Workshop on Modern Nonparametric Methods for Time Series, Reliability & Optimization 2012



Heverlee, Belgium, September 10-12, 2012

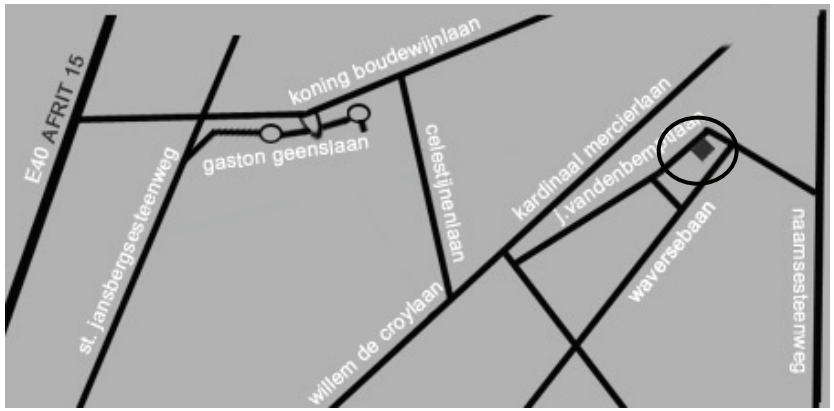
Final program and practical information



Workshop on Modern Nonparametric Methods for Time Series, Reliability & Optimization 2012

September 10-12, 2012

Boardhouse, Jules Vandenbemptlaan 6, B-3001 Heverlee, Leuven
Tel: +32-(0)-16.31.44.44
http://www.boardhouse.be/



Katholieke Universiteit Leuven

Notes

Notes

Dear Participant,

It is a great honor to welcome you to Leuven. It is the first time that we are organizing this workshop. With your enthusiasm and contributions we will undoubtedly make this a successful event.

The workshop has a theoretical as well as a practical flavor. The general theme of the workshop aims at nonparametric methods for regression and time series, reliability analysis and optimization methods which are used in the field of statistics and engineering.

The reason for organizing this workshop is threefold. First, we would like to bring researchers of different groups together so they can exchange ideas and discuss recent developments in their respective field. Second, by organizing two short courses we would like to encourage scientists who are not familiar with nonparametric methods to get a flavor of this broad and interesting domain. Finally, optimization techniques become more and more popular in the field of engineering and statistics. Therefore, we also put an emphasis on this topic in the workshop.

Organizing a workshop requires a huge effort and therefore we would like to thank a number of colleagues. Elsy Vermoesen and Ida Tassens have been crucial in all administrative and practical aspects of the workshop. Liesbeth Van Meerbeek created the workshop website and kept it up to date. My co-organizers, De Brabanter Jos (KUL, KaHoSL), Gijbels Irène (KUL), Van Impe Jan (KUL), Vandewalle Joos (KUL), Veraverbeke Noël (UHasselt) and Antoniadis Anestis (UJF) had a major role in constructing the workshop's technical agenda. Without financial support from various sponsors, the registration fee would have been substantially higher and therefore we gratefully acknowledge the Scores4Chem knowledge platform, LStat, IUAP and the Laboratory of Enzyme and Brewing Technology of KaHoSL. Finally, we also acknowledge Eva Hiripi from Springer for taking interest in our workshop.

Further we would like to thank all speakers for presenting their work at the workshop.

We have no doubt this will be a fruitful and exciting workshop!

On behalf of the organizing committee,

Kris De Brabanter

Table of contents

History Leuven, Belgium …… 3
Access to conference room, lunches, coffee breaks and social activities …… 3
Practical information regarding short courses and talks …… 3
Organization and sponsors …… 4
Social activities …… 4
Internet access …… 5
Short courses …… 6
Invited talks …… 7
Contributed talks …… 10
Contributed session Ia …… 10
Contributed session Ib …… 10
Contributed session II …… 12
Contributed session III …… 13
Contributed session IV …… 14
Conference program …… 17
Organizing and scientific committee …… 18
List of participants …… 19
Notes …… 20
Map …… 22

Participants

Name	Institution	Email
Biau Gérard	Université Pierre et Marie Curie	gerard.biau@upmc.fr
Claeskens Gerda	KU Leuven	gerda.claeskens@econ.kuleuven.be
Croux Christophe	KU Leuven	christophe.croux@econ.kuleuven.be
De Brabanter Jos	KU Leuven/KaHoSL	jos.debrabanter@kahosl.be
De Brabanter Kris	KU Leuven	kris.debrabanter@esat.kuleuven.be
De Rouck Gert	KaHoSL	gert.derouck@kahosl.be
Freyermuth Jean-Marc	KU Leuven	jean-marc.freyermuth@kuleuven.be
Gijbels Irène	KU Leuven	irene.gijbels@wis.kuleuven.be
Gins Geert	KU Leuven	geert.gins@cit.kuleuven.be
Gribkova Svetlana	Université Pierre et Marie Curie	svetlana.gribkova@etu.upmc.fr
Gros Sébastien	KU Leuven	sebastien.gros@esat.kuleuven.be
Györfi László	Budapest Univ. Technology & Economics	gyorfi@szit.bme.hu
Gugushvili Shota	Vrije Universiteit Amsterdam	s.gugushvili@vu.nl
Hohsuk Noh	UCL	word5810@gmail.com
Huyck Bart	KU Leuven	bart.huyk@cit.kuleuven.be
Kauermann Göran	Ludwig-Maximilians-Universität München	goeran.kauermann @stat.uni-muenchen.de
Logist Filip	KU Leuven	filip.logist@cit.kuleuven.be
Pelckmans Kristiaan	Uppsala University	kristiaan.pelckmans@it.uu.se
Puertas Jose	KU Leuven	gpuertas@kuleuven.esat.be
Simar Léopold	UCL	leopold.simar@uclouvain.be
Sznajder Dominik	KU Leuven	dominik.sznajder@wis.kuleuven.be
Telen Dries	KU Leuven	dries.telen@cit.kuleuven.be
Vandewalle Joos	KU Leuven	joos.vandewalle@esat.kuleuven.be
Van Graan Francois	North-West University South Africa	francois.VanGraan@nwu.ac.za
Van Impe Jan	KU Leuven	jan.vanimpe@cit.kuleuven.be
Vanpaemel Dina	KU Leuven	dina.vanpaemel@wis.kuleuven.be
Varron Davit	Université de Franche-Comté	dvarron@univ-fcomte.fr

Local organizing committee

Name	Institution
Atoniadis Anestis	Université Joseph Fourier (FR)
De Brabanter Kris	KU Leuven (BE)
De Brabanter Jos	KaHo Sint Lieven (BE)
Gijbels Irène	KU Leuven (BE)
Van Impe Jan	KU Leuven (BE)
Vandewalle Joos	KU Leuven (BE)
Veraverbeke Noël	UHasselt (BE)

Scientific committee

Name	Institution
Atoniadis Anestis	University Joseph Fourier (FR)
Beirlant Jan	KU Leuven (BE)
Carbonez An	KU Leuven (BE)
Claeskens Gerda	KU Leuven (BE)
Croux Christophe	KU Leuven (BE)
Delaigle Aurore	University of Melbourne (AU)
Dewil Raf	De Nayer Mechelen (BE)
Gijbels Irène	KU Leuven (BE)
Györfi László	Budapest University (HU)
Hallin Marc	Univ. Libre Bruxelles (BE)
Opsomer Jean	Colorado State University (US)
Thas Olivier	RUGent (BE)
Van Impe Jan	KU Leuven (BE)
Van Keilegom Ingrid	Université Catholique de Louvain (BE)

Leuven, Belgium

Leuven is situated in the Flemish-speaking part of Belgium, at about 20 kilometers east of Brussels, in the heart of Western Europe. The city of Leuven appeared in historical documents for the first time in the year 884. In that year, the plundering Vikings settled around an old fortification at the Dijle River, called 'Luvanium' in Latin or 'Lovon' in the local vernacular. Nowadays, the city has 90 000 inhabitants.

The Katholieke Universiteit Leuven (KU Leuven) was founded in 1425 by Pope Martin V. The university bears the double honour of being the oldest extant Catholic university in the world and the oldest university in the Low Countries.

At present the university counts more than 40 000 students, around 12% of whom are international students from more than 120 nations. The university is dynamically integrated in the town (with dozens of historical university buildings). The rich historical tradition continues to serve as a solid foundation for top-level research and centers of academic excellence. KU Leuven carries out its academic activities at various campuses, research parks and hospital facilities in close cooperation with the members of the KU Leuven Association and with its hospital partners.

Access to conference room, lunches, coffee breaks and social activities

Your badge is the key to access the conference room, lunches, coffee breaks and social activities including the conference dinner. Please wear your badge at all times during the conference.

Accompanying persons will receive a ticket to attend the conference dinner. Supplementary tickets for the conference dinner cost 50 euro and had to be booked in advance.

Talks

Each contributed talk is allocated a 20 minute time slot + 5 minutes questions. Speakers are requested

- not to exceed their time slot
- check all hardware details at the break before their session.

Each invited talk is allocated a 40 minute time slot + 5 minutes questions. The two short courses will take approximately 2.5 hours. Slides for the short courses will be provided to the participants of the workshop.

Organization and sponsors

The Workshop on Modern Nonparametric Methods for Time Series, Reliability & Optimization 2012 is organized in collaboration with:

- Katholieke Universiteit Leuven (BE)
- KaHo Sint Leuven (Associatie KULeuven, BE)
- Universiteit Hasselt (BE)
- Université Joseph Fourier (FR)

with support from

- Scores4Chem knowledge platform (http://cit.kuleuven.be/scores4chem/)
- LStat (http://lstat.kuleuven.be/)
- IUAP
- Laboratory of Enzyme and Brewing Technology of KaHoSL

Technical co-sponsor

- Springer

Contributed volume

We have negotiated with Springer-Verlag to publish an edited book mainly based on the talks given at the workshop. Selected authors will be asked to extend their abstract to full research articles which can be based on your talk or other results in your area of research or a closely related field.

Social activities

Welcome reception: September 9, 2012 at 19:00 (Boardhouse)
Discover an ancient belgian tradition: beer brewing and tasting

Surplus is the beer of the Associatie KU Leuven. It combines the Belgian beer brewing tradition with scientific innovation from the KU Leuven and KaHoSL. The beer is produced in a small scale research brewery at KaHoSL under the supervision of Gert De Rouck and Guido Aerts.

Conference dinner: September 11, 2012 at 19:30 (Voltaire, Boardhouse)
Supplementary tickets for the conference dinner had to be booked in advance.
Unfortunately supplementary tickets for on-site requests cannot be provided.

Conference program

	Monday	Tuesday	Wednesday
8:00 – 8:45	Registration & Welcome		
8:45 – 9:00	Scores4Chem/LStat Opening		
9:00 – 11:30	Sébastien Gros	László Györfi	Contributed Session IV
11:30 – 11:45	**break**	**break**	&
11:45 – 12:30	Christophe Croux	Göran Kauermann	Concluding remarks
12:30 – 14:15	Lunch	Lunch	Lunch
14:15 – 15:00	Contributed Session Ia	Léopold Simar	Gert De Rouck
15:00 – 15:45	Contributed Session Ib	Gérard Biau	
15:45 – 16:15	**break**	**break**	
16:15 – 17:15	Contributed Session II	Contributed Session III	
19:30 – …		Conference Dinner	

absolute deviation penalty and a ridgetype penalty. We show that the proposed estimator identifies relevant and irrelevant components consistently and achieves the nonparametric optimal rate of convergence for the relevant components. We also provide some numerical evidence of the estimator, and illustrate its usefulness through a real data example to identify important body measurements to predict percentage of body fat of an individual.

11:30 Jose Gervasio Puertas & Johan Suykens
QP reloaded: an efficient reuse of the kernel matrix eigenvectors

Abstract: In this talk, we present a novel numerical method to address a family of QP optimization problems arising in the context of, for instance, novelty detection and probability density estimation. Theoretically well grounded and relying only on very simple and general assumptions regarding the kernel matrix as well as the inherent problem structure, we show how to derive an efficient approximate (or exact) solution by using a subset (or all) of the kernel matrix eigenvectors. As the most computationally demanding part of the proposed method is that of finding a subset of eigenvectors for the kernel matrix, it is particularly well suited for the QP optimization problems in the described family involving only the calculation of a few eigenvectors from the kernel matrix.

Biking in Leuven

Guests staying in the conference hotel are entitled to request a Boardhouse bike at the reception. This is a free service of the conference hotel for their guests.

Internet access

Wireless access

A wireless network will be available during the workshop

SSID: boardhouse1
Key: 1111111111

Short courses

1. **Sébastien Gros**: Introduction to Nonlinear Programming

   When: Monday, September 10, 2012 at 9:00 – 11:30

   Abstract: in the first part of this talk we will review the major properties of nonlinear programming (NLP): global & local optimality, first and second-order optimality conditions, constraint qualification, geometric interpretation and sensitivity analysis. Convex optimization will be approached in more details in the second part of the talk, reviewing duality, generalized constraints, and some classical forms of convex problems. The last part of the presentation will review the state-of-the-art derivative-based numerical optimization methods: Newton and quasi-Newton methods, convergence properties, globalization techniques, quadratic and successive quadratic programming, interior-point methods. The presentation will attempt to make the connections clear between the theory and its practical consequences.

2. **László Györfi**: Nonparametric and machine learning algorithms for time series prediction

   When: Tuesday, September 11, 2012 at 9:00 – 11:30

   Topics:
   1. Nonparametric regression estimation
   2. Machine learning aggregation for prediction and for squared loss.
   3. Pattern recognition.
   4. Prediction for 0-1 loss.

techniques are verified on real data originating from micro-array expression studies, and from Genome-Wide Association data. Learning how to make a prognosis of a patient is an important ingredient to the task of building an automatic system for personalized medical treatment. A prognosis here is understood as a useful characterization about the (future) time of an interesting event. In cancer research, a typical event is the relapse of a patient after receiving treatment. Since this involves essentially a form of prediction, it is naturally to phrase this problem in a context of machine learning.

10:15 Break

Session chair: Gérard Biau & Kris De Brabanter

10:40 Mia Hubert, Irène Gijbels & Dina Vanpaemel
Reducing the mean squared error of quantile-based estimators by smoothing

Abstract: Many univariate robust estimators are based on quantiles. As already theoretically pointed out by Fernholz (1997), smoothing the empirical distribution function with an appropriate kernel and bandwidth can reduce the variance and mean squared error (MSE) of some quantile-based estimators in small data sets. In this talk we will explain how this idea can be applied on several robust estimators of location, scale and skewness (Hubert *et al.*, 2012). We will explain how to select the bandwidth robustly and how to reduce the bias. We will show trough a simulation study that the use of this smoothing method indeed leads to smaller MSEs, also at contaminated data sets. In particular, better performances for the medcouple are obtained, which is a robust measure of skewness that can be used for outlier detection in skewed distributions (Brys *et al.*, 2004). The smoothed medcouple can be used to adapt the adjusted boxplot (Hubert & Vandervieren, 2008) to detect outliers in small data set of which the underlying distribution is skewed. This will be illustrated on a real data example.

11:05 Hohsuk Noh & Eun Ryung Lee
Component selection in additive quantile regression models

Abstract: Nonparametric additive models are powerful techniques for multivariate data analysis. Although many procedures have been developed for estimating additive components both in mean regression and quantile regression, the problem of selecting relevant components has not been addressed much especially in quantile regression. In this article, we present a doubly-penalized estimation procedure for component selection in additive quantile regression models that combines basis function approximation with a variant of the smoothly clipped

Contributed Session IV: Wednesday, September 12, 2012

Session chair: Gérard Biau

9:00 **Jean-Marc Freyermuth, Florent Autiny, Gerda Claeskens & Rainer von Sachs**
Hyperbolic wavelet thresholding rules: the curse of dimensionality through the maxiset approach

Abstract: In this talk we are interested in nonparametric multivariate function estimation. In Autin *et al.* (2012), we determine the maxisets of several estimators based on thresholding of the empirical hyperbolic wavelet coefficients. That is we determine the largest functional space over which the risk of these estimators converges at a chosen rate. It is known from the univariate setting that pooling information from geometric structures (horizontal/vertical blocks) in the coefficient domain allows to get 'large' maxisets (see e.g Autin *et al.* (2011a,b,c)). We show that in the multivariate context the benefits of information pooling is versatile. In a sense these estimators are much more exposed to the curse of dimensionality while their potential can be preserved under certain conditions. Finally, in the spirit of Schneider and von Sachs (1996), Neumann and von Sachs (1997), we propose an application in the field of time series analysis through the ability of such methods to adaptively estimate an evolutionary spectrum of a locally stationary process by 2-D wavelet smoothing over time and frequency.

9:25 **Francois Van Graan**
Crossvalidation selection for the smoothing parameter in nonparametric hazard rate estimation

Abstract: An improved version of the crossvalidation bandwidth selector is proposed by using the technique of bootstrap aggregation. Simulation results show that this selector compares very favorably to a bootstrap bandwidth selector based on an asymptotic representation of the mean weighted integrated squared error for the kernel-based estimator of the hazard rate function in the presence of right-censored samples.

9:50 **Kristiaan Pelckmans & Liu Yang**
Aggregating methods for high-dimensional inference of survival data

Abstract: This presentation reviews new theoretical and practical results in the inference from high-dimensional data originating from survival studies. The reviewed techniques are closely related to the recently proposed class of aggregating techniques and the study of online learning from experts. The

Invited talks

1. **Christophe Croux**: Robust exponential smoothing

   When: Monday, September 10, 2012 at 11:45 – 12:30

   Abstract: Robust versions of the exponential and Holt-Winters smoothing method for forecasting are presented. They are suitable for forecasting univariate time series in the presence of outliers. The robust exponential and Holt-Winters smoothing methods are presented as recursive updating schemes that apply the standard technique to pre-cleaned data. Both the update equation and the selection of the smoothing parameters are robustified. A simulation study compares the robust and classical forecasts. The presented method is found to have good forecast performance for time series with and without outliers, as well as for fat tailed time series and under model misspecification. The method is illustrated using real data incorporating trend and seasonal effects. We also discuss a multivariate extension. This is joint work with Sarah Gelper and Roland Fried.

2. **Göran Kauermann**: Penalized Estimation with Applications in Economics and Finance

   When: Tuesday, September 11, 2012 at 11:45 – 12:30

   Abstract: The central idea of penalized estimation is that a functional relationship in the model is fitted with a high dimensional spline basis. A penalty imposed on the fitted spline coefficients induces on the one hand numerical feasibility and stability, on the other hand the penalty guarantees a smooth, that is differentiable fit. The idea is well established in smooth regression or functional data analysis with Eilers & Marx (1996, Statistical Science) marking a milestone. The concept of penalized estimation is however quite general and allows to be employed in various other areas of smooth, that is functional estimation. In the talk we give some examples of the flexibility of the idea, in particular we will establish smooth copula estimation. The penalty approach itself mirrors close connections to Bayesian estimation, by comprehending the penalty as prior distribution imposed on the spline coefficients. We develop the idea further towards mixture models.

3. **Léopold Simar**: Methodological Advances and Perspectives in Nonparametric Frontier Analysis

When: Tuesday, September 11, 2012 at 14:15 – 15:00

Abstract: In production theory, economists are interested to determine a production frontier, the boundary of the production set that is the set of all possible combinations of inputs (factor of production like labor, energy, capital,...) and of outputs (the goods that are produced). This boundary represents the locus of optimal combinations of inputs and outputs. Technically, this is equivalent of estimating the boundary of the support of a density of a multivariate random variable where the sample is the observed set of inputs/outputs of firms. The estimated frontier is then used as benchmark against which each firm can be compared and the distance to the optimal frontier serves as a measure of technical inefficiency. Most of the nonparametric estimators are based on the idea of enveloping the cloud of data: this involve**s** the use of advanced statistical tools, but also optimization tools, implying some numerical burden. Recent methodological advances allow today to make inference about the level of inefficiency of firms in fully nonparametric setups. The paper presents the state of the art in this field, showing the technical difficulties linked to the problem where the bootstrap is the only available tool for making practical inference. A lot of challenges are still to be solved and the paper presents the most recent developments and perspectives to address these challenges. This includes: testing issues, robustness to outliers, explaining inefficiencies by environmental factors, introducing noise in the production process...

4. **Gérard Biau**: Random Forests

When: Tuesday, September 11, 2012 at 15:00 – 15:45

Abstract: Random forests are a scheme proposed by Leo Breiman in the 2000's for building a predictor ensemble with a set of decision trees that grow in randomly selected subspaces of data. Despite growing interest and practical use, there has been little exploration of the statistical properties of random forests, and little is known about the mathematical forces driving the algorithm. In this talk, we propose an in-depth analysis of a random forests model suggested by Breiman in 2004, which is very close to the original algorithm. We show in particular that the procedure is consistent and adapts to sparsity, in the sense that its rate of convergence depends only on the number of strong features and not on how many noise variables are present.

propose to build the test statistics as functional violation measures based on the empirical copula estimator. Furthermore, the statistical inference is based on the bootstrapped distribution of the test statistics. This requires resampling scheme under the null hypothesis and we propose a smooth constrained nonparametric copula estimation procedure as a remedy. It is based on the local polynomial smoothing of the initial constrained estimator and on transforming its partial derivatives by rearrangement technique. The proposed methodology is generic and flexible and can be applied to other dependence concepts, which can be expressed as shape constraints on the copula function. This talk is based on the results described in Sznajder (2011) and Gijbels & Sznajder (2011) and refers to previous work on other dependence structures in Gijbels & Sznajder (2011) and Gijbels & Sznajder (2012).

Contributed Session III: Tuesday, September 11, 2012

Session chair: Francois Van Graan

16:15 Davit Varron & Alexis Flesch
Empirical likelihood in some semiparametric models: a computationally feasible approach

Abstract: This talk is about empirical likelihood methods in semiparametric models. More precisely we will consider the setup where the unknown parameter can be written $\theta = T(P)$, with P being the unknown distribution of the sample, and T is sufficiently smooth function. Under some assumptions of empirical processes theory (of Vapnik-Chervonenkis type), we provide a computantionaly feasible method to build asymptotic confidence regions, when θ is either finite dimensional or a functional parameter. Some applications to survival analysis are investigated.

16:40 Shota Gugushvili & Peter Spreij
Non-parametric Bayesian drift estimation for stochastic differential equations

Abstract: In this talk we will discuss posterior consistency for nonparametric Bayesian estimation of a drift coefficient of a stochastic differential equation.

presence of nonlinear distortions. When the system is close to linear, the performance of the LS-SVM is only slightly better than the linear models.

Contributed Session II: Monday, September 10, 2012

Session chair: Göran Kauermann

16:15 Svetlana Gribkova & Olivier Lopez
Nonparametric copula estimation for right censored data

Abstract: In this work, we provide a way to define new nonparametric estimators of the copula function for censored data, adapted to various schemes of bivariate censoring. The constructed piecewise copula estimator can be seen as an extension of the empirical copula, which is not available under the censoring. We consider its smooth versions as well, one of them permitting a bias correction. Their performances are compared by simulations. We investigate the asymptotic properties, and obtain functional central limit theorems for these estimators. The applications to copula density estimation and goodness-of-fit testing for copula models under censoring are considered.

16:40 Dominik Sznajder & Irène Gijbels
Testing for stochastic monotonicity using copulas

Abstract: In this talk we discuss tests for stochastic monotonicity. This dependence concept originates from the work of Tukey (1958) and Lehmann (1959), where it was called complete regression dependence. This name refers to the idea behind the concept, namely that all conditional quantiles are monotonic functions (in the same direction). In particular it implies, what is nowadays called, regression monotonicity, i.e., monotonicity of $E(Y|X=x)$ as a function of x. The stochastic monotonic relation between random variables is of particular interest in financial, insurance and econometric studies. For recent developments in testing for stochastic monotonicity and a broad overview of applications see Lee *et al.* (2009) and Delgado & Escanciano (2012). Stochastic monotonicity is a very strong dependence concept. It implies Tail Monotonicity, which further implies Quadrant Dependence, both studied in Gijbels & Sznajder (2011) and Gijbels & Sznajder (2012) respectively. The general methodology for the testing procedure is based on smooth constrained copula bootstrapping. It is motivated by the fact that often dependence between random variables in a random vector can be expressed as a feature of the underlying copula function. Such a copula function is a linking function between the joint distribution of a random vector and the marginal distributions. We

5. **Gert De Rouck**: Belgian Beer culture: the biotechnological art of beer creation

When: Wednesday, September 12, 2012 at 14:15 – …

Abstract: Belgian beer is well known nowadays in the whole world. More than 1200 different beers in a wide range of beer styles (beers from Trappist Monks, Abbey beers, high alcohol specialty beers, Lagers, and even spontaneous fermented acidic beers) are available in the international market. During this talk, an overview of the history of the Belgian Beer culture will be presented together with the question: what is Belgian Beer culture? Beer is made with only 4 to 6 ingredients. How can Belgian brewers create all these beer styles and varieties with this limited amount of raw materials? Belgian Beer is of high standard and found its roots in out of the box thinking. Also the Belgian Brewing Technology is of outstanding quality. The strength is again in the introduction of innovative processes whereby beer quality and lowest total costs are essential. To conclude: this small country creates great processes and beers that slowly conquer the world. A passion for taste and joy are essential to immerse yourself in our Beer culture.

Contributed talks

Contributed Session Ia and Ib are organized by the chemical engineering department CIT-BIOTEC, KU Leuven.

Contributed Session Ia: Monday, September 10, 2012

Session chair: Jan Van Impe

14:15 Filip Logist
Optimization in Engineering: from theory to practical applications

Abstract: This talk will focus on modeling and optimisation techniques and their importance for the improvement of real-life engineering applications. Several steps will be illustrated starting from process modeling, over process monitoring to process control and optimization. Specific topics include the issue of dealing with multiple objectives in, e.g., experimental design or process optimisation. Also recent interests in optimisation under uncertainty (also called robust optimisation) will be briefly discussed.

14:40 Dries Telen
A Comparison of Two Robust Optimal Experiment Design Methods

Abstract: In dynamic bioprocess models parameters often appear in a nonlinear way. To calibrate these models, the Fisher Information Matrix explicitly depends on the current parameter estimates. Hence, it is advisable to take this parametric uncertainty into account in the design procedure in order to obtain an experiment which is robust with respect to changes in the parameters. In this talk we apply computationally efficient approximate robustifcation strategies based on a worst case scenario. Both methods exploit linearization techniques to avoid the hard to solve max-min optimization problems.

Contributed Session Ib: Monday, September 10, 2012

Session chair: Jan Van Impe

15:00 Geert Gins
Bias and variance of batch-end quality prediction: influence of measurement noise

Abstract: The development of automated monitoring systems to assist human process operators in their decisions is an important challenge for the chemical

and life sciences industries (Venkatasubramanian *et al.*, 2003). Especially for batch processes, which are commonly used for the production of goods with high added value (e.g., medicines, high performance polymers and enzymes), monitoring of the batch quality is of utmost importance due to the high costs associated to lost batches. Today's chemical and biochemical production processes and plants are equipped with numerous sensors that measure various flow rates, temperatures, pressures, pH, concentrations, . . . Partial Least Squares (PLS (Nomikos & MacGregor, 1995)) has been developed to deal with large datasets of correlated measurements and to filter noise from these measurements. However, noise present on both online sensor measurements and offline quality measurements will never be removed completely and will hence negatively influence the predictive performance of the PLS models. Several researchers investigated the influence of noise on PLS predictions to provide prediction intervals for spectroscopy calibration problems (Faber & Kowalski, 1997). However, they assume that the measurements contain all necessary information to predict the output, while in batch-end quality prediction, important explanatory variables are often missing. In addition, a linear relation between spectroscopy measurements and analyte concentration is guaranteed by Lambert-Beer's law for spectroscopy, while the relation between online process measurements and batch product quality might contain (small) non-linear contributions. To the authors' knowledge, up until today, no distinct relation has been identified to describe the influence of measurement noise on PLS batch-end quality predictions, which partly explains the reluctance of companies to accept this technique. This talk aims at investigating the influence of noise on PLS predictions of final batch quality by conducting extensive Monte Carlo simulations on data of an industrial-scale penicillin production process. The research focuses on noise on the quality measurements, which are generally performed offline on small product samples, often using measurement techniques with low accuracies, resulting in high output noise levels.

15:25 Bart Huyck
LS-SVM Based Identification of a Pilot Scale Distillation Column

Abstract: In this talk we describe the identification of a binary distillation column with Least-Squares Support Vector Machines (LS-SVM). It is our intention to investigate whether a kernel based model, particularly an LS-SVM, can be used for the simulation of the top and bottom temperature of a binary distillation column. Furthermore, we compare the latter model with standard linear models by means of mean-squared error (MSE). It will be demonstrated that this nonlinear model class achieves higher performances in MSE than linear models in the

Nonparametric copula estimation for right censored data

Svetlana Gribkova[1] & Olivier Lopez[1]

[1]Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie
email: svetlana.gribkova@etu.upmc.fr
[2]Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie
email: olivier.lopez0@upmc.fr

June 13, 2012

Abstract

In this work, we provide a way to define new nonparametric estimators of the copula function for censored data, adapted to various schemes of bivariate censoring. The constructed piecewise copula estimator can be seen as an extension of the empirical copula, which is not available under the censoring. We consider its smooth versions as well, one of them permitting a bias correction. Their performances are compared by simulations. We investigate the asymptotic properties, and obtain functional central limit theorems for these estimators. The applications to copula density estimation and goodness-of-fit testing for copula models under censoring are considered.

1 Introduction

A copula is a cumulative distribution function $C(x, y)$ on $[0, 1]^2$ with uniform marginals, which "couple" the univariate distribution functions $F_1(x), F_2(y)$ of the variables T_1, T_2 to their joint distribution function $F(x, y)$ by the simple relation

$$F(x, y) = C(F_1(x), F_2(y)).$$

By the theorem of Sklar(see, [1] for more details), if the margins are continuous, then the copula is unique. If the variables of interest are fully observed, then, given a sample $(T_{1i}, T_{2i})_{1\leq n}$, the nonparametric empirical copula estimator can be written as

$$C_n(u, v) = F_n(F_{1n}^{-1}(u), F_{2n}^{-1}(v)),$$

where $F_n(u, v)$ is the empirical distribution function and $F_{1n}(u), F_{2n}(v)$ are its margins. Thus, copula estimation goes back to the estimation of joint distribution function linking two variables. The weak convergence of this copula estimator in the space $l^\infty([0, 1]^2)$ was proved in [4].

In the present work we aim to extend the empirical copula estimator to the case of right censored data. Under censoring instead of observing the variable T_1(resp. T_2) one observes the variable $Y_1 = \min(T_1, C_1)$(resp. Y_2) and $\delta_1 = \mathbb{I}_{T_1\leq C_1}$(resp. δ_2), where C_1, C_2 are censoring random variables. Thus, in the bivariate case the observed sample of censored observations is of the form $(Y_{1i}, Y_{2i}, \delta_{1i}, \delta_{2i})_{1\leq i\leq n}$.

Under random censoring one does not always observe the variables of interest (T_1, T_2), so that the empirical distribution function of these variables is not available from the data, neither the empirical copula. Thus, the idea would be to replace the function $F_n(x, y)$ by its modification adapted to the case of censored data. While marginal distribution functions F_1, F_2 under censoring can be estimated consistently by a well known Kaplan-Meier estimator(see, [3]), in the bivariate case there does not exist an estimator which would be consistent for all censoring configurations and which would define a proper distribution function. Thus, the estimating procedures are different for various schemes of censoring. Nevertheless, in most cases the bivariate estimator takes a general form

$$\hat{F}_n(x, y) = \frac{1}{n}\sum_{i=1}^{n} \hat{W}_{in}\, \mathbb{I}_{Y_{1i}\leq x, Y_{2i}\leq y}, \qquad (1)$$

while random weights $\hat{W}_{in}$ depend on the particular model assumptions.

We will construct a copula estimator which will be consistent for any model with an estimator of the joint distribution function of the form (1) satisfying functional central limit theorem. The form of the estimator will not depend on the exact form of random weights attributed to observations, and thus on the particular type of censoring.

By the following we are giving two examples of censoring configurations with a bivariate joint distribution function estimator satisfying indicated assumptions.

2 Examples

2.1 Censoring acts only on the one of two variables

This configuration occurs, for example, when the censoring affects the lifetime of interest but not the covariable or in the case of the insurance data such as bivariate loss-ALAE described in [6]. The data contains observations on the loss variable T censored by a policy limit C of the contract and the ALAE variable X which is non censored and represents expenses directly associated with individual claims such as lawyer's fees for example. For the censoring mechanism it is supposed that the variables C and T are independent and $\mathbb{P}(T \leq C|X,T) = \mathbb{P}(T \leq C|T)$.

The observed sample consists then of a triple $(X_i, Y_i, \delta_i)_{1\leq i\leq n}$ with $Y_i = \min(T_i, C_i)$ and $\delta_i = \mathbb{I}_{T_i\leq C_i}$. The estimator of the joint distribution function of (X,T), proposed and investigated in [7], is of the form (1) with

$$\hat{W}_{in} = \frac{\delta_i}{1-\hat{G}(Y_i-)}.$$

where $\hat{G}(y)$ is the Kaplan-Meier estimator of the distribution function $G(y)$ of the censoring variable C.

Under some conditions on tails of the distributions we prove the functional central limit theorem for the described bivariate estimator.

2.2 Censoring variables differ only through an additional observed variable

This type of situations appears frequently in insurance contracts subscribed by two persons in a couple. If T_1(resp. T_2) is the lifetime of the husband(resp. his wife), the censoring variables C_1 and C_2 are their ages at a moment of the exit from the study for a reason different from the death. As in the first example, censoring variables (C_1, C_2) are supposed to be independent of the couple (T_1, T_2). The observed variables are then $Y_1 = T_1 \wedge C_1$ and $Y_2 = T_2 \wedge C_2$. Besides these variables the additional variable ε is usually observed which is the age difference $T_2 - T_1$ between two individuals in couple. For many cases, the event which stops observation occurs at the same time for both members of the couple, so censoring variables are also related by $C_2 - C_1 = \varepsilon$. The observations are then formed of $(Y_{1i}, Y_{2i}, \varepsilon_i, \delta_{1i}, \delta_{2i})_{1\leq i\leq n}$. The estimator of the joint distribution function of (T_1, T_2), which generalizes the estimator proposed in [8], is introduced in [9] and takes again the form (1) with

$$\hat{W}_{in} = \frac{\delta_{1i}\delta_{2i}}{1-\hat{G}(\max(Y_{1i}, Y_{2i}-\varepsilon)-)},$$

where $\hat{G}(y)$ is the estimator of the distribution function $G(y)$ of the censoring variable C_1. It was shown(see, [9] for the details) that the functional central limit theorem holds for this estimator as well.

3 Copula estimation

We consider the models such that the bivariate estimator of the joint distribution function of variables (T_1, T_2) is of the form

$$\hat{F}_n(x,y) = \frac{1}{n}\sum_{i=1}^{n} \hat{W}_{in}\mathbb{I}_{Y_{1i}\leq x, Y_{2i}\leq y},$$

and satisfies the functional central limit theorem, i.e. the empirical process $\sqrt{n}(\hat{F}_n(x,y) - F(x,y))$ converges weakly in $l^\infty(\mathbb{R}^2)$ to a tight gaussian process.

Let us define the censored empirical copula estimator as

$$\hat{C}_n(u,v) = \hat{F}_n(\hat{F}_{1n}^{-1}(u), \hat{F}_{2n}^{-1}(v)).$$

and the empirical copula process

$$\alpha_n(u,v) = \sqrt{n}(\hat{C}_n(u,v) - C(u,v)).$$

Under some regularity conditions we prove that $\alpha_n(u,v)$ converges weakly in $l^\infty[0,1]^2$ to a certain gaussian process.

As in [4], we consider the smoothed censored copula estimator, based on the smoothed version of the joint distribution function of the form

$$\hat{\mathbb{F}}_n(x,y) = \frac{1}{n}\sum_{i=1}^{n} \hat{W}_{in} K\left(\frac{x-Y_{1i}}{h_1}\right) K\left(\frac{y-Y_{2i}}{h_2}\right),$$

where

$$K(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} k(u,v)dudv,$$

with $k : \mathbb{R}^2 \mapsto \mathbb{R}$ is a symmetric bivariate kernel verifying $\iint k(x,y) = 1$. We consider the estimator

$$\hat{\mathbb{C}}_n(u,v) = \hat{\mathbb{F}}_n(\hat{\mathbb{F}}_{1n}^{-1}(u), \hat{\mathbb{F}}_{2n}^{-1}(v)),$$

which is the censored generalization of the estimator, proposed in [4], and the transformed estimator of the form

$$\hat{\mathbb{C}}_n^T(u,v) = \frac{1}{n}\sum_{i=1}^{n} \hat{W}_{in} K\left(\frac{\Phi^{-1}(u) - \Phi^{-1}(\hat{F}_n(Y_{1i}))}{h}\right) \times K\left(\frac{\Phi^{-1}(v) - \Phi^{-1}(\hat{F}_n(Y_{2i})}{h}\right),$$

where $\Phi(x)$ is a distribution function such that Φ' and $\frac{\Phi'(x)^2}{\Phi(x)}$ are bounded and $h \to 0$ is a bandwidth. Using this transformation, suggested in [5], permits to avoid a corner bias problem and thus to estimate copulas with the second order derivatives unbounded on the boundaries of the unit square(see, [5] for more details). Under some smoothness assumptions and supposing that $h^2\sqrt{n} \to 0$,

we prove the weak convergence of the smoothed empirical processes $\tilde{\alpha}_n(u,v) = \sqrt{n}(\hat{\mathbb{C}}_n(u,v) - C(u,v))$ and $\tilde{\alpha}_n^T(u,v) = \sqrt{n}(\hat{\mathbb{C}}_n^T(u,v) - C(u,v))$ to some tight gaussian processes. The bandwidth choice is accomplished by bootstrap procedure.

4 Conclusion

In this work we provided new nonparametric estimators of copula function linking two variables under right censoring, considering some type of models coming from the applications. The suggested estimators can be seen as the extensions of the empirical copula and its smoothed version to the censored framework. The smoothed versions of the piecewise estimator are considered. The talk will be illustrated by the simulation results which permit a comparison between these two versions. Some applications to the copula density estimation and goodness-of-fit testing for the censored data will be considered as well.

References

[1] Sklar, M., (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris.*, **8**, 229–231.

[2] Dabrowska, D., (1988) Kaplan-Meier estimate on the plane. *Annals of Statistics*, **16**, 1475–1489.

[3] Kaplan, E.L., Meier, P., (1958). Nonparametric estimation from incomplete observations. *JASA*, **53**, 457–481.

[4] Fermanian, J-D., Radulović, D. and Wegkamp, M., (2004). Weak convergence of empirical copula processes. *Bernouilli*, **10**, 847–860.

[5] Omelka, M., Gijbels, I. and Veraverbeke, N., (2009). Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *Annals of Statistics*, **37**, 3023–3058.

[6] Denuit, M., Purcaru, O. and Van Keilegom, I., (2006). Bivariate archimedean copula models for censored data in non-life insurance. *Journal of Actuarial Practice*, **13**, 5–32.

[7] Stute, W., (1993). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis*, **45**, 89–103.

[8] Lin, D. Y., Ying, Z., A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika*, **80**, 573–581.

[9] Gribkova, S., Lopez O., Saint-Pierre P., (2012). A simplified model for studying bivariate mortality under right-censoring. *Preprint HAL*, http://hal.archives-ouvertes.fr/hal-00683483.

Empirical likelihood in some semiparametric models: a computationally feasible approach.

Davit VARRON[1] & Alexis FLESCH[2]

[1]Université de Franche-Comté
email: alexis.flesch@univ-fcomte.fr
[2]Université de Franche-Comté
email: davit.varron@univ-fcomte.fr

Abstract

This talk is about empirical likelihood methods in semiparametric models. More precisely we will consider the setup where the unknown parameter can be written $\theta = T(\mathbf{P})$, with $\mathbf{P}$ being the unknown distribution of the sample, and T is sufficiently smooth function. Under some assumptions of empirical processes theory (of Vapnik-Chervonenkis type), we provide a computantionaly feasible method to build asymptotic confidence regions, when θ is either finite dimensional or a functional parameter. Some applications to survival analysis are investigated.

1 Introduction

Consider a nonparametric family $\mathcal{P}$ of probability measures on a measurable space $\big(\mathcal{X}, \mathcal{T}\big)$. The well known plug-in method consists in estimating an unknown parameter θ_0, for which we can write $\theta_0 := T(\mathbf{P}_0)$, where T is a specified map from $\mathcal{P}$ to a metric space (E, d), and $\mathbf{P}_0$ is the common law of an independent, identically distributed [i.i.d] sample X_i which we assume based on an underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$. In that case, writing δ_x for the Dirac measure at point x, the plug-in estimator of θ, based on the n first observations is $\hat{\theta} := T(\mathbf{P}_n)$, where $\mathbf{P}_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \in \mathcal{P}_d$ is the empirical measure and $\mathcal{P}_d$ is the set of all discrete probability measures on $\Big(\mathcal{X}, \mathcal{T}\Big)$. In this framework, the empirical likelihood method developed by Owen [3] is heuristically very appealing. Indeed, writing, for $u \geq 1$ and $n \in \mathbb{N}^*$:

$$\mathbb{S}_n := \Big\{(p_1, \dots, p_n) \in [0,1]^n, \sum_{i=1}^{n} p_i = 1\Big\}, \tag{1.1}$$

$$\mathbb{S}_{n,u} := \Big\{(p_1, \dots, p_n) \in \mathbb{S}_n,\ \prod_{i=1}^{n} np_i \geq u\Big\}, \tag{1.2}$$

$$\overline{\mathcal{P}}_{n,u} := \Big\{\widetilde{\mathbf{P}}_n = \sum_{i=1}^{n} p_i \delta_{X_i},\ (p_1, \dots, p_n) \in \mathbb{S}_{n,u}\Big\}, \tag{1.3}$$

one can define a region around θ, with threshold u by

$$\mathcal{R}_{n,u} := \Big\{T(\widetilde{\mathbf{P}}_n),\ \widetilde{\mathbf{P}}_n \in \overline{\mathcal{P}}_{n,u}\Big\}, \tag{1.4}$$

under the assumption that T is properly defined on a set $\mathcal{P}' \supset \mathcal{P} \cup \mathcal{P}_d$. When $(E, ||\cdot||)$ is a vector space, T is linear, and $T(\delta_{X_1})$ is concentrated in a finite dimensional subspace $E_0 \subset E$, the methodology of Owen [2] is sufficient enough to prove the validity of (1.4) since, in that case, one can write $\theta := \mathbb{E}(Y_1)$, with

$Y_i := T(\delta_{X_i})$, and

$$\mathcal{R}_{n,u} = \big\{\theta \in E_0,\ -2\ln \mathbf{EL}_n(\theta) \leq -2\ln u\big\}, \text{ with} \tag{1.5}$$

$$\mathbf{EL}_n(\theta) := \max\Big\{ \prod_{i=1}^n np_i,\ p_i \in \mathbb{S}_n,\ \sum_{i=1}^n p_i Y_i = \theta \Big\}. \tag{1.6}$$

Bertail [1] investigated the validity of the above mentioned empirical likelihood method when T is nonlinear (in which case (1.5) fails to hold in general), but satisfies a Hadamard differentiability property, in the setup of empirical process. More precisely consider a class $\mathcal{F}$ of real Borel functions on $\big(\mathcal{X}, \mathcal{T}\big)$, for which there exists an envelope $H \geq 1$ being square integrable under each $\mathbf{P} \in \mathcal{P}$ (assumption (HG1)) and such that

$(HG2)$ $\displaystyle\int_0^\infty \sqrt{\ln \sup_{Q \in \mathcal{P}_d} \mathcal{N}(\epsilon \,||\, H \,||_{Q,2}, \mathcal{F}, ||\cdot||_{Q,2})}d\epsilon < \infty$. (V-C type condition).

Denoting by $\big(\ell_\infty(\mathcal{F}), ||\cdot||_{\mathcal{F}}\big)$ the space of all real bounded functions on $\mathcal{F}$ endowed with its sup norm,

$$||\,\psi\,||_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |\,\psi(f)\,|, \tag{1.7}$$

each $\mathbf{P} \in \mathcal{P}'$ can be identified to an element of $\ell_\infty(\mathcal{F})$ through the application $\mathbf{P} \to \big\{f \to \int f d\mathbf{P}\big\}$. Also, for each $\mathbf{P} \in \mathcal{P}'$, we shall denote by $\mathcal{C}(\mathcal{F}, \mathbf{P}) \subset \ell_\infty(\mathcal{F})$ the space of functions that are uniformly continuous on the totally bounded set $\big(\mathcal{F}, ||\cdot||_{\mathbf{P},2}\big)$. We will assume that

$(HT1)$ T is Hadamard differentiable at every $\mathbf{P} \in \mathcal{P}$, tangentially to $\mathcal{C}(\mathcal{F}, \mathbf{P})$.

$(HT2)$ T is Gâteaux differentiable at every $\mathbf{P} \in \mathcal{P} \cup \mathcal{P}_d$, along the directions $\delta_x - \mathbf{P}$.

We shall denote by $dT_{\mathbf{P}}$ the (Hadamard or Gâteau) differential at point $\mathbf{P}$. Under assumptions which are slightly weaker than $(HT1)$ and $(HT2)$, Bertail established a convergence result, for fixed $u \geq 1$ between $\mathcal{R}_{n,u}$ and the linearized regions

$$\mathcal{R}^L_{n,u} := \Big\{T(\mathbf{P}_0) + dT_{\mathbf{P}_0}\big(\widetilde{\mathbf{P}}_n - \mathcal{P}\big),\ \widetilde{\mathbf{P}}_n \in \overline{\mathcal{P}}_{n,u}\Big\}. \tag{1.8}$$

That convergence holds in probability, with respect to the followin (Haussdorff) distance :

$$\mathcal{D}(A,B) := \max\Big\{ \sup_{\theta \in A} \inf_{\theta' \in B} ||\,\theta - \theta'\,||, \sup_{\theta \in B} \inf_{\theta' \in A} ||\,\theta - \theta'\,|| \Big\},\ A, B \subset E. \tag{1.9}$$

Despite their asymptotic validities, $\mathcal{R}_{n,u_\alpha}$ and $\mathcal{R}^L_{n,u_\alpha}$ still present serious practical drawbacks. Indeed, the first region is not numerically viable, since, in full generality, the description of $\mathcal{R}_{n,u_\alpha}$ as the image of $\overline{\mathcal{P}}_{n,u_\alpha}$ by T requires a number of operations of order ϵ^{-n} (for a given precision ϵ). On the other hand, the linearized regions $\mathcal{R}^L_{n,u_\alpha}$, which satisfies the dual representation

$$\mathcal{R}^L_{n,u_\alpha} = \Big\{T(\mathbf{P}_0) + dT_{\mathbf{P}_0}\big(P - \mathbf{P}_0\big),\ \mathbf{EL}^L_n(P) \geq u_\alpha\Big\}, \text{ with} \tag{1.10}$$

$$\mathbf{EL}^L_n(P) := \max\Big\{ \prod_{i=1}^n np_i,\ (p_1, \ldots, p_n) \in \mathbb{S}_n,\ dT_{\mathbf{P}_0}\Big(\sum_{i=1}^n p_i \delta_{X_i} - P\Big) = 0 \Big\} \tag{1.11}$$

do not pose the just mentioned computational issue, since the linearity of $dT_{\mathbf{P}_0}$ and its finite rank allow a convex optimization algorithm. However, they are not confidence regions is the statistical sense, as $\mathbf{P}_0$ appears (twice) in (1.8). In this talk, we will first show that, under a the following assumption :

$(HT3)$ On the metric space $\big(\mathcal{P}', ||\cdot||_{\mathcal{F}}\big)$, the application $\mathbf{P} \mapsto dT_{\mathbf{P}}$ is continuous at every $\mathbf{P} \in \mathcal{P}$, under the following seminorm

$$|||\,L\,|||_{\mathbf{P}} := \sup_{\substack{\mathbf{P}' \in \mathcal{P}_d \\ ||\mathbf{P}'||_{\mathcal{F}} \leq 1}} ||\,L(\mathbf{P}' - \mathbf{P})\,||, \tag{1.12}$$

quite elementary arguments to build asymptotic confidence bands when plugging in the empirical measure in $\mathcal{R}^L_{n,u_\alpha}$, and hence defining

$$\mathcal{R}^{LL}_{n,u_\alpha} := \Big\{ T(\mathbf{P}_n) + dT_{\mathbf{P}_n}\Big(\widetilde{\mathbf{P}}_n - \mathbf{P}_n\Big),\ \widetilde{\mathbf{P}}_n \in \overline{\mathcal{P}}_{n,u_\alpha} \Big\}, \tag{1.13}$$

which is computationally feasible thanks to the dual representation

$$\mathcal{R}^{LL}_{n,u_\alpha} = \Big\{ T(\mathbf{P}_n) + dT_{\mathbf{P}_n}\big(P - \mathbf{P}_n\big),\ \mathbf{EL}^{LL}_n(P) \geq u_\alpha \Big\}, \text{ with} \tag{1.14}$$

$$\mathbf{EL}^{LL}_n(P) := \max\Big\{ \prod_{i=1}^n np_i,\ (p_1,\ldots,p_n) \in \mathbb{S}_n,\ dT_{\mathbf{P}_n}\Big(\sum_{i=1}^n p_i \delta_{X_i} - P \Big) = 0 \Big\}. \tag{1.15}$$

Theorem 1 *Denote by $\mathbf{P}_0$ the law of X_1. Then, under $(HT1)$, $(HT2)$ and $(HT3)$ we have, for fixed $\alpha \in (0,1)$*

$$\mathcal{D}\Big(\mathcal{R}^{LL}_{n,u_\alpha}, \mathcal{R}^L_{n,u_\alpha}\Big) \to_{\mathbb{P}}\ 0,\ \textit{as } n \to \infty. \tag{1.16}$$

Our second results aims toward extending the preceding result to functionals T taking values in a space of trajectories, say $\ell_\infty(\mathfrak{T})$. Many statistical problems exhibit an unknown function $\mu_0(\cdot)$ to be estimated, which can be written as a more or less complex functional of $\mathbf{P}_0$, upon which we can often make suitable differentiability and second order continuity. A typical example arises in survival analysis under independent censorship, where the survival function can be expressed through product integrals. As it will be shown, it is nevertheless possible derive a limit law for the processe $t \mapsto \boldsymbol{EL}^{LL}_n(\mathbf{P}_0, t)$ defined by

$$\boldsymbol{EL}^{LL}_n(\mathbf{P}, t) := \max\Big\{ \prod_{i=1}^n np_i,\ \Big[dT_{\mathbb{P}_n}\big(\sum_{i=1}^n p_i(\delta_{X_i} - \mathbf{P})\big)\Big](t) = 0 \Big\},\ \mathbf{P} \in \mathcal{P},\ t \in \mathfrak{T}. \tag{1.17}$$

To state our second result, it seems suitable to introduce the following notations

$$m_i(\mathbf{P}, t) := \Big[dT_{\mathbf{P}_0}\big(\sum_{i=1}^n p_i(\delta_{X_i} - \mathbf{P})\big)\Big](t), \tag{1.18}$$

$$\hat{m}_i(\mathbf{P}, t) := \Big[dT_{\mathbb{P}_n}\big(\sum_{i=1}^n p_i(\delta_{X_i} - \mathbf{P})\big)\Big](t)\Big\},\ \mathbf{P} \in \mathcal{P},\ t \in \mathfrak{T}. \tag{1.19}$$

Theorem 2 *Denote by $\mathbf{P}_0$ the law of X_1, and write $V(t) := \mathrm{Var}\big(m_i(\mathbf{P}_0, t)\big)$ for $t \in \mathfrak{T}$. Let T be an application taking values in $\ell_\infty(\mathfrak{T})$. Assume that $(HT1)$, $(HT2)$ and $(HT3)$ hold, and under other assumptions pertaining to empirical likelihood theory, there exists $\mathfrak{T}$-indexed, almost surely bounded Gaussian process for which*

$$-2\ln \boldsymbol{EL}^{LL}_n(\mathbf{P}_0, \cdot) \to_{\mathcal{L}}\ G^2(\cdot)V^{-1}(\cdot). \tag{1.20}$$

This result allows to build asymptotic confidence regions, which are computationally feasible. Indeed, the computational aspects reduce to a dichotomic search of the endpoint of an interval, for fixed t.

2 Conclusion

We provide a method for building asymptotic confidence regions, based on empirical likelihood, which is computationally feasible, and which we hope to be flexible enough to adapt to a large scale of nonparametric models. Some differentiability assumption need to be checked. They turn out to be satisfied for the first nonparametric models that may come in mind, such as centered moment estimations, or confidence bands for survival functions in survival analysis in the presence of censored data.

References

[1] Bertail, P., (2006). Empirical likelihood in some semiparametric models. *Bernoulli*, **12(2)**, 299–331, .

[2] Owen, A.B., (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.*, **18(1)**, 90–120.

[3] Owen, A.B, (2001). *Empirical Likelihood*. Chapman and Hall/CRC.

Non-parametric Bayesian drift estimation for stochastic differential equations

Shota Gugushvili[1] & Peter Spreij[2]

[1]Department of Mathematics, Vrije Universiteit, Amsterdam, The Netherlands
email: s.gugushvili@vu.nl
[2]Korteweg-de Vries Institute for Mathematics, Universiteit van Amsterdam, Amsterdam, The Netherlands
email: spreij@uva.nl

Abstract

In this talk we will discuss posterior consistency for non-parametric Bayesian estimation of a drift coefficient of a stochastic differential equation.

1 Extended abstract

Stochastic differential equations of the form

$$dX_t = b(X_t)dt + dW_t \tag{1}$$

driven by Brownian motion W play an important role for modelling purposes in numerous branches of science, with applications ranging from physics to mathematical finance. In this talk we assume that the soution X to (1) is initialised at its invariant distribution, so that it is ergodic and strictly stationary, and consider the problem of non-parametric Bayesian estimation of the drift coefficient b based on a discrete time sample $X_0, X_\Delta, \ldots, X_{n\Delta}$ from the solution X to (1) corresponding to the true drift coefficient b_0. This is a well-studied statistical problem in the frequentist literature. However, non-parametric Bayesian estimation of the drift coefficient b_0 is also possible, see e.g. the references in [2]. In particular, employing the Markov property of the solution X to (1), the likelihood corresponding to the observations $X_0, X_\Delta, \ldots, X_{n\Delta}$ can be written as

$$\pi_b(X_0)\prod_{i=1}^{n} p_b(\Delta, X_{(i-1)\Delta}, X_{i\Delta}),$$

where $\pi_b(x)$ denotes the density of X_0, while $p_b(t,x,y)$ are transition densities of X. A Bayesian would put a prior Π on the parameter space $\mathcal{X}$ and next obtain the posterior measure of any measurable subset $B \subset \mathcal{X}$ through Bayes' formula

$$\Pi(B|X_0,\ldots,X_{n\Delta}) = \frac{\int_B \pi_b(X_0)\prod_{i=1}^{n} p_b(\Delta, X_{(i-1)\Delta}, X_{i\Delta})\Pi(db)}{\int_{\mathcal{X}} \pi_b(X_0)\prod_{i=1}^{n} p_b(\Delta, X_{(i-1)\Delta}, X_{i\Delta})\Pi(db)}.$$

It has been argued convincingly in [1] and elsewhere that a desirable property of a Bayes procedure is posterior consistency. In our context this will mean that for every neighbourhood (in a suitable topology) U_{b_0} of b_0,

$$\Pi(U_{b_0}^c|X_0,\ldots,X_{n\Delta}) \to 0, \quad P_{b_0}\text{-a.s.}$$

as $n \to \infty$. That is, roughly speaking, a consistent Bayesian procedure asymptotically puts posterior mass equal to one on every fixed neighbourhood of the true parameter: the posterior concentrates around the true parameter.

In this talk we will show that under some classical assumptions on the class of drift coefficients coupled with a suitable condition on the prior Π, posterior consistency holds true in our model.

References

[1] Diaconis P. and Freedman D., (1986). On the consistency of Bayes estimates. With a discussion and a rejoinder by the authors. *Ann. Statist.*, **14**, 1–67.

[2] Gugushvili S. and Spreij P., (2012). Non-parametric Bayesian drift estimation for stochastic differential equations. arXiv:1206.4981 **[math.ST]**.

Abstract presentation MNM - Heverlee September 2012

Hyperbolic wavelet thresholding rules: the curse of dimensionality through the maxiset approach

Jean-Marc Freyermuth[*], *Florent Autin*[†], *Gerda Claeskens*[*], *Rainer von Sachs*[◇]

[*] ORSTAT and Leuven Statistics Research Center, K.U.Leuven, Belgium, E-mail: Jean-Marc.Freyermuth@econ.kuleuven.be

[†] LATP1, université d'Aix Marseille, 1, France.

[*] ORSTAT and Leuven Statistics Research Center, K.U.Leuven, Belgium.

[◇] ISBA, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

In this talk we are interested in nonparametric multivariate function estimation. In Autin et al. (2012), we determine the maxisets of several estimators based on thresholding of the empirical hyperbolic wavelet coefficients. That is we determine the largest functional space over which the risk of these estimators converges at a chosen rate. It is known from the univariate setting that pooling information from geometric structures (horizontal/vertical blocks) in the coefficient domain allows to get 'large' maxisets (see e.g Autin et al. (2011a,b,c)). We show that in the multivariate context the benefits of information pooling is versatile. In a sense these estimators are much more exposed to the curse of dimensionality while their potential can be preserved under certain conditions. Finally, in the spirit of Schneider and von Sachs (1996), Neumann and von Sachs (1997), we propose an application in the field of time series analysis through the ability of such methods to adaptively estimate an evolutionary spectrum of a locally stationary process by 2-D wavelet smoothing over time and frequency.

References

Autin, F., Claeskens, G., Freyermuth, J-M. (2012). Hyperbolic wavelet thresholding rules: the curse of dimensionality through the maxiset approach. *Discussion Paper*.

Autin, F., Freyermuth, J-M., von Sachs, R. (2011a). Ideal denoising within a family of Tree Structured Wavelets. *Electron. Journ. of Statist.*, vol. 5, 829-855.

Autin, F., Freyermuth, J-M., von Sachs, R. (2011b). Block-threshold-adapted estimators via a maxiset approach. *Submitted.*

Autin, F., Freyermuth, J-M., von Sachs, R. (2011c) Combining thresholding rules: a new way to improve the performance of wavelet estimators. *to appear in the Journal of Nonparametric Statistics*.

Schneider, K., von Sachs, R. (1996). Wavelet smoothing of evolutionary spectra by non-linear thresholding. *Appl. Comput. Harmon. Anal.*, **3**, 268-282.

Neumann, H., von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. of Statist.*, **25**, 38-76.

Crossvalidation selection for the smoothing parameter in nonparametric hazard rate estimation

Francois van Graan[1]

[1]School for Mathematical, Statistical and Computer Sciences, North West University, Potchefstroom Campus, Potchefstroom, 2520
email: Francois.VanGraan@nwu.ac.za

Abstract

An improved version of the crossvalidation bandwidth selector is proposed by using the technique of bootstrap aggregation. Simulation results show that this selector compares very favorably to a bootstrap bandwidth selector based on an asymptotic representation of the mean weighted integrated squared error for the kernel-based estimator of the hazard rate function in the presence of right-censored samples.

1 Introduction

In survival analysis and reliability, the random variable of interest X^0 is called the failure time. It represents the time to some event, called a failure. A curve that provides useful information about the failure time is the hazard rate function (instantaneous failure time),

$$r_0(x) = \lim_{h \to 0} \frac{P(x \leq X^0 < x + h | X^0 > x)}{h}$$

When estimation is done nonparametrically, kernel estimators can be used for estimating r_0. Also, complications arises when complete samples from the failure time population cannot be observed, due to the intervention of another random variable called a censoring variable. In this presentation a model which allows for random right censorship will be considered.

Nonparametric estimation of quantities such as r_0 by kernel methods imply specification of a smoothing parameter . Two methods for estimating the smoothing parameter have been proposed in the literature. A version of crossvalidation for smoothing parameter selection in the censored hazard case was proposed by Patil (1993). This method can be considered as an extension of the method proposed by Sarda and Vieu (1991) in the case of uncensored hazard function estimation. An improved method of bandwidth selection was proposed by Cao et al., (1996) . This method is based on ideas from smoothed bootstrap methodology. Explicit expressions of the asymptotic mean integrated squared error are obtained and minimized to find an estimated bandwidth. It is shown that this method performs better than the method of crossvalidation.

Asymptotic optimality results show that the rate of convergence for crossvalidation is rather slow. This can be ascribed to the fact that crossvalidation shows greater variability than other methods of smoothing parameter selection. Hall and Robinson (1992) show that the stochastic variability of crossvalidation can be reduced by using bootstrap aggregation, or bagging. In the nonparametric density estimation setting, they developed theory which showed that when bagging is implemented using an adaptively chosen sample size, the variability of crossvalidation can be reduced by an order of magnitude.

In this presentation the objective is to extend this method to nonparametric hazard rate estimation when kernel methods that depend on the choice of a smoothing parameter are employed. An algorithm for choosing the smoothing parameter will be introduced and the results from a numerical study will show that the variability of crossvalidation can be reduced by implementing bagged crossvalidation.

References

[1] Cao, R.,Gonzalez-Manteiga, W., and Marron J.S., (1996). Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation. *Journal of the American Statistical Association*, **17**, 1130–1140.

[2] Hall,P. and Robinson, A.P., (2009). Reducing variability of crossvalidation for smoothing-parameter choice. *Biometrika*, **96**, 175–186.

[3] Patil,P.N., (1993). Bandwidth choice for nonparametric hazard rate estimation. *Journal of Statistical Planning and Inference*, **35**, 15–30.

Aggregating Methods for High-dimensional inference of Survival Data

Kristiaan Pelckmans, Liu Yang

July 16, 2012

This presentation reviews new theoretical and practical results in the inference from high-dimensional data originating from survival studies. The reviewed techniques are closely related to the recently proposed class of aggregating techniques and the study of online learning from experts. The techniques are verified on real data originating from micro-array expression studies, and from Genome-Wide Association data.

Learning how to make a prognosis of a patient is an important ingredient to the task of building an automatic system for personalized medical treatment. A prognosis here is understood as a useful characterization about the (future) time of an interesting event. In cancer research, a typical event is the relapse of a patient after receiving treatment. Since this involves essentially a form of prediction, it is naturally to phrase this problem in a context of machine learning.

1 Aggregating Estimators

Present research of the authors extends the class of aggregating methods towards problems of survival analysis. The set of methods of aggregation, averaging and mirror descent goes back to Nemirovski's work [NY1983], and the present derivations go much along the lines as set out in [JRT2008]. A strong feature of this method is that they scale up logarithmic in the dimensionality of the problem, making it possible to process (statistically) efficiently high-dimensional data. The surprising property is that this methods can obtain convergence rates in terms of $O(1/n)$, where n is the number of samples. This is especially startling since the methods are computationally very simple and do not boil down explicitly to an optimization problem. Application of this framework to the setting of survival analysis and use of the Concordance Index (CI) leads us to the formulation of the algorithm which we denote as 'Aggregation for Prognosis Through Exponential weighting' (APTER). In case of adopting the proportional hazard assumption and using a partial likelihood criterion [VH94], a similar averaging scheme is obtained which obtains the fast rates of convergence.

2 An Online Doctor

The techniques underlying the aggregating approaches are similar to the methods employed by researchers in the online learning from experts framework. Here, one is not so much after inference, but aims to design an algorithm which makes good predictions *along the way*. That is, every time a new data sample is presented, a prediction is required. The total quality of the algorithm is judged by the errors of those predictions. Recent interest in this framework adopts a deterministic view, and are based on advances since the introduction of the WINNOW and EG algorithms. It is an interesting feat that the analysis driving this research draws from the set of techniques which are used heavily in nonparametric statistics and empirical process theory.

In recent work we extend this methods towards a context of survival analysis. The idea is that

the doctor has to perform both exploration while treating patients to the best he can. This trade-off is treated extensively in research on (adversary) multi-armed bandits as in [ACFS2003], a theory which relates closely to our framework.

3 Studies in Bio-informatics

The methods above are tuned towards applications of survival analysis in bio-informatics [BBHL2009]. Firstly, we revisit the analysis available to studies of micro-array expression signatures in the context of breast cancer. Therefor, we draw on results presented in [BT2004, BNSABFL2007, VPS2011, K2010]. Those tasks are already challenging as the high-dimensionality of the data renders classical tools impractical or ineffective. Penalization helps us, but to the price of computational complexity. Furthermore, we argue how this technique improves over earlier approaches

In Genome-Wide Association studies (GWAs), those challenges are even more prominent. Here, one has to deal with $O(10^5)$-dimensional data [HS2008]. We point out how the proposed techniques help us to exploit structure as well as variation in the data, and how the results can be interpreted in a biological context.

References

[NY1983] Nemirovsky, A.S. and Yudin, D.B. Problem Complexity and Method Efficiency in Optimization. Wiley, (1983).

[CL2006] Cesa-bianchi N., Lugosi, G. Prediction, learning and games. Cambridge University Press, 2006.

[BBHL2009] Bickel, P., Brown, J., Huang, H. and L. Qunhua, An Overview of Recent Developments in Genomics and the Statistical Methods that Bear on Them? Phil. Trans. Royal Society, A, 4313–4337, 2009.

[JRT2008] A. Juditsky, P. Rigollet, and A. B. Tsybakov Learning by mirror averaging Ann. Statist. Volume 36, Number 5, 2183–2206, 2008.

[HS2008] Hardy J., Singleton A. "Genomewide association studies and human disease". N. Engl. J. Med. 360 (17): 1759–1768, 2009.

[K2010] Koscielny, S., Why Most Gene Expression Signatures of Tumors Have Not Been Useful in the Clinic. Science Translational Medicine 2, 14ps2, 2010.

[BNSABFL2007] Bøvelstad, HM and Nygård, S. and Størvold, H.L. and Aldrin, M. and Borgan, Ø. and Frigessi, A. and Lingjærde, O.C. Predicting survival from microarray data - a comparative study, Bioinformatics, 23(16), 2080–2087, 2007.

[VPS2011] Van Belle, V. and Pelckmans, K. and Suykens, J.A.K. and Van Huffel, S. Learning transformation models for ranking and survival analysis, Journal of Machine Learning Research, pp. 819–862, 2011.

[VH94] Penalized likelihood in cox regression, Verweij, P.J.M. and Van Houwelingen, H.C., Statistics in Medicine, 13(23-24), pp. 2427–2436, 1994.

[BT2004] Bair E., Tibshirani, R., Semi-supervised methods to predict patient survival from gene expression data. PLoS Biol., 2(4), 2004.

[ACFS2003] Auer, P. and Cesa-Bianchi, N. and Freund, Y. and Schapire, R.E. The non-stochastic multiarmed bandit problem. SIAM Journal on Computing, 32(1), 2003.

Reducing the mean squared error of quantile-based estimators by smoothing

Mia Hubert[1] & Irène Gijbels[1] & Dina Vanpaemel[1]

[1]Department of Mathematics & LStat, KU Leuven

email: Mia.Hubert, Irene.Gijbels, Dina.Vanpaemel @wis.kuleuven.be

Abstract

Many univariate robust estimators are based on quantiles. As already theoretically pointed out by [2], smoothing the empirical distribution function with an appropriate kernel and bandwidth can reduce the variance and mean squared error (MSE) of some quantile-based estimators in small data sets. In this talk we will explain how this idea can be applied on several robust estimators of location, scale and skewness [4]. We will explain how to select the bandwidth robustly and how to reduce the bias. We will show trough a simulation study that the use of this smoothing method indeed leads to smaller MSEs, also at contaminated data sets. In particular, better performances for the medcouple are obtained, which is a robust measure of skewness that can be used for outlier detection in skewed distributions [1]. The smoothed medcouple can be used to adapt the adjusted boxplot [3] to detect outliers in small data set of which the underlying distribution is skewed. This will be illustrated on a real data example.

References

[1] Brys, G., Hubert, M., Struyf, A., (2004). A robust measure of skewness. *Journal of Computational and Graphical Statistics*, **13**, 996–1017.

[2] Fernholz, L.T., (1997). Reducing the variance by smoothing. *Journal of Statistical Planning and Inference*, **57**, 29–38.

[3] Hubert, M., Vandervieren, E., (2008). An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis*, **12**, 5186–5201.

[4] Hubert, M., Gijbels, I., Vanpaemel, D., (2012). Reducing the mean squared error of quantile-based estimators by smoothing. *Test*, doi: 10.1007/s11749-012-0293-3.

Component Selection in Additive Quantile Regression Models

Hohsuk Noh[1] & Eun Ryung Lee[2]

[1]Université catholique de Louvain
email: word5810@gmail.com
[2]University of Mannheim
email: silverryuee@gmail.com

Abstract

Nonparametric additive models are powerful techniques for multivariate data analysis. Although many procedures have been developed for estimating additive components both in mean regression and quantile regression, the problem of selecting relevant components has not been addressed much especially in quantile regression. In this article, we present a doubly-penalized estimation procedure for component selection in additive quantile regression models that combines basis function approximation with a variant of the smoothly clipped absolute deviation penalty and a ridge-type penalty. We show that the proposed estimator identifies relevant and irrelevant components consistently and achieves the nonparametric optimal rate of convergence for the relevant components. We also provide some numerical evidence of the estimator, and illustrate its usefulness through a real data example to identify important body measurements to predict percentage of body fat of an individual.

1 Introduction

Suppose that Y is a response of interest which depends on a vector of random covariates $\mathbf{X} = (X^1, \ldots, X^p) \in \mathbb{R}^p$, $p \geq 2$. We are interested in the conditional quantile of Y given $\mathbf{X}$. The relationship between Y and $\mathbf{X}$ can be modeled as

$$Y = Q_\tau(\mathbf{X}) + U_\tau,$$

where $Q_\tau(\cdot)$ is an unknown real-valued function and U_τ is an unobserved random variable whose τth quantile conditional on $\mathbf{X} = \mathbf{x}$ is 0 for almost every $\mathbf{x}$. Many nonparametric methods are available to estimate $Q_\tau(\cdot)$ yet they are all severely affected by the so-called curse of dimensionality when p is large. To tackle this issue, several authors have proposed dimensional reduction methods, a popular one of which is to consider the following model based on the additive structure of $Q_\tau(\mathbf{X})$:

$$Y_i = Q_\tau(\mathbf{X}_i) + U_{i,\tau} = \mu_\tau + \sum_{k=1}^{p} m_{k,\tau}(X_i^k) + U_{i,\tau} \quad (1)$$

where $(Y_i, \mathbf{X}_i, U_{i,\tau}) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}$ for $1 \leq i \leq n$ are independent and identically distributed and $P(U_{i,\tau} \leq 0 | \mathbf{X}_i = \mathbf{x}_i) = \tau$ for almost every $\mathbf{x}_i$. To fit this additive quantile regression model (1), there have been many proposals including [3], [10], [6], [8] and [2]. Most of them have more or less similar implication that it is possible to estimate the additive component functions in multi-dimension settings at the same accuracy as in univariate cases. However, all these works are based on the assumption that the given covariates are all relevant to the conditional quantile of interest, which is not always the case in practice. Motivated by this concern, we consider the problem of selecting the relevant covariates in the model (1).

Although there have been a few works in the literature which deal with problem of covariate selection in additive mean regression framework such as [7], [9] and [4], to the best of our knowledge such works do not exist in the context of quantile regression. The only exception that we have found is the empirical analysis of [5] about child malnutrition data in India. They adapted an idea of [1] (in framework of boosting with squared error loss) for covariate selection in additive mean regression models to their semiparametric quantile regression models with additive structures in order to identify the relevant factors (covariates) in the data, but did not discuss its statistical properties. Contrary to them, in this paper we not only propose a method to identify the relevant covariates in the model (1) and estimate the corresponding additive component functions, but also derive its theoretical properties. Because the dependence pattern between the response and covariates can be different across the quantile levels in quantile regression, our work has quite different implications from the works for covariate selection in additive mean regression models. From a point of view of component selection, in a regression model with a heteroscedastic error, the set of relevant component functions for the conditional quantile of interest

can vary across the quantile levels. This is the characteristic feature of component selection in quantile regression, which differentiate itself from that in mean regression. It is the main motivation for this work.

The main idea of component selection in our work is that the magnitude of the L_2-norm of m_k is closely related to relevance of the covariate X^k. Motivated by this observation, we propose two-stage procedure for estimating $m_{k,\tau}(\cdot)$ simultaneously with the identification of irrelevant covariates. The initial estimator is computed for estimating $\|m_k\|_{L_2}$. Then we obtain the final estimator which automatically discards presumably irrelevant component functions by referencing to the estimates of the L_2 norm of each component, $\|\tilde{m}_k\|_{L_2},\ k = 1, \ldots, p$, in a lasso-type penalized framework with basis approximation. This work can be seen as an extension of the idea of [12] and [11] for parametric linear models to nonparametric additive quantile regression models. Since we focus on the problem of component selection, we use our initial estimator only in the first stage for estimating $\|m_k\|_{L_2}$. However, we found that the initial estimator shows more stable numerical performance than the estimator discussed by [6] for the same model with basis approximation. We will discuss further the comparison between our initial estimator and the one of [6].

2 Conclusion

This article has developed a doubly-penalized estimation for nonparametric additive quantile regression models that can simultaneously perform component selection and estimation of smooth component functions. We have established the theoretical properties of our procedure, including consistency in component selection and the nonparametric optimal rate of convergence in estimation. Different from the works concerned with component selection in the literature, we introduced an extra penalty that controls the L_2-norms of component functions for the stability of the estimation using the idea of P-spline, independently from the penalty for component selection. Finally, we have shown that in the numerical study our doubly-penalized estimator performs more stably when the covariates are fairly correlated than a singly-penalized one.

References

[1] Bühlmann, P. and Yu, B. (2003). Boosting with the l_2 loss:regression and classification. *Journal of the American Statistical Association*, **98**, 324–339.

[2] Cheng, Y., De Gooijer, J. G. and Zerom, D. (2011). Efficient Estimation of an Additive Quantile Regression Model. *Scandinavian Journal of Statistics*, **38**, 46–62.

[3] De Gooijer, J. G. and Zerom, D. (2003). On Additive Conditional Quantiles With High-Dimensional Covariates. *Journal of the American Statistical Association*, **98**, 35–146.

[4] Fan, J., Feng, Y. and Song. R. (2011). Nonparametric Independence Screening in Sparse Ultra-High-Dimensional Additive Models. *Journal of the American Statistical Association*, **106**, 544–557.

[5] Fenske, N., Kneib, T. and Hothorn, T. (2011). Identifying Risk Factors for Severe Childhood Malnutrition by Boosting Additive Quantile Regression. *Journal of the American Statistical Association*, 106, 494–510.

[6] Horowitz, J. L. and Lee, S. (2005). Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association*, **100**, 1238–1249.

[7] Huang, J. Z. and Yang, L., (2004). Identification of non-linear additive autoregressive models. *Journal of the Royal Statistical Society*, **B66**, 463–477.

[8] Lee, Y. K., Mammen, E. and Park, B. U. (2010). Backfitting and smooth backfitting for additive quantile models. *Annals of Statistics*, **38**, 2857–2883.

[9] Xue, L. (2009). Variable selection in additive models. *Statistica Sinica*, **19**, 1281–1296.

[10] Yu, K. and Lu, Z. (2004). Local Linear Additive Quantile Regression. *Scandinavian Journal of Statistics*, **31**, 333–346.

[11] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, **101**, 1418–1429.

[12] Zou, H. and Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Annals of Statistics*, **36**, 1509–1533.