FRT). FRTs are length-biased and subjected to a multiplicative censoring scheme at the cross-section (VanEs et. al., 2000). In presence of covariates, the proportional hazards model may not be valid as the proportional structure is not preserved under length-biasedness and multiplicative censoring. However, it has

been shown that the accelerated failure time (AFT) model is preserved when he disease incidence process can be assumed to be stationary (VanEs et. al. 2000).

This leads to the question as to whether or not existing regression estimators for the AFT model are valid for the FRT case and in particular if efficiency is preserved. The problem however is that under length-biasedness the resulting covariate distribution is not free of the regression parameter ($\theta$). Since existing estimators rely on conditioning on the covariates, a "naive" analysis may result in loss of efficiency. In section 2 we derive the semiparametric efficient score and show that if the core covariate distribution is left completely unspecified then there is no loss of efficiency due to conditioning on the covariates. Infact, an efficient estimator for the core incident cases is also efficient for the FRT cases.

In section 3 we derive a efficient estimator for $\theta$ for the general AFT model under right censoring as the maximizer of the estimated profile log-likelihood. The latter is obtained by writing the log-likelohood in terms of a least-favorable sub-model (Murphy and van der Vaart, 2000) and substituting it by an estimator that converges at a rate faster than $n^{-1/4}$. This estimator can also be used to estimate the core incident-case survival curve from FRT data.

## 2 Score calculus

Let $T$ be the core incident case failure time i.e. the time from a certain initiating event like birth or disease onset to the event of interest like death with corresponding distribution $F_T$. In prevalent cohort study designs only subjects who have experienced the initiating but not the terminating event prior to sampling can be sampled and we observe the time from sampling to the terminating event ($\tilde{T}$). Thus the sample is length-biased i.e. biased towards larger realizations of $T$. The distribution $F_{LB}$ for the length-biased version $T_{LB}$ of $T$ can be written as $F_{LB}(t) = \int_0^t u dF_T(u)/\mu_T$, where, $\mu_T = \int_0^\infty u dF_T(u)$ (Cox, 1969). Further, if a stationary Poisson process can be assumed for the disease incidence then the sampling time can be shown to be uniformly distributed between the initiating and terminating times (Cox, 1969 and VanEs et. al., 2000). Thus $\tilde{T} = T_{LB}V$, where $V \sim$ Uniform(0,1). It follows that $\tilde{T}$ has density

$$g_{\tilde{T}} = \frac{1 - F_T(t)}{\mu_T} \equiv \frac{S_T(t)}{\mu_T}. \tag{1}$$

Suppose that $Z$ is the covariate vector with corresponding to $T$ and with density $h$, then under the AFT model,

$$T = e^{\theta' Z} U, \tag{2}$$

where, $\theta$ is the $p \times 1$ regression parameter and $U$ a non-negative random variable independent of $Z$ and with density $g$, survival function $S$ and hazard $\lambda(u) \equiv g(u)/S(u)$. It follows from (1) that $\tilde{T}$ with covariates ($\tilde{Z}$) collected at the cross-section also follow the AFT model given by

$$\tilde{T} = e^{\theta' \tilde{Z}} \tilde{U}, \tag{3}$$

where $\tilde{Z} \sim e^{\theta' z} h(z) / \int e^{\theta' z} h(z) dz$ and $\tilde{U} \sim S(u) / \int_0^\infty S(v) dv$.

We consider possibly right-censored FRT data $(\tilde{T}_i \wedge \tilde{C}_i, \delta_i, \tilde{Z}_i; i = 1 \cdots n)$, where, $\delta_i \equiv I\{\tilde{T} \leq \tilde{C}\}$ and assume that $\tilde{T}$ and $\tilde{C}$ are conditionally independent given $\tilde{Z}$, $\mu_g =$

$\int S(v)dv < \infty$, $E_g U^2 \lambda(U) < \infty$. Let $\theta_0$ be the true value of $\theta$ belonging to the interior of $\Theta$, which is a compact subset of $\Re^k$. For any $\theta \in \Theta$, let $U(\theta) \equiv \tilde{T} e^{-\theta' \tilde{Z}}$ and $U^C(\theta) \equiv \tilde{C} e^{-\theta' \tilde{Z}}$. The hazard function for $U(\theta)$ is then

$$\lambda_{U(\theta)}(u) = \frac{e^{(\theta-\theta_0)'z} S(e^{(\theta-\theta_0)'z} u)}{\int_{e^{(\theta-\theta_0)'z} u}^{\infty} S(v)dv}.$$

**Theorem 1** *Suppose that the covariate vector $\tilde{Z}$ is almost surely bounded. Then the efficient score for estimating $\theta$ in (3) is*

$$\tilde{l}_{\theta,S} = \int_0^{U^c(\theta_0)} (z - E\{\tilde{Z} | U^c(\theta_0) \geq s\}) R\phi(s) dM(s), \tag{4}$$

*where $\phi(u) = 1 - ug(u)/S(u)$, $Ra(t) = a(t) - \int_t^\infty a(u)S(u)du / \int_t^\infty S(u)du$ for $a \in L_2^0(S)$ and $M(t) = I\{U(\theta) \leq t\} - \int_0^t I\{U(\theta) > s\} \lambda_{U(\theta)}(s)ds$.*

Note that the above efficient score in (4) has been derived unconditionally on $\tilde{Z}$ because of its dependence on $\theta$. However, it does does not use any information in the marginal distribution of $\tilde{Z}$. The reason being that the tangent sets of the two nuisance parameters $S$ and $h$ are orthogonal. Further, (4) is similar to the one for the core incident cases derived in Bickel, Klaassen, Ritov and Wellner, 1993, pg 150, except for the $\phi$ function. The implications are: (i) FRT data can be analyzed conditionally on the covariates as in the core incident case without any loss of information and (ii) Efficient estimators for $\theta$ in (2) will also be efficient for $\theta$ in (3). However, the efficiency bounds may be different.

# 3 Efficient Estimation

Here we derive a asymptotically efficient estimator for $\theta$ in (2) using the profile likelihood theory (Severini and Wong, 1992 and Murphy and van der Vaart, 2000). For arbitrarily fixed $\theta$ define, $\epsilon_\theta = \log T - \theta' Z$ and $\epsilon_\theta^c = \log C - \theta' Z$ and $e \equiv e_\theta = \epsilon_\theta \wedge \epsilon_\theta^c$. The data consists of $n$ independent realizations of $Y = (e, \delta, Z)$. We specify the semiparametric model in terms of the regression parameter $\theta$ and the hazard function $\lambda$ of the errors by $\mathcal{P} = \{P_{\theta,\lambda} : \theta \in \Theta, \lambda \in \Lambda\}$. Let $(\theta_0, \lambda_0)$ be the true value of the parameter and $S_0$ be the survival function corresponding to $\lambda_0$. We further assume the following.

(A1) The covariate vector $Z$ is bounded almost surely with density $h$.

(A2) $\theta_0$ belongs to the interior of an open and bounded set $\Theta \subset \Re^k$. Along with (A1) this gives $\alpha \equiv \text{ess. sup}_{\theta \in \Theta} |(\theta - \theta_0)' Z| < \infty$.

(A3) $\tau = \sup_t \{t : Pr[C > \exp\{t+\alpha\} \mid Z] > 0\}$ exists and is finite and further $S_0(\tau+\alpha) > 0$.

Let $\zeta_{\theta,z}$ and $\gamma_{\theta,z}(t) \equiv \exp\{-\int_{-\infty}^t \lambda_0(s + (\theta - \theta_0)'z)ds\} \zeta_{\theta,z}(t)$ denote the survival function of $\epsilon_\theta^c$ and the at-risk probability function respectively, given $Z = z$ and under $P_{\theta,\lambda_0}$ and consider the submodel $\theta \mapsto \lambda_\theta$ given by

$$\lambda_\theta(t) = \frac{\int \lambda_0(t + (\theta - \theta_0)'z) \gamma_{\theta,z}(t) h(z) dz}{\int \gamma_{\theta,z}(t) h(z) dz} \equiv \frac{g_\theta(t)}{G_\theta(t)}. \tag{5}$$

We show that this is a least-favorable submodel in the sense of minimizing the Kulback-Lieber distance and define the log-likelihood as

$$L_n(\theta, \lambda_\theta) = \sum_{i=1}^{n} \left\{ \delta_i \log \lambda_\theta(e_i) - \int_{-\infty}^{e_i} \lambda_\theta(u) du \right\} \tag{6}$$

An estimator for the above least favorable curve is obtained by solving a kernel-smoothed self-consistent equation for the survival function for censored data (Efron, 1967 and Cosslett, 2004) as

$$\hat{\lambda}_\theta(t) = \frac{h_n^{-1} \int K\left(h_n^{-1}(t-v)\right) dF_{n,\theta}(v)}{\int \bar{K}\left(h_n^{-1}(t-v)\right) dG_{n,\theta}(v)} \equiv \frac{\hat{g}_{n,\theta}(t)}{\hat{G}_{n,\theta}(t)}, \tag{7}$$

where $K$ is suitable kernel with bandwidth $h_n$ and $G_{n,\theta}$ and $F_{n,\theta}$ are the empirical distribution functions of the observed residuals and uncensored residuals respectively.

Define $\hat{\theta} \equiv \hat{\theta}_n$ to be the maximizer of $L_n(\theta, \hat{\lambda}_\theta)$. In order to prove consistency of $\hat{\theta}$, we show the uniform convergence of a trimmed version of $L_n(\theta, \hat{\lambda}_\theta)$ to $L_n(\theta, \lambda_\theta)$ and apply the argmax theorem (corollary 3.2.2 in van der Vaart and Wellner, 1996). For asymptotic normality and efficiency we show the uniform convergence of $\hat{\lambda}_\theta$ to $\lambda_\theta$ at a rate $n^{-\nu_1}$ where $\nu_1 \geq 1/4$ and $\hat{\lambda}'_\theta$ to $\lambda'_\theta$ at a rate $n^{-\nu_2}$ where $\nu_1 + \nu_2 \geq 1/2$ and apply the profile likelihood theorem (Theorem 1 in Murphy and van der Vaart, 2000).

Numerical studies illustrate the estimation procedure and the finite-sample properties of the proposed estimator.

# References


Bickel, P., Klaassen, C., Ritov, Y. and Wellner, J. (1993). *Efficient and adaptive estimation for semiparametric models.* Springer-Verlag, New York.

Cosslett, S. (2004). Efficient semiparametric estimation of censored and truncated regression via a smoothed self-consistency equation. *Econometrica* **72**, 1277-1293.

Cox, D. (1969). Some sampling problems in technology. *New developments in survey sampling*, eds. Johnson and Smith. Wiley, New York.

Efron, B. (1967). The two sample problem with censored data. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* **4**, Berkeley: University of California press.

Murphy, S. and van der Vaart, A. (2000). On profile likelihood. *J. Amer. Statist. Assoc.* **95**, 449-485.

Severini, T. and Wong, W. (1992). Profile likelihood and conditionally parametric models. *Ann. Statis.* **20**, 1768-1802.

van der Vaart, A. and Wellner, J. (1996). *Weak convergence and Empirical processes: With applications in statistics.* Springer-Verlag, New York.

Van Es, B., Klaasen, C. and Oudshoorn, K. (2000). Survival analysis under cross section sampling: length bias and multiplicative censoring. *J. Statist. Plan. Infer.* **91**, 295-312.

# Selected contributed paper 4

# Uniform in bandwidth consistency of the kernel–based Hill estimator

**Julia Dony**
*Department of Mathematics, Free University Brussels (VUB)*

**Abstract.** We consider the kernel–based Hill estimator for the tail index of a Pareto–type distribution and establish its weak consistency, uniformly in a certain range of bandwidths tending to zero at particular rates. This "uniform in bandwidth" result permits to consider estimators of the tail index that are based upon data–dependent bandwidths or bandwidths depending on the location.

## 1 Introduction and statement of the results

Let $(X_i, Y_i), i \geq 1$ be i.i.d. random vectors in $\mathbb{R}^d \times \mathbb{R}$ and let $\mathcal{F}$ be a class of measurable functions $\varphi : \mathbb{R} \to \mathbb{R}$ for which $\mathbb{E}\varphi^2(Y) < \infty$. For $t \in \mathbb{R}^d$ fixed, we denote the regression function by $m_\varphi(t) := \mathbb{E}[\varphi(Y)|X = t]$ and consider the kernel–type estimator

$$\hat{\varphi}_{n,h}(t) := \frac{1}{nh^d} \sum_{i=1}^{n} \varphi(Y_i) K\Big(\frac{t - X_i}{h}\Big),$$

where $K$ is a uniformly bounded kernel function with support contained in $[-M, M]^d$, and where $0 < h < 1$ is a bandwidth. Likewise, let $\hat{f}_{n,h}(t)$ be the kernel density estimator of $f_X(t)$, the density of $X$, which corresponds to the choice $\varphi(y) \equiv 1$ in the above formula. If $h_n$ is a deterministic sequence of positive numbers going to zero and such that $nh_n^d / \log\log n \to \infty$, it is well–known that under some regularity conditions, the Nadaraya–Watson–type estimator $\hat{m}_{n,h_n,\varphi}(t) := \hat{\varphi}_{n,h_n}(t)/\hat{f}_{n,h_n}(t)$ is a (strongly) consistent estimator for $m_\varphi(t)$. Moreover, it was shown in Einmahl and Mason (2005) that under some additional assumptions, the consistency of $\hat{m}_{n,h,\varphi}(t)$ is preserved uniformly

for $t \in I$ where $I \subset \mathbb{R}^d$ is compact, and uniformly in $a_n \leq h \leq b_n$ for appropriate positive sequences $a_n$ and $b_n$ converging to zero.

The uniformity in $h$ makes it possible to choose the bandwidth $h_n$ depending on the data and/or the location. If the function class $\mathcal{F}$ is uniformly bounded, they show that one can choose $h$ from an interval of the form $[c_n \log n/n, b_n]$, where $c_n \to \infty$ and $b_n \to 0$. Assuming in the unbounded case that the envelope function $F$ of $\mathcal{F}$ satisfies the condition

$$\text{(F.}p\text{)} \qquad \mu_p := \sup_{x \in J} \mathbb{E}\left[F^p(Y)|X = x\right] < \infty \quad \text{for some } p > 2,$$

where $J = I^\epsilon$ for some $\epsilon > 0$, it is shown that their result remains valid if one chooses $a_n^d \geq c(\log n/n)^{1-2/p}$. In addition, the corresponding convergence rates of $\hat{\varphi}_{n,h}(t)$ to $m_\varphi(t) f_X(t)$ have been obtained in both cases as well.

In the following we shall describe uniform in bandwidth results for $\hat{\varphi}_{n,h}(t)$ at fixed points $t \in \mathbb{R}^d$, i.e. pointwise and not uniformly over compact subsets. This will allow us to achieve the unifomity in $h$ on larger intervals than in the previously mentioned "uniform on compacts" case, hence improving the result by Einmahl and Mason (2005). In Section 2, our "pointwise" uniform in bandwidth results will turn out to be particularly useful to establish the uniform in bandwidth consistency of a class of kernel tail index estimators.

Towards establishing these consistency results, we impose some additional conditions. In particular, we consider classes $\mathcal{F}$ such that

(F.i) $\mathcal{F}$ is pointwise measurable,
(F.ii) $\mathcal{F}$ has a measurable envelope function $F(y) \geq \sup_{\varphi \in \mathcal{F}} |\varphi(y)|, y \in \mathbb{R}$,
(F.iii) $\mathcal{F}$ is a VC class of functions,

and we let $a_n, n \geq 1$ be a sequence of non–random numbers satisfying

(H.i) $a_n \searrow 0$,
(H.ii) $a_n^d \log\log n \searrow$ and $n a_n^d / \log\log n \nearrow$,

where "$\nearrow, \searrow$" denote non–decreasing and non–increasing respectively. For convenience, we recall the assumptions on the kernel function $K : \mathbb{R}^d \to \mathbb{R}$.

(K.i) $\sup_{x \in \mathbb{R}^d} |K(x)| < \infty$ and $\int K(x)dx = 1$,
(K.ii) $K$ has support contained in $[-M, M]^d$ for some $M > 0$.

The pointwise uniform–in–$h$ result for $\hat{\varphi}_{n,h}(t)$ that we present here holds when $F$ admits a finite moment generating function. Clearly, this is more general than considering a bounded class of functions. Moreover, this extension seems to be new, also for fixed bandwidth sequences.

**Theorem 1** *Suppose that the envelope function of $\mathcal{F}$ has a finite moment generating function in a neighborhood of 0. Then if $f_X$ is bounded on a neighborhood of $t$, and $b_0 < 1$ is a positive constant, it follows from the above mentioned assumptions on $\mathcal{F}$ and $K$ that*

$$\limsup_{n \to \infty} \sup_{a_n \leq h \leq b_0} \sup_{\varphi \in \mathcal{F}} \frac{\sqrt{nh^d}|\hat{\varphi}_{n,h}(t) - \mathbb{E}\hat{\varphi}_{n,h}(t)|}{\sqrt{\log\log n}} < \infty, \quad a.s.,$$

*for all non–increasing sequences $a_n$ that go to zero at rates given by (H.ii).*

It is well known that the condition in (H.ii) is optimal in the bounded case, so that Theorem 1 shows that there is no difference in terms of range of bandwidths between the bounded case and the case where $F$ admits a finite moment generating function. We note that similar extensions under moment–type conditions like (F.$p$) have also been considered in Dony (2008).

Our uniform in bandwidth results are important in establishing the consistency of kernel regression estimators using a *data–dependent* bandwidth sequence $\hat{h}_n := H_n(X_1, \ldots, X_n)$, $n \geq 1$. More particularly to show that almost surely or in probability, $\sup_{\varphi\in\mathcal{F}} |\hat{m}_{n,\hat{h}_n,\varphi}(t) - m_\varphi(t)| \to 0$. Detailed proofs of such results are provided in Dony (2008) and are based upon an empirical process representation of $\hat{\varphi}_{n,h}(t)$, where the index class is a class of functions that depends upon the sample size $n \geq 1$. The main tools are some moment and exponential deviation inequalities for empirical processes. To achieve the consistency uniformly in bandwidth, a blocking is applied and the interval $[a_n, b_0]$ is split into several smaller intervals.

## 2 Application to extreme value statistics

Let $Y_1, \ldots, Y_n$ be independent real valued variables with a Pareto–type distribution with tail index $\tau > 0$, meaning that there exists a number $\tau > 0$ such that

$$(F.\tau) \qquad \lim_{y\to\infty} \frac{1 - F(\lambda y)}{1 - F(y)} = \lambda^{-\tau}, \quad \lambda > 0.$$

The main concern is to provide a consistent estimator for the tail index $\tau > 0$. To do so, Hill (1975) proposed an estimator based upon the $k$ largest observations, namely

$$\hat{\tau}_n(k) := \Big(\frac{1}{k}\sum_{i=1}^{k} \log Y_{n-i+1:n} - \log Y_{n-k:n}\Big)^{-1},$$

where $Y_{i:n}, 1 \leq i \leq n$ denote the order statistics of $Y_1, \ldots, Y_n$. It has been shown in Mason (1982) that consistency of $\hat{\tau}_n(k)$ cannot be achieved without letting $k$ going to infinity. Moreover, necessary and sufficient conditions for $\hat{\tau}_n(k) \to_{\mathbb{P}} \tau$ are that $k = k_n \to \infty$ and $k_n/n \to 0$. Somewhat later, Csörgő, Deheuvels and Mason (1985) proposed the following kernel–based estimator for $\tau$, namely

$$\hat{\tau}_{n,h} := \frac{\frac{1}{nh}\sum_{j=1}^{n} K\big(\frac{j}{nh}\big)}{\sum_{j=1}^{n} \frac{j}{nh} K\big(\frac{j}{nh}\big)\{\log Y_{n-j+1:n} - \log Y_{n-j:n}\}} =: \frac{\kappa_{n,h}}{\phi_{n,h}}.$$

Note that when $K(u) = \mathbb{1}\{0 < u < 1\}$ and $h = k/n$, the kernel–based Hill estimator corresponds to the classical Hill estimator. The advantage of considering an extended kernel–based version of $\hat{\tau}_n(k)$ is that it permits to reduce the expected mean squared error by choosing an appropriate bandwidth and kernel function. The following proposition provides the uniform in bandwidth consistency of $\hat{\tau}_{n,h}$ to $\tau$.

**Proposition 2** *Let $\hat{\tau}_{n,h}$ be the kernel–based Hill estimator based upon i.i.d. variables $Y_1, \ldots, Y_n$ with a Pareto–type distribution of tail index $\tau > 0$, and defined with a kernel function satisfying (K.i)–(K.iv) below. Then it holds for any non–random sequences $a_n \leq b_n$ with $b_n \to 0$ and satisfying $n^{\alpha/2} a_n^{\alpha+1} \to \infty$ that $\sup_{a_n\leq h\leq b_n} |\hat{\tau}_{n,h} - \tau| = o_{\mathbb{P}}(1)$.*

It is worth to notice that $\hat{\tau}_{n,h}$ is expressed in terms of a process similar to $\hat{\varphi}_{n,h}(t)$, though with a fixed design $\{j/nh, 1 \leq j \leq n\}$. Therefore, before handling the estimator as such, we start with an analogue of Theorem 1 for the process $\hat{\beta}_{n,h} := (nh)^{-1} \sum_{i=1}^{n} Y_i K(t_{i,n}/h)$, where $t_{i,n} = i/(n+1)$. (Refer to Dony (2008) for a detailed proof of the following theorems.)

**Theorem 3** *Let $Y_1, Y_2, \ldots$ be i.i.d. variables with mean $\gamma \geq 0$ and satisfying $\mathbb{E}\exp(s|Y_1|) < \infty$ for some $s > 0$. If $K$ is Hölder–continuous with exponent $0 < \alpha \leq 1$, it follows that $\sup_{a_n \leq h \leq b_0} |\hat{\beta}_{n,h} - \mathbb{E}\hat{\beta}_{n,h}| \to_{\mathbb{P}} 0$ provided $n^{\alpha/2} a_n^{\alpha+1} \to \infty$.*

The proof of Proposition 2 will be a consequence of Theorem 4 below, which is based upon a decomposition of $\phi_{n,h}$ and an application of Theorem 3 . To see this, note that $(F.\tau)$ implies that $Y_i =_d (1-U_i)^{-1/\tau} L((1-U_i)^{-1})$, where $U_1, \ldots, U_n$ are independent variables uniformly distributed in $]0,1[$ and $L$ is a slowly varying function at infinity. Hence, by Karamata's representation,

$$\log Y_{n-j+1:n} - \log Y_{n-j:n} \stackrel{d}{=} -\tau^{-1}\{\log(1-U_{n-j+1:n}) - \log(1-U_{n-j:n})\}$$

$$+ \int_{(1-U_{n-j:n})^{-1}}^{(1-U_{n-j+1:n})^{-1}} \frac{b(u)}{u} du + \log \frac{c((1-U_{n-j+1:n})^{-1})}{c((1-U_{n-j:n})^{-1})},$$

where $c(y) \to c_0 > 0$ and $b(y) \to 0$ as $y \to \infty$. Consequently, $\phi_{n,h}$ can be decomposed into three processes as follows :

$$\phi_{n,h} \stackrel{d}{=} \sum_{j=1}^{n} \frac{j}{nh} K\big(\frac{j}{nh}\big)\{\tau^{-1} A_{n,j}^{(1)} + A_{n,j}^{(2)} + A_{n,j}^{(3)}\} =: \tau^{-1}\phi_{n,h}^{(1)} + \phi_{n,h}^{(2)} + \phi_{n,h}^{(3)}.$$

To deal with the asymptotic behavior of $\phi_{n,h}, \kappa_{n,h}$ and thus $\hat{\tau}_{n,h}$, uniformly in $h$, the following (additional) assumptions on the kernel $K$ need to be made :

(K.i) $\sup_{x \in \mathbb{R}^d} |K(x)| < \infty$ and $\int K(x)dx = 1$,
(K.ii)$'$ $K$ has support contained in $[0, M]$ for some $M < \infty$,
(K.iii) $K$ is non–increasing and non–negative on its positivity set,
(K.iv) $K$ is Hölder–continuous on $[0, M]$ with exponent $0 < \alpha \leq 1$.

**Theorem 4** *For any right–continuous kernel function satisfying (K.i)–(K.iii) and any non–increasing sequences $a_n \leq b_n$ such that $b_n \to 0$ and $na_n \to \infty$, it holds that*

*(i)* $\sup_{a_n \leq h \leq b_n} |\kappa_{n,h} - 1| = o(1)$.

*If moreover (K.iv) holds for some $0 < \alpha \leq 1$ which is such that $n^{\alpha/2} a_n^{\alpha+1} \to \infty$,*

*(ii)* $\sup_{a_n \leq h \leq b_n} |\phi_{n,h}^{(1)} - 1| = o_{\mathbb{P}}(1)$,
*(iii)* $\sup_{a_n \leq h \leq b_n} |\phi_{n,h}^{(i)}| = o_{\mathbb{P}}(1)$, $i = 2, 3$.

It follows now readily that $\sup_{a_n \leq h \leq b_n} |\hat{\tau}_{n,h} - \tau| = o_{\mathbb{P}}(1)$, establishing the (weak) uniform in bandwidth consistency of the kernel–based Hill estimator $\hat{\tau}_{n,h}$ to $\tau$.

## References

Csörgő, S., Deheuvels, P. and Mason, D.M. (1985). Kernel estimates of the tail index of a distribution, *Ann. Stat.* **13**, 1050–1077.

Dony, J. (2008). *Nonparametric regression estimation : An empirical process approach to the uniform in bandwidth consistency of kernel estimators and conditional U–statistics.* Ph.D. Thesis, Vrije Universiteit Brussel, Belgium.

Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Stat.* **33**, 1380–1403.

Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Stat.* **3**, 1163–1174.

Mason, D.M. (1982). Laws of large numbers for sums of extreme values. *Ann. Probab.* **10**, 754–764.

# Selected contributed paper 5

# Robust inference in generalized partially linear models

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

**Ricardo Cao**
*Universidade da Coruña*

**Wenceslao González–Manteiga**
*Universidade de Santiago de Compostela*

**Daniela Rodriguez**
*Universidad de Buenos Aires and CONICET*

**Abstract.** In this talk, we introduce a family of robust statistics which allow to decide between a parametric model and a semiparametric one. More precisely, under a generalized partially linear model, i.e., when the observations satisfy $y_i|(\mathbf{x}_i, t_i) \sim F(\cdot, \mu_i)$ where $\mu_i = H(\eta(t_i) + \mathbf{x}_i^{\mathrm{T}}\beta)$ with $H$ a known link function, we want to test $H_0 : \eta(t) = \alpha + \gamma t$ against $H_1 : \eta$ is a smooth function. A general approach which includes a robust version of the deviance and a robustified quasi–likelihood is considered. The asymptotic behavior of the test statistic under the null hypothesis is obtained.

## Summary

Semiparametric models contain both a parametric and a nonparametric component. Sometimes the nonparametric component plays the role of a nuisance parameter. A lot of research has been done on estimators of the parametric component in a general framework, aiming to obtain asymptotically efficient estimators. The aim of this talk is to consider semiparametric versions of the generalized linear models where the response $y$ is to be predicted by covariates $(\mathbf{x}, t)$, where $\mathbf{x} \in I\!R$ and $t \in \mathcal{T} \subset I\!R$ with $\mathcal{T}$ a compact set. Without loss of generality we will assume that $\mathcal{T} = [0, 1]$. It will also

be assumed that the conditional distribution of $y|(\mathbf{x},t)$ belongs to the canonical exponential family $\exp\left[y\theta(\mathbf{x},t) - B\left(\theta(\mathbf{x},t)\right) + C(y)\right]$, for known functions $B$ and $C$. Then, $\mu(\mathbf{x},t) = \mathrm{E}\left(y|(\mathbf{x},t)\right) = B'\left(\theta(\mathbf{x},t)\right)$, with $B'$ as the derivative of $B$. In generalized linear models (McCullagh and Nelder, 1989), which is a popular technique for modeling a wide variety of data, it is often assumed that the mean is modeled linearly through a known link function, $g$, i.e.,

$$g(\mu(\mathbf{x},t)) = \beta_0 + \mathbf{x}^{\mathrm{T}}\beta + \alpha t \ .$$

For instance, an ordinary logistic regression model assumes that the observations $(y_i, \mathbf{x}_i, t_i)$ are such that the response variables are independent binomial variables $y_i|(\mathbf{x}_i,t_i) \sim Bi(1,p_i)$ whose success probabilities depend on the explanatory variables through the relation

$$g(p_i) = \beta_0 + \mathbf{x}_i^{\mathrm{T}}\beta + \alpha t_i \ ,$$

with $g(u) = \log\left(\frac{u}{1-u}\right)$.

In many situations, the linear model is insufficient to explain the relationship between the response variable and its associated covariates. A natural generalization, which suffers from the *curse of dimensionality*, is to model the mean nonparametrically in the covariates. An alternative strategy is to allow most predictors to be modeled linearly while one or a small number of predictors enter the model nonparametrically. This is the approach we will follow, so that the relationship will be given by the semiparametric generalized partially linear model

$$\mu(\mathbf{x},t) = \mathrm{E}\left(y|(\mathbf{x},t)\right) = H\left(\eta(t) + \mathbf{x}^{\mathrm{T}}\beta\right) \tag{1}$$

where $H = g^{-1}$ is a known link function, $\beta \in I\!R$ is an unknown parameter and $\eta$ is an unknown continuous function.

Severini and Wong (1992) introduced the concept of generalized profile likelihood, which was later applied to this model by Severini and Staniswalis (1994). In this method, the nonparametric component is viewed as a function of the parametric component, and $\sqrt{n}-$consistent estimates for the parametric component can be obtained when the usual optimal rate for the smoothing parameter is used. Such estimates fail to deal with outlying observations.

Härdle, Mammen and Müller (1998) considered a test statistic to decide between a linear and a semiparametric model. Their proposal is based on the estimation procedure considered by Severini and Staniswalis (1994) modified to deal with the smoothed and unsmoothed likelihoods. A comparative study of different procedure was performed by Müller (2001).

As it is well known, such procedures fail to deal with outlying observations and so does the test statistic. In a semiparametric setting, outliers can have a devastating effect, since the extreme points can easily affect the scale and the shape of the function estimate of $\eta$, leading to possibly wrong conclusions on $\beta$ and on the hypothesis on $\eta$ to be tested.

Robust procedures for generalized linear models have been considered among others by Stefanski, Carroll and Ruppert (1986), Künsch, Stefanski and Carroll (1989), Bianco and Yohai (1995), Cantoni and Ronchetti (2001), Croux and Haesbroeck (2002) and Bianco, García Ben and Yohai (2005). The basic ideas from robust smoothing and from robust regression estimation have been adapted to deal with the case of independent observations following a partly linear regression model with $H(t) = t$; we refer to Gao

and Shi (1997), He, Zhu and Fung (2002) and Bianco and Boente (2004). Under a generalized partially linear model (1), Boente, He and Zhou (2006) introduced a general profile–based two–step robust procedure to estimate the parameter $\beta$ and the function $\eta$ while Rodriguez (2008) developed a three–step method to improve the computational time of the previous one. On the other hand, robust tests for a given alternative, under a partly linear regression model were studied in Bianco, Boente and Martínez (2006). Besides, a robust approach for testing the parametric form of a regression function versus an omnibus alternative, based on the centered asymptotic rank transformation, was considered by Wang and Qu (2007) when $H(t) = t$ and $\beta = 0$, i.e., under the nonparametric model $y_i = \eta_0(t_i) + \epsilon_i$.

In this talk, we will discuss a procedure to develop a robust test based on these estimators to decide between a linear and a semiparametric model under a generalized partially linear model (1). A bootstrap approach will also be considered.

# References

Bianco, A. and Boente, G. (2004). Robust estimators in semiparametric partly linear regression models. *Journal of Statistical Planning and Inference*, **122**, 229-252.

Bianco, A., Boente, G. and Martínez, E. (2006). Robust tests in semiparametric partly linear models. *Scand. J. Statist.*, **33**, 435-450.

Bianco, A., García Ben, M. and Yohai, V. (2005). Robust estimation for linear regression with asymmetric errors. *Canadian Journal of Statistics*, **33** 1-18.

Bianco, A. and Yohai, V. (1995). Robust estimation in the logistic regression model. *Lecture Notes in Statistics*, **109**, 17-34. Springer–Verlag, New York.

Boente, G., He, X. and Zhou, J. (2006). Robust Estimates in Generalized Partially Linear Models. *Annals of Statistics*, **34**, 2856-2878.

Cantoni, E. and Ronchetti, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association*, **96**, 1022-1030.

Croux, C. and Haesbroeck, G. (2002). Implementing the Bianco and Yohai estimator for logistic regression. *Computational Statistics & Data Analysis*, **44**, 273-295.

Gao, J. and Shi, P. (1997). M–type smoothing splines in nonparametric and semiparametric regression models. *Statistica Sinica*, **7**, 1155-1169.

Härdle, W., Mammen, E. and Müller, M. (1998). Testing parametric versus semiparametric modeling in generalized linear models. *J. Amer. Statist. Assoc.*, **93**, 1461-1474.

He, X., Zhu, Z. and Fung, W. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika*, **89**, 579-590.

Künsch, H., Stefanski, L. and Carroll, R. (1989). Conditionally unbiased bounded-influence estimation in general regression models with applications to generalized linear models. *Journal of the American Statistical Association*, **84**, 460-466.

Mc Cullagh, P. and Nelder, J.A. (1983). *Generalized linear models.* London: Chapman and Hall.

Müller, M. (2001). Estimation and testing in generalized partial linear models. A comparative study. *Statistics and Computing*, **11**, 299-309.

Rodriguez, D. (2008). *Estimación robusta en modelos parcialmente lineales generalizados*. PhD. Thesis. Universidad de Buenos Aires. (in spanish)

Severini, T. and Staniswalis, J. (1994). Quasi-likelihood estimation in semiparametric models. *Journal of the American Statistical Association*, **89**, 501-511.

Severini, T. and Wong, W. (1992). Generalized profile likelihood and conditionally parametric models. *Annals of Statistics*, **20**, 1768-1802.

Stefanski, L., Carroll, R. and Ruppert, D. (1986). Bounded score functions for generalized linear models. *Biometrika*, **73**, 413-424.

Wang, L. and Qu, A. (2007). Robust tests in regression models with omnibus alternatives and bounded influence. *J. Amer. Statist. Assoc.*, **102**, 347-358.

# Selected contributed paper 6

# Time series clustering based on nonparametric forecast

**J.A. Vilar Fernández**
*Department of Mathematics, University of Coruña, Spain*

**A.M. Alonso**
*Department of Statistics, University Carlos III, Madrid, Spain*

**J.M. Vilar Fernández**
*Department of Mathematics, University of Coruña, Spain*

**Abstract.** A new clustering procedure based on nonparametric forecasts is studied. The proposed procedure is valid to deal with a general class of autoregressive processes, including nonlinear processes. Two alternative approaches are considered to measure the dissimilarity between two time series: the $L^1$-distance between their forecast densities at a given horizon and the mean squared difference of the forecasts.

## 1 Introduction

Time series clustering is aimed at classifying the series under study into homogeneous groups in such a way that the within-group-series similarity is minimized and the between-group-series dissimilarity is maximized. This a central problem in many application fields and hence time series clustering is nowadays an active research area in different disciplines including signal processing, finance and economics, medicine, seismology, meteorology and pattern recognition, among others (see, e.g. Liao, 2005).

As with other clustering problem, the metric chosen to assess the similarity/dissimilarity between two data objects plays a crucial role in time series clustering. However, the concept of dissimilarity between two time series is non trivial. Recently, Alonso et al. (2006) argue that, in many practical situations, the real interest of clustering is the long term behavior and, in particular, on how the forecasts at a specific horizon can be grouped. For this kind of situation, they propose a dissimilarity measure based on

comparing the full forecast densities associated to each series in the sample. Note that comparing the forecast densities instead of the point forecasts can help to separate into different clusters times series having similar or equal predictions but different underlying generating models (e.g. models that differ only in the innovations distribution). In practice, the forecast densities are unknown and must be approximated from the data.

In this paper, the clustering procedure proposed by Alonso et al. (2006) is extended to cover the case of nonparametric models of arbitrary autoregressions. Our approach does not assume any parametric model for the true autoregressive structure of the series, which is estimated by using kernel smoothing techniques. In our procedure, the mechanism to obtain bootstrap predictions is based on mimicking the generating process using a nonparametric estimator of the autoregressive function and a bootstrap resample of the nonparametric residuals. In this way, we provide an useful device for classifying nonlinear autoregressive time series, including extensively studied parametric models such as the threshold autoregressive ($TAR$) models, the exponential autoregressive ($EXPAR$) models, the smooth-transition autoregressive ($STAR$) models and the bilinear models, among others (see e.g. Tong (1990) and the references included therein).

## 2 Description of the clustering procedure

Denote by $\Xi$ the class of real valued stationary processes $\{X_t\}_{t\in\mathbb{Z}}$ that admit a general autoregressive representation of the form

$$X_t = m(\boldsymbol{X}_{t-1}) + \varepsilon_t, \tag{1}$$

where $\{\varepsilon_t\}$ is an i.i.d. sequence and $\boldsymbol{X}_{t-1}$ is a $d$-dimensional vector of known lagged variables. The unknown autoregressive function $m(\cdot)$ is assumed to be a smooth function but it is not restricted to any pre-specified parametric model. Hence, both linear and nonlinear autoregressive processes are included in $\Xi$.

Our concern is to perform a cluster analysis on a set $S$ of $s$ partial realizations from time series belonging to $\Xi$, i.e. $S = \left\{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(s)}\right\}$, where, for $i = 1, \ldots, s$, $\mathbf{X}^{(i)} = \left(X_1^{(i)}, \ldots, X_T^{(i)}\right)$. Following the ideas by Alonso et al. (2006), we adopt the criterion of measuring the dissimilarity between two time series objects in terms of the disparity between their corresponding full forecast densities at a specific future time $T + b$ or the mean square difference of the forecasts. Hence our purpose is that the cluster solution captures similarities in the behaviors of the predictions at a prefixed horizon.

Specifically, Alonso et al. (2006) use the following distance between time series $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(j)}$, with $i, j = 1, \ldots, s$,

$$D_{ij}^0 = \int \left(f_{X_{T+b}^{(i)}}(x) - f_{X_{T+b}^{(j)}}(x)\right)^2 dx, \tag{2}$$

where $f_{X_{T+b}^{(i)}}(\cdot)$ denotes the density function of the forecast $X_{T+b}^{(i)}$, with $T+b$ the prefixed prediction horizon.

The distance $D_{ij}^0$ presents a serious drawback to perform cluster analysis. If the sets $\{x : f_{X_{T+b}^{(i)}}(x) > \varepsilon\}$ and $\{x : f_{X_{T+b}^{(j)}}(x) > \varepsilon\}$, are disjoints for an sufficiently small $\varepsilon$, then

$$D_{ij}^0 \approx \int f_{X_{T+b}^{(i)}}^2(x)dx + \int f_{X_{T+b}^{(j)}}^2(x)dx$$

and $D_{ij}^0$ has a poor performance in the clustering task. For it, we propose using the following distances $D_{ij}^1$ and $D_{ij}^2$ defined by

$$D_{ij}^1 = \int \left| f_{X_{T+b}^{(i)}}(x) - f_{X_{T+b}^{(j)}}(x) \right| dx, \tag{3}$$

and

$$D_{ij}^2 = E\left(X_{T+b}^{(i)} - X_{T+b}^{(j)}\right)^2 \tag{4}$$

Direct computation of distances $D_{ij}^u$, $u = 1, 2$, is not feasible in practice because the forecast densities are unknown. To overcome this difficulty, distances $D_{ij}^u$ are consistently approximated by replacing the unknown forecast densities by kernel-type density estimates constructed on the basis of bootstrap predictions. In particular, we have considered a bootstrap procedure based on generating a process

$$X_t^* = \hat{m}_g(\boldsymbol{X}_{t-1}^*) + \varepsilon_t^*, \tag{5}$$

where $\hat{m}_g$ is a nonparametric estimator of $m$ and $\{\varepsilon_t^*\}$ is a conditionally i.i.d. resample from the nonparametric residuals. This bootstrap method, called *autoregression bootstrap*, completely mimics the dependence structure of the underlying process.

A detailed description of the steps involved in generating a set of bootstrap predictions is provided below.

Let $(X_1, \ldots, X_T)$ be a partial realization from a process $X(t) \in \Xi$, i.e. $X(t)$ admits the representation given in (1). The resampling scheme proceeds as follows.

1. Estimate the autoregressive function $m(\cdot)$ using a modified Nadaraya-Watson estimator with bandwidth $g_1$.

2. Compute the nonparametric residuals, $\widehat{\varepsilon}_t = X_t - \hat{m}_{g_1}(\boldsymbol{X}_{t-1})$, $t = d+1, \ldots, T$.

3. Construct a kernel estimate of the density function, $f_{\tilde{\varepsilon}}$, associated to the centered residuals $\tilde{\varepsilon}_t = \widehat{\varepsilon}_t - \widehat{\varepsilon}_\bullet$ with $\widehat{\varepsilon}_\bullet$ the mean of the $\widehat{\varepsilon}_t$.

   Using the Rosenblatt-Parzen estimator with kernel $H(u)$, we obtain $\hat{f}_{\tilde{\varepsilon},h}(u)$

4. Draw a bootstrap-resample $\varepsilon_k^*$ of i.i.d. observations from $\hat{f}_{\tilde{\varepsilon},h}$ as follows

   $$\varepsilon_k^* = \hat{F}_n^{-1}(U) + hZ, \quad k = 1, 2, 3, \ldots$$

   where $U$ is a random value from uniform distribution $U(0, 1)$ and $Z$ is a random value from a variable with density $H(u)$.

5. Define the bootstrap series $X_t^*$, $t = 1, \ldots, T$, by the recursion

   $$X_t^* = \hat{m}_{g_1}(\boldsymbol{X}_{t-1}^*) + \varepsilon_t^*,$$

   where $\hat{m}_{g_1}$ is defined in Step (2).

6. Estimate the bootstrap autoregressive function, $m^*$, on the basis of the bootstrap sample $(X_1^*, \ldots, X_T^*)$ obtained in the previous step. Estimation is carried out using again the modified Nadaraya-Watson estimator with bandwidth $g_2$. The resulting estimator is denoted by $\hat{m}_{g_2}^*$.

7. Compute bootstrap prediction-paths by the recursion

$$X_t^* = \hat{m}_{g_2}^*(\boldsymbol{X}_{t-1}^*) + \varepsilon_t^*,$$

for $t = T+1, T+2, \ldots, T+b$, $b > 0$, where $T+b$ is the horizon pre-selected by the user to carry out the clustering, and $X_t^* = X_t$, for $t \leq T$.

8. Repeat Steps (1)-(7) a large number ($B$) of times to obtain replications of the $b$-step ahead bootstrap future observations.

Now, we come back to the clustering procedure. Applying the resampling method to the $i$th time series in study, $\mathbf{X}^{(i)}$, provides a bootstrap sample $(X_{T+b}^{(i)*1}, X_{T+b}^{(i)*2}, \ldots, X_{T+b}^{(i)*B})$ that allows us to estimate the unknown density of $X_{t+b}^{(i)}$. In particular, we consider the Rosenblatt-Parzen kernel smoother to obtain $\hat{f}_{X_{T+b}^{(i)*}}(x)$, the $b$-step-ahead density estimator at point $x$ for the $i$th time series, $i = 1, \ldots, s$. Then, the $L^1-$distance $D_{ij}^1$ can be approximated by the "plug-in" version

$$\hat{D}_{ij}^{1*} = \int \left| \hat{f}_{X_{T+b}^{(i)*}}(x) - \hat{f}_{X_{T+b}^{(j)*}}(x) \right| dx, \quad i, j = 1, \ldots, s. \tag{6}$$

Alternatively, the distance $D_{ij}^2$ can be approximated by

$$\hat{D}_{ij}^{2*} = \frac{1}{B}\sum_{t=1}^{B}\left(X_{T+b}^{(i)*t}\right)^2 + \frac{1}{B}\sum_{t=1}^{B}\left(X_{T+b}^{(j)*t}\right)^2 - \frac{2}{B}\sum_{t=1}^{B} X_{T+b}^{(i)*t} X_{T+b}^{(j)*t}, \quad i, j = 1, \ldots, s. \tag{7}$$

Once the pairwise dissimilarity matrix $\hat{D}^{u*} = \left(\hat{D}_{ij}^{u*}\right)$, $u = 1, 2$, is obtained, a standard clustering algorithm based on $\hat{D}^{u*}$ is carried out. We consider an agglomerative hierarchical clustering method.

In many cases, the time series are not stationary. In this case, each of the time series is transformed using logarithms (if required) and taking an appropriate number of regular differences. The bootstrap prediction-paths for the transformed series were constructed following Steps (1)-(8). Then, the resulting bootstrap prediction-paths are backtransformed to obtain the bootstrap predictions for the original series.

In this work, we establish the consistency of our dissimilarity measures under appropriate conditions, and hence our clustering procedures asymptotically lead to the clustering using the true generating process. A simulation study is carried out. The results show the good behavior of our procedures for a wide variety of nonlinear autoregressive models and its robustness to non Gaussian innovations. Finally, the proposed methodology is applied to a real data set involving economic time series. More specifically the dataset is formed by a collection of series representing the monthly industrial production indices for European countries.

# References

Alonso, A.M., Berrendero, J.R., Hernández, A., Justel, A. (2006). Time series clustering based on forecast densities. *Comput. Statist. Data Anal.*, **51**, 762-776.

Liao, T.W. (2005). Clustering of time series data: a survey. *Pattern Recognit.*, **38**, 1857-1874.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford Statistical Science Series, 6.

# IV

# Posters

# Poster 1

# On the estimation of the density probability by trigonometric series

**Smail Adjabi**
*Laboratory LAMOS. University of Bejaia, Algeria*

**Nora Saadi**
*Laboratory LAMOS. University of Bejaia, Algeria*

**Abstract.** We present in this paper an estimator based on a new orthogonal trigonometric series. We give its statistical properties, the asymptotic properties and the rate of convergence of the mean integrated square error. The comparison by simulation on a test density between the estimator obtained from a new trigonometric system with Fejer estimator also based on orthogonal trigonometric system, show that our estimator is more performant in the sense of the mean integrated square error.

## 1 Introduction

The nonparametric probability density estimation by orthogonal series is a good alternative to the popular method of the kernel estimator of Parzen-Rosenblatt, when the support of the densities to estimate is contained in a fixed compact interval of the real line, since in this case convenient orthonormal systems are available, such as trigonometric basis. Density estimation using orthogonal functions is a topic which has received considerable attention in the literature in recent years. Some notable examples include work by Cencov [4], Kronmal and Tarter [7], [8], Wahba [9], Hall [5], [6] and Bosq [2].
Assume that $h(t)$ is a continuous and strictly positive density function (with respect to Lebesgue measure) over the compact interval $I = [a, b]$ and suppose that $\{e_k(t)\}_{k=0}^{\infty}$ is a complete orthonormal basis for $L_2([a, b])$. By assumption, the density $h(t)$ can be

represented in an orthogonal series expansion by

$$h(t) = \sum_{k=0}^{\infty} a_k e_k(t), \quad t \in [a,b], \tag{1}$$

where

$$a_k = \int_a^b e_k(x)h(x)dx = \mathbb{E}[e_k(X)], \quad k = 0,1,2\ldots. \tag{2}$$

For convenience, suppose that $\sum_{k=0}^{\infty} a_k e_k(t)$ converges uniformly on $[a,b]$ and each function $e_k(t)$ is continuous. Let $X_1, X_2, \ldots, X_n$ be independent, identically distributed random variables having the common density $h(t)$ on $I = [a,b]$. Since $a_k = \int_a^b e_k(x)h(x)dx = \mathbb{E}[e_k(X)]$, $k = 0,1,2\ldots$ , it is clear that $\hat{a}_k = \frac{1}{n}\sum_{i=1}^{n} e_k(X_i)$, is an unbiased estimator of $a_k$.
An estimate of $h(t)$ is given by

$$\hat{h}_{d_n}(t) = \sum_{k=0}^{d_n} \hat{a}_k e_k(t), \tag{3}$$

where, $(d_n)$ is a sequence of positive numbers chosen so that $d_n \longrightarrow \infty$ when $n \longrightarrow \infty$. The aim of this work is to introduce an estimator based on a new orthogonal trigonometric series and give its statistical properties and the asymptotically properties. To justify choice and quality means of the obtained estimator, one numerically compares it, on the basis of integrated mean square error by simulation, with Fejer estimator also based on orthogonal trigonometric series.

## 2 Estimator based on a new trigonometric series

Let $X_1, \ldots, X_n$ be a sample of independent identically distributed random variables, each with the common (unknown) probability density function $h$, over the compact interval $I = [-\pi, \pi]$. An estimate of $h(t)$ is given by

$$\hat{h}_{d_n}(t) = \sum_{k=0}^{d_n} \hat{a}_k e_k(t)\,, \tag{4}$$

where the new trigonometric basis is the form

$$e_k(t) = \frac{1}{\sqrt{2\pi}}(\cos(kt) + \sin(kt))1_{[-\pi,\pi]}(t), \; k = 0,1,2\ldots, \tag{5}$$

$$\hat{a}_k = \frac{1}{n}\sum_{i=1}^{n} e_k(X_i) = \frac{1}{\sqrt{2\pi}n}\sum_{i=1}^{n}(\cos k(X_i) + \sin k(X_i))\,, \tag{6}$$

and $(d_n)$ is a sequence of positive numbers chosen so that $d_n \longrightarrow \infty$ when $n \longrightarrow \infty$. The estimate $\hat{h}_{d_n}(t)$ of $h(t)$ is then

$$\hat{h}_{d_n}(t) = \frac{1}{4\pi n}\sum_{i=1}^{n}\left[\frac{\sin[\frac{(2d_n+1)(X_i-t)}{2}]}{\sin[\frac{X_i-t}{2}]} + \frac{\sin[\frac{(2d_n+1)(\frac{\pi}{2}-(X_i+t))}{2}]}{\sin[\frac{\frac{\pi}{2}-(X_i+t)}{2}]}\right]. \tag{7}$$

**Theorem 1** *Assume that $h(t)$ is square integrable and that the sequence of positive integers $d_n$ is chosen so that $d_n \longrightarrow \infty$ as $n \longrightarrow \infty$. Then*

$$MISE(\hat{h}_{d_n}(t)) = \int_{-\pi}^{\pi} h^2(t)dt + \frac{d_n+1}{2\pi n} + \frac{1}{\sqrt{2\pi}n}\sum_{k=0}^{d_n}\beta_{2k} - \frac{n+1}{n}\sum_{k=0}^{d_n} a_k^2. \tag{8}$$

**Theorem 2** *If $d_n = o(\sqrt{n}))$ and $d_n \longrightarrow \infty$ as $n \longrightarrow \infty$, then*

$$\lim_{n\longrightarrow\infty} \mathbb{V}(\hat{h}_{d_n}(t)) = 0.$$

**Theorem 3** *If $d_n = o(\sqrt{n}))$ and $d_n \longrightarrow \infty$ as $n \longrightarrow \infty$, then*

$$\lim_{n\longrightarrow\infty} MSE(\hat{h}_{d_n}(t)) = \lim_{n\longrightarrow\infty} \mathbb{E}[\hat{h}_{d_n}(t) - h(t)]^2 = 0.$$

**Theorem 4** *If $d_n = o(\sqrt{n})$ and $d_n \longrightarrow \infty$ as $n \longrightarrow \infty$, then*

$$\lim_{n\longrightarrow\infty} MISE(\hat{h}_{d_n}(t)) = \lim_{n\longrightarrow\infty} \mathbb{E}\int_{-\pi}^{\pi} [\hat{h}_{d_n}(t) - h(t)]^2 dt = 0.$$

**Theorem 5** *If $d_n \longrightarrow \infty$ as $n \longrightarrow \infty$, then*

$$\lim_{n\longrightarrow\infty} P[\sup_{t\in[-\pi,\pi]} \mid \hat{h}_{d_n}(t) - h(t) \mid < \epsilon] = 1,\ \forall \epsilon > 0.$$

**Theorem 6** *If*

$$\sum_{k=d_n+1}^{\infty} a_k^2 = O(d_n^{-r}), \quad r > 0, \tag{9}$$

*and*

$$d_n = \alpha \log n, \quad 0 < \alpha < 1. \tag{10}$$

*Then,*

$$MISE(\hat{h}_{d_n}(t)) = O((\log n)^{-\delta}),\ \delta = \min(1-\alpha, r). \tag{11}$$

# 3 Fejer estimator

Given a sample set $X_1, \ldots, X_n$ of independent identically distributed random variables, each with the common (unknown) probability density function $h$, over the compact interval $I = [a, b]$. The estimator of $h(t)$ is defined in [1] by

$$\hat{h}_{d_n}(t) = \frac{c_0}{2} + \sum_{k=1}^{d_n}(1 - \frac{k}{d_n+1})[\overline{c_k}\cos(kz(t)) + \overline{s_k}\sin(kz(t))], \tag{12}$$

where

$$z(t) = \frac{2\pi(t-a)}{b-a},\ c_0 = \frac{2}{b-a},\ \overline{c_k} = \frac{c_0}{n}\sum_{i=1}^{n}\cos(kz(X_i)),\ \overline{s_k} = \frac{c_0}{n}\sum_{i=1}^{n}\sin(kz(X_i)) \tag{13}$$

The mean integrated square error of Fejer estimator [1] is

$$\begin{aligned} MISE(\hat{h}_{d_n}(t)) &= c_0 \int_a^b h^2(t)dt - \frac{c_0^2}{2} + \sum_{k=1}^{d_n}(1 - \frac{k}{d_n+1})^2[\mathbb{V}ar(\overline{c_k}) + \mathbb{V}ar(\overline{s_k}) \\ &- \frac{d_n+1+k}{d_n+1-k}(c_k^2 + s_k^2)] \end{aligned} \tag{14}$$

# 4 Selection of terms in an orthogonal series density estimator

The performance and smoothness of the orthogonal series density estimate depend on $d$: $0 < d \leq d_n$, the number of terms in the series expansion. Kronmal and Tarter [7] proposed a term by term optimal stopping rule for choosing $d$ by minimizing an estimated $MISE$ criterion.
The rule adopted to determine the optimal value $d^*$ rests on the following algorithm. From $d = 1$, one increases the value of $d$ unit until $MISE$ increases. We give to $d^*$ the preceding value just before the increase of $MISE$. Then, we will add to sum (4) the $d^{th}$ term if and only if

$$\Delta_d = MISE(\hat{h}_d(t)) - MISE(\hat{h}_{d-1}(t)) \leq 0. \tag{15}$$

We have

$$\Delta_d = MISE(\hat{h}_d(t)) - MISE(\hat{h}_{d-1}(t)) = \frac{n+1}{n}\mathbb{V}ar(e_d(X)) - \mathbb{E}(e_d^2(X))$$

Let us set $\theta_i = e_d(X_i), i = 1, ..., n \;\; and \;\; \overline{\theta} = \frac{1}{n}\sum_{i=1}^n \theta_i$. A symmetric unbiased estimator of $\Delta_d$ is defined in [3] by

$$\hat{\Delta}_d = \frac{1}{n}\left[\frac{n+1}{n-1}\sum_{i=1}^n(\theta_i - \overline{\theta})^2 - \sum_{i=1}^n \theta_i^2\right]. \tag{16}$$

Now, let us set a positive integer $d_n$, the optimal $d^*$ is then of the form

$$d^* = \begin{cases} \inf\{d : 1 \leq d \leq d_n\} & \text{if } \hat{\Delta}_d > 0 \\ d_n & \text{otherwise.} \end{cases}$$

In the case of the estimator based on a new trigonometric series.

$$\theta_i = e_d(X_i) = \frac{1}{\sqrt{2\pi}}(\cos(X_i) + \sin(X_i)), \quad i = 1, ..., n$$

and

$$\overline{\theta} = \frac{1}{n}\sum_{i=1}^n \left(\frac{1}{\sqrt{2\pi}}(\cos(X_i) + \sin(X_i))\right).$$

In the case of Fejer Estimator.
The unbiased estimate of $\Delta_d$ is defined in [1] by

$$\hat{\Delta}_d = \frac{1}{n}\left[c_0^2 - \frac{2nd+n+1}{n-1}(n\overline{c}_d^2 + n\overline{s}_d^2 - c_0^2)\right],$$

where $c_0$, $\overline{c}_d$ et $\overline{s}_d$ are defined in (13).

# 5 Simulation studies

To justify the quality of the new estimator, one carries out simulation studies for several values of $n$ on a test density: the normal $\mathcal{N}(0.85, 1)$. The numerical experimentation will be useful for us to compare the performances of the new estimator with the Fejer one, when comparing mean integrated square error of the two estimators. Results of simulation are given in table 1, where

$d^*_{NB}$: Optimal smoothing parameter associated to the new estimator;

$d^*_F$: Optimal smoothing parameter associated to the Fejer estimator;

$MISE^*_{NB}$: Optimal mean integrated square error associated to the new estimator;

$MISE^*_F$: Optimal mean integrated square error associated to the Fejer estimator.

| $n$ | $d^*_{NB}$ | $d^*_F$ | $MISE^*_{NB}$ | $MISE^*_F$ |
|---|---|---|---|---|
| 50 | 2 | 3 | 0.008532792 | 0.01244004 |
| 100 | 3 | 3 | 0.006326546 | 0.01033164 |
| 1000 | 3 | 3 | 0.004498078 | 0.008434078 |
| 1500 | 3 | 3 | 0.004371485 | 0.008363798 |
| 2500 | 3 | 3 | 0.004270211 | 0.008307574 |
| 3000 | 3 | 3 | 0.004244893 | 0.008293518 |
| 5000 | 3 | 3 | 0.004194256 | 0.008265406 |
| 7500 | 3 | 3 | 0.004168937 | 0.00825135 |
| 10000 | 3 | 3 | 0.004156278 | 0.008244321 |

**Table 1:** Optimal smoothing parameter and optimal MISE for the new and Fejer basis.

- The Kronmal Tarter method gives the same values of the smoothing parameter when estimating the density either by the new basis or by the Fejer basis.
- The mean integrated square error values $MISE^*_{NB}$ associated to the new estimator are always lower than the mean integrated square error values $MISE^*_F$ associated to the Fejer estimator. What means that the new estimator is more performant than that the Fejer one.
- The mean integrated square error associated to the new estimator decreases, when increasing the sample size. However, it is noted that the convergence towards 0 is very slow.

# References

[1] Asselin de Beauville, J.P. (1978). Estimation non paramétrique de la densité de probabilité et du mode. Exemple de la distribution gamma. *Revue de Statistique Appliquée*, 26, 47-70.

[2] Bosq, D. (2005). Estimation suroptimale de la densité par projection. *Canadian Journal of Statistics*, 33 (1), 21-37.

[3] Bosq, D. and Lecoutre J. (1987). *Théorie de l'estimation fonctionnelle*. Economica Edition.

[4] Cencov, N. N. (1962). Evaluation of an unknown distribution density from observations. Soviet Mathematics, 3, 1559-1562.

[5] Hall, P. (1983). Orthogonal series distribution function estimation, with applications. *Journal of the Royal Statistical Society*, 45, 81-88

[6] Hall, P. (1986). On the rate of the convergence of orthogonal series density estimation. *Journal of the Royal Statistical Society*, 48, 115-122.

[7] Kronmal, R. and Tarter, M. (1968). The estimation of probability densities and cumulatives by Fourier series methods. *American Statistical Association Journal*, 63, 925-952.

[8] Tarter, M. and Kronmal, R. (1976). An introduction to the implementation and theory of nonparametric density estimation. *The American Statistician*, 30, 105-112.

[9] Wahba, G. (1981). Data based optimal smoothing of orthogonal series density estimates. *The Annals of Statistics*, 9(1), 146-156.

# Poster 2

# Inference for a partly linear autoregressive model with moving average errors

**Ana M. Bianco**
*Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET*

**Graciela Boente**
*Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET*

**Abstract.** In this talk, we generalize the partly linear autoregression model considered in the literature by including moving average errors when we want to allow a large dependence to the past observations. The strong ergodicity of the process is derived. A Fisher–consistent procedure to estimate the parametric and nonparametric components is provided together with a test statistic that allows to check the presence of a moving average component. Also, a Monte Carlo study is carried out to check the performance of the given proposals.

## Summary

When dealing with time series data, autoregressive models with moving average errors (ARMA models) have been extensively used in applications. They correspond to linear autoregressive models where the errors are described by a moving average process. More precisely, an ARMA$(p, q)$ model, is a stationary process $\{y_t : t \geq 1\}$ verifying

$$y_t = \sum_{j=1}^{p} \varphi_j y_{t-j} + \varepsilon_t \,, \tag{1}$$

where $\varepsilon_t = u_t - \sum_{j=1}^{q} \theta_j u_{t-j}$ with $u_t$ independent and identically distributed (i.i.d.) random variables and $u_t$ is independent of $\{y_{t-j}, j \geq 1\}$ with $E|u_t| < \infty$.

It is well known that, when there is large dependence to the past observations, ARMA models have several advantages with respect to autoregressive models. However, the assumption of a linear autoregression function is quite restrictive. As pointed by Bosq (1996) a nonparametric predictor is "in general more efficient and more flexible than the predictor based on Box and Jenkins method and nearly equivalent if the underlying model is truly linear", see also Carbon and Delecroix (1993) for a comparative study on 17 series. Nevertheless, the nonparametric autoregression model $y_t = m(\mathbf{X}_t) + u_t$, where $\mathbf{X}_t = (y_{t-1}, \ldots, y_{t-r})^{\mathrm{T}}$, faces the problem known as the "curse of dimensionality". In order to solve the problem of empty neighborhoods, an approach can be to introduce moving average errors which reduce the dependence to the past in $\mathbf{X}_t$ obtaining, thus, a smaller dimension $r$. This approach was followed by Boente and Fraiman (2002) who introduced nonparametric ARMA models that allow the autoregressive part of the model to be nonparametric, while the moving average part remains linear.

As noted by Gao and Yee (2000), another disadvantage of the fully nonparametric autoregresive model is that it neglects a possible linear relationship between $y_t$ and any lag $y_{t-k}$ . To solve the "curse of dimensionality", following a semiparametric approach, several authors have introduced partly linear models for autoregressive models in order to combine the advantages of both parametric and nonparametric methods. A stochastic process $\{y_t\}$, defined over a probability space $(\Omega, \mathcal{A}, \mathcal{P})$, satisfies a partly linear autoregressive model if it can be written as

$$y_t = \sum_{i=1}^{p_1} \beta_{o,i} y_{t-c_i} + \sum_{j=1}^{p_2} g_{o,j}(y_{t-d_j}) + u_t \ , \tag{2}$$

where $g_{o,j} : I\!R \to I\!R$ are smooth functions and $u_t$ are i.i.d. random variables, independent of $\{y_{t-j}, j \geq 1\}$, $Eu_t = 0$ and $Eu_t^2 < \infty$. However, these models do not take into account a large dependence to the past unless $p_1$ and $p_2$ are large. To reduce the order of the process, we can allow a dependence structure in the errors as in (1). Combining models (1) and (2), one can consider a stationary process $\{y_t : t \geq 1\}$ verifying

$$y_t = \sum_{i=1}^{p_1} \beta_{o,i} y_{t-c_i} + \sum_{j=1}^{p_2} g_{o,j}(y_{t-d_j}) + \varepsilon_t \ , \qquad \varepsilon_t = u_t - \sum_{j=1}^{q} \theta_{o,j} u_{t-j} \tag{3}$$

with $u_t$ i.i.d. random variables and $u_t$ independent of $\{y_{t-j}, j \geq 1\}$, $E|u_t| < \infty$. From now on, we will refer to a stochastic process verifying (3) as a partly linear ARMA$(p_1, p_2, q)$ model and it will be denoted by PARTLIARMA$(p_1, p_2, q)$ model.

For simplicity and convenience, we will focus our attention on the case $p_1 = p_2 = 1$, $c_1 = 1$ $d_1 = 2$, which leads to the PARTLIARMA$(1, 1, q)$ model

$$y_t = \beta_o y_{t-1} + g_o(y_{t-2}) + \varepsilon_t \ , \qquad \varepsilon_t = u_t - \sum_{j=1}^{q} \theta_{o,j} u_{t-j} \tag{4}$$

with $u_t$ i.i.d. and $u_t$ independent of $\{y_{t-j},\ j \geq 1\}$, $E|u_t| < \infty$.

We provide sufficient conditions that allow to establish the strong ergodicity of the process. Besides, we discuss several issues regarding how to define a Fisher–consistent functional for $g_o$, $\beta_o$ and $\boldsymbol{\theta}_o$ which allows to define estimators of parametric and nonparametric components through an iterative procedure.

We present a statistic to test $H_0 : \boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^{\mathrm{T}} = \mathbf{0}$, that is to check the presence of a moving average component and we derive its asymptotic distribution.

Finally, we compare the performance of the given proposals through a Monte Carlo study.

## References

Boente, G. and Fraiman, R. (2002). Ergodicity, geometric ergodicity and mixing conditions for nonparametric ARMA processes. *Bulletin of the Brazilian Mathematical Society*, **33**, 13-23.

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction.* Lectures Notes in Statistics, **110**, Springer–Verlag.

Carbon, M. and Delecroix, M. (1993). Nonparametric forecasting in time series, a computational point of view. *Applied Stoch. Models and Data Analysis*, **9**, 215-229.

Gao, J. and Yee, T. (2000). Adaptive estimation in partly linear autoregressive models. *Canad. J. Statist.*, **28**, 571–586.

# Poster 3

# Logistic and local logistic distance-based regression

**Eva Boj**[1]
*Dep. Matemàtica Econòmica, Financera i Actuarial. Univ. de Barcelona*

**Pedro Delicado**[1]
*Dep. d'Estadística i Investigació Operativa. Univ. Politècnica de Catalunya*

**Josep Fortiana**[1]
*Dep. de Probabilitat, Lògica i Estadística. Univ. de Barcelona*

**Abstract.** This paper introduces Logistic and Local Logistic Distance-Based Regression. Only distances between individuals are required to fit these binary regression models. Therefore they are applicable to mixed (quantitative and qualitative) explanatory variables or when the regressor is of functional type.

## 1 Introduction

Boj, Delicado and Fortiana (2008) introduced a nonparametric regression technique (local linear distance-based regression) as an application of the Weighted Distance-Based Regression (DBR) presented there. In this paper we introduce other applications of the same technique. In general, any statistical technique based on Weighted Least Squares can be adapted to accept data given in the format of an inter-individual distances matrix, just replacing any Weighted Least Squares step by the corresponding Weighted DBR.

There are many statistical techniques based on Iterative Reweighted Least Squares, ranging from Generalized Linear Models fitting algorithms (McCullagh and Nelder 1989) to Robust Regression estimation (see, for instance, Street, Carroll and Ruppert 1988). All of them can be adapted to be used when the available information from covariates is given by a distances matrix. Here we develop Logistic DBR in detail to illustrate how this adaptation has to be done.

[1]Work supported by grant MTM2006-09920.

Distance-Based Regression (DBR) was introduced by Cuadras and Arenas (1990). Let $\Omega = \{\mathcal{O}_1, \ldots, \mathcal{O}_n\}$ be a set of $n$ objects (or individuals or cases) randomly drawn from a population. For individual $\mathcal{O}_i$ we have observed the value $y_i$ of a continuous one-dimensional response variable. A distance function $\delta$ (being a metric or semi-metric) is defined between the elements of $\Omega$, usually based on predictors $\boldsymbol{Z}$ observed for every $\mathcal{O}_i \in \Omega$ as $\boldsymbol{z}_i$. Let $\boldsymbol{\Delta} = (d^2_{i,j})_{i=1..n,j=1..n}$ be the inter-individual squared distances matrix. The available information $\boldsymbol{Z}$ for the elements of $\Omega$ can be a mixture of quantitative and qualitative variables or, possibly, other nonstandard quantities, such as character strings, functions or other kind of non-numerical explanatory variables. The aim of the DBR is to predict the response variable for a new individual $\mathcal{O}_{n+1}$ from the same population, using $(d^2_{n+1,1}, \ldots, d^2_{n+1,n})$, the vector of squared distances from $\mathcal{O}_{n+1}$ to the remaining individuals, as the only available information. Boj, Delicado and Fortiana (2008) introduce the weighted version of DBR, where each response $y_i$ has a weight $w_i \leq 0$.

## 2 Logistic Distance-Based Regression

Let $\Omega = \{\mathcal{O}_1, \ldots, \mathcal{O}_n\}$ be a set of $n$ individuals (or cases) randomly drawn from a population. For individual $\mathcal{O}_i$ we have observed $y_i$, the value of a binary (0-1) response variable. A distance function $\delta$ (being a metric or semi-metric) is defined between the elements of $\Omega$ (possibly based on observed covariates $\boldsymbol{Z}$). We assume that all the relevant information on the relation between $y_i$ and other characteristics of individual $\mathcal{O}_i$, $i = 1, \ldots, n$, is summarized in the relation between the $n$ vector $\boldsymbol{y} = (y_1, \ldots, y_n)$ and the $n \times n$ inter-individuals squared distances matrix $\boldsymbol{\Delta}$. Given a new individual $\mathcal{O}_{n+1}$ from the same population, our objective is to predict $y_{n+1}$ using the squared distances from $\mathcal{O}_{n+1}$ to the remaining individuals, as the only available information.

Observe that we are facing up to a *distance-based binary regression* problem. To fit such a model, we assume that the data are following a logistic model in the following sense. Let the $n \times r$ matrix $\boldsymbol{X}$ be a Euclidean configuration of $\boldsymbol{\Delta}$, and let $\boldsymbol{x}_i'$ be the $i$-th row of $\boldsymbol{X}$. Then we assume that $y_i$ is an observation of

$$Y_i \sim \text{Bern}(p_i),\ \text{logit}(p_i) = \boldsymbol{x}_i'\boldsymbol{\beta},$$

for an unknown $\boldsymbol{\beta} \in \mathbf{R}^r$, where $\text{logit}(p) = \log(p/(1-p))$. Observe that this assumption does not depend of the particular choice of the Euclidean configuration $\boldsymbol{X}$: let $\boldsymbol{V}$ be a different $r$-dimensional Euclidean configuration of $\boldsymbol{\Delta}$, then $\boldsymbol{V} = \boldsymbol{X} \cdot \boldsymbol{T}$, where $\boldsymbol{T}$ is a $r \times r$ orthogonal matrix, and therefore

$$\boldsymbol{X}\boldsymbol{\beta} = (\boldsymbol{X} \cdot \boldsymbol{T} \cdot \boldsymbol{T}')\boldsymbol{\beta} = \boldsymbol{V}\boldsymbol{\gamma},$$

and $\text{logit}(p_i) = \boldsymbol{x}_i'\boldsymbol{\beta} = \boldsymbol{v}_i'\boldsymbol{\gamma}$ with $\boldsymbol{\gamma} = \boldsymbol{T}'\boldsymbol{\beta} \in \mathbf{R}^r$. We conclude that only the relation between $\boldsymbol{y}$ and $\boldsymbol{\Delta}$ determines whether the logistic model is adequate or not.

We propose an algorithm to compute the fitted values $\hat{\boldsymbol{y}}$ and the prediction $\boldsymbol{y}_{n+1}$ in logistic regression model that not need explicit statement of the Euclidean configuration $\boldsymbol{X}$, that is, $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{y}}_{n+1}$ are independent of the particular choice of $\boldsymbol{X}$. It is an adaptation of the standard Reweighted Least Squares Algorithm used to fit standard logistic regression (see, for instance, Section 13.7 in Wasserman 2004 or Appendix 14.A in Peña 2002).

**Algorithm for Logistic DBR.**
Choose starting values $\boldsymbol{p}^0 = (p_1^0, \ldots, p_n^0)$ (for instance, $\boldsymbol{p}^0$ can be the fitted values of $\boldsymbol{y}$ given by a standard DBR if all the $\hat{y}_i$ are in $(0, 1)$). Set $s = 0$ and iterate the following steps until convergence.

1. Set
$$z_i^s = \text{logit}(p_i^s) + \frac{y_i - p_i^s}{p_i^s(1 - p_i^s)},\ i = 1, \ldots, n.$$
2. Let the weight vector $\boldsymbol{\nu}^s = (\nu_1^s, \ldots, \nu_n^s)'$ with $\nu_i^s \propto p_i^s(1 - p_i^s)$.
3. Fit the Weighted DBR (fitting equation for the Weighted DBR; see Boj, Delicado and Fortiana 2008) using the squared distances matrix $\boldsymbol{\Delta}$, the response vector $\boldsymbol{z}^s = (z_1^s, \ldots, z_n^s)$, and the weight vector $\boldsymbol{\nu}^s$. Let $\hat{\boldsymbol{z}}^s = (\hat{z}_1^s, \ldots, \hat{z}_n^s)$ the fitted values. Define
$$p_i^{s+1} = \frac{\exp(\hat{z}_i^s)}{1 + \exp(\hat{z}_i^s)}$$
and $\boldsymbol{p}^{s+1} = (p_1^{s+1}, \ldots, p_n^{s+1})$.
4. Set $s = s + 1$ and go back to the step 1.

The fitted values $\hat{\boldsymbol{y}}$ are
$$\hat{\boldsymbol{y}} = \boldsymbol{p}^{s+1}$$
where $s$ is the last iteration of the algorithm, when convergence is achieved.

This algorithm allows us to give the prediction $\hat{p}_{n+1}$ of the expected response value for a new individual $\mathcal{O}_{n+1}$, with squared distances $\boldsymbol{d}_{[n+1]}$ to the other individuals. In the last iteration of the algorithm we use prediction equation for new cases in the Weighted DBR (see Boj, Delicado and Fortiana 2008) to obtain the predicted value of $z$ for the new individual, $\hat{z}_{n+1}^s$, and then we compute
$$\hat{p}_{n+1} = \frac{\exp(\hat{z}_{n+1}^s)}{1 + \exp(\hat{z}_{n+1}^s)}.$$

Equations defining $\hat{\boldsymbol{y}}$ and $\hat{p}_{n+1}$ are the core of Logistic DBR.

### 2.1 Weighted Logistic Distance-Based Regression

The algorithm for Logistic DBR can be easily modified when original data have different weight. Let $\boldsymbol{w} = (w_1, \ldots, w_n)'$ be these weights. Only two modifications are needed. First, the initial values $\boldsymbol{p}^0$ can be obtained by Weighted DBR with weights $\boldsymbol{w}$. Second, in Step 2, the $i$-th element $\nu_i^s$ of $\boldsymbol{\nu}^s$ now must be proportional to $w_i p_i^s(1-p_i^s)$, $i = 1, \ldots, n$.

## 3 Local Logistic Distance-Based Regression

We consider again the framework stated in the previous section when Logistic DBR was introduced. Our objective is now to fit a *local Logistic DBR*, where *local* refers to the fact that when the Logistic DBR model is used to predict the value of the binary response variable for an object $\mathcal{O}_{n+1}$, we use only the information provided by observed objects $\mathcal{O}_i$,

$i = 1, \ldots, n$, that are *close* to $\mathcal{O}_{n+1}$, giving to $\mathcal{O}_i$ a weight that is a decreasing function of the distance between $\mathcal{O}_i$ and $\mathcal{O}_{n+1}$. The idea is to translate to the Logistic DBR context the principles of *local likelihood*, as stated in Loader (1999). Our approach parallels that used in Boj, Delicado and Fortiana (2008) when local linear DBR is defined.

Let $m(\mathcal{O}_{n+1})$ the expected value of the binary response $y$ corresponding to the object $\mathcal{O}_{n+1}$. This is the value we want to estimate and we do that by using Weighted Logistic DBR. We assume that two distance functions, $\delta_1$ and $\delta_2$, are defined between the elements of $\Omega$ (the set of observable objects). We consider the weights

$$w_i(\mathcal{O}_{n+1}) = K(\delta_1(\mathcal{O}_{n+1}, \mathcal{O}_i)/h) / \sum_{j=1}^{n} K(\delta_1(\mathcal{O}_{n+1}, \mathcal{O}_j)/h)$$

where $h$ is an smoothing parameter (depending on $n$). Let $\mathbf{\Delta}_2$ be the matrix of squared distances between functions defined from $\delta_2$. We fit a Weighted Logistic DBR starting from the initial elements

$$\mathbf{\Delta}_2 = (\delta_2(\mathcal{O}_i, \mathcal{O}_j)^2)_{i=1\ldots n, j=1\ldots n},\ \boldsymbol{y} = (y_i)_{i=1\ldots n},\ \text{ and } \boldsymbol{w} = (w_i(\mathcal{O}_{n+1}))_{i=1\ldots n}$$

as it is stated in Subsection 2.1. We consider the new individual $\mathcal{O}_{n+1}$ and we compute the squared distances from object $\mathcal{O}_{n+1}$ to other individuals $\mathcal{O}_i$:

$$\boldsymbol{\delta}_{2,n+1} = (\delta_2(\mathcal{O}_{n+1}, \mathcal{O}_1)^2, \cdots, \delta_2(\mathcal{O}_{,}\mathcal{O}_n)^2).$$

Then we run the Algorithm for Logistic DBR that takes into account that observations are weighted (Subsection 2.1) and the equation defining $\hat{p}_{n+1}$ is used to obtain the *Local Logistic DBR* estimator of $m(\mathcal{O}_{n+1})$:

$$\hat{m}_{LLDBR}(\mathcal{O}_{n+1}) = \hat{y}_{n+1}.$$

There are two distance functions involved in the local logistic distance-based estimation: one of them, $\delta_1$, is used to compute the weight of observed objects $\mathcal{O}_i$ around the object $\mathcal{O}_{n+1}$ where the response function is estimated, and the other, $\delta_2$, defines the distances between observations for computing the Logistic DBR. The distances $\delta_1$ and $\delta_2$ can coincide or not. In the context of local linear DBR Boj, Delicado and Fortiana (2008) show that using two distance functions provides much more flexibility than using only one (that is, $\delta_1 = \delta_2$).

# References

Boj, E., P. Delicado and J. Fortiana (2008). Local linear functional regression based on weighted distance-based regression. *Submitted.*

Cuadras, C. M. and C. Arenas (1990). A distance-based regression model for prediction with mixed data. *Communications in Statistics A. Theory and Methods*, **19**, 2261–2279.

Loader, C. (1999). *Local Regression and Likelihood*. New York: Springer.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models (2nd ed)*. London: Chapman and Hall.

Peña, D. (2002). *Regresión y Diseño de Experimentos.* Madrid: Alianza Editorial. In Spanish.

Street, J. O., R. J. Carroll and D. Ruppert (1988). A note on computing robust regression estimates via iteratively reweighted least squares. *The American Statistician* **42**, 152–154.

Wasserman, L. (2004). *All of Statistics. A Concise course in Statistical Inference.* New York: Springer.

# Poster 4

# Transformation kernel estimation of insurance risk

**Catalina Bolancé**[1]
*RFA-IREA, Department of Econometrics, University of Barcelona*

**Montserrat Guillén**
*RFA-IREA, Department of Econometrics, University of Barcelona*

**Abstract.** A transformation kernel density estimator that is suitable for heavy-tailed distributions is discussed. Using a truncated Beta transformation, the choice of the bandwidth parameter becomes straightforward. An application to insurance data and the calculation of the value-at-risk are presented.

## 1 Introduction

This paper is about estimating the density function nonparametrically when data are heavy-tailed like in many insurance applications. Other approaches are based on extremes, a subject that has received much attention in the economics literature. Embrechts et al (1999), Coles (2001), Reiss and Thomas (2001) have treated extreme value theory (EVT) in general. Chavez-Demoulin and Embrechts (2004), based on Chavez-Demoulin and Davison (2005), have discussed smooth extremal models in insurance. Their focus is devoted to highlight the nonparametric trends, as a time-dependence is present in many catastrophic risk situations (such as storms or natural disasters) and in the financial marked. A recent work by Cooray and Ananda (2005) combine the lognormal and the Pareto distribution and derive a distribution which has a suitable shape for small claims and can handle heavy tails. Others have addressed this subject with the g-and-h distribution, like Dutta and Perry (2006) for operation risk analysis.

We have analysed claim amounts in a one-dimensional setting and we have realized that a nonparametric approach that accounts for the asymmetric nature of the density is

[1]The Spanish Ministry and /FEDER support SEJ2007-63298 is acknowledged.

preferred for insurance loss distributions (Bolance et al. 2003, Buch-Larsen et al, 2005). Moreover, we have applied the method on a liability data set and compared the nonparametric kernel density estimation procedure to classical methods (Buch-Larsen, 2006). Several authors (Clements et al., 2003) have devoted much interest to transformation kernel density estimation, which was initially proposed by Wand et al. (1991) for asymmetrical variables and based on the shifted power transformation family. Buch-Larsen et al. (2005) proposed an alternative transformations based on a generalization of the Champernowne distribution, simulation studies have shown that it is preferable to other transformation density estimation approaches for distributions that are Pareto-like in the tail. In the existing contributions, the choice of the bandwidth parameter in transformation kernel density estimation is still a problem. One way of undergoing bandwidth choice is to implement the transformation approach so that transformation leads to a beta distribution, then use existing theory to optimize bandwidth parameter choice on beta distributed data and backtransform to the original scale. Our results are particularly relevant for applications in insurance, where the claims amounts are analyzed and usually small claims (low cost) coexist with only a few large claims (high cost).

Given a sample $X_1, ..., X_n$ of independent and identically distributed (iid) observations with density function $f_{\mathbf{x}}$, the classical kernel density estimator is:

$$\hat{f}_{\mathbf{x}}(x) = \frac{1}{n}\sum_{i=1}^{n} K_b(x - X_i), \tag{1}$$

where $b$ is the bandwidth or smoothing parameter and $K_b(t) = K(t/b)/b$ is the kernel (see, Silverman, 1986 or Wand and Jones, 1995).

An error distance between the estimated density $\hat{f}_{\mathbf{x}}$ and the theoretical density $f_{\mathbf{x}}$ that has widely been used in the analysis of the optimal bandwidth $b$ is the mean integrated squared error ($MISE$). It has been shown (see, for example, Silverman, 1986, chapter 3) that the $MISE$ is asymptotically equivalent to $A - MISE$:

$$\frac{1}{4}b^4 (k_2)^2 \int \{f''_{\mathbf{X}}(x)\}^2 dx + \frac{1}{nb}\int K(t)^2 dt, \tag{2}$$

where $k_2 = \int t^2 K(t)\, dt$. If the second derivative of $f_{\mathbf{x}}$ exists (and we denote it by $f''_{\mathbf{X}}$), then $\int \{f''_{\mathbf{x}}(x)\}^2 dx$ is a measure of the degree of smoothness because the smoother the density, the smaller this integral is. From the expression for $A - MISE$ it follows that the smoother $f_{\mathbf{x}}$, the smaller the value of $A - MISE$.

Terrell and Scott (1985, Lemma 1) showed that $Beta(3,3)$ defined on the domain $(-1/2, 1/2)$ minimizes the functional $\int \{f''_{\mathbf{x}}(x)\}^2 dx$ within the set of beta densities with same support. We assume that a transformation exists so that $T(X_i) = Z_i$ $i = 1, ..., n$ is assumed from a $Uniform(0,1)$ distribution. We can again transform the data so that $G^{-1}(Z_i) = Y_i$ $i = 1, ..., n$ is a random sample from a random variable $\mathbf{y}$ with a $Beta(3,3)$ distribution. In this work, we use a parametric transformation $T(\cdot)$, namely the modified Champernowne cdf as proposed by Buch-Larsen et al. (2005) (KMCE method).

Let us define the kernel estimator of the density function for the transformed variable:

$$\hat{g}(y) = \frac{1}{n}\sum_{i=1}^{n} K_b(y - Y_i), \tag{3}$$

| | | | KIBMCE | |
|---|---|---|---|---|
| | Empirical | KMCE | $l = 0.99$ | $l = 0.98$ |
| Young | 1104 | 2912 | 1601 | 1716 |
| Older | 1000 | 1827 | 1119 | 1146 |

**Table 1:** Estimation of $VaR(\mathbf{x}, \mathbf{0.95})$, in thousands.

which should be as close as possible to a $Beta(3,3)$. We can obtain an exact value for the bandwidth parameter that minimizes $A-MISE$ of $\hat{g}$. Finally, in order to estimate the density function of the original variable, since $y = G^{-1}(z) = G^{-1}\{T(x)\}$, the transformation kernel density estimator is:

$$\hat{f}_{\mathbf{x}}(x) = \hat{g}(y)\left[G^{-1}\{T(x)\}\right]' T'(x). \quad (4)$$

The estimator in (3) asymptotically minimizes $MISE$ and the properties of the transformation kernel density estimation (4) are studied in Bolancé et al. (2008). Since we want to avoid the difficulties of the estimator defined in (4), we will construct the transformation so that we avoid the extreme values of the beta distribution domain (KIBMCE method).

## 2 Data study

In this section, we apply our estimation method to a data set that contains automobile claim costs from a Spanish insurance company for accidents occurred in 1997. For small costs, we see that the KIBMCE density in the mode is greater than for the KMCE approach proposed by Buch-Larsen et al. (2005) both for young and older drivers. For both methods, the tail in the estimated density of young policyholders is heavier than the tail of the estimated density of older policyholders. This can be taken as evidence that young drivers are more likely to claim a large amount than older drivers. The KIBMCE method produces lighter tails than the KMCE methods. Based on the results in the simulation study presented in Bolancé et al (2008), we believe that the KIBMCE method improves the estimation of the density in the extreme claims class. Table 1 shows the difference in risk measures.

## 3 Simulation study

This section presents a comparison of our inverse beta transformation method with the results presented by Buch-Larsen, et al. (2005) based only on the modified Champernowne distribution. Our objective is to show that the second transformation, that is based on the inverse of a Beta distribution, improves density estimation. The simulation results can be found in Table 1. For every simulated density and for sample sizes $N = 100$ and $N = 1000$, the results presented here correspond to the following error measures $L_1$, $L_2$ and $WISE$ for different values of the trimming parameter $l = 0.99, 0.98$. The benchmark results are labelled KMCE and they correspond to those presented in Buch-Larsen, et al. (2005).

In general, we can conclude that after a second transformation based on the inverse of a certain Beta distribution cdf the error measures diminish with respect to the KMCE method. In some situations the errors diminish quite substantially with respect to the existing approaches. Note that the KMCE method was studied in Buch-Larsen, et al. (2005) and the simulation study showed that it improved on the error measures for the existing methodological approaches (Clements, et al., 2003 and Wand, et al., 1991).

## References

Bolancé, C., Guillén, M. and Nielsen, J.P. (2003). Kernel density estimation of actuarial loss functions. *Insurance: Mathematics and Economics*, **32**, 19-36.

Bolancé, C., Guillén, M. and Nielsen, J.P. (2008). Inverse Beta transformation in kernel density estimation. *Statistics & Probability Letters*, accepted, to appear.

Dutta, K. and Perry, J. (2006). A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital.*FRB of Boston Working Paper No. 06-13* Available at SSRN: http://ssrn.com/abstract=918880

Buch-Larsen, T., Guillen, M., Nielsen, J.P. and Bolancé, C. (2005). Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. *Statistics*, **39**, 503-518.

Clements, A.E., Hurn, A.S. and Lindsay, K.A. (2003). Möbius-like mappings and their use in kernel density estimation. *Journal of the American Statistical Association*, ´**98**, 993-1000.

Chavez-Denoulin, V. and Davison, P. (2005). Generalized additive modelling of sample extremes. *Applied Statistics*, **54**, 1, 207-222.

Chavez-Denoulin, V. and Embrechts, P. ((2004). Smooth Extremal Models in Finance and Insurance. *The Journal of Risk and Insurance*, **71**, 2, 183-199.

Coles, S. (2001). *An introduction to statistical modeling of extreme values*, Springer, London.

Cooray, K. and Ananda, M.M.A. (2005). Modeling actuarial data with a composite lognormal-Pareto model. *Scandinavian Actuarial Journal*, **5**, 321-334.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1999). *Modelling Extremal Events*, Springer, London.

Reiss, R.D. and Thomas M.(2001). *Statistical Analysis of Extreme Values: with Applications to Insurance, Finance, Hydrology and Other Fields*, 2nd Edition. Birkhauser.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman & Hall.

Terrell, G.R. and Scott, D.W. (1985). Oversmoothed nonparametric density estimates. *Journal of the American Statistical Association*, **80**, 209-214.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing.* London: Chapman & Hall.

Wand, P., Marron, J.S. and Ruppert, D. (1991). Transformations in density estimation. *Journal of the American Statistical Association*, **86**, 414, 343-361.

# Poster 5

# New multi-class and directional clustering tests based on nearest neighbor contingency tables

**Elvan Ceyhan**
*Department of Mathematics, Koç University, Sarıyer, 34450, Istanbul, Turkey*

**Abstract.** Spatial interaction between two or more classes or species has important implications in various fields and causes multivariate patterns such as segregation or association. Segregation occurs when members of a class or species are more likely to be found near members of the same class or conspecifics; while association occurs when members of a class or species are more likely to be found near members of another class or species. The null case for both patterns is either all species or classes exhibit complete spatial randomness (CSR) or random labeling (RL). The clustering tests based on nearest neighbor contingency tables (NNCTs) that are in use in literature are two-sided tests. In this article, I introduce new versions of clustering tests for multiple classes and directional NNCT-tests for the two-class case. I use three examples for illustrative purposes: Pielou's Douglas-fir/penderosa pine data, swamp tree data, and an artificial data set. I also provide guidelines for using these NNCT-tests.

# Poster 6

# Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices

**J.E. Chacón**
*Departamento de Matemáticas. Universidad de Extremadura, Spain*

**T. Duong**
*Computational Imaging and Modeling Group. Institut Pasteur, Paris, France*

**Abstract.** We propose the first plug-in bandwidth selector with unconstrained parametrizations of both the final and pilot bandwidth matrix. This new selector shows the most improvement over the existing methods for target densities whose structure is corrupted by pre-sphering.

## 1 Multivariate kernel density estimation

For a $d$-variate random sample $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ drawn from a density $f$ the kernel density estimator is

$$\hat{f}_{n\mathbf{H}}(\mathbf{x}) = n^{-1} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x} - \mathbf{X_i}) \tag{1}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_d)^T$ and $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{id})^T, i = 1, 2, \ldots, n$. Here $K(\mathbf{x})$ is the multivariate kernel, which we assume to be a spherically symmetric probability density function having a finite second order moment, i.e. there exists $m_2(K) \in \mathbb{R}$ such that $m_2(K)\mathbf{I}_d = \int \mathbf{x}\mathbf{x}^T K(\mathbf{x}) d\mathbf{x}$, where $\mathbf{I}_d$ is the $d \times d$ identity matrix. The parameter $\mathbf{H}$ is the bandwidth matrix, which is symmetric and positive-definite; and $K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2}\mathbf{x})$.

In common with the majority of researchers in this field, we use the Mean Integrated Squared Error (MISE) as our optimality criterion:

$$\mathrm{MISE}(\mathbf{H}) \equiv \mathrm{MISE}\big(\hat{f}_{n\mathbf{H}}\big) = \mathbb{E} \int \Big(\hat{f}_{n\mathbf{H}} - f(\mathbf{x})\Big)^2 \, d\mathbf{x}.$$

The ideal MISE-optimal bandwidth selector is

$$\mathbf{H}_{\mathrm{MISE}} = \arg\min_{\mathbf{H}\in\mathcal{F}} \mathrm{MISE}(\mathbf{H})$$

where $\mathcal{F}$ is the set of all symmetric and positive-definite $d \times d$ matrices.

## 2 Unconstrained plug-in bandwidth selection

Plug-in methods for choosing the bandwidth matrix rely on an asymptotic form of the MISE, known as the AMISE. Wand (1992) shows that $\mathrm{MISE}(\mathbf{H}) = \mathrm{AMISE}(\mathbf{H}) + o(n^{-1}|\mathbf{H}|^{-1/2} + \|\mathrm{vec}\ \mathbf{H}\|^2)$, with

$$\mathrm{AMISE}(\mathbf{H}) = n^{-1}|\mathbf{H}|^{-1/2}R(K) + \frac{\mu_2(K)^2}{4}(\mathrm{vec}^T\ \mathbf{H})\mathbf{R}(\mathsf{D}^{\otimes 2}f)(\mathrm{vec}\ \mathbf{H}). \tag{2}$$

Here, vec $\mathbf{A}$ stands for the vector operator, respectively, applied to a symmetric matrix vec $\mathbf{A}$ (see Wand and Jones, 1995) and if $\mathsf{H}f = \partial^2 f/(\partial\mathbf{x}\partial\mathbf{x}^T)$ denotes the Hessian matrix of $f$, we are writing $\mathsf{D}^{\otimes 2}f = \mathrm{vec}\ \mathsf{H}f$. Besides, for a function $g\colon \mathbb{R}^d \to \mathbb{R}^p$ we introduce the notation $\mathbf{R}(g) = \int g(\mathbf{x})g(\mathbf{x})^T d\mathbf{x}$ and we will omit the bold font if $R(g) \in \mathbb{R}$ (i.e., when $p = 1$).

To select the bandwidth matrix we must estimate the AMISE function first. In the AMISE function the only unknown term is $\mathbf{R}(\mathsf{D}^{\otimes 2}f)$, so a plug-in procedure consists of replacing this quantity with an estimator $\mathbf{R}(\mathsf{D}^{\otimes 2}\hat{f}_{n\mathbf{G}})$, where $\mathbf{G}$ is another bandwidth matrix (called pilot bandwidth matrix) and then of selecting the plug-in matrix $\hat{\mathbf{H}}_{\mathrm{PI}}$ as the bandwidth minimizing

$$\mathrm{PI}(\mathbf{H}) = n^{-1}|\mathbf{H}|^{-1/2}R(K) + \frac{\mu_2(K)^2}{4}(\mathrm{vec}^T\ \mathbf{H})\mathbf{R}(\mathsf{D}^{\otimes 2}\hat{f}_{n\mathbf{G}})(\mathrm{vec}\ \mathbf{H})$$

But this raises the new problem of how to choose the pilot bandwidth $\mathbf{G}$.

All the previous approaches, as Wand and Jones (1994) or Duong and Hazelton (2003), put some constraints on the pilot bandwidth $\mathbf{G}$ to simplify both the computation of the estimator and the study of its asymptotic properties. Moreover, to choose the pilot bandwidth they need to pre-transform the data, and this pre-transformation may cause troubles to the final estimator. Here we present the first plug-in bandwidth selector which uses unconstrained bandwidth matrices at all stages of the plug-in methodology, therefore not needing any pre-transformation of the data.

First, the estimator $\mathbf{R}(\mathsf{D}^{\otimes 2}\hat{f}_{n\mathbf{G}})$ can be explicitly computed by noting that

$$\mathrm{vec}\ \mathbf{R}(\mathsf{D}^{\otimes 2}\hat{f}_{n\mathbf{G}}) = n^{-2}\sum_{i,j=1}^{n} \mathsf{D}^{\otimes 4}L_{\mathbf{G}}(\mathbf{X}_i - \mathbf{X}_j),$$

where $\mathsf{D}^{\otimes 4}L_{\mathbf{G}}$ denotes the vector containing all the partial derivatives of $L_{\mathbf{G}}$ of order 4.

The mean square error of this estimator is defined as

$$\mathrm{MSE}(\mathbf{G}) = \mathbb{E}\|\mathrm{vec}\ \mathbf{R}(\mathsf{D}^{\otimes 2}\hat{f}_{n\mathbf{G}}) - \mathrm{vec}\ \mathbf{R}(\mathsf{D}^{\otimes 2}f)\|^2,$$

with $\|\cdot\|$ standing for the Euclidean norm. Let us denote $\mathbf{G}_{\mathrm{AMSE}}$ the bandwidth matrix that asymptotically minimizes $\mathrm{MSE}(\mathbf{G})$. The following result describes the order of this optimal pilot bandwidth and that of the minimal MSE.

**Theorem 1** *Under some smoothness assumptions, every entry of* $\mathbf{G}_{\mathrm{AMSE}}$ *is of order* $n^{-2/(d+6)}$ *and the MSE obtained when this bandwidth matrix is used is of order* $n^{-4/(d+6)}$.

When the pilot bandwidth matrix is of this precise of order, we also obtain the convergence rate of the plug-in bandwidth selector.

**Theorem 2** *Under some smoothness assumptions, the relative rate of convergence of* $\hat{\mathbf{H}}_{\mathrm{PI}}$ *to* $\mathbf{H}_{\mathrm{MISE}}$ *is* $n^{-2/(d+6)}$.

# 3 Simulations

In this section, we undertake a numerical simulation study to compare the finite sample performance of the following selectors:

- Wand and Jones's plug-in selector (1994) with individual pilot selectors parametrized by $\mathbf{G} = g^2\mathbf{I}_d$; labelled WJ
- Duong and Hazelton's plug-in selector (2003) with a single selector parametrized by $\mathbf{G} = g^2\mathbf{I}_d$; labelled DH
- our proposed plug-in selector with unconstrained pilot selectors; labelled CD.

We consider samples of size $n = 1000$ for 500 simulation runs. For each simulation, we compute the Integrated Squared Error (ISE) between the resulting kernel density estimate and the target density.

The bivariate target densities which we consider are shown in Figure 1 and the ISE boxplots are shown in Figure 2. A more exhaustive study is given in Chacón and Duong (2008).

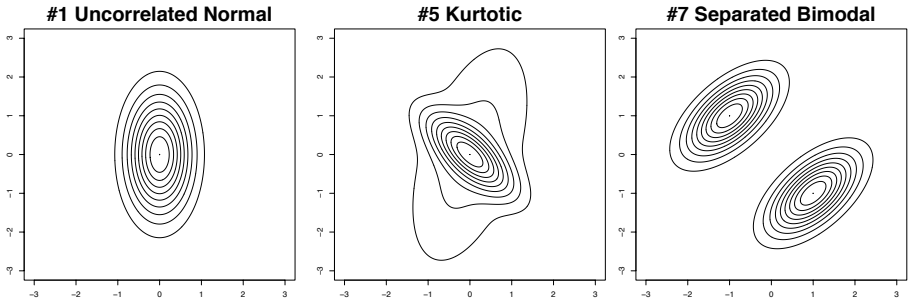


**Figure 1:** Contour plots for the 3 target densities.

Two clear conclusions can be drawn:

1. For densities for which the use of pre-transformations do not represent a serious handicap, namely #1 in this study, our selector performs as well as the other two. In fact, all the three selectors have an entirely similar behaviour. This means that, although our proposal is more general than the existing ones (because of the use of an unconstrained pilot), it does not lose power against the other methods even if we have a situation where a pilot bandwidth matrix with a single smoothing parameter is appropriate.

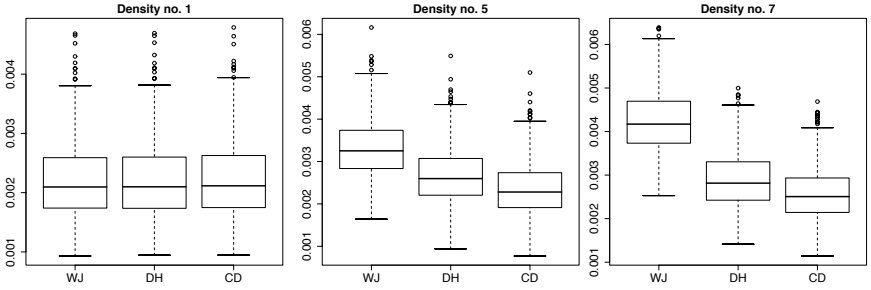


**Figure 2:** Box-plots for the ISEs of the plug-in methods WJ, DH and CD (from left to right) and $n = 1000$.

2. However, if the density is such that the single-parameter parametrization of the pilot bandwidth matrix is not suitable for the transformed data, then the plug-in selector with unconstrained pilot bandwidth clearly outperforms the other two methods. This occurs for densities #5 and #7 here.

# References

Chacón, J.E. and Duong, T. (2008) Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices. *Submitted*.

Duong, T. and Hazelton, M.L. (2003) Plug-in bandwidth matrices for bivariate kernel density estimation. *J. Nonparametr. Stat.* **15**, 17–30.

Wand, M.P. (1992) Error analysis for general multivariate kernel estimators. *J. Nonparametr. Stat.* **2**, 2–15.

Wand, M.P. and Jones, M.C. (1994) Multivariate plug-in bandwidth selection. *Comput. Statist.* **9**, 97–117.

Wand, M.P. and Jones, M.C. (1995) *Kernel Smoothing*, Chapman & Hall.

# Poster 7

# On nonparametric predictive inference with incomplete data

**Frank P.A. Coolen**
*Dept of Mathematical Sciences, Durham University, Durham DH1 3LE, UK*

**Tahani A. Maturi**
*Dept of Mathematical Sciences, Durham University, Durham DH1 3LE, UK*

**Abstract.** Coolen (1998) introduced nonparametric predictive lower and upper probabilities for $m$ future Bernoulli random quantities, based on the number of successes in $n$ trials. We explore the generalization of this approach if data are only available in the form of a set of values for the number of successes.

## 1 Introduction

Statistical inference with incomplete data is important in many applications. The manner in which inferential methods based on different foundations deal with incomplete data differs substantially. Methods in which uncertainty is quantified by lower and upper probabilities do not require any further assumptions on the incomplete data. Coolen (1998) presented an inferential approach for Bernoulli quantities, which uses lower and upper probabilities to quantify uncertainty. This approach, which fits in the more general framework of 'nonparametric predictive inference' (NPI) (Coolen (2006)), only considers events of the form '$A$ is followed by $B$', without any assumptions on the single events $A$ or $B$. NPI for Bernoulli quantities has been presented for events of the form '$Y_1^n = s$ is followed by $Y_{n+1}^{n+m} \in R$', with $Y_i^j$ the number of successes in trials $i$ to $j$, and $R \subset \{0, 1, \ldots, m\}$. This raises the question how NPI for Bernoulli quantities deals with incomplete information, in particular of the form $Y_1^n \in S \subset \{0, 1, \ldots, n\}$. This is currently being investigated, initial insights are reported here. Explicit formulae, detailed study of properties and a discussion of principles of such inferences, will be presented elsewhere.

## 2 NPI for Bernoulli random quantities

NPI for Bernoulli quantities (Coolen (1998)) defines direct predictive lower and upper probabilities for future observations, based on available data. This fits in the wider framework of NPI, with strong internal consistency and frequentist properties (Augustin and Coolen (2004), Coolen (2006)). Suppose that we have a sequence of $n+m$ exchangeable Bernoulli trials, each with success and failure as possible outcomes, and data consisting of $n$ trials with $s$ successes, with sufficient data representation $Y_1^n = s$ and interest in $Y_{n+1}^{n+m}$. Let $R = \{r_1, \ldots, r_t\}$, with $1 \leq t \leq m+1$ and $0 \leq r_1 < r_2 < \ldots < r_t \leq m$, and let $\binom{s+r_0}{s} = 0$. The NPI upper probability (Coolen (1998)) for the event $Y_{n+1}^{n+m} \in R$, given data $Y_1^n = s$, for $s \in \{0, \ldots, n\}$, is

$$\overline{P}(Y_{n+1}^{n+m} \in R | Y_1^n = s) = \binom{n+m}{n}^{-1} \sum_{j=1}^{t} \left[ \binom{s+r_j}{s} - \binom{s+r_{j-1}}{s} \right] \binom{n-s+m-r_j}{n-s}$$

The lower probability can be derived via the conjugacy property,

$$\underline{P}(Y_{n+1}^{n+m} \in R | Y_1^n = s) = 1 - \overline{P}(Y_{n+1}^{n+m} \in \{0, 1, \ldots, m\} \backslash R | Y_1^n = s)$$

In this approach, past observations are related to future random quantities via an assumed underlying latent variable representation on the real line, with a threshold such that all points to one side represent 'successes' and all to the other side represent 'failures'. No knowledge about this threshold is assumed (Thomas Bayes used a similar representation, but added a distributional assumption on the threshold - the prior distribution). In NPI, with the latent variable representation, past observations are related to future observations via Hill's assumption $A_{(n)}$ (Hill (1968)). Suppose that the ordered values of the latent variables corresponding to the $n$ observations are $u_{(1)} < u_{(2)} < \ldots < u_{(n)}$. These $n$ values define a partition of the real line, consisting of $n+1$ intervals. Hill's $A_{(n)}$ states that a future random quantity $U_{n+1}$ has equal probability $1/(n+1)$ to be in each of these intervals. In NPI this $U_{n+1}$ is the latent variable corresponding to the first future observation, which will again be a success or failure, depending on which side of the threshold $U_{n+1}$ is.

When interested in $m$ future observations (Coolen (1998)), the same assumption needs to be made for each future observation consecutively, so one needs to assume $A_{(n)}, \ldots, A_{(n+m-1)}$. Under these assumptions, all $\binom{n+m}{n}$ different orderings of the underlying latent variables on the real line, which represent the first $n$ observations and the $m$ future observations, have equal probability, also after information about the number of successes in the first $n$ observations has become available. Denoting these $\binom{n+m}{n}$ different orderings by $O_j$ for $j = 1, \ldots, \binom{n+m}{n}$, the above lower and upper probabilities are derived by counting orderings: for the lower probability, only those orderings are included for which $Y_1^n = s$ *must* be followed by $Y_{n+1}^{n+m} \in R$, while for the upper probability all orderings are included for which $Y_1^n = s$ *can* be followed by $Y_{n+1}^{n+m} \in R$. An example of an application of NPI for Bernoulli quantities is comparison of proportions (Coolen and Coolen-Schrijner (2007)).

# 3 Incomplete data

We now explore the generalization of NPI for Bernoulli random quantities with incomplete data, assuming that the data information is $Y_1^n \in S \subset \{0, 1, \ldots, n\}$. It is important to derive general expressions for the lower and upper probabilities for the events that $Y_1^n \in S$ is followed by $Y_{n+1}^{n+m} \in R$, for any sets $S$ and $R$. We explore this generalization by discussing the derivation of such lower and upper probabilities and some of their properties, and by a basic example.

We consider again the $\binom{n+m}{n}$ different orderings $O_j$ of the $n+m$ latent variables, which are all equally likely under the assumptions discussed above. The reasoning that leads to the lower and upper probabilities for the event $(Y_{n+1}^{n+m} \in R | Y_1^n \in S)$ is the same as before: The lower probability for this event is derived by counting all orderings $O_j$ for which $Y_1^n \in S$ *must* be followed by $Y_{n+1}^{n+m} \in R$, while the upper probability is derived by counting all orderings $O_j$ for which $Y_1^n \in S$ *can* be followed by $Y_{n+1}^{n+m} \in R$. It is important to emphasize that, for the lower probability, $O_j$ is only included in the count if, for this particular $O_j$, for *each* $s \in S$, $Y_1^n = s$ *must* be followed by $Y_{n+1}^{n+m} \in R$, whereas for the upper probability an $O_j$ is already included if there is *at least one* $s \in S$ for which $Y_1^n = s$ *can* be followed by $Y_{n+1}^{n+m} \in R$. Hence, the actual events that correspond to the lower and upper probabilities for $(Y_{n+1}^{n+m} \in R | Y_1^n \in S)$ differ substantially, in a way that could be described as 'most conservative', and which also ensures that the conjugacy property remains valid,

$$\underline{P}(Y_{n+1}^{n+m} \in R | Y_1^n \in S) = 1 - \overline{P}(Y_{n+1}^{n+m} \in \{0, 1, \ldots, m\} \backslash R | Y_1^n \in S)$$

Basic logic and set theory imply some important general properties for these lower and upper probabilities. Let $S_1 \subset S_2$, then for all $R$,

$$\begin{aligned} \underline{P}(Y_{n+1}^{n+m} \in R | Y_1^n \in S_1) &\geq \underline{P}(Y_{n+1}^{n+m} \in R | Y_1^n \in S_2) \\ \overline{P}(Y_{n+1}^{n+m} \in R | Y_1^n \in S_1) &\leq \overline{P}(Y_{n+1}^{n+m} \in R | Y_1^n \in S_2) \end{aligned}$$

For all $R$ that are strict subsets of $\{0, 1, \ldots, m\}$, we have

$$\underline{P}(Y_{n+1}^{n+m} \in R | Y_1^n \in \{0, 1, \ldots, n\}) = 0 \text{ and } \overline{P}(Y_{n+1}^{n+m} \in R | Y_1^n \in \{0, 1, \ldots, n\}) = 1,$$

reflecting that $S = \{0, 1, \ldots, n\}$ provides no information. For any given set $S$, the lower and upper probabilities are increasing in $R$.

As mentioned before, the lower probability $\underline{P}(Y_{n+1}^{n+m} \in R | Y_1^n \in S)$ and upper probability $\overline{P}(Y_{n+1}^{n+m} \in R | Y_1^n \in S)$ are conservative, as is clear from the way they are derived. They can be used without further assumptions about reasons for reporting $S$ instead of a specific unique value for $Y_1^n$, which is in line with alternative approaches for dealing with set-valued data in imprecise probability theory, but which cannot be achieved with precise probabilities. An interesting case where such set-valued data naturally occur is statistical quality control, if batches are accepted if at least a certain number of tested products function satisfactorily; NPI methods for this are in development.

# 4 Example

We consider $n = 4$ available observations and $m = 2$ future observations, so the underlying assumed data representation has $\binom{6}{2} = 15$ different orderings of past and future observations, each having probability $1/15$ under the inferential assumptions in this paper. Table 1 gives a variety of lower and upper probabilities, represented as pairs $(15\underline{P}(Y_5^6 \in R|Y_1^4 \in S), 15\overline{P}(Y_5^6 \in R|Y_1^4 \in S))$.

| $\times 1/15$ | $R = \{0\}$ | $\{1\}$ | $\{2\}$ | $\{0,1\}$ | $\{0,2\}$ | $\{1,2\}$ |
|---|---|---|---|---|---|---|
| $S = \{0\}$ | $(10,15)$ | $(0,5)$ | $(0,1)$ | $(1,15)$ | $(10,15)$ | $(0,5)$ |
| $\{1\}$ | $(6,10)$ | $(3,8)$ | $(1,3)$ | $(12,14)$ | $(7,12)$ | $(5,9)$ |
| $\{2\}$ | $(3,6)$ | $(4,9)$ | $(3,6)$ | $(9,12)$ | $(6,11)$ | $(9,12)$ |
| $\{3\}$ | $(1,3)$ | $(3,8)$ | $(6,10)$ | $(5,9)$ | $(7,12)$ | $(12,14)$ |
| $\{4\}$ | $(0,1)$ | $(0,5)$ | $(10,15)$ | $(0,5)$ | $(10,15)$ | $(14,15)$ |
| $\{1,2\}$ | $(3,10)$ | $(2,11)$ | $(1,6)$ | $(9,14)$ | $(4,13)$ | $(5,12)$ |
| $\{1,3\}$ | $(1,10)$ | $(1,12)$ | $(1,10)$ | $(5,14)$ | $(3,14)$ | $(5,14)$ |
| $\{2,3\}$ | $(1,6)$ | $(2,11)$ | $(3,10)$ | $(5,12)$ | $(4,13)$ | $(9,14)$ |
| $\{1,2,3\}$ | $(1,10)$ | $(1,13)$ | $(1,10)$ | $(5,14)$ | $(2,14)$ | $(5,14)$ |
| $\{0,1,2,3\}$ | $(1,15)$ | $(0,14)$ | $(0,10)$ | $(5,15)$ | $(1,15)$ | $(0,14)$ |
| $\{1,2,3,4\}$ | $(0,10)$ | $(0,14)$ | $(1,15)$ | $(0,14)$ | $(1,15)$ | $(5,15)$ |
| $\{0,1,3,4\}$ | $(0,15)$ | $(0,14)$ | $(0,15)$ | $(0,15)$ | $(1,15)$ | $(0,15)$ |

**Table 1:** NPI lower and upper probabilities ($\times 15$), $n = 4$ and $m = 2$.

The case $S = \{0,1,3,4\}$ only leads to non-trivial lower and upper probabilities for two sets $R$ (due to conjugacy), which is due only to the specific ordering of the 4 past and 2 future observations in which the latter two are in between the second and third ordered past observation, in the underlying assumed data representation. for that specific ordering none of the values in this $S$ can be followed by precisely one future success, for all other orderings this is possible for at least one of the values in $S$. These results illustrate clearly the decreasing (increasing) nature of the lower (upper) probabilities if $S$ becomes larger. Imprecision is pretty large, which is due to the specific manner in which the lower and upper probabilities are derived, and their conservative nature, yet it should not be too surprising. For example, if one gets information that, out of 4 trials, the number of successes was either 1 or 3, clearly this information does not reveal much, in particular if one has no idea why this specific information was given. Most remarkable, perhaps, are some of these inferences for $R = \{1\}$. For example, when one compares the values corresponding to $S = \{1\}$, $S = \{3\}$ and $S = \{1,3\}$, one might perhaps be surprised that, for the latter case, the lower and upper probabilities are not also equal to $3/15$ and $8/15$, respectively. This is a feature of the NPI approach where it differs fundamentally from other approaches. Such behaviour of these NPI lower and upper probabilities as function of $S$ are currently being studied in more detail.

## References

Augustin, T. and Coolen, F.P.A. (2004). Nonparametric predictive inference and interval probability. *Journal of Statistical Planning and Inference*, *124*, 251–272.

Coolen, F.P.A. (1998). Low structure imprecise predictive inference for Bayes' problem. *Statistics and Probability Letters*, *36*, 349–357.

Coolen, F.P.A. (2006). On nonparametric predictive inference and objective Bayesianism. *Journal of Logic, Language and Information*, *15*, 21–47.

Coolen, F.P.A. and Coolen-Schrijner, P. (2007). Nonparametric predictive comparison of proportions. *Journal of Statistical Planning and Inference*, *137*, 23–33.

Hill, B.M. (1968): Posterior distribution of percentiles: Bayes theorem for sampling from a finite population. *Journal of the American Statistical Association*, *63*, 677–691.

# Poster 8

# Robust nonparametric estimation for functional data: $\mathbb{L}^p$ errors and applications

**Christophe Crambes**
*Institut de Mathématiques, Université Montpellier II, France*

**Laurent Delsol**
*Institut de Statistique, Université Catholique de Louvain, Belgium*

**Ali Laksaci**
*Laboratoire de Mathématiques, Université Djillali Liabès, Algery*

**Abstract.** Recently, robust estimators have been considered in regression models where the covariate is functional. Almost sure convergence and asymptotic normality results have been obtained. From this asymptotic normality result together with uniform integrability conditions, the asymptotic expressions of $L^p$ errors of such estimators can be obtained.

## 1 Introduction

In many practical situations, one is faced to functional-type phenomena. It is now possible to take into account their functional nature thanks to technological improvements that allow to collect data discretized on thinner grids. Functional Statistics is an important topic of modern statistics (see [16], [17], [4], and [11] for a review). In this talk, we consider the issue of predicting the value of a real variable of interest from the observation of a functional explanatory variable (i.e. a random variable taking values in a semimetric space $(\mathcal{E}, d)$). one often considers a regression model of the form

$$Y = r(X) + \epsilon$$

and want to estimate the value of the regression operator $r$ at a point $x$ of $\mathcal{E}$ that corresponds to the conditional mean of $Y$ given $X = x$. Kernel methods have been used to make its estimation in nonparametric functional regression models (see [11] for a review). However, in presence of outliers it may be more relevant to consider robust estimators. In the classical case where the explanatory variable is multivariate, the robust methods has been considered by many authors (see for instance, [15], [7], [3], and [14] for recent references) following the pioneer work [13]. In the more general case where the explanatory variable is functional, there are few results (see [5], [6], [12], [8], [10]). Recently, [2] have obtained the almost complete convergence of robust estimators based on kernel methods. In the same context, [1] have considered the asymptotic normality of these estimators.

The aim of this talk is to present these robust estimators and complete the existing literature. We quickly recall the asymptotic normality given by [1]. Then we explain how a uniform integrability result can be used to derive from this result the explicit expression of the $\mathbb{L}^p$ errors of these robust estimators (extending the work of [9]). We conclude our talk with the application of robust methods in the study of time series or curve discrimination.

# 2 The Model

One focuses in this talk on the way a real-valued variable of interest $Y$ depends on a variable $X$ that takes values in a semimetric space $(\mathcal{E}, d)$. In this work we are interested in estimating a function $\theta : x \mapsto \theta_x$, where for each $x \in \mathcal{E}$, $\theta_x$ is defined as the solution (with respect to $t$), assumed to be unique, of the following equation:

$$\Psi(x,t) := \mathbb{E}\left[\psi_x\left(Y,t\right)|X = x\right] = 0, \tag{1}$$

where $\psi_x$ is a known function chosen by the statistician. The function $\Psi\left(x,t\right)$ is unknown. We propose to consider the following kernel estimator of $\Psi\left(x,t\right)$ constructed from a dataset of $n$ pairs $(X_i, Y_i)_{1\leq i\leq n}$ identically distributed as $(X, Y)$:

$$\widehat{\Psi}(x,t) = \frac{\sum_{i=1}^{n} K\left(h^{-1}d(x,X_i)\right)\psi_x\left(Y_i,t\right)}{\sum_{i=1}^{n} K\left(h^{-1}d(x,X_i)\right)}, \qquad \forall t \in \mathbb{R}, \tag{2}$$

where $K$ is a kernel and $h = h_n$ is a sequence smoothing parameters. Then, a natural estimator of $\theta_x$ is $\widehat{\theta}_n = \widehat{\theta}_n(x)$ that is the solution of the empiral equation:

$$\widehat{\Psi}(x,\widehat{\theta}_n) = 0. \tag{3}$$

# 3 Asymptotic results

## 3.1 Convergence in probability and asymptotic normality

The results we recall in the present section have been obtained by [1] for independent datasets:

$$\widehat{\theta}_n - \theta_x \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

and

$$\left(\frac{nF(h_n)}{V_n(x)}\right)^{1/2}\left(\widehat{\theta}_n-\theta_x-B_n(x)\right)\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}(0,1),$$

with explicit expressions for $V_n(x)$ and $B_n(x)$.

### 3.2 A uniform integrability result

We now state a uniform integrability result that will be used together with previous asymptotic normality result to get the explicit expression of dominant terms in $\mathbb{L}^p$ errors. This result has been obtained for arithmetically $\alpha$-mixing pairs $(X_i, Y_i)$. We introduce the quantities

$$F(h)=\mathbb{P}\left(d(X,x)\leq h\right),$$

known in the litterature as the *small balls probabilities*. Under some hypotheses, for $0\leq q<k$, ($k$ has an explicit expression not given here) we show that the quantity

$$\left|\sqrt{nF(h_n)}\left(\Psi_n(x,t)-\mathbb{E}\left[\Psi_n(x,t)\right]\right)\right|^q,$$

is uniformly integrable, where $\Psi_n(x,t)=\dfrac{1}{nF(h_n)}\displaystyle\sum_{i=1}^{n}K\left(h^{-1}d(x,X_i)\right)\psi_x\left(Y_i,t\right)$.

### 3.3 Moments convergence

We now can obtain the convergence of the moments of $\hat{theta}_n-\theta_x$ from the asymptotic normality of our estimators and the uniform integrability result given in the previous paragraph. We have, for all $q<q'$ (we have an explicit definition of $q'$, not given here)

$$\mathbb{E}\left[\left|\widehat{\theta}_n-\theta_x\right|^q\right]=\mathbb{E}\left[\left|B_n(x)+\sqrt{\frac{V_n(x)}{nF(h_n)}}W\right|^q\right]+o\left(\frac{1}{\sqrt{nF(h_n)}^q}\right).$$

From the explicit expressions of $B_n(x)$ and $V_n(x)$ given in [1] one can derive a more explicit expression of these $\mathbb{L}^p$ errors following the same approach as in [9]. The previous result give a $\mathbb{L}^p$ convergence result for general robust estimators and opens interesting prospects with respect to the choice of the smoothing parameter.

## 4 Applications

We will quickly illustrate the usefulness of robust estimators in an application example concerning the prediction of energetic cunsumption during one year from the values collected the previous years (see Figure 1).

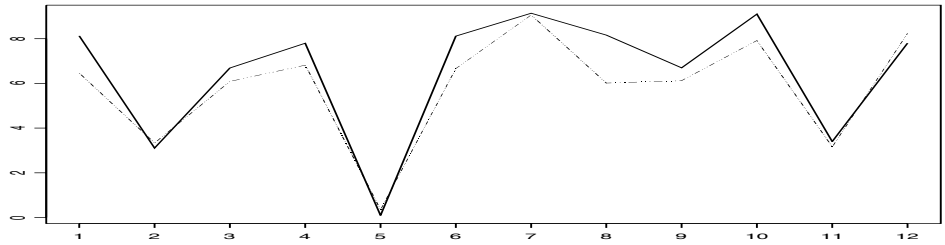


**Figure 1:** Prediction of the total petroleum consumption for electricity generation by a robust approach.

# References

[1] M. Attouch, A. Laksaci, E. Ould-Saïd. (2007). Asymptotic distribution of robust estimator for functional nonparametric models, Prépublication, LMPA No 314, . *Submitted*.

[2] N. Azzeddine, A. Laksaci, E. Ould-Saïd. (2007). On the robust nonparametric regression estimation for functional regressor, *Statist. and Probab. Lett.*. *Accepted under minor revision*.

[3] G. Boente, R. Fraiman. (1990). Asymptotic distribution of robust estimators for nonparametric models from mixing processes. *Ann. Statist.*, **18**, 891-906.

[4] D. Bosq, (2000). *Linear processes in function spaces.* Lecture Notes in Statistics, **149**, Springer-Verlag, New York.

[5] B. Cadre. (2001). Convergent estimators for the $L_1$-median of a Banach valued random variable. *Statistics*, **35** , 509-521.

[6] H. Cardot, C. Crambes, P. Sarda. (2005). Quantiles regression when the covariates are functions. *J. of Nonparam. Stat.*, **17**, 841-856.

[7] G. Collomb, W. Härdle. (1986). Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations. *Stoch. Proc. Appl.*, **23**, 77-89.

[8] S. Dabo-Niang, F. Ferraty and P. Vieu. (2006). Mode estimation for functional random variable and its application for curves classification. *Far East J. Theor. Stat.* **18** (1) 93-119.

[9] L. Delsol. (2007) Régression non-paramétrique fonctionnelle: Expressions asymptotiques des moments. *Annales de l'I.S.U.P.*, **LI** (3) 43-67.

[10] M. Ezzahrioui. (2007). Prévision dans les modèles conditionnels en dimension infinie. *PhD Thesis (in French)* .

[11] F. Ferraty and P. Vieu. (2006). *Nonparametric functionnal data analysis.* Springer-Verlag, New York.

[12] F. Ferraty, A. Laksaci and P. Vieu. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Stat. Inference Stoch. Process*, **9** (1), 47-76.

[13] P.J. Huber. (1964). Robust estimation of a location parameter. *Ann. of the Math. Statist.*, **35**, 73-101.

[14] N. Laïb and E. Ould-Saïd. (2000). A robust nonparametric estimation of the autoregression function under an ergodic hypothesis. *Canad. J. Statist.*, **28**, 817-828.

[15] R. Robinson. (1984). *Robust Nonparametric Autoregression.* Lecture Notes in Statistics,Springer-Verlag, New York, **26**, 247-255.

[16] J. Ramsay and B.W. Silverman. (2002) *Applied functional data analysis.* Springer-Verlag, New York.

[17] J. Ramsay and B.W. Silverman. (2005). *Functional data analysis* (2nd Ed.). Springer-Verlag, New York.

# Poster 9

# Additive models for testing separability

**Rosa M. Crujeiras**[1]
*Institut of Statistics. Université catholique de Louvain*

**Rubén Fernández-Casal**[2]
*Department of Mathematics. University of Coruña*

**Wenceslao González-Manteiga**
*Department of Statistics and OR. University of Santiago de Compostela*

**Abstract.** We propose a testing technique for assessing separability of spatio-temporal processes. Our approach is based on the representation of the log-periodogram as the response variable in a regression model. Within this context, separability can be interpreted as additivity in spatial and temporal frequency components. We also provide some simulation results comparing the performance of the test with a marginal integration and a Backfitting approach. The testing method is also applied to a real-data case.

## 1 Introduction

In order to describe the behaviour of spatio-temporal processes, a model for the dependence structure must be assessed. Since the direct extension of spatial covariance models to the spatio-temporal situation may not be adequate, a great effort has been made in order to obtain new dependence models that allow for space-time interactions (see, for instance, Cressie and Huang (1999)). However, if the spatial and temporal components in the variability of the process are proved to be independent, inference becomes simpler. This situation is known as separability.

[1]The authors acknowledge the support of Ministerio de Educación y Ciencia project MTM2005-00820 and Xunta de Galicia project PGIDIT06PXIB207009PR.
[2]R. Fernández-Casal acknowledge the support of Xunta de Galicia Project PGIDIT05TIC00701CT.

Let $Z(\mathbf{s},t)$ denote a zero-mean second-order stationary spatio-temporal process, observed at spatial locations on a regular grid $\mathbf{s} \in D = \{1,\ldots,n_1\} \times \{1,\ldots,n_2\}$, $N = n_1 n_2$ and at time moments $t \in \{1,\ldots,T\}$. The spatio-temporal covariance function associated with $Z$ is defined as $C(\mathbf{u},v) = \mathrm{cov}(Z(\mathbf{s}+\mathbf{u},t+v), Z(\mathbf{s},t)), \quad (\mathbf{u},v),(\mathbf{s},t) \in R^3$. Assuming that $\int_{R^2}\int_R C(u,v)dvdu < \infty$, by Khinchin's theorem, the covariance function can be written as:

$$C(\mathbf{u},v) = \int_{\Pi^2}\int_{\Pi} e^{i(\mathbf{u},v)'(\lambda,\omega)} f(\lambda,\omega)d\lambda d\omega, \quad \Pi = [-\pi,\pi], \Pi^2 = [-\pi,\pi]\times[-\pi,\pi],$$

where $'$ denotes the transpose operator, $f(\lambda,\omega)$ is the spectral density and $\lambda$ and $\omega$ denote the frequencies associated with the spatial and the temporal components. Under the separability assumption, the covariance can be decomposed as the tensor product of a spatial and a temporal covariances, namely $C_S$ and $C_T$, such that $C(\mathbf{u},v) = C_S(\mathbf{u})C_T(v)$. Therefore, the spectral density can be also written in this way, considering the corresponding spatial and temporal spectral densities $f_S$ and $f_T$, that is: $f(\lambda,\omega) = f_S(\lambda)f_T(\omega)$. The most well-known nonparametric estimator of the spectral density is the periodogram, which is given by:

$$I(\lambda_{\mathbf{k}},\omega_k) = \frac{1}{(2\pi)^3 NT}\left|\sum_{(\mathbf{s},t)} Z(\mathbf{s},t)e^{-i(\mathbf{s},t)'(\lambda_{\mathbf{k}},\omega_k)}\right|^2, \quad \mathbf{k} = (k_1,k_2)$$

and it is usually evaluated at the Fourier frequencies $(\lambda_{\mathbf{k}},\omega_k) \in \Pi^2 \times \Pi$, $(\lambda_{\mathbf{k}},\omega_k) = \left(\frac{2\pi k_1}{n_1}, \frac{2\pi k_2}{n_2}, \frac{2\pi k}{T}\right)$, $k_i = 0,\pm1,\ldots,\pm\lfloor\frac{n_i-1}{2}\rfloor, i = 1,2 \quad k = 0,\pm1,\ldots\pm\lfloor\frac{T-1}{2}\rfloor$. For a wide class of processes, in particular for Gaussian stationary processes, the periodogram can be obtained as the response variable in a multiplicative regression model:

$$I(\lambda_{\mathbf{k}},\omega_k) = f(\lambda_{\mathbf{k}},\omega_k)V_{\mathbf{k},k} + R(\lambda_{\mathbf{k}},\omega_k),$$

where $V_{\mathbf{k},k}$ are i.i.d. standard exponential random variables and the remainder term $R(\lambda_{\mathbf{k}},\omega_k)$ can be uniformly bounded. Taking logarithms, the log-periodogram at $(\lambda_{\mathbf{k}},\omega_k)$, denoted by $Y_{\mathbf{k},k}$ is given by:

$$Y_{\mathbf{k},k} = m(\lambda_{\mathbf{k}},\omega_k) + z_{\mathbf{k},k} + r_{\mathbf{k},k}, \tag{1}$$

where $z_{\mathbf{k},k}$ are i.i.d. Gumbel(0,1) distributed, $m = \log f$ and $r_{\mathbf{k},k}$ is asymptotically negligible.

## 2 The test

Considering the logarithm of the spectral density $m$, under the assumption of separability, this function is given by the sum of the spatial and temporal log-spectral densities: $m(\lambda,\omega) = m_S(\lambda) + m_T(\omega)$. Then, the problem of assessing a separable covariance can be formulated in terms of the log-spectral density as:

$$H_0 : m = m_S + m_T, \quad \text{vs.} \quad H_a : m \neq m_S + m_T.$$

From expression (1), the testing problem can be viewed as a test for additivity in a regression model, but regarding the particular features of the remainders $z_{\mathbf{k},k}$ and $r_{\mathbf{k},k}$.

The two most popular methods for the estimation of an additive regression function are the Backfitting algorithm and the marginal integration estimator. In Dette *et al.* (2005), the authors revise four different test statistics, based on the marginal integration approach for estimation, in order to assess additivity. We consider an $L_2$-distance test statistic:

$$Q = \frac{1}{NT} \sum_{\mathbf{k},k} \left( \widehat{m}(\lambda_{\mathbf{k}}, \omega_k) - \widehat{m_0}(\lambda_{\mathbf{k}}, \omega_k) \right)^2, \tag{2}$$

where $\widehat{m}$ is given by $\widehat{m}(\lambda,\omega) = \frac{1}{NT} \sum_{\mathbf{k},k} K_{ST,G}((\lambda_{\mathbf{k}}, \omega_k) - (\lambda,\omega)) Y_{\mathbf{k},k}, \quad G \in M_{3\times 3}$, with kernel $K_{ST}$ and bandwidth matrix $G$ and $\widehat{m_0}(\lambda,\omega) = \widehat{m_S}(\lambda) + \widehat{m_T}(\omega) + \hat{c}$, where $\hat{c} = \overline{Y}$. Classical marginal integration estimation is applied for the spatial and temporal components in the frequency domain. Under some regularity conditions, the asymptotic normal behaviour of this test statistic can be obtained following Dette *et al.* (2005).

However, a large sample size is needed for an accurate performance of the asymptotic distribution, which may not be the case in a real situation. Therefore, we propose an algorithm in order to apply this test in practice, which allows also to compare the behaviour of Backfitting vs. marginal integration in this particular setting:

Step 1. Compute $\widehat{m^I}$ and compute $\widehat{m_0^I}$ using the marginal integration estimator or the Backfitting algorithm.

Step 2. Compute the observed test statistic $Q^{obs}$.

Step 3. Draw a random sample of the process $Z$ applying the Inverse Fourier Transform to $e^{\widehat{m_0^I}(\lambda_{\mathbf{k}},\omega_k)}$ and obtain $\hat{C}(\mathbf{u}, v)$, for $\mathbf{u} \in \{1-n_1, \ldots, n_1 - 1\}$, $v \in \{1-T, \ldots, T-1\}$. Obtain a realization of the process in a grid $\{1, \ldots, n_1\} \times \{1, \ldots, n_2\} \times \{1, \ldots, T\}$ based on $\hat{C}$.

Step 4. Obtain the test statistic for this generated sample, $Q^b$.

Step 5. Repeat steps 3 and 4 and get $Q^1, \ldots, Q^B$. Approximate the $p$-value of the test statistic as the percentage of bootstrap replicates that exceed the observed value $Q^{obs}$.

Consider the following spatio-temporal (spatially isotropic) covariance model:

$$C(\mathbf{u}, v) = \sigma^2 \exp\left( -\frac{3}{a} \left( \beta \sqrt{\|\mathbf{u}\|^2 + bv^2} + (1-\beta)(\|\mathbf{u}\| + b|v|) \right) \right),$$

where $\sigma^2$ denotes the sill, $a$ is the practical range and $b$ is a spatial-temporal scale parameter. $\beta = 0$ corresponds with a separable model, whereas $\beta = 1$ is a geometrically anisotropic non-separable model. Results of the algorithm for $21 \times 21$ observations of a Gaussian process can be seen in Table 1. In all cases, the behaviour of the test under the null hypothesis of separability is good. However, the test with Backfitting presents less power than the one with marginal integration in distinguishing from separability. We have run more simulations with larger sample sizes, obtaining better results in terms of power, but the same conclusions about BF against CMIE. Our methods has been also compared with the one proposed by Scaccia and Martin (2005) for lattice processes, with

| Method | $\beta = 0.0$ | $\beta = 0.25$ | $\beta = 0.50$ | $\beta = 0.75$ | $\beta = 1.00$ |
|---|---|---|---|---|---|
| | | | $\alpha = 0.01$ | | |
| CMIE | 0.016 | 0.025 | 0.058 | 0.130 | 0.316 |
| BF | 0.006 | 0.012 | 0.028 | 0.075 | 0.187 |
| SM | 0.013 | 0.016 | 0.065 | 0.075 | 0.171 |
| | | | $\alpha = 0.05$ | | |
| CMIE | 0.060 | 0.085 | 0.147 | 0.270 | 0.494 |
| BF | 0.046 | 0.061 | 0.103 | 0.192 | 0.379 |
| SM | 0.052 | 0.079 | 0.102 | 0.188 | 0.331 |
| | | | $\alpha = 0.10$ | | |
| CMIE | 0.105 | 0.146 | 0.222 | 0.374 | 0.608 |
| BF | 0.093 | 0.117 | 0.178 | 0.276 | 0.496 |
| SM | 0.112 | 0.133 | 0.173 | 0.253 | 0.458 |

**Table 1:** Percentage of rejections for $Q$ with marginal integration estimation (CMIE), Backfitting (BF) and Scaccia and Martin's test (SM) for different $\beta$. Significance levels: $\alpha = 0.01$, $\alpha = 0.05$ and $\alpha = 0.10$. Grid size: $21 \times 21$, $a = 5$, $\sigma^2 = 1$ and $b = 1$.

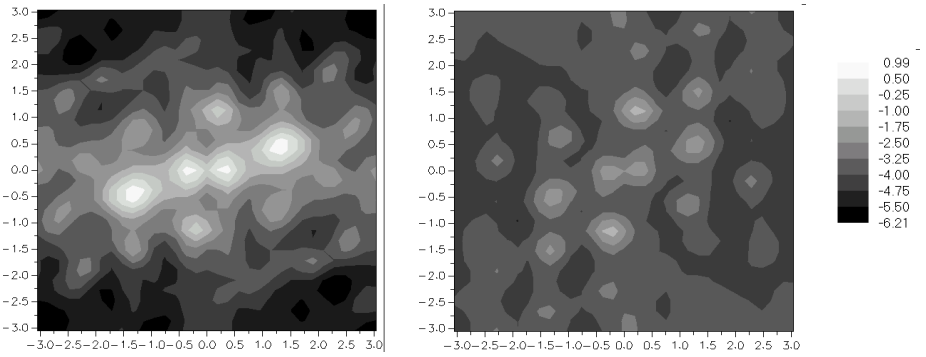


**Figure 1:** Estimated surfaces of the log-spectral density for temporal frequency 0. Left plot: $\widehat{m}(\lambda_{\mathbf{k}}, 0)$. Right plot: $\widehat{m_0}(\lambda_{\mathbf{k}}, 0)$ (under separability).

better results.

The method has been also applied to the wind-speed data studied in Cressie and Huang (1999). The observed value for the test statistic was 41681.74 and the $p$-value approximation (using the previous algorithm with marginal integration) was $p < 0.001$. The non-separability of this data set has been also assessed by other authors. This result is not surprising, given that the difference between the estimations under the null and the alternative hypothesis is quite relevant (see Figure 1).

## References

Cressie, N. and Huang, H. (1999). Classes of nonseparable spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, **94**, 1330-1340

Dette, H., von Lieres und Wilkau, C. and Sperlich, S. (2005) A comparison of different nonparametric methods for inference on additive models. *Nonparametric Statistics*, **17**, 57-81.

Scaccia, L. and Martin, R. (2005). Testing axial symmetry and separability of lattice processes. *Journal of Statistical Planning and Inference*, **131**, 19-39.

# Poster 10

# Functional quantile regression estimation: Application to functional times series prediction

**Sophie Dabo-Niang**
*Laboratory GREMARS-EQUIPPE. University Charles De Gaulle*

**Ali Laksaci**
*Department of Mathematics. University Djillali Liabès*

**Abstract.** We study a kernel estimator of conditional quantiles of a scalar response variable given a random variable taking values in a semi-metric space. Asymptotic behavior of the estimate are obtained. We apply this estimation to functional times series prediction problem, to construction of confidence prediction bands and determination of conditional confidence bands. A real data application where we are interested in forecasting hourly ozone concentration is considered.

## 1 Introduction

Let us introduce $n$ pairs of random variables $(X_i, Y_i)_{i=1,\ldots,n}$ that we suppose drawn from the pair $(X, Y)$, valued in $\mathcal{F} \times R$, where $\mathcal{F}$ is a semi-metric space. Let $d$ denotes the semi-metric. Assume that there exists a regular version of the conditional probability of $Y$ given $X$ that is absolutely continuous with respect to Lebesgue measure on $R$ and has a bounded density. For $x \in \mathcal{F}$, we will denote the conditional cumulative distribution function, (*cdf.*) of $Y$ given $X = x$ by $F^x$ and by $f^x$ the conditional density of $Y$ given $X = x$.
Let $\alpha \in ]0, 1[$, the $\alpha^{th}$ conditional quantile noted $t_\alpha(x)$ is defined by

$$F^x(t_\alpha(x)) = \alpha.$$

To insure existence and unicity of $t_\alpha(x)$, we assume that $F^x$ is strictly increasing. This last is estimated by

$$\hat{F}^x(y) = \begin{cases} \dfrac{\sum_{i=1}^n K(h_K^{-1}d(x,X_i))H(h_H^{-1}(y-Y_i))}{\sum_{i=1}^n K(h_K^{-1}d(x,X_i))} & \text{if } \sum_{i=1}^n K(h_K^{-1}d(x,X_i)) \neq 0 \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

where $K$ is a kernel, $H$ is a *cdf* and $h_K = h_{K,n}$ (*resp.* $h_H = h_{H,n}$) is a sequence of positive real numbers.
Because of the definition of $t_\alpha(x)$, the kernel estimate of this conditional quantile is related to the conditional distribution estimator (1). Then, a natural estimator of $t_\alpha(x)$ is defined by:

$$\hat{F}^x(\widehat{t_\alpha}(x)) = \alpha. \quad (2)$$

An estimate similar to $\hat{F}^x(y)$ has been introduced in the special case where $X$ is a real random variable by Roussas (1969). This last work has been extended in different ways by many authors in univariate or multivariate cases. See Gannoun *et al.* (2003), Yu *et al.* (2003) for recent advances and more exhaustive list of references.
The main goal of this paper is to study the nonparametric estimation of quantile regression when the explanatory variable is functional. The nonparametric study of the conditional quantile estimation is relatively restricted if the explanatory variable is functional, see Ferraty *et al.* (2006), Ferraty and Vieu (2006), Ezzahrioui and Ould-said (2006a, 2006b), Dabo-Niang & Laksaci (2006).
We are interested in nonparametric estimation of the conditional quantile when the data are dependent and of functional nature. Our nonparametric model is quite general compare with those of Ezzahrioui & *al.* (2006a, 2006b) and Ferraty & Vieu (2006) in the sense that we use less restrictive assumptions. We prove under general conditions, the $L^p$-norm convergence (with rates) and we establish the asymptotic normality of the kernel estimator $\widehat{t}_\alpha(x)$.

Our results are applied to build predictive intervals, confidence intervals. The most important application of our results is the prediction of a real characteristic of a functional variable. A particular case of functional data is when observations come from a continuous time series. Let $(Z_t)_{t\in[0,b[}$ be a random real valued continuous time process. From $Z_t$ we construct: $N$ functional random variables $(X_i)_{i=1,\ldots,N}$ defined by: $\forall t \in [0,b[$, $X_i(t) = Z_{N^{-1}((i-1)b+t)}$ and a real characteristic $Y_i = G(X_{i+1})$. Our goal is to predict $Y_N$ given the whole past $(X_i,Y_i)_{i=1,\ldots,N-1}$ and $X_N$. To this aim, we use an alternative approach to regression method: that is the conditional median estimate $\widehat{Y}_N = \widehat{t_{0.5}}(X_N)$ which gives also an approximation (derived from asymptotic normality result) of an $1-\eta$ confidence interval $I_\eta$ of $Y_N$. This confidence interval $I_\eta$ can also be approximated by using a predictive interval, based on estimated quantiles of order $\eta/2$ and $1-\eta/2$.

Finally, our model has been implemented and applied to some environmentally data. These data concern ozone concentration and are collected around the city of "Le Casset" in the south-est of France, defined by the geographical coordinates (45° 00'N, 06° 28'E). The data are hourly ozone measurements from January to December 2003. We are interested in the prediction of this ozone concentration one day ahead.

The obtained predicted values $(\widehat{Y_{363}^1},\ldots,\widehat{Y_{363}^{15}})$ corresponds to the hourly ozone concentration prediction of the last day of December 2003, between 6h AM to 8h PM. The

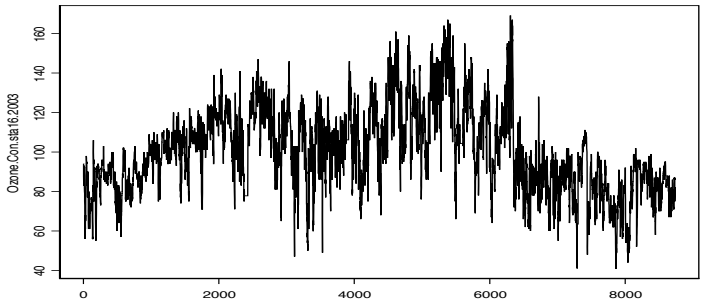


**Figure 1:** Hourly ozone concentration of year 2003.

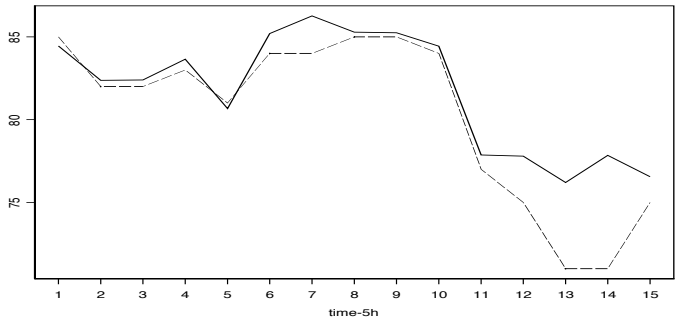


**Figure 2:** Prediction by the conditional quantiles.

results are given in Fig.2, where we draw two curves corresponding to the observed values (dashed curve) and the predicted values (solid curve). Clearly, Fig. 2 shows the global good behavior of our functional forecasting procedure.

# References

Dabo-Niang, S. and Laksaci, A., (2008). Nonparametric estimation of conditional quantiles when the regressor is valued in a semi-metric space, *Submitted*.

Ezzahrioui, M., Ould-Said, E. (2006a). Asymptotic normality of the kernel estimators of the conditional quantile in the normed space, *LMPA n° 258*.

Ezzahrioui, M., Ould-Said, E. (2006b). Asymptotic results of the kernel estimator of the conditional quantile in the normed space under $\alpha$-mixing hypothesis, *Preprint*.

Ferraty, F., Laksaci, A. and Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models, *Statist. Inf. for Stoch.*

*Proc*, **9**, No.2, 47-76.

Ferraty, F. and Vieu, Ph., (2006), *Nonparametric functional data analysis.* Springer-Verlag, New-York.

Gannoun, A., Saracco, J., Yu, K., (2003). Nonparametric prediction by conditional median and quantiles, *J. Stat. Plann. Inference*, **117**, No.2, 207-223.

Roussas, G. G., (1968), On some properties of nonparametric estimates of probability density functions. *Bull. Soc. Math. Greece (N.S.)*, **9**, fasc. 1, 29–43.

Yu, K., Lu, Z. and Stander, J., (2003). Quantile regression: applications and current research areas, *The Statistician*, **52**, 3, 331-350.

# Poster 11

# Local polynomial quantile regression with parametric features

**Anouar El Ghouch**
*Department of Econometrics. University of Geneva*

**Marc G. Genton**
*Department of Econometrics. University of Geneva*

**Abstract.** In this paper, we propose a new approach to conditional quantile function estimation that combine both parametric and nonparametric techniques. The new quantile regression estimator behaves like a parametric one when this latter is correct and converges to the nonparametric solution as the parametric start deviates from the true underlying model.

## 1 Introduction

Unlike parametric techniques, nonparametric kernel smoothing are well known flexible methods that can be used without making restrictive assumptions about the form of the unknown target function. In general, their performances depend on the smoothness of the regression function, the sample size $n$, the selected kernel $K$ (typically a symmetric density function), and the bandwidth $h_n$ that describes the degree of smoothing applied to the data. Many kernel smoothers, such as Nadarya-Watson, local linear and nearest neighbour, share the same form for the asymptotic mean squared error (MSE), namely $(h^2a)^2 + (nh)^{-1}b$, where the first term is the squared asymptotic bias and the second term is the asymptotic variance. The quantities $a$ and $b$ depend on the unknown data generating procedure and on the chosen kernel but neither on $n$ nor on $h_n$. From this formula it is clear that, for a fixed $n$, one can reduce the bias just by choosing a small bandwidth ($0 < h_n \to 0$) however, this will inevitably increase the variance of the estimator. In the literature, there are at least two approaches that have been proposed in

order to reduce the bias without increasing the instability of the resulting estimator. The first one, aims at improving the bias rate from $O(h^2)$ to $O(h^4)$ by using, for example, higher-order kernels or variable kernel methods. As noted by Jones and Signorini (1997) in the context of density estimation, the merit of such an approach is not clear for finite (small to moderate) sample size. The second approach, that will be used here, attempts to remove the bias asymptotically by acting only on the leading constant term $a$ without changing the variance of the estimator. This is particularly interesting since the decrease of the bias allows to increase the bandwidth, and therefore to use more data in the local fit which will be beneficial also for reducing the variance. To achieve that goal, one of the most used techniques is to guide the nonparametric regression function by a parametric pilot estimate. See Einsporn (1987), Fan and Ullah (1999), Mays et al. (2001), and Hagmann and Scaillet (2007) for more details.
The literature cited above focuses on the mean regression or density estimation. However, to the best of our knowledge, except for the recent work of Su and Ullah (2007), there is no available literature about such technique in the context of quantile regression. Su and Ullah (2007) propose a double smoothing estimator where a local linear primary fit is multiplicatively adjusted by a another local linear fit. Our approach and motivations here are completely different. We are particularly interested in the case where a naive parametric estimator (that may be completely misspecified) is available but either fails to fit adequately the observed data or casts some doubts about its accuracy and efficacy. The parametric estimator is then corrected additively through a $p$'th order local polynomial quantile regressor. This considerations yield a consistent and substantial better estimate of the underlying conditional quantile function and its derivatives with a single bandwidth. The latter not only controls the local window size, as it is the case for the classical kernel methods, but also locally adapts to the global parametric model. Another advantage of the proposed method is that our estimator can be seen as a generalization of the classical local polynomial fit. It shares with the well known local linear smoother ($p = 1$) some good properties such as small boundary effects, adaptive design, and high minimax efficiency. However, it has typically a smaller mean squared error and a faster rate of convergence. In this paper, we give a Bahadur-type representation of the proposed estimator from which consistency and asymptotic normality are derived under $\alpha$-mixing assumption. We also discuss the numerical implementation and investigate the performance of the estimator via some simulations. Finally, we propose and numerically study a practical bandwidth selector based on the plug-in principle.

## 2 The procedure

The objective function is given by $Q_\pi(x) = \arg\min_a \mathbb{E}_x\left(\varphi_\pi(Y-a)\right)$. The local polynomial (LP) estimator $\hat{\beta} = (\hat{\beta}_0, \ldots, \hat{\beta}_p)^T$ is defined by

$$\arg\min_b \sum_{i=1}^{n} \varphi_\pi(Y_i - \tilde{X}_i^T b) K_{h,i}, \tag{1}$$

where $\tilde{X}_i = (1, X_i - x, \ldots, (X_i - x)^p)^T$ and $K_{h,i} = K((X_i - x)/h)$, with $K$ being a kernel function.
To motivate our approach, let assume that instead of $Q_\pi(x)$, we are interested in $Q_\pi(x) -$

$q(x) \equiv \arg\min_a \mathbb{E}_x(\varphi_\pi(Y - q(X) - a))$, for some given function $q$. One can first estimate $Q_\pi$ and then subtract from it $q$ or directly search the argument that minimizes $\sum_{i=1}^n \varphi_\pi(Y_i - q(X_i) - \tilde{X}_i^T b)K_{h,i}$ with respect to $b \in \mathbb{R}^{p+1}$. An obvious way to get back to $Q_\pi(x)$ is by minimizing $\sum_{i=1}^n \varphi_\pi(Y_i - (q(X_i) - q(x)) - \tilde{X}_i^T b)K_{h,i}$. It can be shown that this is a valid estimator for $Q_\pi(x)$ and its derivatives up to the $p$'th order. However, for reasons that will become clear later, here we suggest to replace in the last equation $q(X_i) - q(x)$ by $r_q(X_i) := q(X_i) - \sum_{j=0}^p \frac{q^{(j)}(x)}{j!}(X_i - x)^j$, provided that $q^{(p)}(x)$ exists. This consideration leads to a new class of local polynomial semiparametric estimators given by

$$\arg\min_b \sum_{i=1}^n \varphi_\pi(Y_i - r_i(\hat{\theta}) - \tilde{X}_i^T b)K_{h,i}, \tag{2}$$

$r_i(\hat{\theta}) \equiv r_{X_i}(\hat{\theta})$, with $r_x(\theta)$ being a shortcut for $r_{q_\theta}(x)$, $q_\theta(x) \equiv q_\pi(x, \theta)$ being a parametric model for $Q_\pi(x)$ and $\hat{\theta} = \arg\min_{\theta\in\Theta} n^{-1}\sum_{i=1}^n \varphi_\pi(Y_i - q_\theta(X_i))$.

## 3 Asymptotic theory

Under very weak assumptions, see Komunjer (2005), $\hat{\theta}$ converges in probability to $\theta^* = \arg\min_{\theta\in\Theta} \mathbb{E}(\varphi_\pi(Y - q_\theta(X)))$. The latter parameter is the best possible value of $\theta \in \Theta$ with respect to the "distance" $\varphi_\pi$. We assume that $\hat{\theta} - \theta^* = O_p(\delta_n)$ with $\delta_n \to 0$.

**Theorem 1 (Bahadur-type representation)** *Under some regularity assumptions,*

$$\begin{aligned} &H_n(\hat{\beta} - \beta) - \frac{h^{p+1}}{(p+1)!}\Lambda^{-1}u[Q_\pi^{(p+1)}(x) - q_\pi^{(p+1)}(x, \theta^*)] \\ =&\frac{a_n^2}{f(x, Q_\pi(x))}\Lambda^{-1}\sum_{t=1}^n e_t\tilde{X}_{h,t}K_{h,t} + r_n, \end{aligned}$$

*where* $e_t = \pi - I(Y_t < Q_\pi(X_t))$, $\tilde{X}_{ht} = (1, (X_t - x)/h, \ldots, ((X_t - x)/h)^p)^T$, $H_n = diag(1, h, \ldots, h^p)$ *and* $r_n = o_p(a_n) + h_n^{p+1}(O_p(\delta_n) + o_p(1))$, *with* $a_n^{-1} = \sqrt{nh_n}$. $f(x, y)$ *is the density of* $(X, Y)$. *The matrix* $\Lambda$ *and the vector* $u$ *depend only on the kernel* $K$

**Theorem 2 (Asymptotic normality)** *Under some regularity assumptions,*

$$\begin{aligned} &\sqrt{nh_n}\left\{H_n(\hat{\beta} - \beta) - \frac{h^{p+1}}{(p+1)!}\Lambda^{-1}u[Q_\pi^{(p+1)}(x) - q_\pi^{(p+1)}(x, \theta^*)] + h_n^{p+1}(O_p(\delta_n) + o_p(1))\right\} \\ &\xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma_\pi^2(x)\Sigma\right), \end{aligned}$$

*where* $\sigma_\pi^2(x) = \dfrac{\pi(1-\pi)}{f_x^2(Q_\pi(x))f_0(x)}$. $\Sigma$ *is a matrix that depent only on* $K$, $f_x(y)$ *and* $f_0$ *are the density of* $Y|X$ *and the density of* $X$ *respectivally.*

| | | Model **M1** | | | | Model **M2** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Method | $10^2\times$ $\mathbb{B}ias^2$ | $10^2\times$ $\mathbb{V}ar$ | $10^2\times$ $MSE$ | $RE\%$ | $10^2\times$ $\mathbb{B}ias^2$ | $10^2\times$ $\mathbb{V}ar$ | $10^2\times$ $MSE$ | $RE\%$ |
| $\lambda = 0$ | LL | 1.839 | 10.07 | 11.91 | 47.24 | 1.839 | 10.07 | 11.91 | 47.24 |
| | $PQ_1$ | 0.003 | 5.624 | **5.627** | 100.0 | 0.003 | 5.624 | **5.627** | 100.0 |
| | $GLL_1$ | 0.006 | 5.817 | 5.823 | 96.64 | 0.006 | 5.817 | 5.823 | 96.64 |
| | $PQ_2$ | 0.003 | 5.672 | 5.675 | 99.15 | 0.003 | 5.672 | 5.675 | 99.15 |
| | $GLL_2$ | 0.005 | 6.050 | 6.055 | 92.93 | 0.005 | 6.050 | 6.055 | 92.93 |
| $\lambda = 2$ | LL | 2.920 | 10.30 | 13.22 | 70.24 | 2.396 | 11.42 | 13.82 | 81.14 |
| | $PQ_1$ | 9.049 | 6.239 | 15.29 | 60.75 | 25.77 | 9.415 | 35.18 | 31.87 |
| | $GLL_1$ | 1.077 | 8.210 | **9.287** | 100.0 | 1.695 | 9.518 | **11.21** | 100.0 |
| | $PQ_2$ | 2.444 | 11.56 | 14.00 | 66.32 | 4.461 | 15.85 | 20.31 | 55.22 |
| | $GLL_2$ | 0.439 | 10.30 | 10.73 | 86.52 | 0.355 | 11.05 | 11.40 | 98.32 |
| $\lambda = 10$ | LL | 2.611 | 15.15 | 17.76 | 87.49 | 3.414 | 21.90 | 25.31 | 65.32 |
| | $PQ_1$ | 228.0 | 31.80 | 259.8 | 5.982 | 597.3 | 139.0 | 736.3 | 2.245 |
| | $GLL_1$ | 2.535 | 15.53 | 18.07 | 86.02 | 3.213 | 21.68 | 24.90 | 66.40 |
| | $PQ_2$ | 1.381 | 16.98 | 18.36 | 84.63 | 2.962 | 24.32 | 27.28 | 60.60 |
| | $GLL_2$ | 0.502 | 15.04 | **15.54** | 100.0 | 0.463 | 16.07 | **16.53** | 100.0 |
| $\lambda = 20$ | LL | 3.080 | 20.81 | 23.89 | 75.77 | 5.331 | 35.45 | 40.78 | 45.29 |
| | $PQ_1$ | 951.3 | 138.1 | 1089 | 1.661 | 2410 | 568.7 | 2979 | 0.620 |
| | $GLL_1$ | 3.356 | 23.81 | 27.17 | 66.62 | 4.799 | 35.77 | 40.57 | 45.52 |
| | $PQ_2$ | 3.079 | 18.55 | 21.63 | 83.67 | 0.874 | 28.15 | 29.02 | 63.64 |
| | $GLL_2$ | 0.842 | 17.26 | **18.10** | 100.0 | 0.141 | 18.33 | **18.47** | 100.0 |

**Table 1:** $\mathbb{B}ias$, $\mathbb{V}ar$, $MSE$ and relative efficiency percentage. Sample size $n = 100$ and $N = 1000$ replications.

# 4 Simulation study

Here we focused on the median function with i.i.d. data. Our objective is to compare our hybrid local linear estimator with both the fully parametric and the fully nonparametric competitors. To do so, we need first to specify a parametric model that will be adjusted to the data. The two situations to be considered are: (1) a fixed parametric model which is likely incorrect is available; (2) no candidate model is known to the analyst. In this study, we use a very simple approach that consists to employ the Akaike's information criterion (AIC) to select the model that seems to best fit the data among the set of all polynomial regression models of degree equal to or less than 20, i.e. $\{\sum_{j=0}^{1}\theta_j X^j, \ldots, \sum_{j=0}^{20}\theta_j X^j\}$. Five conditional quantile estimating methods are compared: (**LL**) the standard local linear quantile estimator; (**$PQ_1$**) the parametric estimator $q(x,\hat{\theta})$ based on the estimating equation $\arg\min_\theta \varphi_\pi(Y_i - q(X_i,\theta))$, where $q(x,\theta)$ is a given (fixed) model; (**$GLL_1$**) the LL quantile estimator guided by $q(x,\hat{\theta})$; (**$PQ_2$**) similar to $PQ_1$ but now the parametric model is chosen using the data by the mean of the AIC criterion as described above; (**$GLL_2$**) the LL quantile estimator guided by a data-driven parametric model selected using the AIC criterion. $PQ_1$ and $GLL_1$ correspond to the situation (1) while LL, $PQ_2$ and $GLL_2$ correspond to situation (2). Two data generating procedures are used : (**M1**) $Y_i = 10 - 6X_i^2 + 2.8X_i^3 + \lambda\exp(-4(X_i-1)^2) + \epsilon_i$; (**M2**) $Y_i = 10 - 6X_i^2 + 2.8X_i^3 + \lambda\sin((\pi/2.25)(X_i-1)^2) + \epsilon_i$. As the fixed parametric start, we used $q(x,\theta) = \theta_0 + \theta_1 X + \theta_2 X^2 + \theta_3 X^3$ so that the parameter $\lambda$ in M1 and M2 is a misspecification parameter that controls the deviation from the specified model. Table **??** reports the overall squared bias, the overall variance, and the overall mean squared

error. We also give a column with the relative efficiency percentage ($RE\%$) which is $(MSE_{best}/MSE_{\hat{Q}}) \times 100$, where $MSE_{best}$ is the minimum observed value of the mean squared error (shown in bold).

# References

Jones, M. C. and Signorini, D. F. (1997). A comparison of higher-order bias kernel density estimators.*J. Amer. Statist. Assoc.*, **92**, 1063-1073.

Einsporn, R. (1987). A link between least squares regression and nonparametric curve estimation. *Ph. D. Dissertation, Virginia Tech, Blacksburg, VA.*

Fan, Yanqin and Ullah, Aman. (1999). Asymptotic normality of a combined regression estimator. *J. Multivariate Anal.*, **71**, 191-240.

Mays, James E. and Birch, Jeffrey B. and Starnes, B. Alden. (2001). Model robust regression: combining parametric, nonparametric, and semiparametric methods. *J. Nonparametr. Statist.*, **13**, 245-277.

Hagmann, M and Scaillet, O. (2007). Local multiplicative bias correction for asymmetric kernel density estimators. *J. Comput. Graph. Statist.*, **141**, 213-249.

Su, L. and A. Ullah. (2007). Nonparametric Prewhitening Estimators for Conditional Quantiles. Forthcoming in *Statistica Sinica.*.

Komunjer, Ivana. (2005). Quasi-maximum likelihood estimation for conditional quantiles. *J. Econometrics*, **128**, 137-164.

# Poster 12

# Nonparametric tools for dependence assessment and prediction in the spatial setting

**Pilar García-Soidán**
*Department of Statistics and Operations Research, University of Vigo, Spain*

**Raquel Menezes**
*Department of Mathematics for Science and Technology, University of Minho, Portugal*

**Abstract.** The nonparametric techniques provide us with a variety of tools for prediction at a particular spatial location. These methods may be directly applied to produce nonparametric predictors or in an indirect way, when used to supply an adequate estimation of the dependence structure. This work focuses on the referred procedures, giving particular emphasis to those approaches developed more recently.

## 1 Introduction

The need to reconstruct a phenomenon over the whole observation region, from a finite set of data, can be found in a broad spectrum of areas, such as geostatistics, hydrology, atmospheric science, etc. In most cases, only a single realization is available at each of the spatial locations observed and this fact conveys o demand further assumptions about the random process involved, in order to make inference possible in this setting. Stationarity is a typical requirement and, under this condition, different techniques have been developed to provide us with spatial predictors.

On one hand, the nonparametric approaches may be directly applied for the latter purpose, such as that proposed in Menezes et al. (2007), based on the kernel method, where the mean-squared error of the referred predictor tends to be negligible as the sample size increases. Moreover, bandwidth selectors are also suggested, which are constructed

by applying cross-validation techniques. Nevertheless, an additional option, widely used in practice, is that of proceeding via the kriging techniques, which produce accurate predictions at any location of the observation region, under several assumptions; see, for instance, Christakos (1992) and Cressie (1993). In particular, implementation of kriging predictors demands the second-order structure to be appropriately estimated.

Typically, the variogram of the covariance functions make available a measure of the spatial dependence, when assuming intrinsic or second-order stationarity of the random process, respectively. In a first step, the nonparametric methods may be applied for approximation of the latter functions, including the traditional estimator proposed in Matheron (1963) or more robust alternatives, given in Cressie and Hawkins (1980) and in Genton (1998). More recently, kernel proposals have been suggested in García-Soidán et al. (2003) and in Menezes et al. (2008), for uniformly random data and clustered data, respectively. However, validity of the latter estimators is not necessarily satisfied and, therefore, they cannot be used directly for prediction by using the kriging equations.

We can cope with the problem mentioned above by selecting an appropriate parametric model and then deriving optimal estimates of the parameters, although the model misspecification problem is one of its main drawbacks. The adequateness of a parametric model may be analyzed by graphical diagnostics, although they are often difficult to assess. Goodness of fit tests has been proposed by Maglione and Diblasi (2004) or in García-Soidán (2008), under gaussianity or stationarity of the random process, respectively, to check whether or not a suitable model for the variogram or the covariance functions has been selected.

Another alternative to produce valid estimation could be that of proceeding in a nonparametric way for transformation of a given estimator into a valid one. Hall, Fisher and Hoffman (1994) propose to combine truncation and inversion of the estimator considered, although no strategy about selection of the truncation term is supplied. In García-Soidán and Menezes (2007), an orthonormal series approach is suggested so as to obtain a valid estimator, under several hypotheses, such as selection of valid bases and also of those terms corresponding to positive Fourier coefficients.

This manuscript overviews the referred procedures, paying particular attention to those approaches developed recently and, more specifically, to those provided by the authors of the current work.

## References

Christakos, G. (1992). *Random field models in earth sciences.* Academic Press, San Diego.

Cressie, N. (1993). *Statistics for spatial data.* Wiley, New York.

Cressie, N.; Hawkins, D. (1980). Robust estimation of the variogram. *Mathematical Geology*, **12**, 115-125.

García-Soidán, P. (2008). Goodness of fit test for variogram models based on the kernel method. *Unpublished manuscript.*

García-Soidán, P.; Menezes, R. (2007). Covariance estimation via the orthonormal series. Proceedings of the 56th Session of the International Statistics Institute 2007, Lisboa.

García-Soidán, P.; González-Manteiga, W.; Febrero, M. (2003). Local linear regression estimation of the variogram. *Statistics and Probability Letters*, **64**, 169-179.

Genton, M. (1998). Highly robust variogram estimation. *Mathematical Geology*, **30**, 213-221.

Hall, P.; Fisher, I.; Hoffmann, B. (1994). On the nonparametric estimation of covariance functions. *Annals of Statistics*, **22**, 2115-2134.

Maglione, D. S.; Diblasi, A. M. (2004). Exploring a valid model for the variogram of an intrinsic spatial process. *Stochastic Environmental Research and Risk Assessment*. **18**, 366-376.

Matheron, G. (1963). Principles of Geostatistics. *Economic Geology*, **58**, 1246-1266.

Menezes, R.; Ferreira, C.; García-Soidán, P. (2007). A nonparametric approach for spatial prediction. Proceedings of the International Conference of Statistics and Probability in Science and Technology 2007, Porto.

Menezes, R.; García-Soidán, P.; Febrero, M. (2008). A kernel variogram estimator for clustered data. *Scandinavian Journal of Statistics*, **35**, 1, 18-37.

# Poster 13

# Nonparametric inference for branching processes

**M. González**[1]
*Department of Mathematics. University of Extremadura. Spain*

**C. Gutiérrez**
*Department of Mathematics. University of Extremadura. Spain*

**R. Martínez**
*Department of Mathematics. University of Extremadura. Spain*

**M. Mota**
*Department of Mathematics. University of Extremadura. Spain*

**Abstract.** Discrete-time branching processes are stochastic growth population models in which the individuals (or couples) with reproductive capacity give rise new individuals in each generation. The behaviour of these populations is strongly related to the main parameters of the offspring distribution. In practice these values are unknown and their estimation is necessary. Usually it must be observed the whole family tree up to a given generation in order to estimate the offspring distribution. In this work, we deal with the problem of estimating the main parameters of the model assuming that the only observable data are the total number of individuals in each generation. We set out the problem in a nonparametric framework and obtain the maximum likelihood estimator of the offspring distribution using the EM algorithm. Finally, we show the accuracy of the algorithm by way of simulated examples.

[1]The research was supported by the Ministerio de Educación y Ciencia and the FEDER through the Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica, grant MTM2006-08891.

# Poster 14

# Testing for polynomial versus nonparametric regression using mixed model penalized splines

**Sonja Greven and Ciprian Crainiceanu**
*Department of Biostatistics, Johns Hopkins University*

**Helmut Küchenhoff**
*Department of Statistics, Ludwig-Maximilians-Universität München*

**Annette Peters**
*Institute of Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health*

**Abstract.** We consider the non-standard problem of testing for random effects in linear mixed models. This problem, while more general, is often encountered when testing for polynomial regression against a general alternative, modeled nonparametrically using mixed model penalized splines. We propose two approximations to the null distribution of the restricted likelihood ratio test statistic.

## 1 Introduction

Linear mixed models are commonly used to model longitudinal or clustered data. More recently, they have been employed in nonparametric regression to estimate the smoothing parameters for penalized splines using (restricted) maximum likelihood. We focus on linear mixed models of the form

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{Z_1 b_1} + \ldots + \boldsymbol{Z_S b_S} + \boldsymbol{\varepsilon}, \tag{1}$$

with random effects $\boldsymbol{b_s} \sim N(\boldsymbol{0}, \sigma_s^2 \boldsymbol{I_{K_s}})$ independent of $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma_\varepsilon^2 \boldsymbol{I_n})$, $K_s$ columns in $\boldsymbol{Z_s}$, $\boldsymbol{I_\nu}$ the identity matrix of size $\nu$, and $n$ the sample size. In the case of mixed model

penalized splines, $\boldsymbol{X}$ includes polynomial terms for a covariate $x$ of interest, while the corresponding $\boldsymbol{Z_s b_s}$ models general smooth deviations from this polynomial.

We are interested in testing one of the variance components

$$H_{0,s} : \sigma_s^2 = 0 \quad \text{versus} \quad H_{A,s} : \sigma_s^2 > 0, \tag{2}$$

corresponding e.g. to testing a random intercept or testing for linearity against a general alternative. This problem is non-standard due to the parameter on the boundary of the parameter space. For i.i.d. data, the Likelihood Ratio Test (LRT) statistic for (2) has an asymptotic $0.5\chi_0^2 : 0.5\chi_1^2$ null distribution (Self and Liang, 1987; Stram and Lee, 1994). However, for mixed model penalized splines responses are correlated under the alternative. For $S = 1$, Crainiceanu and Ruppert (2004) derived the finite sample and asymptotic null distribution of the (restricted) LRT for testing (2). For $S > 1$, they recommend a parametric bootstrap, which can be computationally expensive. We develop two faster approximations to the finite sample null distribution of the restricted LRT.

# 2 Two Approximations to the RLRT Null Distribution

## 2.1 Fast Finite Sample Approximation

Our first approximation is inspired by pseudo-likelihood estimation, where nuisance parameters are replaced by consistent estimators. Liang and Self (1996) showed that under certain regularity assumptions the asymptotic distribution of the pseudo-LRT is the same as if the nuisance parameters were known. For our problem, we assume that under regularity conditions the prediction of $\sum_{i \neq s} \boldsymbol{Z_i b_i}$ is good enough to allow the distribution of the restricted LRT (RLRT) in model (1) to be closely approximated by that in

$$\widetilde{\boldsymbol{Y}} = \boldsymbol{X\beta} + \boldsymbol{Z_s b_s} + \boldsymbol{\varepsilon}, \tag{3}$$

with $\widetilde{\boldsymbol{Y}} = \boldsymbol{Y} - \sum_{i \neq s} \boldsymbol{Z_i b_i}$ assumed known. As model (3) has only one variance component $\sigma_s^2$, the exact null distribution of the RLRT for (2) is known (Crainiceanu and Ruppert, 2004) and can be simulated from efficiently. This approach is implemented in the R-package `RLRsim` by Fabian Scheipl, available from CRAN.

## 2.2 Mixture Approximation to the Bootstrap Distribution

If a parametric bootstrap is preferred, but computationally intensive, we propose the following parametric approximation to the RLRT distribution

$$RLRT \overset{d}{\approx} aUD, \tag{4}$$

where $U \sim \text{Bernoulli}(1-p)$, $D \sim \chi_1^2$, and $\overset{d}{\approx}$ denotes approximate equality in distribution. $p$ and $a$ can be estimated from a bootstrap sample, while (4) stabilizes quantile estimation and reduces the necessary bootstrap sample size. We propose estimation using the

method of moments, as ML estimation of $p$ is very sensitive to numerical imprecisions (`MIXED` in SAS, `lme` in R).

Note that both our proposed approximations are asymptotically identical to the $0.5\chi^2_0 : 0.5\chi^2_1$ approximation when the i.i.d. assumption holds.

# 3 Simulation Study

We conducted an extensive simulation study, covering a range of important situations. Eight settings combined one or two of the following: random intercepts, random slopes, smooth uni- and bivariate functions. Variance components not tested for were assumed to be positive. 10,000 samples each were simulated from the RLRT null distribution, and our two approximations were compared to a bootstrap and the $0.5\chi^2_0 : 0.5\chi^2_1$ approximation. The fast finite sample approximation produced empirical type I error rates close to the nominal level, comparable to the exact distribution when $S = 1$. The approximation was usually good even for $n = 30$; the necessary sample size increased somewhat when random effects were highly correlated. The $aUD$ approximation reduced the necessary bootstrap sample size by about 10-20% for $\alpha = 0.05$ and 50% for $\alpha = 0.001$. The $0.5\chi^2_0 : 0.5\chi^2_1$ approximation was always very conservative.

# 4 An Application from Epidemiology

The Airgene study investigates inflammatory responses to ambient air pollution concentrations in myocardial infarction survivors. Three inflammatory blood markers were measured monthly up to 8 times in 1,003 patients from six European cities. Air pollution and weather variables were recorded concurrently in each city, and patient information was collected at baseline.

Analyses had to account for longitudinal data structure and non-linearity of pollutant and weather dose-response functions. Smooth functions were estimated using mixed model P-splines, penalizing deviations from linearity (Greven et al., 2006). A typical model for the $j^{th}$ marker value of the $i^{th}$ patient, $y_{ij}$, is

$$y_{ij} = u_i + f(poll_{ij}) + \sum_{l=2}^{L} \beta_l x_{l,ij} + g(temp_{ij}) + \cdots + \varepsilon_{ij}, \tag{5}$$

with $u_i \overset{iid}{\sim} N(0, \sigma^2_u)$ a random patient intercept, $f(.)$ and $g(.)$ smooth functions of air pollutant and temperature, $x_l$ e.g. patient's gender, and $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2_\varepsilon)$.

To illustrate, Figure 1 shows the estimated smooth associations between log(IL-6) and two different sizes of particles in two cities. We want to know if these functions are different from linear, as commonly assumed. This corresponds to testing (2) in (5), where $\sigma^2_s$ controls the smoothness of $f$. Note that the i.i.d. assumption is violated and $S \geq 2$. Test results for the ultrafine particles in Athens are given in Table 1. The fast finite sample approximation reduces computation time from hours to seconds, with results similar to a bootstrap, indicating nonlinearity. For the bootstrap, computation time can be reduced by using the $aUD$ approximation with less samples. The $0.5\chi^2_0 : 0.5\chi^2_1$ approximation is clearly conservative. The $PM_{2.5}$-log(IL-6) dose-response function in Rome was found to be not significantly different from linear ($p = 0.1$).

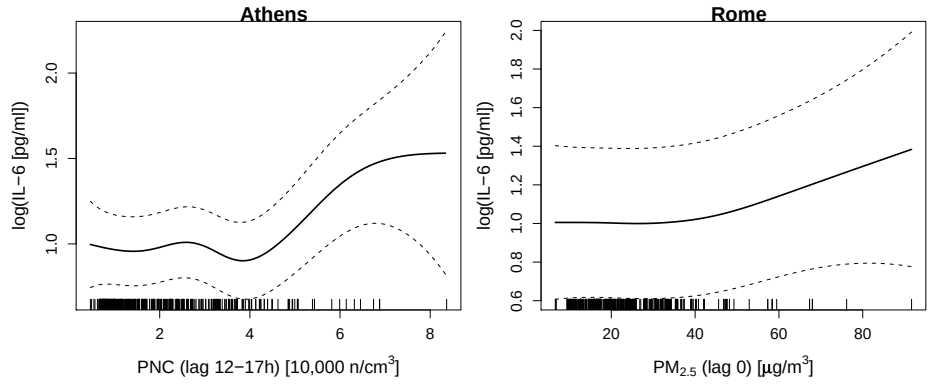


**Figure 1:** Estimated dose-response functions with approximate pointwise 95% confidence intervals for log(IL-6) and a) ultrafine particles (diameter $< 0.1\mu$g) in Athens b) $PM_{2.5}$ (diameter $< 2.5\mu$g) in Rome.

| Approximation | Samples | Computation time | p-value |
|---|---|---|---|
| Fast Finite Sample (Matlab) | 100,000 | 33sec | 0.014 |
| Fast Finite Sample (R) | 100,000 | 88sec | 0.016 |
| $aUD$ | 1,000 | 19min | 0.017 |
| $aUD$ | 10,000 | 3.4h | 0.019 |
| $0.5\chi^2_0 : 0.5\chi^2_1$ | - | - | 0.031 |
| Parametric Bootstrap | 10,000 | 3.4h | 0.017 |

**Table 1:** Testing the association between log(IL-6) and ultrafine particles in Athens for linearity. The test statistic was $RLRT = 3.5$.

# 5 Summary

We have discussed testing for random effects in linear mixed models. An important special case is testing for polynomial versus nonparametric regression using mixed model penalized splines. We have proposed two approximations to the finite sample null distribution of the RLRT. Extensive simulations showed superiority of both approximations to the $0.5\chi^2_0 : 0.5\chi^2_1$ approximation and parametric bootstrap currently used. Our results extend existing methodology to linear mixed models with more than one random effect and correlated responses, such as for mixed model penalized splines. We applied our approach to testing for linearity of air pollution dose-response functions.

# References

Crainiceanu, C., and Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B*, **66**(1), 165-185.

Greven, S., Küchenhoff, H., and Peters, A. (2006). Additive mixed models with P-Splines. *Proceedings of the 21st International Workshop on Statistical Modelling*, 201-207. Eds.: Hinde J, Einbeck J and Newell J.

Liang, K.-Y. and Self, S.G. (1996). On the Asymptotic Behaviour of the Pseudolikelihood Ratio Test Statistic. *Journal of the Royal Statistical Society: Series B*, **58**(4), 785-796.

Self, S.G. and Liang, K.-Y. (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. *Journal of the American Statistical Association*, **82**(398), 605-610.

Stram, D.O. and Lee, J.-W. (1994). Variance Components Testing in the Longitudinal Mixed Effects Model. *Biometrics*, **50**(3), 1171-1177.

# Poster 15

# CLT in nonlinear wavelet regression with left-truncated $\alpha$-mixing observations

**Han-Ying Liang**[1]
*Department of Mathematics, Tongji University, and Department of Statistics and OR, University of Vigo*

**Jacobo de Uña-Álvarez**[2]
*Department of Statistics and OR, University of Vigo*

**Abstract.** In this paper we establish the asymptotic normality of a new nonlinear wavelet estimator for the regression function, in the context of left-truncated, dependent data.

## 1 Introduction

The importance of wavelets in curve estimation is well known since the nineties. Unlike other commonly used estimation techniques (e.g. kernel estimators), wavelets adapt (in the minimax sense) to the smoothness degree of the underlying curve. Hall and Patil (1995, 1996) established the asymptotic mean integrated squared error (MISE) of the nonlinear wavelet estimator of the density and the regression function, including the case of discontinuities (which do not influence the rate of convergence).

In Survival Analysis and other fields, censoring and truncation are phenomena typically encountered in the recording of the data. Several papers have considered adaptation of nonlinear wavelets to the censored scenario (e.g. Li et al., 2008). However, for the best

[1]Supported by the National Natural Science Foundation of China (10571136) and by the Grants MTM2005-01274 (FEDER support included) and PGIDIT07PXIB300191PR
[2]Supported by the Grants MTM2005-01274 (FEDER support included) and PGIDIT07PXIB300191PR

of our knowledge, literature on nonlinear wavelets for left-truncated data is non-existent. Besides, results on wavelets for survival data have been restricted to the independent framework; however, dependent survival data are found when sampling clusters of individuals (family members, or repeated measurements taken on the same subject; see Cai and Kim, 2003, for further motivation). For $\alpha$-mixing (complete) observations, nonlinear wavelet density and regression estimation was considered in Liang et al. (2005) and Truoug and Patil (2001) respectively. Interestingly, the MISE rate of convergence is unchanged (w.r.t. the independent case) under this type of short-range dependence. For left-truncated, $\alpha$-mixing data, de Uña-Álvarez and Liang (2008) introduced a new nonlinear wavelet estimator of the regression function (and of the covariate's density), and they derived the asymptotic MISE formula. In this paper we establish the asymptotic normality of these estimators.

## 2 The nonlinear wavelet estimators

In the sequel, $\{(X_k, Y_k, T_k) =: \xi_k,\, k \geq 1\}$ is assumed to be a stationary $\alpha$-mixing sequence of random vectors from $(X, Y, T)$, where $Y$ is a continuous lifetime, $X$ is a continuous one-dimensional covariate, and $T$ is the left-truncation (continuous) variable (see e.g. Woodroofe, 1985). We assume throughout that $T$ and $(X, Y)$ are independent. Recall that the sequence $\{\xi_k, k \geq 1\}$ is said to be $\alpha$-mixing if the $\alpha$-mixing coefficient

$$\alpha(n) := \sup_{k\geq 1} \sup\{|P((A \cap B) - P(A)P(B)| : A \in \mathcal{F}_{n+k}^{\infty}, B \in \mathcal{F}_1^k\}$$

converges to zero as $n \to \infty$, where $\mathcal{F}_l^m$ denotes the $\sigma$-algebra generated by $\xi_l, \xi_{l+1}, \ldots, \xi_m$ with $l \leq m$. Among various mixing conditions used in the literature, $\alpha$-mixing is reasonably weak; see Cai and Kim (2003) for applications in Survival Analysis.

Introduce the regression function

$$\mathbb{E}(Y|X = x) := m(x) \quad x \in \mathbb{R}, \tag{1}$$

which can be written as $m(x) = \frac{h(x)}{v(x)}$, where $v(x)$ stands for the covariate's density, and $h(x) = \int_{\mathbb{R}} y f(x, y)dy$ with $f(\cdot, \cdot)$ being the joint density function of $(X, Y)$.

For all function $v(\cdot)$ in $L_2(\mathbb{R})$, we have the following wavelet expansion:

$$v(x) = \sum_{j=-\infty}^{\infty} a_{mj}\phi_{mj}(x) + \sum_{i=m}^{\infty}\sum_{j=-\infty}^{\infty} a_{ij}\psi_{ij}(x), \tag{2}$$

where $a_{mj} = \int v(x)\phi_{mj}(x)dx$ and $a_{ij} = \int v(x)\psi_{ij}(x)dx$ are the wavelet coefficients of the function $v(\cdot)$ and the series in (2) converges in $L_2(\mathbb{R})$. See Daubechies (1992) for more details on wavelets. We define the non-linear wavelet estimator of $v(x)$ as

$$\hat{v}_n(x) = \sum_{j=-\infty}^{\infty} \hat{a}_{mj}\phi_{mj}(x) + \sum_{i=m}^{\pi}\sum_{j=-\infty}^{\infty} \hat{a}_{ij} I(|\hat{a}_{ij}| > \delta)\psi_{ij}(x), \tag{3}$$

where $\delta > 0$ is a "threshold" and $\pi \geq 1$ is another smoothing parameter, and the wavelet coefficients $\hat{a}_{mj}$ and $\hat{a}_{ij}$ are defined as follows:

$$\hat{a}_{mj} = \frac{\theta_n}{n}\sum_{k=1}^{n} \frac{1}{G_n(Y_k)}\phi_{mj}(X_k), \qquad \hat{a}_{ij} = \frac{\theta_n}{n}\sum_{k=1}^{n} \frac{1}{G_n(Y_k)}\psi_{ij}(X_k). \tag{4}$$

Here, $G_n$ denotes the product-limit estimator of the truncation distribution $G$, while $\theta$ is an estimator of the probability of no truncation, namely

$$\theta_n = \frac{G_n(y)[1 - F_n(y-)]}{C_n(y)}.$$

where $F_n$ stands for the product-limit estimator of the lifetime distribution $F$.

Similarly, if the function $h$ is square-integrable then its wavelet expansion is given by

$$h(x) = \sum_{j=-\infty}^{\infty} b_{mj}\phi_{mj}(x) + \sum_{i=m}^{\infty}\sum_{j=-\infty}^{\infty} b_{ij}\psi_{ij}(x), \tag{5}$$

where $b_{mj} = \int h(x)\phi_{mj}(x)dx$ and $b_{ij} = \int h(x)\psi_{ij}(x)dx$. Note that

$$H_n(x) = \frac{\theta_n}{n}\sum_{k=1}^{n}\frac{Y_k}{G_n(Y_k)}I(X_k \le x)$$

is an estimator of $H(x) = \int_{u\le x} h(u)du$. So, the proposed non-linear wavelet estimator of $h(x)$ is

$$\hat{h}_n(x) = \sum_{j=-\infty}^{\infty} \hat{b}_{mj}\phi_{mj}(x) + \sum_{i=m}^{\pi}\sum_{j=-\infty}^{\infty} \hat{b}_{ij}I(|\hat{b}_{ij}| > \delta)\psi_{ij}(x), \tag{6}$$

where $\hat{b}_{mj} = \frac{\theta_n}{n}\sum_{k=1}^{n}\frac{Y_k}{G_n(Y_k)}\phi_{mj}(X_k)$, $\hat{b}_{ij} = \frac{\theta_n}{n}\sum_{k=1}^{n}\frac{Y_k}{G_n(Y_k)}\psi_{ij}(X_k)$. Further, from (3) and (6), a wavelet estimator of $m(x)$ is given by $\hat{m}(x) = \hat{h}_n(x)/\hat{v}_n(x)$.

## 3 Main results

In the sequel, let $C, C_0, C_1, \cdots$ and $c$ denote generic finite positive constants, whose values are unimportant and may change from line to line, $A_n = O(B_n)$ stand for $A_n \le CB_n$, $a_n \asymp b_n$ mean $0 < \liminf a_n/b_n \le \limsup a_n/b_n < \infty$. Throughout this paper, we assume that

$$a_G < a_F \quad b_G \le b_F < \infty. \tag{7}$$

Let $x_m = [2^m x]/2^m$ for $x \in \mathbb{R}$ and

$$\Sigma(x) = \begin{pmatrix} \Sigma_0(x) & \Sigma_1(x) \\ \Sigma_1(x) & \Sigma_2(x) \end{pmatrix},$$

$$\Sigma_i(x) = \int\int \frac{y^{2-i}f(x,y)}{G(y)}\Big[\sum_l \phi(u+l)\phi(l)\Big]^2 dudy \quad (i = 0, 1, 2).$$

In order to formulate the main results, we need to impose the following assumptions.

(A1) For all integers $j \ge 1$, the joint conditional density $v_j^*(\cdot,\cdot)$ of $X_1$ and $X_{j+1}$ exists on $R \times R$ and satisfies $v_j^*(t_1, t_2) \le C$ for all $t_1, t_2 \in \mathbb{R}$ with $|t_1 - t_2| \le \delta_0$ for some $\delta_0 > 0$.

(A2) (i) The density $v(\cdot)$ satisfies $0 < v(x) \leq C_2$ for $x \in \mathbb{R}$;

(ii) $f(x, y)$ is bounded and continuous with respect to the first component.

(A3) The smoothing parameters $\pi$ and $\delta$ are functions of $n$. Suppose that $\pi \to \infty$ as $n \to \infty$ in such a manner that $2^{\pi}\delta^2 = O(n^{-\epsilon})$ for some $0 < \epsilon < 1$, $\delta \geq C_3(n^{-1}\ln(n))^{1/2}$.

(A4) The sequence $\alpha(n)$ satisfies

(i) there exist positive integers $p := p(n)$ and $q := q(n)$ such that $p + q \leq n$, and as $n \to \infty$, $p/n \to 0$, $qp^{-1} \to 0$ and $(n/p)\alpha(q) \to 0$;

(ii) there exist $\gamma > 2$ and $\eta > 1 - 2/\gamma$ such that $\sum_{l=1}^{\infty} l^{\eta}[\alpha(l)]^{1-2/\gamma} < \infty$.

**Theorem 1** *In addition to the usual conditions on the wavelet functions and the assumptions (A1)-(A4) and (7). Assume that the $r$-th derivative $v^{(r)}(\cdot)$ of $v(\cdot)$ is continuous and bounded. Let $\alpha(k) = O(k^{-\lambda})$ for some*

$$\lambda > \max\{3, d(d+\mu)/(2\mu), 1 + 4r/[\epsilon(2r+1)], (\tau-1)(2\tau+1)(2-\epsilon)/(2\epsilon(\tau-2))\}, \quad (8)$$

*where $\tau > 2$, $d > 2$, $\mu > 0$, and*

$$\epsilon(\lambda + 1 + 2b) + 2b/(2r+1) \geq 2(b+1) \quad \textit{for} \quad b > 1. \quad (9)$$

*If $2^m \asymp n^{1/(2r+1)}$ and $(p2^m/n)^{d/2-1}2^{m\mu/(d+\mu)} \to 0$, then*

$$\sqrt{n2^{-m}}(\hat{v}_n(x_m) - v(x_m) - a(x_m)) \xrightarrow{\mathcal{D}} N(0, \sigma^2(x)) \quad x \in \mathbb{R},$$

*where $a(x) = (r!)^{-1}v^{(r)}(x)2^{-rm}\int u^r[\sum_l \phi(u+l)\phi(l)]du$ and $\sigma^2(x) = \theta\Sigma_2(x)$. Further, if $n2^{-(2r+1)m} \to 0$, then $\sqrt{n2^{-m}}(\hat{v}_n(x_m) - v(x_m)) \xrightarrow{\mathcal{D}} N(0, \sigma^2(x))$.*

**Theorem 2** *Suppose that the assumptions in Theorem 2 are satisfied, and that the $r$-th derivative $h^{(r)}(\cdot)$ of $h(\cdot)$ is continuous and bounded. If*

$$2^m \asymp (n\ln(n))^{1/(2r+1)} \quad \textit{and} \quad (p2^m/n)^{d/2-1}2^{m\mu/(d+\mu)} \to 0,$$

*then $\sqrt{n2^{-m}}(\hat{m}_n(x_m) - m(x_m)) \xrightarrow{\mathcal{D}} N(0, \Delta^2(x))$ $x \in \mathbb{R}$, where*

$$\Delta^2(x) = \frac{\theta[\Sigma_0(x)v^2(x) + \Sigma_2(x)h^2(x) - 2v(x)h(x)\Sigma_1(x)]}{v^4(x)}.$$

**Remark.** In Theorems 1 and 2, if we replace $\alpha(k) = O(k^{-\lambda})$ by $\alpha(k) = O(\rho^k)$ for some $0 < \rho < 1$, then (8) and (9) are automatically satisfied.

# References

Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia.

Hall, P. and Patil, P. (1995). Formulae for mean integrated squared error of non-linear wavelet-based density estimators. *Ann. Statist.* **23**, 905-928.

Hall, P. and Patil, P. (1996). On the choice of smoothing parameter, threshold and truncation in nonparametric regression by non linear wavelet methods. *J. Roy. Statist. Soc. Ser. B* **58(2)**, 361-377.

Li, L., MacGibbon, B. and Valenta, C. (2008). On the optimality of wavelet-based nonparametric regression with censored data. Journal of Applied Probability and Statistics, to appear.

Liang, H. Y., Mammitzsch, V. and Steinebach, J. (2005). Nonlinear wavelet density and hazard rate estimation for censored data under dependent observations. *Statist. Decisions.* **23**, 161-180.

Truoug, Y. and Patil, P. (2001). Asymptotics for wavelet based estimates of piecewise smooth regression for stationary time series. *Ann. Inst. Statist. Math.* **53**, 159-178.

de Uña-Álvarez, J. and Liang, H.-Y. (2008) Nonlinear wavelet regression for left-truncated, dependent data. Unpublished manuscript.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* **13**, 163-177.

# Poster 16

# Heteroscedastic unbalanced two-fold nested model when the numbers of classes and subclasses are both large

**Shu-Min Liao**
*Department of Statistics. The Pennsylvania State University*

**Michael G. Akritas**
*Department of Statistics. The Pennsylvania State University*

**Abstract.** In the context of the heteroscedastic and unbalanced two-fold nested design, we consider testing for the subclass effect. The proposed test procedure pertains to cases having a large number of classes and a large number of subclasses, while the cell sizes are small. The asymptotic theory of the proposed test statistic is obtained under both the null and alternative hypotheses. Its performance is compared numerically with that of the classical F-test.

## 1 The general two-fold nested model

In the *general* unbalanced two-fold fixed-effects model, we observe

$$Y_{ijk} = \mu_{ij} + \sigma_{ij} \cdot e_{ijk}, \quad i = 1, \cdots, r; \quad j = 1, \cdots, c_i; \quad k = 1, \cdots, n_{ij}, \tag{1}$$

where the $\mu_{ij}$ and $\sigma_{ij}$ are bounded and $e_{ijk}$ are independent with

$$E(e_{ijk}) = 0, \quad Var(e_{ijk}) = 1. \tag{2}$$

Note that the general model (1), (2) does not assume that the errors $e_{ijk}$ are normally, or even indentically, distributed. Let

$$C = \sum_{i=1}^{r} c_i, \quad n_{ic} = \sum_{j=1}^{c_i} n_{ij}, \quad N_c = \sum_{i=1}^{r}\sum_{j=1}^{c_i} n_{ij} = \sum_{i=1}^{r} n_{ic}. \tag{3}$$

The means $\mu_{ij}$ are typically decomposed as

$$\mu_{ij} = \mu + \alpha_i + \delta_{ij}, \tag{4}$$

where we assume that

$$\sum_{i=1}^{r} n_{ic}\alpha_i = 0 \text{ and } \sum_{j=1}^{c_i} n_{ij}\delta_{ij} = 0, \forall i.$$

In this paper, we are mainly interested in testing $H_0$: $\delta_{ij} = 0$ (no subclass effect). Let

$$MS\delta = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c_i} n_{ij}(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot})^2}{C-r}, \tag{5}$$

$$MSE = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c_i}\sum_{k=1}^{n_{ij}}(Y_{ijk} - \bar{Y}_{ij\cdot})^2}{N_c - C}, \tag{6}$$

where $\bar{Y}_{ij\cdot}$ and $\bar{Y}_{i\cdot\cdot}$ are the corresponding unweighted means of $Y_{ijk}$ within each subclass and within each class, i.e.

$$\bar{Y}_{ij\cdot} = \frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}} Y_{ijk}, \quad \bar{Y}_{i\cdot\cdot} = \frac{1}{n_{ic}}\sum_{j=1}^{c_i}\sum_{k=1}^{n_{ij}} Y_{ijk} = \frac{1}{n_{ic}}\sum_{j=1}^{c_i} n_{ij}\bar{Y}_{ij\cdot\cdot}.$$

Then, the usual F-test statistic for testing $H_0$: $\delta_{ij} = 0$ is

$$F_C^{\delta} \equiv \frac{MS\delta}{MSE}. \tag{7}$$

Under the *normal* model, i.e. if $e_{ijk}$ are assumed to be iid $N(0,1)$, we have that

$$F_C^{\delta} \sim F_{C-r,\, N_c - C}, \text{ under } H_0 : \delta_{ij} = 0. \tag{8}$$

## 2 Proposed test statistic and the main theorem

In this section, we first propose a test statistics for the hypothesis $H_0 : \delta_{ij} = 0$, as $r \to \infty$, $\min(c_i) \to \infty$, but $n_{ij}$ remain fixed, and then summarize its limiting distributions, both under the null and the alternative hypotheses, using a theorem in the end.

It is not hard to verify that, in the homoscedastic case, $E(MSE) = E(MS\delta)$ under the null hypothesis. In the heteroscedastic case, this is no longer true, but the similar equality could be achieved by replacing $MSE$ with

$$MSE^* = \frac{1}{C-r}\sum_{i=1}^{r}\sum_{j=1}^{c_i}\left(1 - \frac{n_{ij}}{n_{ic}}\right) S_{ij}^2, \text{ where } S_{ij}^2 \equiv \frac{1}{n_{ij}-1}\sum_{k=1}^{n_{ij}}(Y_{ijk} - \bar{Y}_{ij\cdot})^2. \tag{9}$$

We then define our test statistic as

$$F_C^* \equiv \frac{MS\delta}{MSE^*}. \tag{10}$$

It is easy to verify that, in the balanced case, $F_C^* = F_C^\delta$, where $F_C^\delta$ is the classical $F$-statistic given in (7). Then, the asymptotic distribution of the proposed test statistic, $F_C^*$, defined in (10), is given by the following theorem.

**Theorem 1 (The asymptotic distribution of $F_C^*$)** *Consider the model and assumptions given in (1), (2), and the decomposition of the means given in (4). In addition, we assume that there exist $\kappa_{ij}$, $\lambda_i$, $a_1$, $b_1$, $b_2$ and $b_3$ such that as $r \to \infty$ and $\min(c_i) \to \infty$,*

$$E(e_{ijk}^3) = 0, \quad E(e_{ijk}^4) = \kappa_{ij}, \text{ and } E|e_{ijk}|^{4+2\epsilon} < \infty \text{ for some } \epsilon > 0;$$

$$\sqrt{C}\left(\frac{r \cdot c_i}{C} - \lambda_i\right) \longrightarrow 0, \text{ where } \lambda_i < 1, \text{ and } \frac{1}{r}\sum_{i=1}^{r} \lambda_i = 1,$$

$$\sqrt{C}\left(\frac{1}{C}\sum_{i=1}^{r}\sum_{j=1}^{c_i} \sigma_{ij}^2 - a_1\right) \longrightarrow 0, \quad \frac{1}{C}\sum_{i=1}^{r}\sum_{j=1}^{c_i} \sigma_{ij}^4 \longrightarrow b_1,$$

$$\frac{1}{C}\sum_{i=1}^{r}\sum_{j=1}^{c_i} \frac{\sigma_{ij}^4}{n_{ij}-1} \longrightarrow b_2, \quad \frac{1}{C}\sum_{i=1}^{r}\sum_{j=1}^{c_i} \frac{\sigma_{ij}^4(\kappa_{ij}-3)}{n_{ij}} \longrightarrow b_3.$$

*Then, under alternatives $\delta_{ij}$ which satisfy*

$$\sqrt{C}\left(\frac{1}{C}\sum_{i=1}^{r}\sum_{j=1}^{c_i} n_{ij}\delta_{ij}^2 - \theta_1\right) \longrightarrow 0, \quad \frac{1}{C}\sum_{i=1}^{r}\sum_{j=1}^{c_i} n_{ij}\delta_{ij}^2\sigma_{ij}^2 \longrightarrow \theta_2,$$

*as $r \to \infty$, $\min(c_i) \to \infty$ while $n_{ij}$ stay fixed, we have*

$$\sqrt{C}\left(F_C^* - (1+\theta^*)\right) \xrightarrow{d} N\left(0, \frac{1}{a_1^2}\left[2(b_1+b_2) + 4(\theta_2 + b_2\theta^*) + (2b_2+b_3)\theta^{*2}\right]\right),$$

*where $\theta^* = \theta_1/a_1$.*

*Under the null hypothesis $H_0 : \delta_{ij} = 0$, which results in $\theta^* = 0$, we then have*

$$\sqrt{C}\left(F_C^* - 1\right) \xrightarrow{d} N\left(0, \frac{2b_1+2b_2}{a_1^2}\right). \tag{11}$$

*Proof:* Omitted for now.

# 3 Simulations

In this section, simulations are used to compare two test procedures: the classical F-test procedure and the proposed test procedure of (11). Let *CF* and *HET* denote them respectively. The achieved simulated sizes, based on 10,000 simulation runs, are shown in Table 1. (Due to the space limit, the details of the simulation settings are omitted.)

| | r=5,C=27 | | r=30,C=165 | | r=50,C=275 | |
|---|---|---|---|---|---|---|
| | CF | HET | CF | HET | CF | HET |
| Normal | 0.3682 | 0.1318 | 0.5596 | 0.0864 | 0.2269 | 0.0832 |
| Exponen | 0.3550 | 0.1134 | 0.5458 | 0.0761 | 0.2182 | 0.0726 |
| LogNorm | 0.3254 | 0.0996 | 0.5470 | 0.0639 | 0.2386 | 0.0596 |
| Mixture | 0.3218 | 0.1028 | 0.5471 | 0.0635 | 0.2194 | 0.0636 |
| | r=5,C=151 | | r=30,C=915 | | r=50,C=1524 | |
| | CF | HET | CF | HET | CF | HET |
| Normal | 0.5997 | 0.0772 | 0.4674 | 0.0627 | 0.0567 | 0.0551 |
| Exponen | 0.5900 | 0.0649 | 0.4748 | 0.0612 | 0.0655 | 0.0526 |
| LogNorm | 0.5926 | 0.0590 | 0.4630 | 0.0523 | 0.1169 | 0.0454 |
| Mixture | 0.5912 | 0.0608 | 0.4643 | 0.0553 | 0.0697 | 0.0520 |
| | r=5,C=502 | | r=30,C=3015 | | r=50,C=5025 | |
| | CF | HET | CF | HET | CF | HET |
| Normal | 0.7600 | 0.0668 | 0.0007 | 0.0578 | 0.0138 | 0.0569 |
| Exponen | 0.7626 | 0.0558 | 0.0015 | 0.0543 | 0.0185 | 0.0530 |
| LogNorm | 0.7453 | 0.0450 | 0.0195 | 0.0424 | 0.0595 | 0.0425 |
| Mixture | 0.7607 | 0.0510 | 0.0015 | 0.0534 | 0.0165 | 0.0506 |

**Table 1:** Sizes over $10,000$ simulation runs under heteroscedasticity and unbalanced design ($\alpha = 0.05$).

# 4 Application

Some more applications will be presented in the talk/poster at ISNI2008.

# 5 Short concluding comments

We have shown, via theoretical derivation and numerical evidence, that the classical F-test procedure is sensitive to departures from homoscedasticity, even under normality and/or under the balanced design (simulation results for the latter case are not shown here). In order to accommodate heteroscedasticity and the unbalanced design, we propose a different test statistic, $F_C^*$ as defined in (10), and then establish its asymptotic distribution, both under the null and alternative hypotheses. Simulation results demonstrate the accuracy of the proposed test procedure as $r \to \infty$ and $min(c_i) \to \infty$.

# References

Akritas, M. G., and Arnold, S. F. (2000), "Asymptotics for Analysis of Variance when the Number of Levels is Large", *Journal of the American Statistical Association*, **95**, 212–226.

Akritas, M. G., and Papadatos, N. (2004), "Heteroscedastic One-Way ANOVA and Lack-of-Fit tests", *Journal of the American Statistical Association*, **99**, 368–382.

Arnold, S. F. (1980), "Asymptotic Validity of F Tests for the Ordinary Linear Model and the Multiple Correlation Model", *Journal of the American Statistical Association*, **75**, 890–894.

Liao, S.-M., and Akritas, M. G. (2008), "Homoscedastic Balanced Two-fold Nested Model when the number of sub-classes is large", *Recent Advances In Linear Models and Related Areas*, Springer, 279–292.

Wang, L., and Akritas, M. G. (2006), "Two-way Heteroscedastic ANOVA when the Number of Levels is Large", *Statistica Sinica*, **16**, 1387–1408.

# Poster 17

# Nonparametric $k$-sample tests based on kernel density estimators

**Pablo Martínez-Camblor**
*CIBER de Epidemiología y Salud Pública de Gipuzkoa*

**Jacobo de Uña-Álvarez**
*Department of Statistics and OR. University of Vigo*

**Abstract.** In this work new k-sample tests based on the comparison of kernel density estimators are introduced and investigated. The impact of the smoothing parameter in the performance (power) of the tests is discussed. An automatic bandwidth selector based on the bootstrap is proposed. The performance of the tests is illustrated in a simulated scenario. Comparison with other tests based on empirical distributions is included.

## 1 Introduction

The k sample problem is an old topic devoted to the comparison of k different populations independently sampled. Traditionally, parametric methods as the One-Way ANOVA F-test or nonparametric location tests such as the Kruskal-Wallis test have been used to that end. In the general nonparametric setup, omnibus k-sample tests were introduced *via* proper generalizations of Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling tests (see Kiefer, 1959 and Scholz and Stephens, 1987). However, such methods may behave poorly (low power) in practice, specially in problems in which the underlying populations have similar locations but different shapes (Zhang and Wu, 2007; Martínez-Camblor et al., 2008).

Comparison of the kernel estimators $f_{n_1}$,... ,$f_{n_k}$ pertaining to the k population densities (assumed to exist) is a possible alternative approach. For the 2-sample problem,

Anderson et al. (1994), Louani (2000), and Cao and Van Keilegom (2006) proposed and investigated such type of smooth tests. For the k-sample problem (with arbitrarily large k), a smooth test statistic was first proposed by Martínez-Camblor et al. (2008). The authors based the comparison on the area under the kernel density estimators which is shared by all of them. This "common area" statistic is explicitly defined by

$\mathcal{AC}_k = \int min\{f_{n_1}, ..., f_{n_k}\}$

In general, a small value of $\mathcal{AC}_k$ indicates that the populations may differ. This smooth test may behave much better than other tests based on the comparison of empirical distribution functions, in situations in which the k populations are different in their shapes (Martínez-Camblor et al., 2008). Besides, these authors proposed a data-driven bandwidth selector based on the bootstrap which, while being computationally feasible, leads to a quasi-optimal level of smoothing (in the sense of maximizing the power of the test).

In this work we introduce news test statistics based on the comparison of the kernel density estimators. The convergence to a normal under the null hypothesis is established for these new tests. In a simulation study, we illustrate the influence of the level of smoothing in the power of the tests. As for the "common area" measure, the bootstrap is used to define an automatic bandwidth selector. Finally, the different tests based on the automatic bandwidth are compared. We include in the comparison the Kruskal-Wallis test and the k-sample versions of the classical Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling tests, as well as three likelihood-ratio tests introduced by Zhang and Wu (2007).

## 2 New k-sample smooth tests

We introduce three new smooth test statistics by using the $L_1$, $L_2$ and $L_\infty$ measures,

$$L_{k,1} = \frac{1}{n}\sum_{i=1}^{k} n_i \int |f_{n_i}(t) - f_n(t)|dt,$$

$$L_{k,2} = \frac{1}{n}\sum_{i=1}^{k} n_i \int (f_{n_i}(t) - f_n(t))^2 dt,$$

$$S_k = \frac{1}{n}\sum_{i=1}^{k} n_i \sup_{t\in} |f_{n_i}(t) - f_n(t)|,$$

where $f_n$ stands for the kernel density estimator (e.g. Wand and Jones, 1995) computed from the pooled sample with $n = n_1 + ... + n_k$ observations. The asymptotic null distribution of the test statistics may be obtained following lines similar to those in Martínez-Camblor et al. (2008) for $\mathcal{AC}_k$; the involved arguments basically refer to Lemma 6 in Horváth (1991). In practice, we suggest using the bootstrap to evaluate the significance of the test. Note that $f_{n_i}$ and $f_n$ need a bandwidth or smoothing parameter. Let $h_{n_i}$ the bandwidth used for the estimator $f_{n_i}$; we take $h_{n_i} = S\widehat{\sigma}_{n_i} n_i^{-1/5}$ where $\widehat{\sigma}_{n_i}$ is the standard deviation of the $i$-th sample and $S$ is a parameter chosen (on a given grid) in order to maximize the power of the test. This is done by using the smoothed bootstrap in the following way. Put $T$ for any test statistic among $\mathcal{AC}_k$, $L_{k,1}$, $L_{k,2}$, and

$S_k$, and let $T_r$ denote the test statistic when using the specific value $S_r$ taken from the grid of $S$-values $\{S_1, ..., S_t\}$.

Step 1. Draw $B$ times $k$ independent resamples $\left\{x^*_{ij,b}\right\}_{j=1}^{n_i}$, $1 \leq i \leq k$, $b = 1, ..., B$, from $f_n$ (use a pilot bandwidth $g$ for $f_n$). Compute the bootstrap p-value of $T_r$: $p^*_r = \frac{1}{B}\sum_{b=1}^{B} I(T^*_{r,b} > T_r)$, where $T^*_{r,b}$ is the bootstrap version of $T_r$ (based on the $b$-th bootstrap resample)

Step 2. Compute the minimum p-value over the grid: $p^*_{\min} = \min\{p^*_1, ..., p^*_t\}$

Step 3. Draw new $B'$ bootstrap resamples as in Step 1 and compute

$$p^{**} = \frac{1}{B'}\sum_{b'=1}^{B'} I(p^{**}_{\min,b'} < p^*_{\min}),$$

where $p^{**}_{\min,b'} = \min\left\{p^{**}_{1,b'}, ..., p^{**}_{t,b'}\right\}$ and $p^{**}_{r,b'} = \frac{1}{B}\sum_{b=1}^{B} I(T^*_{r,b} > T^{**}_{r,b'})$

Step 4. Reject the null hypothesis (of equality of populations) if $p^{**} < \alpha$, where $\alpha$ is the nominal level of the test

This bandwidth selector, which we call *double minimum* bandwidth, was first used for the "common area" measure $\mathcal{AC}_k$, leading to powers close to the optimal (Martínez-Camblor et al., 2008). In this work we explore the performance of the double minimum bandwidth also for the other three test statistics.

# 3 Main conclusions

From the simulations performed so far, several interesting conclusions arise. First, tests based on the comparison of kernel density estimators are more powerful than their competitors (based on empirical distribution functions) at least when the underlying densities are different in their shapes. Among these smooth tests, the $L_{k,1}$ statistic shows the best overall performance. Secondly, the power of the smooth tests is strongly influenced by the bandwidth. The double minimum automatic bandwidth selector seems to provide a quasi-optimal power; at the same time, its computational cost is not as high as in other resampling plans recently proposed for related problems (Cao and Van Keilegom, 2006). Finally, the idea behind the double minimum bandwidth could be adapted to testing scenarios other than the k-sample problem, and this seems to be a promising field of research and applications.

# References

Anderson NH, Hall P & Titterington DM, (1994) Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *Journal of Multivariante Analysis*, **50**, 41–54.

Cao R & Van Keilegom I, (2006) Empirical likelihood tests for two-sample problems via nonparametric density estimation, *Canad. J. Statist.*, **34**, 61-77.

Horvath L, (1991) On $L_p$-norms of multivariate density estimations, *Annals of Statistics,* **19**, 4, 1933–1949.

Kiefer J, (1959) $k$-Sample analogues of the Kolmogorov-Smirnov, Cramér-von Mises test, *Ann. Math. Statist.*, **30**, 420–447.

Martínez-Camblor P, de Uña-Álvarez J & Corral N, (2008) $k$-Sample test based on the common area of kernel density estimator, *Journal of Statistical Planning and Inference*, doi:10.1016/j.jspi.2008.02.008.

Scholz FW & Stephens MA, (1987) $k$-Samples Anderson-Darling test. *J. Amer. Statist. Assoc.* **82**, 918–924.

Wand, MP & Jones, MC (1995) Kernel Smoothing. Chapman & Hall.

Zhang J & Wu Y, (2007) $k$-Sample test based on the likelihood ratio, *Computational Statistics & Data Analysis*, **51**, 9, 4682–4691.

# Poster 18

# Lifetime comparisons with early termination of experiments

**Tahani A. Maturi**[1]
*Dept of Mathematical Sciences, Durham University, Durham DH1 3LE, UK*

**Pauline Coolen-Schrijner**[2]
*Dept of Mathematical Sciences, Durham University, Durham DH1 3LE, UK*

**Frank P.A. Coolen**
*Dept of Mathematical Sciences, Durham University, Durham DH1 3LE, UK*

**Abstract.** We consider lifetime experiments to compare units from different groups, which can be ended before all units have failed. Units' lifetimes may also be right-censored at any time. Nonparametric predictive inference for selecting the best group is presented, with uncertainty quantified by lower and upper probabilities, and with attention to the effect of early termination of the experiment.

## 1 Introduction

We consider comparison of lifetimes of units from different groups, simultaneously placed on an experiment, and we focus on the effect of early termination of the experiment before event times of all units have been observed, where 'event' is either the 'failure' of interest or right-censoring, with the censoring mechanism assumed to be independent of the failure process. This scenario occurs in 'precedence testing' (Balakrishnan and Ng (2006)). We present Nonparametric Predictive Inference (NPI) (Coolen (2006)) for

[1]I am grateful to Pauline Coolen-Schrijner for her guidance towards my independent research career. Pauline, you were my supervisor and my lovely friend.

[2]Pauline died on 23.4.2008, this paper is dedicated to her.

such situations, with uncertainty quantified by lower and upper probabilities for events that compare the failure times of one further unit from each group. Lower and upper probabilities generalize classical probabilities, and a lower (upper) probability for event $A$, denoted by $\underline{P}(A)$ ($\overline{P}(A)$), can be interpreted in several ways (Coolen (2006)): as supremum buying (infimum selling) price for a gamble on the event $A$, or as the maximum lower (minimum upper) bound for the probability of $A$ that follows from the assumptions made. Informally, $\underline{P}(A)$ ($\overline{P}(A)$) can be considered to reflect the evidence in favour of (against) event $A$.

## 2 Main results

Nonparametric predictive inference (NPI) is based on Hill's assumption $A_{(n)}$ (Hill (1968)), which implies direct (lower and upper) probabilities for a future observable random quantity, based on observed values of $n$ related random quantities (Coolen (2006)). NPI is suitable if there is little knowledge about random quantities of interest, other than the $n$ observations, or if one does not want to use such information. Coolen and Yan (2004) presented rc-$A_{(n)}$ as a generalization of $A_{(n)}$ for right-censored data, using the extra assumption that, at a moment of censoring, the residual time-to-failure of a right-censored unit is exchangeable with the residual time-to-failure of all other units that have not yet failed or been censored.

We consider a life-testing experiment to compare units of $k \geq 2$ groups, which are assumed to be fully independent, with the experiment starting on all units at time 0. The experiment can be terminated before all units have failed, say at time $T_0$, which is assumed not to hold any information on residual time-to-failure for units that have not yet failed. We also allow right-censoring to occur before the experiment is stopped, due to a censoring process that is independent of the failure process. So we consider both right-censored observations in the original data and right-censoring due to stopping the experiment at $T_0$. For group $j$, $j = 1, \ldots, k$, $n_j$ units are in the experiment, of which $u_j$ units fail before (or at) $T_0$, with ordered failure times $0 < x_{j,1} < x_{j,2} < \ldots < x_{j,u_j} \leq T_0$, and $c_{j,1} < c_{j,2} < \ldots < c_{j,v_j} < T_0$ are right-censoring times (we assume no tied observations for ease of notation, generalization is straightforward by considering limits, and the example in Section 3 includes ties). Let $x_{j,0} = 0$ and $x_{j,u_j+1} = \infty$ ($j = 1, \ldots, k$). Let $s_{j,i_j}$ be the number of right-censored observations in the interval $(x_{j,i_j}, x_{j,i_j+1})$, with $x_{j,i_j} < c_{j,1}^{i_j} < c_{j,2}^{i_j} < \ldots < c_{j,s_{j,i_j}}^{i_j} < x_{j,i_j+1}$ and $\sum_{i_j=0}^{u_j} s_{j,i_j} = v_j$, so $n_j - (u_j + v_j)$ units from group $j$ are right-censored at $T_0$.

To compare these $k$ groups, we consider a hypothetical further unit from each group which would also have been involved in this experiment, with $X_{j,n_j+1}$ the random failure time for the further unit from group $j$, assumed to be exchangeable with the failure times of the $n_j$ units of the same group included in the experiment. The assumption rc-$A_{(n_j)}$ implies the lower and upper probabilities presented below, which are optimal bounds under the assumptions made (justifications will be presented elsewhere). We restrict attention to the events $X_{l,n_l+1} = \max_{1 \leq j \leq k} X_{j,n_j+1}$, for $l = 1, \ldots, k$. The lower and

upper probabilities below are specified with the use of the following functions:

$$M_{i_j}^j = M_{X_{j,n_j+1}}(x_{j,i_j}, x_{j,i_j+1}) = \frac{1}{n_j+1} \prod_{\{r:c_r<x_{j,i_j}\}} \frac{\tilde{n}_{j,c_r}+1}{\tilde{n}_{j,c_r}}$$

$$M_{i_j,t_j}^j = M_{X_{j,n_j+1}}(c_{j,t_j}^{i_j}, x_{j,i_j+1}) = \frac{1}{(n_j+1)}(\tilde{n}_{j,c_{j,t_j}^{i_j}})^{-1} \prod_{\{r:c_r<c_{j,t_j}^{i_j}\}} \frac{\tilde{n}_{j,c_r}+1}{\tilde{n}_{j,c_r}}$$

$$M_{T_0}^j = M_{X_{j,n_j+1}}(T_0, \infty) = \frac{n_j-(u_j+v_j)}{n_j+1} \prod_{\{r:c_r<T_0\}} \frac{\tilde{n}_{j,c_r}+1}{\tilde{n}_{j,c_r}}$$

where $i_j = 0, \ldots, u_j$, $t_j = 1, \ldots, s_{j,i_j}$, and $\tilde{n}_{j,c_r}$ and $\tilde{n}_{j,c_{j,t_j}^{i_j}}$ are the number of units from group $j$ in the risk set just prior to time $c_r$ and $c_{j,t_j}^{i_j}$, respectively. Also

$$P_{i_j}^j = P(X_{j,n_j+1} \in (x_{j,i_j}, x_{j,i_j+1})) = \frac{1}{n_j+1} \prod_{\{r:c_r<x_{j,i_j+1}\}} \frac{\tilde{n}_{j,c_r}+1}{\tilde{n}_{j,c_r}}$$

The lower probabilities are

$$\underline{P}^{(l)} = \underline{P}\left(X_{l,n_l+1} = \max_{1\le j\le k} X_{j,n_j+1}\right) = \sum_{i_l=0}^{u_l} \left\{ \prod_{\substack{j=1 \\ j\neq l}}^{k} \left[ \sum_{i_j=0}^{u_j} 1(x_{j,i_j+1} < x_{l,i_l}) P_{i_j}^j \right] M_{i_l}^l \right.$$

$$\left. + \sum_{t_l=1}^{s_{l,i_l}} \prod_{\substack{j=1 \\ j\neq l}}^{k} \left[ \sum_{i_j=0}^{u_j} 1(x_{j,i_j+1} < c_{l,t_l}^{i_l}) P_{i_j}^j \right] M_{i_l,t_l}^l \right\} + M_{T_0}^l \prod_{\substack{j=1 \\ j\neq l}}^{k} \sum_{i_j=0}^{u_j} 1(x_{j,i_j+1} < T_0) P_{i_j}^j$$

and the upper probabilities are

$$\overline{P}^{(l)} = \overline{P}\left(X_{l,n_l+1} = \max_{1\le j\le k} X_{j,n_j+1}\right) = \sum_{i_l=0}^{u_l} P_{i_l}^l \prod_{\substack{j=1 \\ j\neq l}}^{k} \left\{ \sum_{i_j=0}^{u_j} 1\left(x_{j,i_j} < x_{l,i_l+1}\right) M_{i_j}^j \right.$$

$$\left. + \sum_{i_j=0}^{u_j} \sum_{t_j=1}^{s_{j,i_j}} 1(c_{j,t_j}^{i_j} < x_{l,i_l+1}) M_{i_j,t_j}^j + 1\left(T_0 < x_{l,i_l+1}\right) M_{T_0}^j \right\} + M_{T_0}^l$$

Note that $T_0$ influences the lower and upper probabilities only through the $u_j$. If $u_l = 0$ then $\overline{P}^{(l)} = 1$, while if $u_j = 0$ for all $j \neq l$ then $\underline{P}^{(l)} = 0$. If the experiment is terminated before a single unit has failed, then $\underline{P}^{(l)} = 0$ and $\overline{P}^{(l)} = 1$ for all groups. These extreme cases illustrate an attractive feature of lower and upper probabilities in quantifying the strength of statistical information, in an intuitive manner that is not possible with precise probabilities. If $T_0$ increases, $\underline{P}^{(l)}$ never decreases and $\overline{P}^{(l)}$ never increases, and they can only change if further events are observed.

If the experiment is ended when all units have been observed (whether the units have been failed or observed as right censoring), the terms including $T_0$ in the above (lower and upper probabilities) formulas will vanish.

# 3 Example

We use data from Desu and Raghavarao (2004)[p.263], representing the recorded times (months) until promotion at a large company, for 19 employees in $k = 3$ departments ('groups'): for group 1: 15, 20+, 36, 45, 58, 60 ($n_1 = 6$), for group 2: 12, 25+, 28, 30+, 30+, 36, 40, 45, 48 ($n_2 = 9$), for group 3: 30+, 40, 48, 50 ($n_3 = 4$), where "+" indicates that the employee left the company at that length of service before getting promotion (right-censoring). We consider at which department one needs to work the longest to get a promotion. In this example the 'best group', in terminology from Section 2, actually represents the worst department for promotions. This data set contains tied observations, in NPI these are dealth with by assuming that they differ a very small amount, such that the lower (or upper) probability of interest is smallest (largest), which is an attractive manner for dealing with ties.

For some ranges of values of $T_0$ the lower probabilities $\underline{P}^{(l)}$ and upper probabilities $\overline{P}^{(l)}$, for $l = 1, 2, 3$, are presented in Table 1. At any value of $T_0$, we can indicate that we have a strong evidence that group $l$ is the best if $\underline{P}^{(l)} > \overline{P}^{(j)}$ for all $j \neq l$, which is not the case in this example. For larger values of $T_0$, such that most units have been observed, group 3 has most imprecision remaining, reflecting that there are only few observations for group 3.

| $T_0$ | $u_1$ | $u_2$ | $u_3$ | $\underline{P}^{(1)}$ | $\overline{P}^{(1)}$ | $\underline{P}^{(2)}$ | $\overline{P}^{(2)}$ | $\underline{P}^{(3)}$ | $\overline{P}^{(3)}$ |
|---|---|---|---|---|---|---|---|---|---|
| 11 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| 17 | 1 | 1 | 0 | 0 | 0.8629 | 0 | 0.9029 | 0.0114 | 1 |
| 27 | 1 | 1 | 0 | 0 | 0.8629 | 0 | 0.9029 | 0.0114 | 1 |
| 33 | 1 | 2 | 0 | 0 | 0.8629 | 0 | 0.7974 | 0.0243 | 1 |
| 38 | 2 | 3 | 0 | 0 | 0.7140 | 0 | 0.6591 | 0.0887 | 1 |
| 42 | 2 | 4 | 1 | 0.0678 | 0.7140 | 0.0248 | 0.5398 | 0.1135 | 0.8332 |
| 47 | 3 | 5 | 1 | 0.0813 | 0.6148 | 0.0315 | 0.4341 | 0.1969 | 0.8332 |
| 52 | 3 | 6 | 3 | 0.2393 | 0.6148 | 0.0315 | 0.3542 | 0.2161 | 0.6618 |
| 59 | 4 | 6 | 3 | 0.2393 | 0.6148 | 0.0315 | 0.3542 | 0.2161 | 0.6618 |
| 61 | 5 | 6 | 3 | 0.2393 | 0.6148 | 0.0315 | 0.3542 | 0.2161 | 0.6618 |

**Table 1:** The best group: lower and upper probabilities.

# References

Balakrishnan, N. and Ng, H.K.T. (2006). *Precedence-Type Tests and Applications.* John Wiley & Sons.

Coolen, F.P.A. (2006). On nonparametric predictive inference and objective Bayesianism. *Journal of Logic, Language and Information*, *15*, 21-47.

Coolen, F.P.A. and Yan, K.J. (2004). Nonparametric predictive inference with right-censored data. *Journal of Statistical Planning and Inference*, *126*, 25–54.

Desu, M. M. and Raghavarao, D. (2004). *Nonparametric Statistical Methods for Complete and Censored Data.* Boca Raton, Chapman & Hall/CRC.

Hill, B.M. (1968): Posterior distribution of percentiles: Bayes theorem for sampling from a finite population. *Journal of the American Statistical Association*, *63*, 677–691.

# Poster 19

# The multiresolution criterion - a geometric interpretation

**Thoralf Mildenberger** [1]
*Fakultät Statistik, TU Dortmund*

**Abstract.** One approach to choosing the amount of smoothing in nonparametric regression is to select the simplest estimate for which the residuals "look like white noise". This can be checked with the so-called multiresolution criterion introduced by Davies and Kovac (2001). We show that this criterion is related to a norm, the *multiresolution norm* (MR-norm), and discuss some remarkable properties.

## 1 The multiresolution criterion

Consider the nonparametric regression model

$$y(t_i) = f(t_i) + \varepsilon(t_i) \qquad (i = 1, ..., N) \tag{1}$$

with fixed design points $0 \leq t_1 < \cdots < t_N \leq 1$, where $\varepsilon(t_1), \ldots, \varepsilon(t_N)$ are i.i.d. normally distributed with variance $\sigma^2$. Methods for estimating $f$ usually require the choice of a smoothing parameter. A recent approach to this, introduced by Davies and Kovac (2001) in connection with their taut-string method, is to choose the simplest estimate for which the residuals "look like white noise". We will consider the vector of evaluations of some estimate $\hat{f}$ at the design points $(\hat{f}(t_1), \ldots, \hat{f}(t_N))$, which we will also denote by $\hat{f}$. Similarly, $y := (y(t_1), \ldots, y(t_N))$ denotes the vector of observations and $r = (\hat{f}(t_1) - y(t_1), \ldots, \hat{f}(t_N) - y(t_N))$ the vector of residuals. The residuals are now examined using the *multiresolution criterion* (Davies and Kovac 2001):

$$\max_{I \in \mathcal{I}} \frac{1}{\sqrt{|I|}} \left| \sum_{t \in I} r_t \right| \leq C, \tag{2}$$

[1] Research supported by SFB 475 "'Reduction of Complexity in Multivariate Data Structures"' funded by the German Research Foundation (DFG).

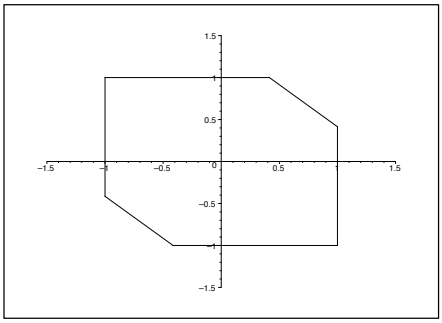


**Figure 1:** The MR-norm unit ball in $\mathbf{R}^2$.

where $\mathcal{I} = \{I | I = \{i, i+1, \ldots, l\}, 1 \leq i \leq l \leq N\}$ is the system of all discrete intervals in $\{1, \ldots, N\}$. $|I|$ denotes the number of points in $I$ and $C$ a suitably chosen constant. The criterion (2) captures the idea that residuals should behave like white noise, that is they should not be too large, nor should they exhibit any structures that indicate that $f$ has not been estimated properly. Since sums of residuals over intervals are evaluated, longer runs of residuals of the same sign make (2) large, even if they are moderate in size.

**Definition 1** *Let $N \in \mathbf{N}$ and $\mathcal{I} = \{I | I = \{i, i+1, \ldots, l\}, 1 \leq i \leq l \leq N\}$. The* multiresolution norm *(*MR-norm *for short) on $\mathbf{R}^N$ is defined by:*

$$\|(x_1, \ldots, x_N)\|_{\mathsf{MR}} := \max_{I \in \mathcal{I}} \frac{1}{\sqrt{|I|}} \left| \sum_{t \in I} x_t \right|.$$

The unit ball in $\mathbf{R}^2$ is shown in Figure 1.

It is clear that the residuals satisfy (2) iff $\|r\|_{\mathsf{MR}} \leq C$, i.e. they are contained in a ball around the origin or - equivalently - $\hat{f}$ is contained in a ball around the data $y$. The ability to detect structure in the residuals is mainly due to the fact that the MR-norm lacks basic invariance properties that e.g. $p$-norms do possess. In the following, we highlight some of these (Mildenberger 2008):

Consider the $p$-norms defined by

$$\|(x_1, \ldots, x_N)\|_p = \begin{cases} \left( \sum_{t=1}^N |x_t|^p \right)^{1/p} & (1 \leq p < \infty) \\ \max\{|x_1|, \ldots, |x_N|\} & (p = \infty) \end{cases}.$$

Let $\mathcal{S}_n$ denote the group of permutations of $\{1, \ldots, N\}$ with

$$\pi x := (x_{\pi(1)}, \ldots, x_{\pi(N)}) \text{ for } \pi \in \mathcal{S}_n, x \in \mathbf{R}^N.$$

Furthermore, we consider the sign group $\mathcal{T}_n := \{-1, +1\}^N$ with component-wise multiplication. We define:

$$sx := (s_1 x_1, \ldots, s_n x_n) \text{ for } s \in \mathcal{T}_n, x \in \mathbf{R}^N.$$

From the definition of $\|\cdot\|_p$ it is clear that $\|\pi x\|_p = \|x\|_p$ and $\|sx\|_p = \|x\|_p$ for all $x \in \mathbf{R}^N, \pi \in \mathcal{S}_N, s \in \mathcal{T}_n$ and $p \in [1, \infty]$, since the $p$-norms depend only on the absolute values of the components. The multiresolution norm of a vector is in general not invariant under these transformations. Consider

$$\begin{aligned} \|(1,-1,1)\|_{\mathsf{MR}} &= 1 \\ \|(1,1,-1)\|_{\mathsf{MR}} &= \sqrt{2}. \end{aligned}$$

This represents a counter-example for both transformations as the second vector can be obtained from the first one by either exchanging the second and third components or by changing their signs.

Let $\mathsf{id}$ denote the identity of $\mathcal{S}_N$ or $\mathcal{T}_N$, respectively. Let $\rho \in \mathcal{S}_N$ denote the reverse-ordering-permutation such that

$$\rho(x_1, \ldots, x_N) = (x_N, \ldots, x_1)$$

for all $x \in \mathbf{R}^N$. Furthermore, let $\nu := (-1, \ldots, -1)$ denote the element of $\mathcal{T}_N$ that flips all signs simultaneously. The MR-norm is invariant under these transformations, and this is a desired property. There are no other transformations in these groups that leave the multiresolution norm invariant for all $x \in \mathbf{R}^N$:

**Theorem 2** *1. For $\pi \in \mathcal{S}_N$, $\|\pi x\|_{\mathsf{MR}} = \|x\|_{\mathsf{MR}}$ for all $x \in \mathbf{R}$ iff $\pi = \mathsf{id}$ or $\pi = \rho$.*

*2. For $s \in \mathcal{T}_N$, $\|sx\|_{\mathsf{MR}} = \|x\|_{\mathsf{MR}}$ for all $x \in \mathbf{R}$ iff $s = \mathsf{id}$ or $s = \nu$.*

The dependence of $\|x\|_{\mathsf{MR}}$ on the sign pattern of $x$ shows up in particular when considering vectors that consist of components with the same absolute size:

**Proposition 3** *Consider the set of all $x = (x_1, \ldots, x_N)$ with $|x_1| = \cdots = |x_N| =: m > 0$.*

1. *$\|x\|_{\mathsf{MR}}$ is maximal iff all components have the same sign. Then $\|x\|_{\mathsf{MR}} = \sqrt{N}m$.*
2. *$\|x\|_{\mathsf{MR}}$ is minimal iff the signs are alternating. Then $\|x\|_{\mathsf{MR}} = m$.*
3. *$\|x\|_{\mathsf{MR}} \geq \sqrt{\ell}m$, where $\ell$ is the length of the longest run of components with the same sign.*

# References

Davies, P.L. and Kovac, A. (2001). Local extremes, runs, strings and multiresolution (with discussion and rejoinder). *Annals of Statistics*, **29**, 1-65.

Mildenberger, T. (2008). A geometric interpretation of the multiresolution criterion in nonparamteric regression. Revised version submitted to *Journal of Nonparametric Statistics*.

# Poster 20

# Bootstrapping the NPMLE for doubly truncated data

**Carla Moreira**[1]
*Department of Statistics and OR, University of Vigo*

**Jacobo de Uña-Álvarez**[1]
*Department of Statistics and OR, University of Vigo*

**Abstract.** In this paper we propose the simple bootstrap as a method to approximate the distribution of the Efron-Petrosian nonparametric maximum likelihood estimator for doubly truncated data. We use the bootstrap to construct pointwise confidence bands for the lifetime distribution. Simulations and application to real data are included.

## 1 Introduction

Truncated data appear in a number of fields, including Astronomy, Economics and Survival Analysis. In a seminal paper, Turnbull (1976) introduced a substitute for the ordinary empirical distribution function for arbitrarily truncated data, considering also the case of grouping and censoring. Later, statistical methods were more deeply investigated for specific situations involving truncation. For example, nonparametric techniques for left-truncated data were developed in the eighties and early nineties (Woodroofe, 1985; Stute, 1993), and then they were properly adapted to the presence of right-censoring (Tsai, Jewell and Wang, 1987; Zhou and Yip, 1999). These methods can be used (after a suitable re-definition of the variables) to deal also with data which are truncated from the right, and so the problem of one-sided truncation is well understood. Literature is much more scarce however for two-sided truncation.

[1]We thank Rosa Crujeiras for her assistance with the programming in R. Work supported by the research Grant MTM2005-01274 (FEDER funding included) of the Spanish Ministerio de Educación y Ciencia, and by the Grant PGIDIT07PXIB300191PR of the Xunta de Galicia.

Assume that one is interested in a positive random variable (or lifetime) $X$ which is observed if and only if $T_0 \leq X \leq T$, where $(T_0, T)$ is a pair of truncation times assumed to be independent of $X$. The sample information is represented by $(T_{0i}, X_i, T_i)$, $i = 1, ..., n$, iid data with the same distribution as $(T_0, X, T)$ given $T_0 \leq X \leq T$. Examples of this situation arise in Astronomy (Efron and Petrosian, 1999) and in medical sciences (Navarro and Ruiz, 1996), among other fields. The nonparametric maximum likelihood estimator (NPMLE) of the distribution function (df) $F$ of $X$ was introduced in Efron and Petrosian (1999). However, for the best of our knowledge, no much formal theory is available for this NPMLE, which is implicitly defined as the maximizer of a conditional likelihood (see Section 2). The application of bootstrap methods seems to be promising in this area where little is known about the convergence properties of the estimator.

In Section 2, a bootstrap method for approximating the finite-sample distribution of the Efron-Petrosian NPMLE is introduced. This bootstrap is then used to defined a pointwise confidence band for the df of the lifetime. We illustrate the performance of the bootstrap in a small simulation study. Section 3 is an application of the proposed methods to some data coming from a recent study on childhood cancer in Portugal (Moreira and de Uña-Álvarez, 2007).

## 2 The bootstrapped Efron-Petrosian NPMLE

Following Efron and Petrosian (1999), put (for $i = 1, ..., n$) $F_i = P(T_{0i} \leq X \leq T_i \mid T_{0i}, T_i)$ and $f_i = P(X = X_i \mid T_{0i}, T_i)$. Note that, given the $(T_{0i}, T_i)$'s, the contribution of $X_i$ to the conditional likelihood is given by $f_i/F_i$. Starting with a initial estimator for $f_i$ (*e.g.* $f_i = 1/n$ for $i = 1, ..., n$), the EM-algorithm suggested by Efron and Petrosian (1999) proceeds by updating the estimator via the maximum likelihood equations (see formula (25) in their paper), until the convergence criterion is reached. We put $F_n$ for the Efron-Petrosian NPMLE, which is a step function with jump $f_i$ at the observed lifetime $X_i$. Note that, unlike for one-sided truncation, doube truncation implies that both large and small values of $X$ are observed with a relatively small probability. In the left-truncated setup, for example, it is reasonable to expect $f_i \geq 1/n$ for small $X_i$'s at the optimum, and this will be indeed the case. However, with doubly truncated data such type of intuition disappears.

In order to approximate the distribution of the Efron-Petrosian NPMLE, we propose a simple bootstrap resampling plan (details about simple and obvious bootstrap resampling under left truncation are given in Gross and Lai, 1996). The plan is as follows: each datum $(T_{0i}, X_i, T_i)$ is re-sampled (with replacement) with a probability of $1/n$, until a resample of size $n$, say $\left\{\left(T_{0i}^{*1}, X_i^{*1}, T_i^{*1}\right), \quad i = 1, ..., n\right\}$, is constructed. The Efron-Petrosian estimator $F_n^{*1}$ is computed with this resample. The procedure is repeated a large number $B$ of times; then, the distribution of $F_n$ is approximated by the empirical distribution of $F_n^{*1}, ..., F_n^{*B}$. In order to construct a $100(1-\alpha)\%$ (pointwise) confidence interval for a specific value $F(x)$, as usual we take the $B\alpha/2$ and the $B(1-\alpha/2)$ empirical percentiles of $F_n^{*1}(x), ..., F_n^{*B}(x)$.

In order to illustrate the performance of the simple bootstrap, we have performed a small simulation study. The lifetime $X$ was simulated according to a $Uniform\,[0, 15]$. The right-truncation time $T$ followed a $Uniform\,[0, 20]$ model, and (given the $T$) the left-truncation time was computed as $T_0 = T - 5$. Then, the vector $(T_0, X, T)$ is maintained if $T_0 \leq X \leq T$, being rejected otherwise. This simulated scenario was designed in order

to imitate the real sampling problem in Section 3. Table 1 reports the results achieved by the 95% bootstrap confidence interval for the nine deciles of $X$ ($F(x) = 0.1, ..., 0.9$) when taking $B = 500$. The sample size was $n = 50$ and the number of trials was 500. The coverages were computed as the proportion of intervals (among the 500) containing the true value of the cumulative distribution. We also computed the mean length of the confidence intervals along the 500 replicates, as a measure of efficiency. Note that the intervals are wider around the median, similarly as for the ordinary empirical distribution. This suggests that the effects of the left and right truncation compensates each other in the simulated example. We have also considered other sample sizes: $n = 100$, 150, and 250. The results achieved by the bootstrap (not shown) indicate that the method could suffer from a serious bias at both tails of the distribution (deciles 0.1, 0.2, 0.8, and 0.9), with actual coverages decreasing to about 88% (or 75% in the extreme tails). More investigation is needed to find proper explanations and corrections for this issue.

| Deciles | Coverage | Mean Length |
|---|---|---|
| 1 | 0.932 | 0.1678888 |
| 2 | 0.934 | 0.2100830 |
| 3 | 0.958 | 0.2375419 |
| 4 | 0.952 | 0.2575485 |
| 5 | 0.946 | 0.2625792 |
| 6 | 0.954 | 0.2562005 |
| 7 | 0.944 | 0.2364633 |
| 8 | 0.938 | 0.2088095 |
| 9 | 0.936 | 0.1722716 |

**Table 1:** Coverages and mean length of Bootstrap Confidence bands along 500 trials from the simulated model. The sample size was 50, and we took B=500.

# 3 Application to real data

The childhood cancer data information was gathered from the IPO (Instituto Português de Oncologia) of Porto by RORENO service (Cancer Registry of the North Region of Portugal). Data include all children (i.e. people aged below 15 years old) diagnosed from cancer between 1999 until 2003, resident in any of the five districts in northern region of Portugal. Children were followed until April 30th 2006.

The data correspond to 409 children diagnosed from cancer, 180 female and 229 male. The most precocious diagnosis corresponded to a 6 days old baby, while the largest age at diagnosis was almost 15 years (Moreira and de Uña-Álvarez, 2007).Let X be the age at diagnosis, let T be the time from birth to end of recruitment (December 31st, 2003). Note that X is observed only when $T - 5 \leq X \leq T$, and hence the sampling of X is doubly truncated.

The Efron-Petrosian NPMLE of the age at diagnosis, together with a 95% bootstrap pointwise confidence band, is reported in Figure 1. For the bootstrap we took B=500. For comparison purposes, we have included in Figure 1 the ordinary empirical df of the ages at diagnosis. The similarity between this estimator and the Efron-Petrosian NPMLE suggests that the sampling bias due to double truncation is not very important in this case.

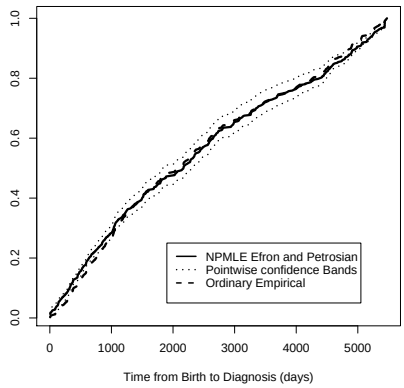


**Figure 1:** Efron-Petrosian NPMLE (continuous line) for the age at diagnosis of childhood cancer (North Potugal data) and 95% bootstrap pointwise confidence band (dotted lines). For comparison, the ordinary empirical distribution is also included (dashed line).

# References

Efron, B. and Petrosian, V. (1999) Nonparametric methods for doubly truncated data. Journal of the American Statistical Association 94, 824-834.

Gross, T. and Lai, T. L. (1996) Bootstrap methods for truncated and censored data. Statistica Sinica 6, 509-530.

Moreira, C. and de Uña-Álvarez, J. (2007) Childhood cancer in North Portugal: incidence, survival and methodological issues (M. I. Gomes, D. Pestana and P. Silva Eds.) Book of Abstracts of the 56th Session of the ISI.

Navarro, J. and Ruiz, J. M. (1996) Failure-rate functions for doubly-truncated random variables. IEEE Transaction on Reliability 45, 685-690.

Stute, W. (1993) Almost sure representation of the product-limit estimator for truncated data. Annals of Statistics 21, 146-156.

Tsai, W.-Y., Jewell, N. P. and Wang, M.-C. (1987) A note on the product-limit estimator under right censoring and left truncation. Biometrika 74, 883-886.

Turnbull, B. W. (1976) The empirical distribution function with arbitrarily grouped, censored and truncated data. Journal of the Royal Statistical Society Series B 38, 290-295.

Woodroofe, M. (1985) Estimating a distribution function with truncated data. Annals of Statistics 13, 163-177.

Zhou, Y. and Yip, P. S. F. (1999) A strong representation of the product-limit estimator for left truncated and right censored data. Jounral of Multivariate Analysis 69, 261-280.

# Poster 21

# $U$-statistical Kolmogorov-type tests: large deviations and asymptotic efficiency

**Nikitin Ya.Yu.**[1]
*Department of Mathematics and Mechanics. St.Petersburg University*

**Abstract.** We consider $U$-statistical analogues of Kolmogorov-Smirnov statistics for goodness-of-fit and symmetry testing and describe their large deviations. This enables to calculate their efficiencies and study the conditions of local asymptotic optimality.

Let $X_1, \ldots, X_n$ be i.i.d. observations with continuous d.f. $F$, and let $F_n$ be their empirical d.f. Consider the Kolmogorov statistic $D_n = \sup_t |F_n(t) - F(t)|$. Large deviations (LD) of $D_n$ were described long ago in Abrahamson (1967).

**Theorem 1** *For any $a \in (0,1)$ it holds true that*

$$\lim_{n\to\infty} n^{-1} \ln P(D_n > a) = f_0(a),$$

*where the function $f_0$ is continuous on $(0,1)$, and $f_0(a) = -2a^2(1+o(1)),\ a \to 0$.*

This result turned out to be very useful when calculating the Bahadur efficiency of numerous variants of Kolmogorov-Smirnov tests, see Bahadur (1971).

We are interested in $U$-statistical generalizations of Theorem 1 and in their applications to testing. Let $h(x_1, ..., x_m)$ be a real symmetric kernel of degree $m \geq 1$. Consider the $U$-statistical e.d.f.

$$G_n(t) = \binom{n}{m}^{-1} \sum_{1\leq i_1<\ldots<i_m\leq n} \mathbf{I}\{h(X_{i_1}, ..., X_{i_m}) < t\}, \quad t \in R^1.$$

[1]Supported by RFBR grants No. 06-01-00179 and 07-01-00159.

Let $G(t) := P(h(X_1, ..., X_m) < t)$ be continuous. Then the $U$-statistical generalization of Kolmogorov statistic $D_n$ is the statistic $DU_n = \sup_t |G_n(t) - G(t)|$. Many goodness-of-fit and symmetry tests are also based on the Smirnov-type statistic $SU_n = \sup_t |G_n(t) - F_n(t)|$ under appropriate choice of the kernel $h$.

Consider the following example. Suppose we are testing exponentiality using the Desu's characterization, see Desu(1971): *Let $X$ and $Y$ be independent non-negative rv's having d.f. $F$. Then the rv's $2\min(X, Y)$ and $X$ are equidistributed iff $F(x) = 1 - \exp(-\lambda x), x \geq 0$ with some $\lambda > 0$.*

Let compare the $U$-statistical e.d.f.

$$\bar{G}_n(t) = \binom{n}{2}^{-1} \sum_{1 \leq j < k \leq n} \mathbf{I}\{2\min(X_j, X_k) < t\}$$

with the usual e.d.f. $F_n(t)$, considering the statistic $DE_n = \sup_t |\bar{G}_n(t) - F_n(t)|$ and assuming that its large values are critical. The limiting distribution of $\sqrt{n}DE_n$ is non-normal but one can find the critical points via simulation. If we know the logarithmic LD asymptotics of $DU_n$, then we can calculate its Bahadur efficiency after relatively simple calculations, see Bahadur (1971) and Nikitin (1995).

The description of LD asymptotics for $U$-statistics for the time being is studied insufficiently. The LD principle established in Eichelsbacher and Löwe (1995) is not sufficient for statistical applications. In Nikitin and Ponikarov (1999) the explicit form of LD's was found for bounded non-degenerate and weakly degenerate kernels. However, the particularity of $U$-statistical Kolmogorov-Smirnov tests consists in that we have to do with the *supremum of a family* of $U$-statistics.

When studying the LD asymptotics of statistics $DU_n$ and $SU_n$ we can limit oneself to one-sided statistics $DU_n^{\pm}$ and $SU_n^{\pm}$. Each of the one-sided statistics is a supremum in $t \in T$, where $T$ is some interval in $R^1$, of a family of $U$-statistics with a kernel $\Psi(\cdot\ ;\ t)$ depending on $t$. For instance, in the case of statistic $DU_n^+$ we have

$$\Psi(x_1, \ldots, x_m; t) = \mathbf{I}\{h(x_1, ..., x_m) < t\} - G(t), \quad t \in R^1,$$

while for the statistic $SU_n^+$ one has for $t \in R^1$

$$\Psi(x_1, \ldots, x_m; t) = \mathbf{I}\{h(x_1, ..., x_m) < t\} - m^{-1}(\mathbf{I}\{x_1 < t\} + \cdots + \mathbf{I}\{x_m < t\}).$$

For each $t$ the kernels $\Psi(\cdot\ ; t)$ are centered and bounded. Consider their projections

$$\psi(s; t) = E(\Psi(X_1, ..., X_m; t) | X_1 = s),$$

their variance functions $\sigma_\psi^2(t) = E\psi^2(X_1; t)$ and the maxima of these variance function $\psi_0^2 := \sup_t \sigma_\psi^2(t)$. Let introduce the following condition of non-degeneracy of these families.

*We call the family of $U$-statistics $\{U_n(t), t \in T\}$ with the family of kernels $\Psi(\cdot\ ; t)$ non-degenerate, if its variance function $\sigma_\psi^2(t) = E\psi^2(X_1; t)$ is positive for any $t \in T$ with possible exception of the ends of the interval $T$.*

For instance, in the case of the statistic $DE_n^+$ the variance function is $\frac{1}{4}\exp(-t)(1 - \exp(-t)), t \geq 0$, while $\psi_0^2 = 1/16$.

We impose also the following *monotonicity condition* on the family $U_n(t)$. Let suppose that there exists the partition of the interval $T$ in $N$ parts $t_0 < t_1 < \cdots < t_N$, so that for any $k = 0, \ldots, N-1$

$$\sup_{t_k \le t < t_{k+1}} U_n(t) \le U_n(t_{k+1}) + \Delta_n(N),$$

where the rv's $\Delta_n(N)$ decrease fastly as $n$ and $N$ grow in the following sense: there exists such sequence $\{\tau_N\}, \tau_N \to \infty$ as $N \to \infty$, that

$$\lim_{N\to\infty} \lim_{n\to\infty} n^{-1} \ln P(\Delta_n(N) > \tau_N) = -\infty.$$

It is easy to prove that the statistics $DU_n^{\pm}$ and $SU_n^{\pm}$ satisfy the condition of monotonicity.

**Theorem 2** *Suppose that the family of $U$-statistics $U_n(t)$ is based on centered, bounded and non-degenerate kernels and satisfies the monotonicity condition. Then there exists such continuous function $v_0$ that for sufficiently small $a > 0$ one has*

$$\lim_{n\to\infty} n^{-1} \ln P(\sup_t U_n(t) > a) = v_0(a).$$

*Moreover, as $a \to 0$, one has $v_0(a) = -a^2/2m^2\psi_0^2 + O(a^3)$.*

The proof is involved and uses variational methods related to Sanov theorem and some ideas from nonlinear functional analysis together with exponential inequalities for the tails of $U$-statistics.

Consider some examples and statistical applications.

1. *Desu test of exponentiality.* It follows easily from Theorem 2 that for some continuous function $v_1$ one has

$$\lim_{n\to\infty} n^{-1} \ln P(DE_n > a) = v_1(a) = -2a^2 + O(a^3), a \to 0. \tag{1}$$

2. One more characterization of exponentiality belongs to Puri and Rubin (1970): *let $X$ and $Y$ be independent rv's with the same absolutely continuous d.f. $F$ on $R^+$. Then d.f. $F$ is exponential iff $|X - Y|$ is distributed as $X$.*

$U$-statistical Kolmogorov-type test statistic $PR_n$ corresponds to the family of $U$-statistics with the kernel

$$\Psi(x_1, x_2; t) = \mathbf{I}\{|x_1 - x_2| < t\} - \frac{1}{2}(\mathbf{I}\{x_1 < t\} + \mathbf{I}\{x_2 < t\}),\ t \ge 0,$$

while the variance function has the form $\sigma_\psi^2(t) = \frac{1}{12}e^{-t}(1 + e^{-t} - 2e^{-2t})$, $t \ge 0$, with the maximum $\frac{10+7\sqrt{7}}{648}$. Hence there exists such continuous function $v_2$, that

$$\lim_{n\to\infty} n^{-1} \ln P(PR_n > a) = v_2(a) = -\frac{7\sqrt{7} - 10}{3} a^2 + O(a^3),\ a \to 0.$$

Next example is connected with testing of symmetry.

3.Let use the characterization of symmetry due to Baringhaus and Henze (1992): *the d.f.*

*of two i.i.d. rv's $X$ and $Y$ is symmetric with respect to zero iff $|X|$ and $|\max(X,Y)|$ are equally distributed.*

Let $L_n(x) = n^{-1}\sum_{j=1}^{n}\mathbf{I}(|X_j| \le x)$, and let

$$G_n^*(x) = \binom{n}{2}^{-1} \sum_{1\le j<k\le n} \mathbf{I}(|\max(X_j, X_k)| \le x),\ x \ge 0.$$

Following to Baringhaus and Henze (1992), consider the statistic

$$BH_n = \sup_{x\ge 0} |L_n(x) - G_n^*(x)|$$

which is fit for the theory developed above. Hence for some continuous function $v_3$ we have

$$\lim_{n\to\infty} n^{-1}\ln P(BH_n \ge a) = v_3(a) = -\frac{27}{8}a^2 + O(a^3),\ a \to 0.$$

Similarly are studied the statistic for testing of exponentiality based on the simplified "loss-of-memory" property from Angus (1982), the statistic for symmetry testing from Abbakumov and Nikitin (1993), the statistic for testing of normality based on Polya characterization, and some others.

$U$-statistical Kolmogorov-Smirnov tests are consistent against a very broad class of alternatives. In the same time, their local Bahadur efficiency for standard alternatives is not very high and is usually inferior to integral tests. Simulation results support this observation.

We are able to describe the special set of alternatives for which the considered tests are locally asymptotically optimal in the Bahadur sense using the variational techniques developed in Nikitin (1995). For such alternatives the use of $U$-statistical Kolmogorov-Smirnov tests is just well-founded.

# References

Abbakumov, V. L., Nikitin, Ya. Yu. (1993). Bahadur efficiency and local optimality of a new nonparametric test of symmetry. In: *Statistics and Probability. A R.R. Bahadur Festschrift*, Wiley Eastern Ltd., New Delhi, 1-12.

Abrahamson, I. G. (1967). Exact Bahadur efficiencies for the Kolmogorov - Smirnov and Kuiper one- and two-sample statistics. *Ann. Math. Stat.*, **38**, 1475–1490.

Angus, J. E. (1982). Goodness-of-fit tests for exponentiality based on a loss-of-memory type functional equation. *J. Statist. Planning Infer.*, **6**, 241-251.

Bahadur, R. R. (1971). *Some limit theorems in statistics*, SIAM, Philadelphia, 1971.

Baringhaus, L., Henze, N., (1992). A characterization of and new consistent tests of symmetry, *Commun.Statist.-Theor.Meth.*,**21**, 1555-1566.

Desu, M.M. (1971). A characterization of the exponential distribution by order statistics. *Ann. Math. Stat.*, **42**, 837 - 838.

Eichelsbacher, P., Löwe, M. (1995). A Large Deviation Principle for m-variate von Mises-Statistics and U-Statistics. *J.Theoret. Prob.*, **8**, 807–823.

Nikitin, Ya. (1995). *Asymptotic efficiency of nonparametric tests.* Cambridge University Press, 1995.

Nikitin, Ya. Yu., Ponikarov, E. V. (1999). Rough large deviation asymptotics of Chernoff type for von Mises functionals and U-statistics. *Proceedings of St.Petersburg Mathematical Society*, **7**, 124–167. English translation in *AMS Transl.*, ser.2, 203(2001), 107 - 146.

Puri, P. S., Rubin, H. (1970). A characterization based on the absolute difference of two i.i.d. random variables. *Ann. Math. Stat.*, **41**, 2113-2122.

# Poster 22

# Power behaviour of permutation tests for fixed sample size and increasing number of variables

**Fortunato Pesarin**
*Department of Statistics. University of Padova*

**Luigi Salmaso**
*Department of Management and Engineering. University of Padova*

**Abstract.** In several applicational fields, it may happen that the number of variables is very much larger than that of subjects. It can be proven that, for a given and fixed number of subjects, when the number of variables diverges and the noncentrality parameter of the underlying population distribution increases with respect to each added variable, then power of multivariate permutation tests based on Pesarin's combining functions (Pesarin, 2001) is monotonically increasing. These results allow us to introduce the notion of "variable-based consistency" for permutation tests based on combination. Sufficient conditions are given in order that the rejection rate converges to one, for fixed sample sizes and attainable $\alpha$-values, when the number of variables diverges.

## References

Blair, R.C., Higgins, J.J, Karniski, W. and Kromrey, J.D. (1994) A study of multivariate permutation tests which may replace Hotelling's $T^2$ test in prescribed circumstances. *Multivariate Behavioral Research*, 29, 141−163.

Pesarin F. (2001) *Multivariate permutation tests with applications in biostatistics.* Wiley, Chichester.

# Poster 23

# A comparison between two multivariate survival models

**Quesada-Rubio, José-Manuel**
**Navarrete-Álvarez, Esteban**
**García-Leal, Julia**
**Martínez-Álvarez, Fernando**
**Rosales-Moreno, María-Jesús**
**Del-Moral-Ávila, María-José**
*Department of statistics and O.R. Granada University (Spain)*

**Abstract.** We present a comparison between a joint proportional hazards frailty model with gap times and survival and a similar joint frailty model for survival where the covariates are introduced assuming an accelerated life model. The models are evaluated by simulation and illustrated with a study of patients with lung cancer.

The heterogeneity is modelled by a group specific quantity which can be interpreted as an unobserved covariate common to the individuals, the frailty. It is usual to consider that survival times depend on the frailty by a proportional hazard model which induce the association between observed times.

There are a limited number of publications dealing with the joint modelling of recurrent events and terminal event such as death. Lancaster and Intrator (1998) and Liu, Wolfe and Huang (2004) consider each event time measured from the beginning of the study. Xueling Huang and Lei Liu (2007) consider the distributions of gap times between consecutive recurrent events:

$$\lambda_j(t_j/z_j, W, T_k, k < j) = \lambda_{0j}(t_j)\, W\, exp(\beta_j^T z_j), j > 1$$

$$\psi(t/z_0, W) = \psi_0(t)\, W^\gamma exp(\alpha^T z_0)$$

where $W$ is the frailty, $Z_j$ the covariates, is the $jth$ gap-specific covariate effect and $j$ survival times.

We consider a joint proportional frailty model, Cox (1972), with a gamma distribution for frailty W and we introduce the covariates assuming an accelerated life model. We call that mixed model:

$$\lambda_j(t_j/z_j, W, T_k, k<j) = \lambda_{0j}(t_j\, exp(\beta_j^T\, z_j))\; W\; exp(\beta_j^T z_j), j>1$$

$$\psi(t/z_0, W) = \psi_0(t\, exp(\beta_j^T z_j))W^{\gamma} exp(\alpha^T z_0)$$

It is possible no specify the marginal distributions, this is a semiparametric model and the regression parameters are updated by a similar method to the maximization of the partial likelihood for the Cox (1972) proportional hazards model. The R package is used for data generation and to fit the model and we illustrate this study with lung cancer patient data.

# References

Cox, D.R. (1972) Regression models and life tables (with discussion). *Journal of the royal statistical society. Series B*, **34**, 187-220.

Huang, X. and Liu, L. (2007). A joint frailty model for survival and gap times between recurrent events. *Biometrics*, **63**, 389-397

Lancaster, A. and Intrator, O. (1998). Panel data with survival: hospitalization of HIV-positive patients. *Journal of the American statistical association*, **93**, 46-53

Liu, L., Wolfe, R.A. and Huang, X. (2004). Shared frailty models for recurrent events and a terminal event. *Biometrics*, **60**, 747-756

# Poster 24

# Estimation and testing non linearity effects in additive models in censored regression

**Javier Roca-Pardiñas**
*Departamento de Estadística e I.O. Universidad de Vigo*

**Jacobo de Uña-Álvarez**
*Departamento de Estadística e I.O. Universidad de Vigo*

**Abstract.** In this paper we consider partially linear additive models in censored regression. We propose a randomly weighted version of the backfitting algorithm that allows for the nonparametric estimation of the effects of the covariates on the response. Moreover a procedure for testing nonlinearity effects is proposed. Backfitting theory is difficult in this context, and a bootstrap procedure is therefore provided for estimating the distribution of the test statistics. Given the high computational cost involved, binning techniques are used to speed up the computation in the estimation and testing process. Simulation results and the application to real data reveal that the predictor obtained with the additive model performs well, and that it is a convenient alternative to the linear predictor when some nonlinear effects are suspected.
**Keywords**: additive models, backfitting, bootstrap.

## 1 Introduction

Let $Y$ be a lifetime which is observed under censoring from the right and Let $\mathbf{X} = (X_1, ..., X_p)'$ be a vector of $p$ covariates. Put $f(\mathbf{x}) = E[\psi(Y) \mid \mathbf{X} = \mathbf{x}]$ for the regression function of $\psi(Y)$ on $\mathbf{X}$, so the model becomes

$$\psi(Y) = f(\mathbf{X}) + \varepsilon = f(X_1, ..., X_p) + \varepsilon \tag{1}$$

where the error term satisfies $E[\varepsilon \mid \mathbf{X}] = 0$ and $\mathbf{X}$ and $\varepsilon$ are independent. Taking $\psi(y) = \ln y$ is useful in regression analysis because $\psi(Y)$ is no longer restricted to $(0, \infty)$. This is the so-called accelerated failure time model, widely used to analyze survival data in the regression framework.

In the censored setup, we observe $(\mathbf{X}_1, Z_1, \delta_1), ..., (\mathbf{X}_n, Z_n, \delta_n)$ independent observations with the same distribution as $(\mathbf{X}, Z, \delta)$, where $Z = \min(Y, C)$, $C$ is the right-censoring variable assumed to be independent of $Y$, and $\delta = \mathbb{I}(Y \leq C)$. Unlike in the "iid" scenario, the weight associated to the $i$-th observation $(\mathbf{X}_i, Z_i, \delta_i)$ under censoring will be typically the jump of the Kaplan-Meier estimator at each point $Z_i$ $(i = 1, ..., n)$, namely

$$W_i = \frac{\delta_i}{n - RankZ_i + 1} \prod_{RankZ_j < RankZ_i} \left[1 - \frac{\delta_j}{n - RankZ_j + 1}\right]$$

where $RankZ_i$ is the rank of $Z_i$ among the ordered $Z$'s and where (in case of ties) uncensored observations are assumed to preceed the censored ones.

In this regression context we consider a flexible approach to estimate the regression function $f(\mathbf{X})$ through a partailly linear additive model (Hastie and Tibshirani, 1990) given by

$$f(\mathbf{X}) = \alpha + \sum\nolimits_{j=1}^{p} (\alpha_j X_j + f_j(X_j)) \tag{2}$$

being $\alpha, \alpha_1, \ldots, \alpha_p$ coefficients and $f_1, \ldots, f_p$ begin one dimensional functions. In the previous moddel, the effect of each covariate $X_j$ is descomposed in a parametric component, $\alpha_j X_j$, and a purely nonparametric component, $f_j(X_j)$. Note that this model nests linear models, additive models and partial additive models as particular cases.

## 2 Testing for Non Linearity Effects

We propose the bootstrap resampling techniques to test for nonlinearity effects in the semiparametric additive model specified in (2). For a given subset of covariates $X_{j_1}, \ldots, X_{j_q}$ with $1 \leq j_1 < \ldots < j_q \leq p$, interest centers on the null hypothesis $\mathbf{H}_0 : f_{j_1} = \ldots = f_{j_q} = 0$, namely, that the effects of the given subset of covariates are linear. For this purpose $L_1$ norm is considered yielding the test statistic

$$T = \sum\nolimits_{l=1}^{q} \sum\nolimits_{i=1}^{n} |\hat{f}_{j_l}(X_{ij_l})|$$

It must be remarked that, if the null hypothesis is verified, then $T$ should be close to zero, but it will generally be positive. Thus, the test rule for checking $\mathbf{H}_0$ with significance level $\alpha$ is that the null hypothesis is rejected if $T$ is larger than its upper $\alpha$- percentile. It is well know that in these kind of tests asymptotic theory is little helpful to determine that percentile, and resampling methods like bootstrap are applied instead.

# 3 Estimation. Bandwidth selection and computational aspects

We have developed an algorithm that enables model in (2) to be estimated. The proposed algorithm is a modified version of the weighted backfitting algorithm (Opsomer, 2000). At each loop of the algorithm are obtained the linear coefficients $\hat{\alpha}, \hat{\alpha}_1, \ldots, \hat{\alpha}_p$ by fitting a linear weighted model. Moreover the algoritm cycles through each of the combinations $X_j$ for $1 \leq j \ldots \leq p$, and the nonparametric estimates $f_j$, are obtained by applyin local linear kernel smoothers (Ruppert and Wand, 1994) to the corresponding partial residuals. These residuals are obtained by removing the estimated effects of the other covariates.

It is well known that the nonparametric estimates $f_1, \ldots, f_p$ heavily depend on the bandwidths $h_1, \ldots, h_p$ used in the local linear kernel estimates. Various proposals for an optimal selection have been suggested for the additive models, yet the difficulty of asymptotic theory in a backfitting context means that nowadays optimal selection is still a challenging open problem. Moreover, a distinction should be drawn between the bandwidth choice for estimation and for testing. Our computational experience has shown that, whereas nonlinear terms tend to be smoothed out and the null hypothesis of no interaction never rejected in the case of large bandwidths, in the case of small bandwidths the nonlinear terms tend to prove significant. Cross-validation was used for the automatic choice of bandwidths.

The bootstrap resampling techniques are time-consuming processes, because it is necessary to estimate the model a great number of times. Moreover, the use of cross-validation technique (explained below) for the choice of the bandwiths implies a high computational cost, inasmuch as it is necessary to repeat the estimation operations several times in order to select the optimal bandwidths. To speed up this process, we used binning-type acceleration techniques (Fan and Marron, 1994) to obtain the binning approximations of $\hat{f}_j$ in each of the iterations of the estimation algorithm.

# 4 Simulations and application to real data

To assess the validity of this estimation procedure, a simulation study is performed. Simulation results have shown that the proposed algorithm works well in practice, and that the additive model is a convenient alternative to linear regression in the presence of nonlinear effects.

Finally, an application to real data has served for illustrating the potential advantages of our censored additive model when compared to more classical regression approaches.

# References

Fan, J. and Marron, J.S. (1994). Fast implementation of non-parametric curve estimators. *Journal of Computational and Graphical Statistics* , **3**, 35-56.

Hastie, T.J. and Tibshirani R.J. (1990). *Generalized Additive Models.* Chapman & Hall.

Opsomer, J.D. (2000). Asymptotic properties of backfitting estimators. *J. Multivar. Anal.* **73**, 166-179.

Ruppert, D. and Wand, M.P. (1995). Multivariate locally weighted least squares regression. *Annals of Statistics*, **22**, 1346-1370.

# Poster 25

# Robust nonparametric regression estimation on Riemannian manifolds

**Daniela Rodriguez**
*Universidad de Buenos Aires and CONICET*

**Guillermo Henry**
*Universidad de Buenos Aires and CONICET*

**Abstract.** In this paper, we propose two families of robust kernel–based regression estimators when the regressors are random objects taking values in a Riemannian manifold. We consider estimators based on kernel methods and estimators based on k-nearest neighbor kernel methods. Strong uniform consistency results as well as the asymptotically normality of both families are established.

## 1 Summary

Nonparametric inference has gained a lot of attention, in recent years, in order to explore the nature of complex nonlinear phenomena. The idea of nonparametric inference is to leave the data to show the structure lying beyond them, instead of imposing one. Nadaraya (1964) and Watson (1964), introduced kernel–based estimators for the regression function $r(\mathbf{x}) = E(y|\mathbf{x})$, when dealing with independent observations $\{(y_i, \mathbf{x}_i)\}_{i=1}^n$ such that $y_i \in I\!R$, $\mathbf{x}_i \in I\!R^d$. Nearest neighbor with kernel methods for the regression function were introduced by Collomb (1980).

Both of them are a weighted average of the response variables and thus, they are highly sensitive to large fluctuations of the responses. As mentioned by several authors, the treatment of outliers is an important step in highlighting features of a data set since extreme points can affect the scale and the shape of any estimate of the regression function

based on local averaging, leading to possible wrong conclusions. This has motivated the interest in combining the ideas of robustness with those of smoothed regression, to develop procedures which will be resistant to deviations from the central model in nonparametric regression models. As it is well known, robust estimators can be obtained via local $M-$estimates. The first proposal of robust estimates for nonparametric regression was given by Cleveland (1979) who adapted a local polynomial fit by introducing weights to deal with large residuals. See also, Tsybakov (1982) and Härdle (1984), who studied pointwise asymptotic properties of a robust version of the Nadaraya–Watson method. These results were extended to $M-$type scale equivariant kernel estimates by Härdle and Tsybakov (1988) and by Boente and Fraiman (1989) who also considered robust equivariant nonparametric estimates using nearest neighbor weights. A review of several methods leading to robust nonparametric regression estimators can be seen in Härdle (1990).

The proposals mentioned above assume that the predictors $\mathbf{x}$ belong to a subset of $I\!R^d$ with non empty interior and therein, the euclidean structure of $I\!R^d$ is considered. However, in many applications, the predictors $\mathbf{x}$ take values on a Riemannian manifold more than on $I\!R^d$ and this structure of the explanatory variables needs to be taken into account when considering neighborhoods around a fixed point $\mathbf{x}$. Several authors such as, Mardia and Jupp (2000), Hall *et al.* (1987) and Fischer *et al.* (1993) discussed methods for spherical and circular data analysis while generalizations to different types of manifolds have been studied by Lee and Ruymgaart (1996), Hendriks (1990) and Hendriks *et al.* (1993). An approach based on the Riemannian geodesic distance on the manifold was considered by Pelletier (2005) for the problem of estimating the density of random objects on a compact Riemannian manifold and also by Pelletier (2006) for that of estimating the regression function which is the aim of our paper. More precisely, let $(M, g)$ be a closed Riemannian manifold of dimension $d$ and let $(y, \mathbf{x})$ be a random vector such that $y \in I\!R$ and $\mathbf{x} \in M$. The classical nonparametric setting assumes that the response variables have finite expectation and focusses on the estimation of the regression function $r(p) = E(y|\mathbf{x} = p)$. Pelletier's (2006) idea was to build an analogue of a kernel on $(M, g)$ by using a positive function of the geodesic distance on $M$, which is then normalized by the volume density function of $(M, g)$ to take into account for curvature. Under classical assumptions on the kernel and the bandwidth sequence, Pelletier (2006) derives an expression for the asymptotic pointwise bias and variance as well as an expression for the asymptotic integrated mean square error.

As in the Euclidean setting, the estimators introduced by Pelletier (2006) are a weighted average of the response variables $y_i$ with weights depending on the distance between $\mathbf{x}_i$ and $p$ implying that they will suffer from the same lack of robustness that the Nadaraya–Watson estimators with carriers in the Euclidean space $I\!R^d$. In this paper we consider two families of robust estimators for the regression function when the explanatory variables $\mathbf{x}_i$ take values on a Riemannian manifold $(M, g)$. The first family combine the ideas of robust smoothing in Euclidean spaces with the kernel weights introduced in Pelletier (2005). The second family generalises the proposal given by Boente and Fraiman (1989) who considered robust nonparametric estimates using nearst neighbor weights when the predictors $\mathbf{x}$ are on $I\!R^d$. Local $M-$estimators adapted to regressors lying on a $d-$dimensional Riemannian manifold will be introduced and their asymptotic properties will be studied. It is worth mentioning that robust estimators for directional data were considered among others by He (1992), Ko and Guttorp (1988) and also by

Agostinelli (2007) who studied robust methods for circular data analysis.

As in Pelletier (2006), our two families of estimators will be consistent with standard kernel or k-nearest with kernel estimators on Euclidean spaces, i.e., it reduces to the local $M-$estimator based on standard kernel weights when $M$ is $I\!R^d$. Moreover, they converge at the same rate as the Euclidean kernel estimators. This result generalizes the conclusions obtained by Pelletier (2006) from the pointwise mean square error.

This aim of this paper is to present two versions of robust local $M-$estimators of the regression function adapted to the fact that the explanatory variables $\mathbf{x}_i$ take values on a Riemannian manifold. Uniform consistency and asymptotic distributions of the proposed estimators are obtained under regular assumptions. For small samples, the behavior of the classical approach is compared through a Monte Carlo study under normality and contamination.

## References

Agostinelli, C. (2007). Robust estimation for circular data. *Computational Statistics and Data Analysis* **51,** 5867-5875.

Boente, G. and Fraiman, R. (1989). Robust nonparametric regression estimation. *Journal of Multivariate Analysis*, **29**, 180-198.

Cleveland, W. (1979). Robust locally weighted regression and smoothing scatter plots. *Jornal of the American Statistical Association* **74**, 829-836.

Collomb, G. (1980). Estimation de la regression par la méthode des k points les plus proches avec noyau: Quelques propiétés de convergence ponctuelle. *Lecture Notes in Mathematics* **821**, 159-175.

Fischer, N.I., Lewis, T. and Embleton, B.J.J. (1993). *Statistical Analysis of Spherical Data.* Cambridge University Press, Cambridge.

Ko, D. and Guttorp, P. (1988). Robustness of estimators for directional data. *Annals of Statistics* **16**, 609-618.

Hall, P. , Watson, G.S. and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika* **74**, 751-762.

Härdle, W. (1984). Robust regression function estimation. *Journal of Multivariate Analysis* **14**, 169-180.

Härdle, W. (1990). *Applied nonparametric regression.* Cambridge University Press.

Härdle, W. and Tsybakov, A.B. (1988). Robust nonparametric regression with simultaneous scale curve estimation. *Annals of Statistics* **16**, 120-135.

He, X. (1992). Robust statistics of directional data: a survey. *Nonparametric Statistics and Related Topics*, eds A. M. Saleh, pp. 87-95. North Holland, New York.

Hendriks, H. (1990). Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions. *Annals of Statistics* **18**, 832-849.

Hendriks, H., Janssen, J.H.M. and Ruymgaart, F.H. (1993). Strong uniform convergence of density estimators on compact Euclidean manifolds. *Statistics and Probabability Letters* **16**, 305-311.

Lee, J.M. and Ruymgaart, F.H. (1996). Nonparametric curve estimation on Stiefel manifolds. *Journal of Nonparametric Statistics* **6**, 57-68.

Mardia, K. and Jupp, P. (2000). *Directional Data*. Wiley, New York.

Nadaraya, E. A. (1964). On estimating regression. *Theory Probality and Applications*, **9**, 141-142.

Pelletier, B. (2005). Kernel density estimation on Riemannian maniflods. *Statistics and Probability Letters* **73**, 297-304.

Pelletier, B. (2006). Nonparametric regression estimation on closed Riemannian manifolds. *Journal of Nonparametric Statistics* **18**, 57-67.

Tsybakov, A.B. (1982). Robust estimates of a function. *Peredaci Informacii* **18**, 39-52. English translation in Problems Inform. *Transmission* **18**, 190-201.

Watson, G.S. (1964). Smooth regression analysis. *Sankhya Ser. A* **26**, 359-392.

# Poster 26

# Asymmetric returns and portfolio selection by using a local correlation measure

**Bård Støve**
*Department of Finance and Management Science. Norwegian School of Economics and Business Administration*

**Karl Ove Hufthammer**
*Department of Mathematics. University of Bergen*

**Dag Tjøstheim**
*Department of Mathematics. University of Bergen*

**Abstract.** A number of studies have provided evidence that correlations in stock returns are asymmetric, e.g. higher correlations during bear markets. In this paper we confirm such results and present an approach for mean-variance portfolio selection, both by using a new measure for local correlation introduced by Hufthammer and Tjøstheim (2008).

## 1 Introduction

Correlation is an important parameter in modern portfolio theory. It is used here to measure the dependence between returns of different assets. The work of Markowitz (1952) introduced the mean-variance portfolio theory, which is one of the most widely used approaches in portfolio selection. The idea is simple; low correlated assets are good for diversification, while highly correlated assets should be avoided.

However, several studies have documented asymmetric characteristics of asset returns, in particular asymmetric correlations in which stock returns tend to have higher correlations with the market when it goes down than when it goes up, see Campbell et al (2002) and Lognin and Solnik (2001). Further, assessing asymmetric correlation requires care,

as a correlation computed conditional on some variables being high or low is a biased estimator of the unconditional correlation, see Forbes and Rigobon (2002) and Hong et al (2007). Other papers have documented that correlations across stock markets change over time, see e.g. Ramchand and Susmel (1998).

Thus the benefit of diversification will erode if the correlations are asymmetric. Several proposals have been made to correct for these shortcomings, e.g. portfolio optimization models that allow correlation of asset returns to vary over time and the use of conditional mean-variance portfolio optimization, see e.g. Campell et al (2002).

In this paper we introduce the use of a local correlation measure, proposed by Hufthammer and Tjøstheim (2008), which is able to capture the state-varying correlations and can be used for portfolio selection. The local correlation estimator is for completeness presented in section 2. In section 3 empirical evidence of the existence of asymmetric correlations in market returns is provided, while section 4 presents a portfolio selection method based on mean-variance portfolio theory and the local correlation estimator.

## 2 Local correlation

Several attempts have been made to construct local dependence measures taking into account the values of the variables involved. For example, for large values of two stochastic variables $X$ and $Y$ the local correlation may be higher than for small values of $X$ and $Y$. In Hufthammer and Tjøstheim (2008) the authors present a new approach to this problem by approximating the joint distribution of $X$ and $Y$ by a family of bivariate normal distributions and by using local likelihood to obtain an estimate of local correlation. The asymptotic analysis poses challenges that are quite different from those of traditional nonparametric regression, and the results imply slower rates of convergence. For a complete introduction to the local dependence measure, see Hufthammer and Tjøstheim (2008). In this paper we use this new local correlation estimator to examine local correlations in market returns.

## 3 Empirical analysis of financial returns dependence

The data is daily international equity price index from DataStream for the United States (S&P 500), the United Kingdom (FTSE 100), France (CAC 40) and Germany (DAX 30). Further, we calculate returns as $100 \times (\ln(p_t) - \ln(p_{t-1}))$, where $p_t$ is the price index from DataStream. These markets have earlier been studied by Longin and Solnik (2001) and Cambell et al (2002), but we now extend the observation span from May 1990 to 31 December 2007. In total the data set consists of 4588 daily observations. Summary statistics for the international equity returns are shown in table 3.

The daily returns are in the range from over $-9\%$ to around $7.5\%$. All series have a negative skewness, and the kurtosis are generally high, between 6 and 7. This implies a deviation from the normal distribution.

In table 2 the correlation estimates for the returns are given. The correlations between the European indices are around 0.7, while the correlations between the European and the S&P 500 are much lower, just above 0.4. This is in line with the previous mentioned studies.

| Statistics | S&P 500 | FTSE 100 | CAC 40 | DAX 30 |
|---|---|---|---|---|
| Mean | 0.031 | 0.022 | 0.021 | 0.032 |
| Variance | 0.965 | 1.013 | 1.659 | 1.903 |
| Maximum return | 5.573 | 5.903 | 7.002 | 7.552 |
| Minimum return | -7.112 | -5.885 | -7.678 | -9.871 |
| Skewness | -0.122 | -0.138 | -0.116 | -0.271 |
| Kurtosis | 7.019 | 6.269 | 6.054 | 7.377 |

**Table 1:** Summary index statistics, May 1990 - December 2007.

| | S&P 500 | FTSE 100 | CAC 40 | DAX 30 |
|---|---|---|---|---|
| S&P 500 | 1 | | | |
| FTSE 100 | 0.412 | 1 | | |
| CAC 40 | 0.426 | 0.773 | 1 | |
| DAX 30 | 0.455 | 0.675 | 0.766 | 1 |

**Table 2:** Correlation of index returns, May 1990 - December 2007.

We next turn to estimates of the local correlation. Figure 1 shows the estimated local correlation between the returns from S&P 500 and FTSE 100 in several gridpoints. The local correlations are close to the classical correlation estimate in interior points, but larger in the bottom left of the plot, that is, when the returns are negative. The estimator of local correlation depends on two smoothing parameters (bandwidths), they are for now choosen by a simple method (i.e. the standard deviation times a constant). If the bandwidths are large, the local correlation will be approximately similar to the correlations in table 2.

Similar plots are obtained for the other pairwise market returns. These plots show different local correlations across different gridpoints, in particular high local correlation when market returns are negative.

The local correlations are also computed on different sub-periods (i.e. five years intervals), and they show that the local correlations may also be time-varying, consistent with similar results for classical correlation, see e.g. Longin and Solnik (1995).

# 4 Portfolio selection

Mean-variance based portfolio construction is the most common approach for asset management. Introduced by Markowitz (1952), the measures of return and risk are the mean and variance of the portfolios' returns, respectively. Portfolios are considered mean-variance efficient if they minimize the variance for a given mean return or maximize the return for a given level of variance.

Assume $N$ risky assets with mean premium return vector $\mu$ and covariance matrix $\mathbf{\Sigma}$. Let $\mathbf{X}$ be a vector of portfolio weights for the $N$ assets. The optimization problem is as follows; an efficient portfolio is determined by minimizing the portfolio variance, subject to a target expected portfolio premium return ($\mu_m$) and that all wealth is invested in the

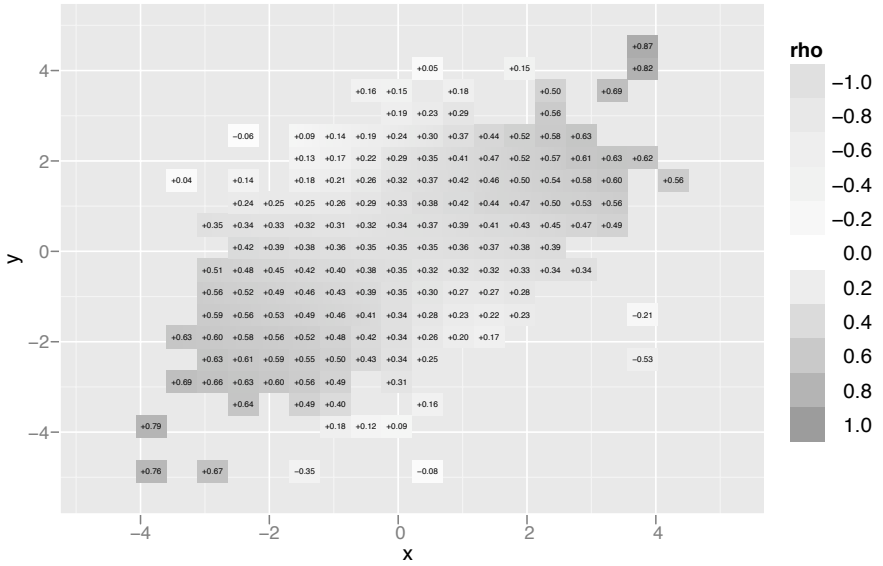


**Figure 1:** Local correlation - S&P 500 (y) and FTSE 100 (x).

risky assets, i.e.

$$\min_{\mathbf{x}} \mathbf{X}'\mathbf{\Sigma}\mathbf{X} \text{ s.t. } \mathbf{X}'\mu = \mu_m \text{ and } \mathbf{X}'\mathbf{1} = 1. \tag{1}$$

The optimal portfolio weights are calculated by setting up the Lagrangian and solving the corresponding first-order condition. Since we find evidence of larger local correlations when markets fall, the benefits from diversification will be severely eroded. This has implications for portfolio selection. We therefore suggest using the local correlation (covariance) for portfolio construction, and we show that this will result in other portfolio weights. Some preliminary results concering the asymptotic distributions of these estimators are provided.

# References

Campbell, R., Koedijk, K. and Kofman, P. (2002). Increased correlation in bear markets. *Financial Analysts Journal* **58**, 87-94.

Forbes, K. and Rigobon, R. (2002). No contagion, only interdependence: Measuring stock market co-movemens. *Journal of Finance* **57**, 2223-2261.

Hong, Y., Jun, T. and Zhou, G. (2007). Asymmetries in stock returns: Statistical tests and economic evaluation. *Review of Financial Studies* **20**, 1547-1581.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets. *Journal of Finance* **56**, 649-676.

Longin, F. and Solnik, B. (1995). Is correlation in international equity returns constant: 1960-1990? *Journal of Finance* **14**, 3-26.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance* **7**, 77-97.

Ramchand, L. and Susmel, R. (1998). Volatility and cross correlation across major stock markets. *Journal of Empirical Finance* **5**, 397-416.

Hufthammer, K. O. and Tjøsteim, D. (2008). Local correlation and local Gaussian likelihood. *Working paper*.

# Poster 27

# Goodness-of-fit tests for conditional models with time-dependent coefficients under censoring and truncation - a case study

**Bianca Teodorescu** [1]
*Institut of Statistics. Université catholique de Louvain*

**Ingrid Van Keilegom**
*Institut of Statistics. Université catholique de Louvain*

**Abstract.** The survival of Gastric Cancer Patients depends on various initial risk factors such as age, delay time between first simptoms and diagnosis, type of tumor, type of treatment and the effects of these risk factors may change with time: some effects may vanish whereas other remain constant. An omnibus goodness-of-fit test is developed to test whether the time-varying Cox model fits the data.

## 1 Introduction

In the context of semiparametric regression with censored data, a lot of models are available to account for the relationship between the survival time and a set of explanatory variables: Cox proportional hazards model, log-logistic model, etc. In a general way, these models can be written as linear regression models, where the dependent variable

[1]The authors would like to thank S. Pita Fernández and E. Casariego Vale for the Gastric Cancer Data

is a monotone transformation ($\phi$) of the survival function ($S$). Particular choices of $\phi$ give well known models in survival analysis.The choice $\phi(u) = \log(\frac{u}{1-u})$ gives the logistic model, $\phi(u) = -\log(u)$ gives the additive risk model and $\phi(u) = \log(-\log(u))$ leads to the proportional hazards model.

In these models, the regression coefficients are often supposed to be constants. But in practice, the structure of the data might be more complex, in the sense that it is better to consider coefficients that can vary over time and also data that can be left-truncated.

More precisely, let $Y$ denote the survival time, $T$ the truncation time and $C$ the censoring time. When data are left-truncated and right-censored we observe $(Z, T, \delta)$ only if $Z \geq T$, where $Z = \min\{Y, C\}$ and $\delta = I_{\{Y \leq C\}}$. Let $(Z_i, T_i, \delta_i, X_i)$, $i = 1, \ldots, n$ be an iid sample from $(Z, T, \delta, X)$, where $X$ is a (one-dimensional) covariate. We are interested in the relationship between the survival function of $Y$, $S(z|X) = P(Y > z|X)$ and $X$. We like to test whether this relationship is of polynomial type, via a known monotone transformation $\phi : [0, 1] \to R$ of the survival function, i.e.:

$$\phi(S(z|X)) = \beta_0(z) + \beta_1(z)X + \ldots + \beta_p(z)X^p, \tag{1}$$

for some known $p$. No assumption is made on the form of the survival function $S(z|X)$, except for the usual smoothness assumptions.

The appropriateness of the parametric modelling of regression data may be judged by comparison with a semi-parametric estimator of the response. For this purpose one may use a squared deviation measure between the two fits. The sum of the squared deviation over all the values of the covariates may be used as a test statistic for testing the parametric model where the critical value is determined from the asymptotic distribution of this statistic.

To be more precise, for a given $\phi$, $H_0 : \exists\beta(z)$ such that (1) holds is tested against $H_A$ : (1) does not hold for any $\beta(z)$. We compare a semiparametric estimator of the response, $\phi(\hat{S}(z|\mathbf{X}))$, with its parametric counterpart, $\hat{\beta}(z)\mathbf{X}$ (same technique as in Cao & González-Manteiga (2007)). Where $\hat{S}$ is a conditional Kaplan-Meier type estimator with truncated data for S, see Iglésias-Pérez & González-Manteiga (1999), and $\hat{\beta}(z)$ is the least-squares estimator proposed in Teodorescu et al (2008). A large deviation between them indicates the lack of fit of the parametric form and thus the rejection of the null.

To measure this deviation, we study a kind of $L_2$ - distance between these two:

$$\hat{\Phi}_n(\hat{\beta}(z)) = \frac{1}{n}\sum_{r=1}^{n}\left(\phi(\hat{S}(z|X_r)) - (\hat{\beta}(z)\mathbf{X})\right)^2.$$

Hence, it is reasonable to reject $H_0$ when $\hat{\Phi}_n(\hat{\beta}(z))$ is large. We have to measure this distance over the whole observed time-period, thus we shall take the integral over an interval $[a, b]$, for some given $a$ and $b$:

$$T_n = \int_a^b \hat{\Phi}_n(\hat{\beta}(z))dz \tag{2}$$

We should also multiply this quantity by a normalizing sequence in order to have a limiting distribution. This leads to the following test statistic: $nh^{1/2}T_n$ that has asymptotic normal distribution under some regularity conditions, see Teodorescu & Van Keilegom (2008).

The convergence of the distribution of the test statistic $nh^{1/2}T_n$ to a normal distribution is quite slow, so that it seems more appropriate not to use the asymptotic critical values in practice. We therefore compute the critical values based on a bootstrap method. The procedure is as folows:

1. Choose a bandwidth $h$ in the interval $(0, \mu(\text{supp}(X)))$ and a pilot bandwidth $g$ (larger than $h$), where $\mu$ is the Lebesgue measure.

2. For all $z \in \{Z_1, \ldots, Z_n\}$ and $x \in \{X_1, \ldots, X_n\}$:

   **a)** Estimate $S(z|x)$, $G(z|x)$ and $L(z|x)$ by the conditionals estimators $\hat{S}_g(z|x)$, $\hat{G}_g(z|x)$ and $\hat{L}_g(z|x)$, respectively, see Iglesias-Pérez & González-Manteiga (1999).

   **b)** Replace $S(z|x)$ by $\hat{S}_g(z|x)$ in (1) and estimate $\beta_0(z), \ldots, \beta_p(z)$ by the least squares estimator proposed by Teodorescu et al. (2008) to obtain $\hat{\beta}_{0,g}(z), \ldots, \hat{\beta}_{p,g}(z)$. Plug-in these estimators into (1) and re-estimate $S(z|x)$ by

$$\tilde{S}_g(z|x) = \phi^{-1}(\hat{\beta}_{0,g}(z) + \hat{\beta}_{1,g}(z)x + \ldots + \hat{\beta}_{p,g}(z)x^p).$$

3. For $b = 1, \ldots, B$:

   **a)** For every $i = 1, \ldots, n$ draw random observations $Y_i^*$, $C_i^*$ and $T_i^*$ from $\tilde{S}_g(\cdot|X_i)$, $\hat{G}_g(\cdot|X_i)$ and $\hat{L}_g(\cdot|X_i)$, respectively. Compute $Z_i^* = \min\{Y_i^*, C_i^*\}$, $\delta_i^* = 1_{\{Y_i^* \leq C_i^*\}}$ and simulate new values $Y_i^*$, $C_i^*$ and $T_i^*$ if $T_i^* > Z_i^*$.

   **b)** Use this resample $\{(T_1^*, Z_1^*, \delta_1^*, X_1), \ldots, (T_n^*, Z_n^*, \delta_n^*, X_n)\}$ to estimate a bootstrap version of the conditional survival function, $\hat{S}_h^*(z|X_i)$ $(i = 1, \ldots, n)$ using the bandwidth $h$. This bootstrap version is used to obtain the bootstrap vector of coefficients $\hat{\beta}_h^{*(b)} = \left(\hat{\beta}_{0,h}^{*(b)}(z), \ldots, \hat{\beta}_{p,h}^{*(b)}(z)\right)$ using the least squares estimator, and to obtain the bootstrap version $\hat{\Phi}_n^*\left(\hat{\beta}_h^{*(b)}(z)\right)$ of $\hat{\Phi}_n\left(\hat{\beta}(z)\right)$.

   **c)** Compute the bootstrap version of the test statistic $nh^{1/2}T_n$, which is given by:

$$\begin{aligned} nh^{1/2}T_{n,b}^* &= nh^{1/2}\int_a^b \hat{\Phi}_n^*\left(\hat{\beta}_h^{*(b)}(z)\right)dz \\ &\simeq n\sqrt{h}\sum_{i=1}^{n-1}\hat{\Phi}_n\left(\hat{\beta}_{h,b}^*(Z_i^*))(Z_{i+1}^* - Z_i^*\right) \end{aligned}$$

4. Order the obtained test statistics and take $nh^{1/2}T_{n,[(1-\alpha)B]}^*$ which approximates the $(1-\alpha)$-quantile of the distribution of $nh^{1/2}T_n$ under $H_0$.

5. If $nh^{1/2}T_n > nh^{1/2}T_{n,[(1-\alpha)B]}^*$, then reject $H_0$, otherwise do not reject $H_0$.

## 2 Gastric Cancer Data - a case study

These procedures are applied to data on people suffering from gastric adenocarcinoma. The survival of gastric cancer patients depends on various initial risk factors such as age

at diagnosis, delay time between first simptoms and diagnosis, extention of tumor (TNM classification), location of tumor, presence of metastasis, type of treatment, sex of the person and others. These risk factors determine the chance of survival for the patients. However, the effects of these risk factors may change with time. Some effects may vanish whereas other remain constant. We analyse the survival of 945 patients treated in the Hospital Xeral-Calde in Lugo, Spain. Previous studies on these data (see Casariego-Vales et al (2001), Rabuñal et al. (2004)) have considered the Cox proportional hazards model with constant coefficients as being appropriate for the data without really checking the hypothesis. The variable of interest was considered to be the time from diagnostic until death and censoring occured when the patients were lost to follow up, died from another reason not related to the desease or were still alive at the end of the study.

These studies showed that the variables with the most impact on the survival of a patient were: age, extention of tumor, presence of metastasis, type of surgery and diagnostic delay.

Here we are only taking into consideration these variables that showed the most impact on the survival time and we are going to make a different approach on the data, by considering as the target population, not only the patients diagnosed in the hospital as previously done, but all the people suffering of gastric adenocarcinoma in the region. This is done by taking as the variable of interest the time from the first symptoms to death and by considering that not only censoring can occur (by the reasons stated above), but also truncation: those people that die before getting a diagnose, are considered as being truncated.

The considered model is the Cox proportional hazards model with time dependent coefficients, which can be written in the following form:

$$\phi(S(z|\mathbf{X})) = \beta_0(z)+\beta_1(z)X_1+\beta_2(z)X_2+\beta_3(z)X_3+\beta_4(z)X_4+\beta_5(z)X_5+\beta_6(z)X_6, \quad (3)$$

where $\phi(u) = \log(-\log(u))$, $X_1$ and $X_2$ are the indicators of having TNM classification 2 and 3, $X_3$ and $X_4$ are the indicators for the type of surgery, $X_5$ is the idicator for the presence of metastasis and $X_6$ is the age at diagnosis minus its mean (67.71606 years).

So we will test:

$$H_0 : \forall z \quad \exists \beta(z) \in R^7 \text{ such that (3) holds,}$$

$$\text{against } H_a : \text{(3) does not hold for any } \beta(z).$$

Under $H_0$, the coefficients $\beta_i(z)$ $(i = 0, \ldots, 6)$ were estimated via the least squares estimator proposed in Teodorescu et al. (2008) and the optimal bandwidth $h = 45$ was selected by means of the bootstrap method proposed above among $\{25, 30, 35, 40, 45, 50\}$. We took $g = 1.5h$, $\alpha = 0.05$ and we conducted 500 bootstrap simulations. The estimated p-value was found to be 0.7528. Hence, we do not reject $H_0$, i.e. we conclude that the proportional hazards model with time varying coefficients is appropriate for these data. An estimator for the coefficients $\beta_i(z)$ as well as their 95 % pointwise confidence intervals are given.

## References

Casariego Vales, E., Pita Fernández, S., Rigueiro Veloso, M.T. et al. (2001) - Supervivencia en 2.334 pacientes con cáncer gástrico y factores que modifican el prognóstico, *Med. Clin. (Barc)*, 117, 361-365

Cao, R. and González-Manteiga, W. (2008). Goodness-of-fit tests for conditional models under censoring and truncation. *Journal of Econometrics*, 143, 166-190.

Iglesias-Pérez, C. and González-Manteiga, W. (1999). Strong representation of a generalized product-limit estimator for truncated and censored data with some applications. *Journal of Nonparametric Statistics*, 10, 213-244.

Rabuñal, R.R., Pita F.S., Rigueiro, V.M.T., Casariego, E.V., Pértega, S.D., García-Rodeja, E. and Abraira, V. (2004). Presentation and Prognosis of Gastric Cancer in Patients Aged 80 Years and Older. *World J. Surg.*, 28, 155-159.

Teodorescu, B., Van Keilegom, I. and Cao, R. (2008). Generalized conditional linear models under left truncation and right censoring. *The Annals of the Institute of Statistical Mathematics* (to appear)

Teodorescu, B. and Van Keilegom, I. (2008). Goodness-of-fit Test for Generalized Conditional Linear Models under Left Truncation and Right Censoring (under revision).

# Index of authors