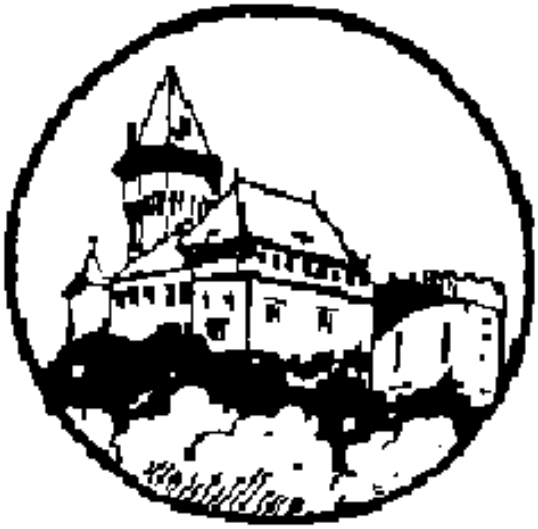
PROBASTAT 2015

The Seventh International Conference on Probability and Statistics



ABSTRACTS

June 29 - July 3, 2015, Smolenice Castle, Slovak Republic

PROBASTAT 2015
The Seventh International Conference on Probability and Statistics
June 29 – July 3, 2015
Smolenice Castle, Slovak Republic

ORGANIZERS
Institute of Measurement Science of the Slovak Academy of Sciences
Faculty of Mathematics, Physics and Informatics, Comenius University
Mathematical Institute of the Slovak Academy of Sciences

IN COOPERATION WITH
The Union of Slovak Mathematicians and Physicists

PROBASTAT 2015
Abstracts

SPONSORED BY
SAS Institute, Slovakia

EDITORS
Viktor Witkovský & Ján Somorčík
PUBLISHER
Institute of Measurement Science
Slovak Academy of Sciences, Bratislava, Slovak Republic
Bratislava, June 2015

Contents

Invited Talks

MARGINAL AND JOINT REGRESSION MODELS FOR CLUSTERED DATA INFERENCE WHEN THE CLUSTER SIZE IS POTENTIALLY INFORMATIVE

SOMNATH DATTA
University of Louisville, USA

We discuss how to extend parametric and nonparametric inference procedures when the classical assumption of independence is violated due to clustering. Clustered data arise in a number of practical applications where observations belonging to different clusters are independent but observations within the same cluster are dependent. While making adjustments for possible cluster dependence, one should also be aware of the "informative cluster size" phenomenon which occurs when the size of the cluster is a random variable that is correlated to the outcome distribution within a cluster, often through a cluster specific latent factor. We demonstrate and compare two correct inference procedures in a longitudinal linear model setting with temporally varying cluster size.

UNOBSERVED HETEROGENEITY IN INCOME DYNAMICS: AN EMPIRICAL BAYES PERSPECTIVE

ROGER KOENKER and JAILING GU
University of Illinois, Urbana–Champaign, USA

Empirical Bayes methods for Gaussian compound decision problems involving longitudinal data are considered. The new convex optimization formulation of the nonparametric (Kiefer–Wolfowitz) maximum likelihood estimator for mixture models is employed to construct nonparametric Bayes rules for compound decisions. The methods are first illustrated with some simulation examples and then with an application to models of income dynamics. Using PSID data we estimate a simple dynamic model of earnings that incorporates bivariate heterogeneity in intercept and variance of the innovation process. Profile likelihood is employed to estimate an AR(1) parameter controlling the persistence of the innovations. We find that persistence is relatively modest, $\hat{\varrho} \approx 0.48$, when we permit heterogeneity in variances. Evidence of negative dependence between individual intercepts and variances is revealed by the nonparametric estimation of the mixing distribution, and has important consequences for forecasting future income trajectories.

GOODNESS-OF-FIT TESTS FOR SEMIPARAMETRIC AND PARAMETRIC HYPOTHESES BASED ON THE PROBABILITY WEIGHTED CHARACTERISTIC FUNCTION

SIMOS MEINTANIS
University of Athens, Greece
and
JAMES SAMUEL ALLISON and LEONARD SANTANA
North–West University, Potchefstroom, South Africa

We suggest certain procedures for testing symmetry and multinormality in the context of structured data based on the novel notion of the probability weighted characteristic function and investigate their finite sample behavior.

DATA DEPTH FOR AUTOREGRESSION WITH APPLICATION TO CRACK GROWTH

CHRISTINE HILDEGARD MÜLLER
Technical University of Dortmund, Germany

A deterministic model of crack growth is given by the Paris–Erdogan equation. The stochastic version of this equation leads to a stochastic differential equation which can be approximated by an $AR(1)$ model. Since crack growth has typically jumps which are innovation outliers, the $AR(1)$ model shall be analyzed by an outlier robust method. We propose to use data depth for this analysis and consider in this talk four types of data depth based on simplicial depth.

PROJECTIVE SHAPE ANALYSIS OF 3D CONTOURS

VIC PATRANGENARU
Florida State University, Tallahassee, USA

Projective shapes of planar curves in images have been studied in the context of functional data analysis, using projective frame representations of projective shapes of planar curves. Tests have been derived for testing neighborhood hypotheses for the one-and multi-sample problem for functional data by Munk et al. (2008), where the one-sample test was applied to the problem of identifying the projective shape of a planar curve. It would be useful to pursue testing and estimation using the Hartley and Zisserman (2004) approach, also to analyze mean 3D projective shapes of spacial contours, which is the objective of our presentation. Our approach follows the methodology of data analysis on Hilbert manifolds in Ellingson et al. (2013).

Acknowledgements

The authors acknowledge support from National Security Agency and from the National Science Foundation.

A MEASURE OF DISPERSION BASED ON AVERAGE EXPONENTIATED VOLUMES, WITH APPLICATION IN EXPERIMENTAL DESIGN

LUC PRONZATO
University of Nice Sophia Antipolis, France
HENRY PHILIP WYNN
The London School of Economics and Political Science, UK
and
ANATOLY A. ZHIGLJAVSKY
Cardiff University, UK

We consider a measure of dispersion in dimension d which is based on the mean exponentiated volume of k-dimensional simplices formed by $k+1$ independent copies, with k less than or equal to d. Mean squared volume is related to the $(d-k)$-th coefficient of the characteristic polynomial of the covariance matrix and forms an extension of the notion of Wilk's generalised variance. We prove its concavity when raised at power $1/k$, and some properties of dispersion-maximising distributions are derived, including a necessary and sufficient condition for optimality. The application of this measure of dispersion to the design of optimal experiments for parameter estimation is considered, with A and D-optimal design coinciding with the particular cases

obtained for $k = 1$ and $k = d$ respectively. Mean of volumes raised to some power different from two, including negative values, will be briefly considered too, with application to space-filling design for computer experiments.

SOME PROPERTIES OF LINEAR SUFFICIENCY

SIMO PUNTANEN
University of Tampere, Finland

A linear statistic Fy, where F is an $f \times n$ matrix, is called linearly sufficient for the Xb under the model $A = \{y, Xb, V\}$, if there exists a matrix A such that AFy is the BLUE for Xb. In this talk we review some important aspects of the linear sufficiency and consider the relations of BLUE of the estimable parametric functions under the linear model A and its counterpart $B = \{Fy, FXb, FVF'\}$, obtained by transforming A by the matrix F. Some possible misunderstandings will also be discussed.

References

[1] BAKSALARY, J. K. and KALA, R. (1981). Linear transformations preserving best linear unbiased estimators in a general Gauss–Markoff model. *Annals of Statistics* **9** 913–916.
[2] BAKSALARY, J. K. and KALA, R. (1986). Linear sufficiency with respect to a given vector of parametric functions. *Journal of Statistical Planning and Inference* **14** 331–338.
[3] DRYGAS, H. (1983). Sufficiency and completeness in the general Gauss–Markov model. *Sankhyā Ser. A* **45** 88–98.
[4] KALA, R., PUNTANEN, S. and TIAN, Y. (2015). Some notes on linear sufficiency. *Statistical Papers*, available online.

SOME ASPECTS OF HIGH DIMENSIONAL AND FUNCTIONAL REGRESSION ANALYSIS

PASCAL SARDA
University of Toulouse, France

We present several aspects of high dimensional regression analysis with a special focus on Functional Linear Regression. In the general setting the number of predictor variables is larger than the sample size and in special (functional) cases these regressors represent discretizations (for instance at different observations times) of a same curve. We present in a first step some selected results on Functional Linear Regression. Coming back to the general case, we compare, always for linear regression, functional concepts to variable selection based on the hypothesis

of sparseness of the coefficient of parameters and which is the main alternative modelization for this situation. Roughly speaking the first approach, respectively the second one, analyses influence of common, respectively specific, variabilities of the predictors on the response. It is shown that both ideas can be combined in the framework of factor models where the predictor vector is decomposed in a sum of two components reflecting specific and common variabilities of explanatory variables. An augmented linear model is thus proposed using principal components as additional explanatory variables. Then usual estimators of true principal components and variable selection procedures are used to estimate parameters of the augmented model. In the last part of the talk, we consider another point of view. Indeed, in the context of functional linear regression, we show that the functional explanatory variable may posses specific local variations. Provided that this hypothesis is satisfied, is is possible to consider that some points of impact of the predictive curve have specific influence on the response. A method for estimating the number and location of points of impact is then proposed.

MIXED EFFECTS MODELS WITH CENSORED RESPONSE FOR LONGITUDINAL HIV STUDIES

FLORIN VAIDA
University of California, San Diego, USA

In this talk we discuss inference for linear and non-linear mixed effects models with censored response. In a parametric framework, the computation of the maximum likelihood and restricted maximum likelihood for linear and non-linear mixed effects models with censored response is frustrated by the double integral in the likelihood over the random effects and over the censored response. We will present a Monte Carlo EM algorithm and a closed-form EM algorithm. For the latter, the E-step expressions rely on formulas for the mean and variance of a truncated multinormal distribution, and can be computed using available software. This leads to an improvement in the speed of computation of up to an order of magnitude.

A wide class of mixed effects models is considered, including the Laird-Ware model, and extensions to different structures for the variance components, heteroscedastic and autocorrelated errors, and multilevel models. We apply the methodology to three case studies from our own biostatistical practice, involving the analysis of longitudinal HIV viral load in recent AIDS studies.

A BALLOONED BETA-LOGISTIC MODEL WITH A BIOASSAY APPLICATION

MIN YI and NANCY FLOURNOY
University of Missouri, Columbia, USA

The beta distribution is a simple and flexible model in which responses are naturally confined to the finite interval $(0, 1)$. The parameters of the distribution can be related to covariates such

as concentration and gender through a regression model. The Ballooned Beta-logistic model, with expected responses equal to the Four Parameter Logistic model, is introduced. It expands the response boundaries of the beta regression model from $(0, 1)$ to (L, U), where L and U are unknown parameters. Under the Ballooned Beta-logistic model, expected responses follow a logistic function, but it differs from the classical Four Parameter Logistic model, which has normal additive normal errors, with positive probability of response from $-\infty$ to $+\infty$. In contrast, the Ballooned Beta-logistic model may have skewed responses with smaller response variances at more extreme covariate values and symmetric responses with relative large variance at central values of the covariate. It may also have monotone increasing or decreasing variances depending on parameter values. These features are common in bioassay data at different concentrations. The asymptotic normality of maximum likelihood estimators is proved even though the support of this non-regular regression model depends on unknown parameters. Given enzyme-linked immunosorbent assay data from different plates, the motivating validation objective is to set boundary criteria for estimates of L and U, after which plates with boundary estimates outside these limits would be considered "reference failures". We find maximum likelihood and least squares estimates converge faster to L and U than do extreme values at the minimum and maximum concentrations. We also find maximum likelihood estimators perform better than least squares estimators when the covariate range is not sufficiently wide.

Short Contributions & Posters

A DISCRIMINATIVE STUDY OF THE MODELS FOR THE FLOW OF PARTICLES

MARIANO AMO–SALAS and ELVIRA DELGADO–MÁRQUEZ
University of Castilla–La Mancha, Ciudad Real, Spain
LENKA FILOVÁ and RADOSLAV HARMAN
Comenius University, Bratislava, Slovakia
and
JESÚS LÓPEZ–FIDALGO
University of Castilla–La Mancha, Ciudad Real, Spain

During the discharge of a two-dimensional silo, the flow of grains through an opening is arrested if the size of the outlet is not large enough. In the outpouring of grains, jamming occurs due to the formation of an arch at the outlet. After breaking the arch, the grains fall until a new arch is formed.

Several models have been proposed to explain this process. In this work, D-optimal designs have been computed for four models, two with 2 unknown parameters and two with 3 unknown parameters, proposed by Janda et al. (2008) and To (2005). In addition, using the methodology of KL-optimality developed by López–Fidalgo et al. (2007), optimal designs for discriminating between models have been computed.

References

[1] JANDA, A., ZURIGUEL, I., GARCIMARTÍN, A., PUGNALONI, L. A. and MAZA, D. (2008). Jamming and critical outlet size in the discharge of a two-dimensional silo. *EPL* **84** 44002.

[2] LÓPEZ–FIDALGO, J., TOMMASI, C. and TRANDA, P. C. (2007). An optimal experimental design criterion for discriminating between non-normal nodels. *Journal of the Royal Statistical Society, Series B* **69** 231–242.

[3] TO, K. (2005). Jamming transition in two-dimensional hoppers and silos. *Physical Reviews E* **71** 060301.

ESTIMATION OF VOLUME OF NON-CONVEX LEVEL-SETS

JAROMÍR ANTOCH and DANIEL HLUBINKA
Charles University, Prague, Czech Republic

Consider a function $\psi : \mathbb{R}^d \times \mathcal{P} \to [0, \infty)$, where $\mathcal{P}$ is a given class of probability distributions on $\mathbb{R}$, and the volume $\lambda(L(P, a, e))$ (λ is the Lebesgue measure) of its level sets

$$L(P, a, e) = \{x : \psi(x, P) \in [a - e, a + e]\}.$$

Let us assume that we are given a random sample $X_1, \ldots, X_n$ from distribution P. The problem is to estimate the volume of $L(P, a, e)$ based on this random sample.

Examples of function ψ may be data depth function, probability density function, test function or some utility function. The need for estimating the volume of the level sets arise, e.g., in classification, for test purposes, in minimum volume sets problem and other.

AN ALTERNATIVE ESTIMATION METHOD FOR SHARED FRAILTY MODELS

ÖYKÜM ESRA AŞKIN
Yıldız Technical University, Istanbul, Turkey
DENIZ İNAN
Marmara University, Istanbul, Turkey
and
ALI HAKAN BÜYÜKLÜ
Yıldız Technical University, Istanbul, Turkey

Standard survival techniques such as proportional hazards model are lack of overcome the unobserved heterogeneity in study population. Frailty models provide an alternative way in order to account for heterogeneity caused by unobservable risk factors by adding random effects. Although vast studies have been done on estimation procedures, Evolutionary Algorithms (EAs) haven't received much attention in both parametric and semi-parametric frailty studies. In this study, the performance of maximum likelihood estimation (MLE) with PSO is compared with the performances of MLE with quasi-Newton and conjugate gradient methods. An extensive simulation study is conducted. Also, the proposed estimation method is implemented to a well-known real data set.

SCHUR-OPTIMAL AUGMENTATION OF BLOCK DESIGNS WITH THE BLOCK SIZE TWO

ALENA BACHRATÁ
Comenius University, Bratislava, Slovakia

Consider a situation, where we already made an experiment using optimal design and then we get more resources to use, and therefore, more observations to make. Our aim is to find the Schur-optimal design (optimal with respect to the class of all Schur isotonic criteria), which is an augmentation of the initial design.

We focus on block experiments with the block size two, which can be represented by concurrence graphs. Due to that, we can use knowledge from graph theory, specifically two methods by Kelmans [1] and Constantine [2]. We find new classes of Schur-optimal augmented designs using extensions of these methods. Namely, we find optimal augmentations of designs, whose concurrence graphs are complete graphs, regular complete multipartite graphs, and star graphs.

References

[1] KELMANS, A. K. (1965). The number of trees of a graph I (in Russian). *Avtomatika i Telemekhanika* **26** 2194–2204.

[2] CONSTANTINE, G. M. (1983). Schur convex functions on the spectra of graphs. *Discrete Mathematics* **45** 181–188.

BARYCENTRIC ALGORITHM FOR COMPUTING D-OPTIMAL SIZE-AND-COST CONSTRAINED DESIGNS OF EXPERIMENTS

EVA BENKOVÁ
Johannes Kepler University, Linz, Austria
and
RADOSLAV HARMAN
Comenius University, Bratislava, Slovakia

In this paper, we study the problem of D-optimal experimental design under two linear constraints, which can be interpreted as simultaneous restrictions on the size and on the total cost of the experiment. For computing a size-and-cost constrained approximate D-optimal design, we propose a "barycentric" algorithm with sequential removal of redundant design points. We analytically prove convergence results about the barycentric algorithm and numerically demonstrate its favourable properties compared to competing methods.

OPTIMAL DESIGN OF EXPERIMENTS VIA LINEAR PROGRAMMING

KATARÍNA BURCLOVÁ and ANDREJ PÁZMAN
Comenius University, Bratislava, Slovakia

We investigate the possibility of extending some results of Pázman and Pronzato (2014) to a larger set of optimality criteria. By a direct use of algebraic relations we show, that the criterion of D- and A-optimality and the class of E_k-criteria can be rewritten in a form:

$$\phi(\xi) = \min_{\mu \in \Xi} \sum_{x \in \mathcal{X}} H(\mu, x)\xi(x)$$

with given $H(\cdot, \cdot)$ for any $\xi \in \Xi$, where Ξ is the set of all design measures on a design space $\mathcal{X}$. Therefore, the task of finding the optimal (maximum) design can be reformulated as a linear programming (LP) problem with infinitely many constraints. Moreover, by simple modification, we can also reformulate the "criterion-robust" problem of maximizing the minimum efficiency in the class of all orthogonally invariant criteria (cf. Harman (2004)).

For computation, the design space $\mathcal{X}$ is supposed to be finite. The proposed algorithms are iterative and solve an LP problem at each step (a modification of the cutting-plane algorithm, cf. Kelley (1960)). The algorithm also allows us to add some supplementary (cost) constraints linear in ξ. Further, we can easily combine optimality criteria, i.e. construct, say, D-optimal design under the condition that the A-optimality criterion attains a prescribed value. So, we obtain reasonable values of both D- and A- efficiency. An extension of the approach to AVE optimality criteria in nonlinear model is straightforward. The ability to perform the computation of approximate optimum designs has been checked on examples.

References

[1] PÁZMAN, A. and PRONZATO, L. (2014). Optimum design accounting for the global nonlinear behavior of the model. *Annals of Statistics* **42**, 1426–1451.

[2] HARMAN, R. (2004). Minimal efficiency of designs under the class of orthogonally invariant information criteria. *Metrika* **60**, 137–153.

[3] KELLEY, J. (1960). The cutting plane method for solving convex programs. *Journal of the Society for Industrial and Applied Mathematics* **8**, 703–712.

A NEW STRUCTURAL EQUATION MODELING BASED METHOD FOR CONDITIONAL PROCESS ANALYSIS

WAI CHAN
The Chinese University of Hong Kong, China
and
JOYCE KWAN and CHERRY CHOI
The Hong Kong Institute of Education, Hong Kong, China

In social science research, the analysis of moderation, mediation, and conditional process is an important methodological question that goes beyond multiple regression. Generally speaking, moderation occurs when the relationship between the independent variable (X) and the dependent variable (Y) varies as a function of a third variable (moderator). Mediation occurs when the effect of X on Y is transmitted through an intervening variable (mediator). Conditional process, collectively, refers to models that combine both mediation and moderation processes. These models are important because they allow researchers to understand, describe, and explain complex social phenomenon and human behavior in a more accurate way. Despite their theoretical usefulness, applied researchers always find it difficult to execute the analysis appropriately due to both technical and conceptual reasons. In this study, a new structural equation modeling (SEM) based method for conditional process analysis is proposed. This method has several advantages as compared to the traditional regression-based approach in terms of its flexibility in model specification, efficiency in parameter estimation, and ability in assessing overall model goodness-of-fit. Computationally, a new computer software program, VS, for implementing the proposed method is introduced. Real examples will be given to demonstrate how VS can be used to examine models that involve conditional process.

Acknowledgements

The work described in this presentation was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administration Region, China (Project No.: CUHK 441113).

MULTIPLE USE CONFIDENCE INTERVALS

MARTINA CHVOSTEKOVÁ
Palacký University, Olomouc, Czech Republic

Based on a fitted regression line it is required to construct interval estimates for unobserved unknown values of an explanatory variable corresponding to a sequence of future observations of a response variable. In the case of an unknown or an unlimited number of future observations the marginal property of multiple use confidence intervals is that at least γ proportion of them contain the true value of an explanatory variable with confidence $1 - \alpha$. We consider a linear regression model with independent identically normally distributed errors. The last results in Mee and Eberhardt (1996), Lee and Mathew (2002) indicate that the tolerance bounds for the model quite well satisfy the condition.

We proposed new method for computing the multiple use confidence intervals. The confidence of the suggested intervals and of the based on the tolerance bounds is compared in a simulation study for a simple linear regression.

Acknowledgements

The work was supported by the project CZ.1.07/2.3.00/30.0041 of the European Social Fund of the Ministry of Education, Youth and Sports of the Czech Republic.

References

[1] LEE, Y. and MATHEW, T. (2002). *Advances on Theoretical and Methodological Aspects of Probability and Statistics*, Taylor & Francis.

[2] MEE, R.W. and EBERHARDT, K.R. (1996). A comparison of uncertainty criteria for calibration. *Technometrics* **38** 221–229.

GENDER DIFFERENCES IN PERFORMANCE OF PROPHYLACTIC IMPLANTABLE CARDIOVERTER DEFIBRILLATORS (ICD): A PRACTICAL EXAMPLE OF A META-ANALYSIS

DAVID CONEN
University Hospital, Basel, Switzerland
BARBORA ARENDACKÁ, CHRISTIAN RÖVER, LEONARD BERGAU and PASCAL MUNOZ
University Medical Center Göttingen, Germany
SOFIEKE WIJERS
Universitair Medisch Centrum, Utrecht, Netherlands
and
MARKUS ZABEL and TIM FRIEDE
University Medical Center, Göttingen, Germany

What are the benefits and risks of primary prophylactic ICD implantation in women as compared to men? This is one of the questions addressed by EU-CERT-ICD project that is aimed at collecting evidence potentially helpful in selecting patients benefiting from primary prophylactic ICD implantation. As a first step in addressing the gender differences, published evidence on the effect of gender on the risk of receiving an appropriate shock after a prophylactic ICD implantation has been collected and the extracted sex-specific hazard ratios from multivariable time-to-event models have been combined in a meta-analysis. We will use this practical example to illustrate steps involved in conducting a meta-analysis and to discuss several methodological issues and our choice of the applied statistical methods. For the pooling, we used a random effects model in combination with the Mandel-Paule estimate for the between study variance and the Knapp-Hartung correction for the confidence interval for the common hazard ratio. A short outlook of further steps and use of the obtained result will be given as well.

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement No. 602299, EU-CERT-ICD.

CORRELATION COEFFICIENT AND MEASURE OF DETERMINATION

HILMAR GERHARD DRYGAS
University of Kassel,Germany

After some general considerations about correlation and partial correlation the regression model $y = a + bx + e$ is investigated. The square of the usual empirical correlation coefficient is in general smaller or equal than the measure of determination. This gives rise to the definition of a new correlation coefficient in the continuous case. An empirical example will also be presented.

SUPERSATURATED DESIGNS FOR SAMPLING FINITE POPULATIONS

KENT M. ESKRIDGE, XIAOJUAN HAO,
JUAN DIEGO HERNANDEZ–JARQUIN and GEORGE L. GRAEF
University of Nebraska, Lincoln, USA

Supersaturated designs are two-level factorial designs useful for screening a large number of factors (p) with a limited number of runs (n) where $n < p$. Substantial work has been done on constructing supersaturated designs for various optimality criteria when there are no restrictions on the treatment combinations to be considered. However, little work has been done on identifying optimal supersaturated designs when candidate design points must be selected from a restricted set of treatment combinations. Constructing supersaturated designs for such finite populations is an important problem in genetics. For example, a large number of p genetic markers may be available on N ($< p$) individuals where the interest is to assess how the markers are related to some end point variable such as disease level or yield. Often the individuals' end point variables are not available and collecting such data on all N individuals is too costly resulting in the need to use the genetic information to identify the n ($< N$) most informative individuals for which the endpoint data should be obtained. The objective of this work is to develop and compare several different criteria in identifying supersaturated designs comprised of the n most informative individuals from a finite population of N individuals where $N < 2^p$ and to compare these measures in identifying important markers and predicting end point variables based on the resulting models. The methods are applied to the USDA Germplasm Resource Information System (GRIN) database of $\sim$20,000 soybean accessions using 50K SNP genetic markers. Preliminary evidence indicates substantial differences between the methods in terms of genetic information and the predictive ability of the resulting models.

COCKTAIL ALGORITHM FOR A-OPTIMAL DESIGNS

LENKA FILOVÁ and RADOSLAV HARMAN
Comenius University, Bratislava, Slovakia

We present a so-called cocktail algorithm for computing A-optimal designs of experiments. This is a modification of the algorithm in Yu (2011) which was originally constructed for D-optimality and which efficiently combines the vertex direction method (Fedorov, 1972), the vertex exchange method (Böhning, 1986), and the multiplicative algorithm (Silvey et al., 1978). Additionally, we have implemented the result of Pronzato (2013) to delete nonsupporting points in the design space. To illustrate the performance of the computation, we present several numerical examples that show improved speed compared to the traditional multiplicative algorithm.

References

[1] BÖHNING, D. (1986). A vertex-exchange-method in D-optimal design theory. *Metrika* **33** 337–347.
[2] FEDOROV, V. V. (1972). *Theory of optimal experiments*. Elsevier.
[3] PRONZATO, L. (2013). A delimitation of the support of optimal designs for Kiefer's Φ_p-class of criteria. *Statistics & Probability Letters* **83** 2721–2728.
[4] SILVEY S. D., TITTERINGTON D. M. and TORSNEY B. (1978). An algorithm for optimal designs on a finite design space. *Communications in Statistics – Theory and Methods* **7** 1379–1389.
[5] YU, Y. (2011). D-optimal designs via a cocktail algorithm. *Statistics and Computing* **21** 475–481.

GEOMETRIC FIT OF CONICS BY LINEAR REGRESSION MODEL WITH NONLINEAR CONSTRAINTS

EVA FIŠEROVÁ and SANDRA DONEVSKA
Palacký University, Olomouc, Czech Republic

For the least squares methodology can be distinguished two main approaches of fitting conics (circle, ellipse, parabola, hyperbola), the algebraic and geometric fit (Chernov, 2010). The algebraic fit corresponds with the problem of minimization squares of algebraic distances from each given point to the curve being described by the implicit equation. The geometric fit is based on minimization of the sum of orthogonal (geometric) distances from the observed data points to the fitted curve.

In the contribution, we focus on solving the problem of the geometric fit by the linear regression model with nonlinear constraints. The constraints are represented by the implicit equation of the certain conics. In order to obtain approximate linear regression model, the nonlinear constraints are being linearized by the first-order Taylor expansion. According to the iterative estimation procedure being proposed in (Köning et al., 2014) will be gained locally best linear unbiased estimators of the unknown algebraic conic parameters and also estimators of their uncertainties. Subsequently, the geometric conic parameters like the centre, angle of rotation, and lengths of the semi-axes, and their uncertainties can be also estimated. The results of performed simulation study of the algebraic and geometric conics parameters estimators accuracy will be also presented.

References

[1] CHERNOV, N. (2010). *Circular and Linear Regression: Fitting Circles and Lines by Least Squares.* Chapman & Hall/CRC.

[2] KÖNING, R., WIMMER, G. and WITKOVSKÝ, V. (2014). Ellipse fitting by linearized nonlinear constraints to demodulate quadrature homodyne interferometer signals and to determine the statistical uncertainty of the interferometric phase. *Measurement Science and Technology* **25** 115001.

DIMENSION REDUCTION OF DIHEDRAL ANGLES DATA USING PRINCIPAL GEODESIC ANALYSIS

MOUSA GOLALIZADEH and ANAHITA NODEHI
Tarbiat Modares University, Tehran, Iran

One of the challenging issues in biological sciences is related to prediction of the protein structures with a great emphasize on their geometrical features. If the internal angles among particular atoms are used to represent the geometrical structure of protein, then it sounds the problem can be tackled using a new branch of statistics, known as directional statistics. On the other hand, the data on hand for determining the protein structure are usually of high dimensions. For example, the common biological technique, known as molecular dynamics simulations, produces massive molecular structure data. The well known principal component analysis might be employed to reduce the dimensions of involved variables leading to both save the computational cost and simplify the statistical inference. However, direct implementation of this technique on dihedral angles are not feasible because those angles are elements of a non-Euclidean space, torus in this case. Instead, one of the new technique, named as the principal geodesic analysis, sounds to perform better in compare to usual methods in evaluating the variability on non-linear spaces. We describe how this new technique is used to reduce the dimension of a set of dihedral angles. A comparison between the aforementioned methods is also provided based on both simulation study and a real data set.

THE EQUATION OF BLUP FOR A MIXED EFFECT UNDER GENERAL MIXED LINEAR MODEL

NESRIN GÜLER
Sakarya University, Sakarya, Turkey
and
SIMO PUNTANEN
University of Tampere, Finland

In this study, a general mixed linear model $\mathcal{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu}, \mathbf{D}, \mathbf{R}, \mathbf{K}\}$ with singular covariance matrix is considered under the assumption that the random effect and the fixed effect assumed to be correlated. The equation of the best linear unbiased predictor (BLUP) for a mixed effect $\mathbf{L}\boldsymbol{\beta}+\mathbf{Su}$, which is a linear combination of the random effect and the fixed effect, is obtained under the model $\mathcal{M}$. By using the BLUP equation for mixed effect, some results related to the equations of the best linear unbiased estimator (BLUE) for the fixed effect and of the BLUP for the random effect are given under the general mixed linear model $\mathcal{M}$.

A SIMULATION STUDY OF BOOTSTRAP METHODS FOR KRIGING IN TIME SERIES FORECASTING

MARTINA HANČOVÁ, JOZEF HANČ and ANDREJ GAJDOŠ
Pavol Jozef Šafárik University, Košice, Slovakia

One of approaches for forecasting future values of a time series or unknown spatial data is kriging. This prediction theory is based on modeling in a sufficiently general class of linear regression models and on finding the best linear unbiased predictor (BLUP), which minimizes the mean squared error of prediction among all linear unbiased predictors.

Our contribution deals with design and efficiency of computer simulations in the form of bootstrap methods for the general class of time series linear regression models, called FDSLRM, whose mean values are given by linear regression and error terms are characterized by a purely finite discrete spectrum and white noise.

We have created seven bootstrap algorithms suitable for FDSLRM. Algorithms are founded on model, block and transformation based bootstrap and simultaneously belong to semi-parametric methods requiring no assumptions about time series probability distribution. Using a simulation study based on real econometric data, we compared our algorithms in the frame of kriging predictions, in different models of the FDSLRM class and with different OLS estimators of variance parameters.

The results of simulation study showed that our algorithms did not really depend on applied types of variance-parameters estimators. As for the point value of the BLUP, the results of

all algorithms (except controlling i.i.d. bootstrap) can be considered as unbiased. From the viewpoint of coverage rate of 95% confidence interval we may say that transformation based bootstrap algorithms have good performance and their efficiency does not depend on a chosen transformation. Algorithms based on block bootstrap are less computationally demanding, but their coverage rates are worse. However the coverage can be improved by a suitable correction.

Acknowledgements

This work is supported by the Scientific Grant Agency of the Slovak Republic (VEGA), grants VEGA 1/0073/15 and VEGA 1/0344/14.

TWO-SAMPLE GRADUAL CHANGE-POINT ANALYSIS

ZDENĚK HLÁVKA and MARIE HUŠKOVÁ
Charles University, Prague, Czech Republic

Applications of change-point analysis are numerous: many authors studied change-points in mean, variance, regression coefficients, or changes occurring in time series. In contrast to the mainstream of existing literature concerning only one-sample problems, we assume that two independent samples are observed and we concentrate on the resulting two-sample gradual change-point problem.

Our study is motivated by a real data set containing jumping speeds observed for 432 girls and 364 boys. Applying the proposed estimator, we study gender differences observed in various age groups. We argue that, in this setup, the proposed two-sample gradual change-point analysis is more reasonable and leads to more precise estimators than application of standard two-sample t-tests.

Apart of establishing the asymptotic distribution of the two-sample gradual change-point estimator, we also investigate its small sample properties in a simulation study.

ELLIPTICAL QUANTILES IN PARAMETRIC REGRESSION

DANIEL HLUBINKA
Charles University, Prague, Czech Republic

We consider parametric model of multivariate (and multiple) regression. The natural extension of classical quantile regression is to consider some version of multivariate quantiles. In the poster we show a possible use of general elliptical quantiles; the ellipitcal quantiles have natural probabilitstic interpretation, can be made flexible for many forms of heteroscedasticity, can work without any moment assumption or with quite weak moment assumptions, can be computed quite quickly and they have other nice properties. On the other hand their shape is limited to

ellipses and, naturally, the quality of fit depends on correct parametrization as for any parametric regression models.

CHANGE POINT ESTIMATORS IN PANEL DATA

LAJOS HORVÁTH
University of Utah, Salt Lake City, USA
MARIE HUŠKOVÁ
Charles University, Prague, Czech Republic
and
GREGORY RICE and JIA WANG
University of Utah, Salt Lake City, USA

The talk will concerns of estimators of common time of a change in the mean parameters of panel data. A CUSUM-like type estimators will be constructed, results on their asymptotic behavior will be presented. Also estimators of the norming constants will be discussed. Dependence between panels is modeled in the form of common factors.

Quite weak assumptions are imposed on the model errors (fulfilled by most of stationary time series of interest). It is assumed that both the number of panels and the number of observations in each panel are large, i.e., both are tending to infinity.

Results of simulation study and an application to real data sets will be presented.

PARTICLE SWARM OPTIMIZATION BASED LIU-TYPE ESTIMATOR

DENIZ İNAN, EROL EĞRİOĞLU, MÜJGAN TEZ and BUSENUR SARICA
Marmara University, Istanbul, Turkey

In this study, a new method for the estimation of shrinkage (k) and biasing (d) parameters of Liu-type estimator was proposed. Because k is kept constant and d is optimized in Liu's method, (k, d) pair is not guaranteed to be the optimal point for the mean square error of the parameters. Optimum (k, d) pair that minimizes mean square error which is the function of k and d parameters should be estimated by a simultaneous optimization process rather than a two-stage process. In this study, by utilizing a different objective function, k and d parameters are optimized simultaneously with particle swarm optimization technique and obtained results were analyzed by a simulation study and two different real-life data application.

CONVEX METHOD FOR VARIABLE SELECTION IN HIGH-DIMENSIONAL LINEAR MIXED MODELS

JOZEF JAKUBÍK

Slovak Academy of Sciences, Bratislava, Slovakia

The analysis of high-dimensional data is currently a popular field of research, thanks to many applications e.g. in genetics. At the same time, the type of problems that tend to arise in genetics, can often be modelled using linear mixed models (LMM) in conjunction with high-dimensional data.

The LMM allows us to specify the covariance structure of the model, which enables us to capture relationships in data, for example population structure, family relatedness etc. Therefore, LMMs are often preferred to linear regression models. In genome-wide association studies in genetics, one studies the dependence of phenotype on the genotype. Genetic information can consist of up to 10^6 variables, but only information about the genotype of a small group of subjects is available. Variable selection in high-dimensional data refers to the selection of a small group of variables which influence observations.

Methods for variable selection in high-dimensional LMMs from Rohart et al. (2014) and Schelldorfer (2011) are based on non-convex optimization problems with one penalty parameter. For problems of dimension higher than 10^4 are methods based on non-convex optimization problems almost unusable, because their computational complexity is beyond the capabilities of current computers.

In this work we introduce a convex method and its simplified variant for variable selection in LMM and we will show that in most cases the simplified method performs comparably well to the full model while its running time is considerably lower.

Acknowledgements

The work was supported by the Scientific Grant Agency VEGA of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences, by the projects VEGA 2/0047/15 and VEGA 2/0043/13.

References

[1] ROHART, F., SAN CRISTOBAL, M. and LAURENT, B. (2014). Selection of fixed effects in high dimensional linear mixed models using a multicycle ECM algorithm. *Computational Statistics & Data Analysis* **80** 209–222.

[2] SCHELLDORFER, J., BÜHLMANN, P. and VAN DE GEER, S. (2011). Estimation for High-Dimensional Linear Mixed-Effects Models Using ℓ_1-Penalization. *Scandinavian Journal of Statistics* **38** 197–214.

STATISTICAL MODELS FOR MANIFOLD DATA WITH APPLICATIONS TO THE HUMAN FACE

STANISLAV KATINA
Masaryk University, Brno, Czech Republic
and
ADRIAN BOWMAN and LIBERT VITTERT
The University of Glasgow, UK

One of the data structures generated by medical imaging technology is high resolution point clouds representing anatomical surfaces. Stereo-photogrammetric and laser images are a widely available source of this kind of data. Raw images are in the form of triangulated surfaces. The first step is to create a standardised representation of surface shape which provides a meaningful correspondence across different images which therefore provides a basis for statistical analyses across samples of images. Point locations with anatomical definitions, referred to as landmarks, have provided the traditional starting point, with analysis and interpretation which are widely used and well understood. This can also be taken as the starting point for more general surface representations, often using templates which are ¢warped✓on to an observed surface by matching landmark positions and subsequent local adjustment of the surface. The aim of the present paper is to use ridge and valley curves to capture the principal features of the surface of interest. Landmarks are used as anchoring points, as usual, but curvature information across the surface is used to guide the curve locations. Once the ridges and valleys have been located, the intervening surface patches are relatively flat and can be represented in a standardised manner by appropriate surface transects, to give a complete surface model. However, the intermediate curve representation is of considerable interest in its own right, as it captures the principal features of the surface and therefore embodies much of the shape information on the object of interest. Methods of identifying these curves are discussed and then applied to investigate the shape variability in a control population and in populations of patients having some psychotic disorders such as schizophrenia and bipolar disorder. The curve representation is shown to capture the majority of the relevant information, in addition to providing a basis of a full surface model.

SEVERAL ESTIMATING METHODS IN THE GROWTH CURVE MODEL

VERONIKA KOPČOVÁ
Pavol Jozef Šafárik University, Košice, Slovakia

We study growth curve model with patterned correlation matrices under the assumption of multivariate normality for the response variables. Estimation of the regression, correlation and scale parameters are discussed in the model using several methods. Estimates are compared for

three important cases: serial correlation structure, uniform correlation structure and connection between uniform and serial correlation structure.

Acknowledgements

The support of the grant VEGA MŠSR 1/0344/14 is kindly announced.

References

[1] KLEIN, D. and ŽEŽULA, I. (2009). The maximum likelihood estimators in the growth curve model with serial covariance structure. *Journal of Statistical Planning and Inference* **139** 3270–3276.

[2] RAO CHAGANTY, N. (2003). Analysis of growth curves with patterned correlation matrices using quasi-least squares. *Journal of Statistical Planning and Inference* **117** 123–139.

[3] RUSNAČKO, R. and ŽEŽULA, I. (2015). Connection between uniform and serial correlation structure in the growth curve model. *To appear.*

PARTIAL SUMMATIONS DIVIDE DISCRETE DISTRIBUTIONS INTO TWO CLASSES

MICHAELA KOŠČOVÁ and JÁN MAČUTEK

Comenius University, Bratislava, Slovakia

and

GEJZA WIMMER

Matej Bel University, Banská Bystrica, Slovakia
Slovak Academy of Sciences, Bratislava, Slovakia

Consider the generalized partial summation

$$P_x(a) = \sum_{j=x}^{\infty} g(j) P_j^*(a), \quad x = 0, 1, 2, \dots , \tag{1}$$

with $\{P_j^*(a)\}_{j=0}^{\infty}$ being called the parent distribution and $\{P_j(a)\}_{j=0}^{\infty}$ the descendant distribution (see Mačutek, 2003, Wimmer and Mačutek, 2012), both defined on nonnegative integers, a is parameter. For the sake of simplicity we limit ourselves to discrete distributions with one parameter only. In Mačutek (2003) a necessary and sufficient condition of invariance under summation (1) is provided,

$$g(x) = 1 - \frac{P_x^*(a)}{P_{x+1}^*(a)}. \tag{2}$$

Under this condition, the parent distribution remains unchanged under (1), i. e. $P_x^* = P_x$, $x = 0, 1, 2, \dots$ In order to emphasize the role of the parameter, we can use the notation $g(x) = g(x, a)$.

Now let us consider a modification of summation (1)

$$P_x = c \sum_{j=x}^{\infty} g(j, \lambda) P_j^*(a), \quad x = 0, 1, 2, \ldots \tag{3}$$

where the formula (2) is kept but parameter a was replaced by λ, c is an appropriate constant which ensures that $\{P_j\}_{j=0}^{\infty}$ is a proper distribution (i.e., it sums to 1).

The distribution $\{P_j\}_{j=0}^{\infty}$ either depends on two parameters, λ and a, for $a \neq \lambda$, or the second parameter λ is "cancelled" by the normalization constant c. With respect to this property, it is possible to categorize every discrete distribution with one parameter into one of the two abovementioned classes. For example, the Poisson distribution generates under summation (3) a new distribution with two parameters, therefore it belongs to the first class. On the other hand, e.g. the geometric distribution belongs to the second class, i. e. it remains unaltered under summation (3).

Acknowledgements

Supported by VEGA grant 2/0047/15.

References

[1] MAČUTEK, J. (2003). On two types of partial summation. *Tatra Mountains Mathematical Publications* **26** 403–410.

[2] WIMMER, G. and MAČUTEK, J. (2012). New integrated view at partial-sums distributions. *Tatra Mountains Mathematical Publications* **51** 183–190.

LIU ESTIMATION IN GENERALIZED LINEAR MODELS

FİKRIYE KURTOĞLU and M. REVAN ÖZKALE
Çukurova University, Adana, Turkey

Multicollinearity among the explanatory variables seriously effects the maximum likelihood estimator in linear regression models that are are too large in absolute value and resulting in large variance-covariance matrix. The adverse effects of multicollinearity on parameter estimation in generalized linear models are also explored by various authors in the case of maximum likelihood estimation. In this study, we introduce a first-order approximated Liu estimator to combat multicollinearity in generalized linear models which is an extension of the Liu estimator in the linear regression model. We also obtain necessary and sufficient condition for the superiority of the first-order approximated Liu estimator over the iteratively reweighted least squares estimator at final iteration by the approximated mean squared error criterion. An estimator of the biasing parameter of the first-order approximated Liu estimator is proposed by minimizing the approximated mean squared error. The results are illustrated by conducting a numerical and simulation study.

FUNCTIONAL SPACES WITH A FINITE DIMENSIONAL CRITERION FOR WEAK CONVERGENCE

PETR LACHOUT
Charles University, Prague, Czech Republic

Application of weak convergence of random processes is limited by existence of an useable convergence criterion. We are able to handle with Skorokhod spaces of functions generally developed and introduced in Straf (1972). Unfortunately, practical criteria for weak convergence are available in finite dimension, only; see Bickel and Wichura (1971), Neuhaus (1971). The criterion is tightness plus convergence of finite dimensional distributions. Our contribution considers product of spaces of functions and shows that tightness of coordinates plus convergence of finite dimensional distributions is a criterion for weak convergence.

References

[1] BICKEL, P.J. and WICHURA, M.S. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Annals of Mathematical Statistics* **42** 1656–1670.
[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*, Wiley.
[3] NEUHAUS, G. (1971). On weak convergence of stochastic processes with multidimensional time parameter. *Annals of Mathematical Statistics* **42** 1285–1295.
[4] STRAF, M. L. (1972). Weak convergence of stochastic processes with several parameters. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* **2** 187–221.

MIXED INTERPOLATION MODELS FOR INVERSE PREDICTION

LYNN ROY LAMOTTE
Louisiana State University Health Sciences Center – School of Public Health, New Orleans, USA

Insects visit a dead body left outdoors. Their characteristics (measurements of size and development and combinations of species) can provide a biological clock useful in estimating the time since death. Given the multivariate measurement $\boldsymbol{y}_*$ from a mystery specimen sampled at the scene, the objective is to devise reasonable and defensible statistical methodology to support an estimate of the age of the specimen. For that purpose, training data are available from rearing experiments for the species in question. They comprise independent observations on $\boldsymbol{y}$ at ages spanning the development cycle, under controlled (principally temperature) conditions. Central features of such data are that the $\boldsymbol{y}$ - age relation is not linear and the variance-covariance matrix evolves steeply with age.

Inverse prediction, also known as calibration, has a reputation of being computationally difficult, particularly with a multivariate response. Methods for heteroscedastic multivariate responses are practically unknown. In the forensic sciences literature, most developments have modeled age as a function of $\boldsymbol{y}$, in reverse cause-effect order, with multiple regression, and ignored the inconstant variance.

These relations can be modeled within the context of mixed models, with separate models for the mean vector and the variance-covariance matrix in terms of age and temperature. At each potential age, comparison of $\boldsymbol{y}_*$ to the model fit to the training data gives a p-value for the test of $\boldsymbol{y}_*$ as a multivariate outlier at that age.

The methodology and computations of mixed models are well-developed and widely available in standard statistical computing packages. In this talk, I shall illustrate the formulation and implementation of multivariate inverse prediction in terms of mixed models.

Acknowledgements

Research reported in this talk was supported by Award 2013-DN-BX-K042, U. S. Department of Justice, National Institute of Justice.

DIFFERENT REGULARIZATION TECHNIQUES IN CHANGE-POINT ESTIMATION

MATÚŠ MACIAK
Charles University, Prague, Czech Republic
and
IVAN MIZERA
University of Alberta, Edmonton, Canada

Classical regression techniques usually rely on some shape or smoothness requirements. Such assumptions are convenient as they make theoretical justification and proofs a lot easier. On the other hand, in many practical situations such nice qualities can not be achieved, technically not even expected. More complex dependence structures with no shape restrictions together with non-smooth or even discontinuous segments need to be properly modelled. In this matter, statisticians usually refer to change-point models and change-point detection and there are various techniques available to solve these problems. Unfortunately, many existing methods require either a prior knowledge for possible locations of change-points or they are multiple-stage estimation techniques. Both can be quite limiting in some situation.

We will discuss an alternative approach to change-point estimation in regression: the main advantage of the method is that it presents an automatic approach with no prior knowledge for change-point locations. The method combines nonparametric regression estimation techniques with difeerent concepts of an L1-norm regularization. Different alternatives are discussed, some statistical inference and other theoretical results are derived. Finite sample performance is also investigated using simulated and real data.

SENSITIVITY APPROACH TO A TEST IN ANOVA MODELS

JAROSLAV MAREK and LUBOMÍR KUBÁČEK
University of Pardubice, Czech Republic

One-way analysis of variance (abbreviated one-way ANOVA) is a technique used to compare means of two or more samples (using the F distribution). The results of a one-way ANOVA can be considered reliable as long as the assumptions of normality, independence and homoscedasticity of samples are met. In this paper we will analyze a test for situation with a different variances in different populations. In such situation the sensitivity analysis is used.

SOME GOODNESS OF FIT TESTS BASED ON U-EMPIRICAL LAPLACE TRANSFORM

BOJANA MILOŠEVIĆ and MARKO OBRADOVIĆ
University of Belgrade, Serbia

In this paper we present some new integral-type tests based on U-empirical Laplace transform. We base our test statistics on a characterization. We calculate the Bahadur efficiencies against some common alternatives. We compare their performances against some other similar tests.

SOME CHARACTERIZATION BASED EXPONENTIALITY TESTS AND THEIR BAHADUR EFFICIENCIES

BOJANA MILOŠEVIĆ and MARKO OBRADOVIĆ
University of Belgrade, Serbia

In this paper we propose new exponentiality tests based on a recent characterization. We construct integral and Kolmogorov-type statistics, derive their asymptotics and calculate the Bahadur efficiency against some common alternatives. We also obtain a class of locally optimal alternatives for each test.

ON MIXING PRIORS IN EMPIRICAL BAYES PREDICTION

IVAN MIZERA
University of Alberta, Edmonton, Canada

An interesting aspect in fitting mixture models is the question how much the quality of the empirical Bayes prediction is affected by a common simplification of the form of the mixing/prior distribution: the latter is considered a product of its marginals – so called "naive Bayes" – as opposed to working in full generality. We discuss advantages of the former, consequences and feasibility of the latter – in the light of a possible adaptive extension of the Kiefer–Wolfowitz strategy proposed by Koenker and Mizera (2014) – and offer some comparisons, mostly of experimental nature.

ESTIMATING THE EFFECT SIZE OF A STATISTICALLY SIGNIFICANT EFFECT

MÄRT MÖLS
University of Tartu, Estonia

Selecting results based on their statistical significance is a known way to introduce bias (see for example publication bias). Sometimes, unfortunately, only estimates for statistically significant results are available and one has to use those biased results for making conclusions and predictions. For example in genome-wide association studies it is usual to test tens of millions genetic mutations but only those few showing statistically significant association with a phenotype are selected for publication and their estimated effect sizes will be available to the wider audience. If one wishes to predict the phenotype of a new individual one can only use those biased estimates as their input for calculations. To use these available but biased (due to selecting statistically significant results) estimates to derive still meaningful (bias-reduced) predictions or effect size measures one can use a mixed model for truncated normal response variable. The model is introduced and its behaviour in practical applications will be evaluated.

CONSISTENCY OF h-MODE DEPTH

STANISLAV NAGY
Katholieke Universiteit Leuven, Belgium
Charles University, Prague, Czech Republic

Consistency results for the sample h-mode depth in the general case of Banach-valued data are established. The rate of convergence is provided, which is linked to the rate at which the sample sequence of bandwidths converges to its population version. The robustness of the h-mode depth, as well as the convergence of the associated modes of the distribution, is also studied.

ITERATIVE ALGORITHMS OF BIASED ESTIMATION METHODS IN BINARY LOGISTIC REGRESSION

M. REVAN ÖZKALE
Çukurova University, Adana, Turkey

Logistic regression is a widely used method to model categorical response data and maximum likelihood (ML) estimation is widespread use in logistic regression. Although ML method is the most used method to estimate the regression coefficients in logistic regression model, multicollinearity seriously effects the ML estimator. To remedy the undesirable effects of multicollinearity, estimators alternative to ML are proposed. Drawing on the similarities between the multiple and logistic regressions, ridge, liu and two parameter estimators are proposed which are based on the ML estimator. On the other hand, first-order approximated ridge estimator is proposed in logistic regression. This study will present further solutions to the problem in the form of alternative estimators which reduce the effect of the collinearity. For this aim, first-order approximated and iterative liu and two parameter estimators are proposed. The ML based, first-order and iterative biased estimators are compared in a simulation study to see effect of sample size and degree of multicollinearity. Graphical representations are presented which support the effect of the shrinkage parameter on the mean square error and prediction mean square error of the biased estimators.

References

[1] LeCessie, S. and Van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. *Applied Statistics* **41** 191–201.
[2] Masson, K., Kibria, B. M. G. and Shukur, G. (2012). On Liu estimators for the logit regression model. *Economic Modelling* **29** 1483-1488.
[3] Schaefer, R. L. (1986). Alternative estimators in logistic regression when the data are collinear. *Journal of Statistical Computation and Simulation* **25** 75–91.
[4] Schaefer, R. L., Roi, L. D. and Wolfe, R. A. (1984). A ridge logistic estimator. *Communications in Statistics – Theory and Methods* **13** 99–113.

STATISTICAL INFERENCE FOR LINE-SEGMENT PROCESSES

ZBYNĚK PAWLAS
Charles University, Prague, Czech Republic

Line-segment processes are stochastic models that have been studied for a number of applications, where systems of segments randomly scattered in space occur. The segments may represent geological faults, textile structure or stress fibres in cells. It is assumed that a single realization of a stationary line-segment process is observed within a bounded sampling window. We are interested in nonparametric estimation of different summary characteristics.

OPTIMAL DESIGNS FOR COPULA MODELS

ELISA PERRONE and WERNER GÜNTHER MÜLLER
Johannes Kepler University, Linz, Austria

Copula modelling has in the past decade become a standard tool in many areas of applied statistics. However, a largely neglected aspect concerns the design of related experiments. Particularly the issue of whether the estimation of copula parameters can be enhanced by optimizing experimental conditions and how robust all the parameter estimates for the model are with respect to the type of copula employed. In this paper an equivalence theorem for (bivariate) copula models is provided that allows formulation of efficient design algorithms and quick checks of whether designs are optimal or at least efficient. Some examples illustrate that in practical situations considerable gains in design efficiency can be achieved. A natural comparison between different copula models with respect to design efficiency is provided as well.

UNITARILY INVARIANT EIV ESTIMATION

MICHAL PEŠTA
Charles University, Prague, Czech Republic

Linear relations, containing measurement errors in input and output data, are considered. Parameters of these so-called errors-in-variables (EIV) models can be estimated by minimizing the total least squares (TLS) of the input-output disturbances, i.e., penalizing the orthogonal squared misfit. This attitude leads into the minimizing the Frobenius norm of the error matrix.

Immediate doubts arise whether this criterion is suitable and in which sense. Such an estimate for a linear model is highly non-linear. What happens if we do not consider squared distances?

An extension of the TLS estimate in the EIV model—the EIV estimate—is proposed, which is scale invariant, interchange, direction, and rotation equivariant. Moreover, it is shown that the EIV estimate coincides with any unitarily invariant penalizing solution to the EIV problem.

TESTING STRUCTURAL CHANGES IN PANEL DATA WITH SMALL FIXED PANEL SIZE AND BOOTSTRAP

BARBORA PEŠTOVÁ and MICHAL PEŠTA
Charles University, Prague, Czech Republic

Panel data of our interest consist of a moderate or relatively large number of panels, while the panels contain a small number of observations. This work establishes testing procedures to detect a possible common change in means of the panels. To this end, we consider a ratio type test statistic and derive its asymptotic distribution under the no change null hypothesis. Moreover, we prove the consistency of the test under the alternative. The main advantage of such an approach is that the variance of the observations neither has to be known nor estimated. On the other hand, the correlation structure is required to be calculated. To overcome this issue, a bootstrap technique is proposed in the way of a completely data driven approach without any tuning parameters. The validity of the bootstrap algorithm is shown. As a by-product of the developed tests, we introduce a common break point estimate and prove its consistency. The results are illustrated through a simulation study. An application of the procedure to actuarial data is presented.

MULTIVARIATE HOMOGENEITY TESTING BASED ON L-MOMENTS

JAN PICEK and TEREZA ŠIMKOVÁ
Technical University, Liberec, Czech Republic

The contribution is devoted to a study of the extension of univariate regional homogeneity tests based on L-moments to the multivariate case (Chebana and Ouarda, 2007). As a tool, multivariate L-moments are used to define the test statistic and copula models to describe the statistical behavior of analyzed dependent variables. Considering the diversity of copula families developed in statistical literature the suitable model is chosen using the goodness-of-fit tests and the Akaike Information Criterion. The methodology is illustrated on multivariate regional frequency analysis of extreme precipitation events in the Czech Republic.

ON A DATA-DEPENDENT CHOICE OF THE TUNING PARAMETER APPEARING IN CERTAIN GOODNESS-OF-FIT TESTS

CHARL PRETORIUS, JAMES SAMUEL ALLISON and LEONARD SANTANA
North–West University, Potchefstroom, South Africa

We propose a data-dependent method for choosing the tuning parameter appearing in many recently developed goodness-of-fit test statistics. The new method, based on the bootstrap, is applicable to a class of distributions for which the null distribution of the test statistic is independent of unknown parameters. No data-dependent choice for this parameter exists in the literature; typically, a fixed value for the parameter is chosen which can perform well for some alternatives, but poorly for others. The performance of the new method is investigated by means of a Monte Carlo study, employing three tests for exponentiality. It is found that the Monte Carlo power of these tests, using the data-dependent choice, compares favourably to the maximum achievable power for the tests calculated over a grid of values of the tuning parameter.

PREDICTIVE DENSITY FOR THE PHASE II SHEWART-TYPE P-CHART

LIZANNE RAUBENHEIMER
Rhodes University, Grahamstown, South Africa

The usual operation of the p-chart will be extended by introducing a Bayesian approach. A beta type I prior distribution with six different parameter combinations will be considered. Control chart limits, average run lengths and false alarm rates will be determined by using a Bayesian method. These results will be compared to the results obtained when using the classical method. A predictive density based on a Bayesian approach will be used to derive the control limits in Phase II. The proposed method gives wider control limits than those obtained from the classical method.

MODEL-BASED DESIGN OF EXPERIMENTS UNDER CONSIDERATION AND COMBINATION OF DIFFERENT APPROACHES

INA REICHERT, PETER OLNEY and TOM LAHMER
Bauhaus University, Weimar, Germany

Design of experiments is a helpful tool in order to reduce measurement errors. A proper design can also diminish the costs for an experiment. In this case the model-based design of experiments is used to find the optimal sensor positions for a given number of sensors. A numerical simulation of a civil engineering structure serves as the model and three different approaches are used to find the optimal setup of sensors. The first method uses the Fisher Information Matrix to reduce the parameter uncertainties. Secondly, the best setup is found by minimizing the mean-squared error between the estimated and the assumed true solution. In the third approach, the sigma-point method, biased estimators for the mean-squared error are used to identify the parameters.

SOME NOTES ON C-OPTIMAL DESIGNS FOR HETEROSCEDASTIC AND CORRELATED OBSERVATIONS

JUAN MANUEAL RODRIGUEZ–DIAZ
University of Salamanca, Spain

For a given model, c-optimality is the criterion that looks for the minimization of the variance of a linear combination of the parameters' estimators. In particular it is needed for checking how good a specific design is for the estimation of each of the model parameters. There is a geometric procedure for computing c-optimal designs for independent observations, specially used for the two-parameter case. However, the studies for heteroscedastic and/or correlated observations are less frequent, and this is the framework on which this work will focus. Efficiency of the designs obtained assuming homoscedasticity and independence when this is not the actual situation will be computed and analyzed in terms of the variance structure.

OPTIMAL DESIGNS RESISTANT TO NUISANCE EFFECTS

SAMUEL ROSA and RADOSLAV HARMAN
Comenius University, Bratislava, Slovakia

Consider an experiment consisting of a set of independent trials. The value of the response of each trial depends on the effect of the treatment selected for the trial, and on some nuisance effects, e.g., the effect of a time trend, or blocking. In this model, we give conditions of optimality for approximate designs for estimating a system of treatment contrasts: the conditions of optimal treatment proportions and of resistance to nuisance effects. We demonstrate our results on the experiment for comparing treatments with a control. We show that the obtained results may be used to obtain optimal approximate designs with a small support by means of linear programming. Efficient exact designs can be constructed from these optimal approximate designs by a simple heuristic.

DIFFERENCES IN SLEEP MICROSTATE CURVES AMONG PATIENTS AFTER STROKE AND HEALTHY SLEEPERS

ZUZANA ROŠŤÁKOVÁ and ROMAN ROSIPAL
Slovak Academy of Sciences, Bratislava, Slovakia

Sleep deprivation, whether from disorder or lifestyle, poses a significant risk in daytime performance. Ischemic stroke resulting in cerebral lesions is a well-known acute disorder that leaves affected patients strongly vulnerable to sleep disturbances that often lead to the above-mentioned impairments. The aim of this study is to identify objective sleep patterns being potential sources of disturbed sleep in stroke patients. To overcome the well-known limits of the standardized sleep scoring into several discrete sleep stages we employed an EEG data-based probabilistic model of sleep with an arbitrary number of different sleep stages – sleep microstates – and a high time resolution. The probabilistic sleep model (PSM) characterizes sleep by posterior probabilities curves. On a wide collection of sleep recordings from healthy subjects and stroke patients we applied functional data clustering methods to sleep microstate curves of the PSM. We found differences between stroke patients and healthy subjects in sleep microstates associated with slow wave sleep. Considering weighted combinations of microstates a better separation of the two groups was obtained. We observed a connection between sleep structure and sleep quality questionnaires as well as a set of tests reecting subjects' daytime cognitive performance.

A TEST FOR SYMMETRY BASED ON THE EMPIRICAL PROBABILITY WEIGHTED CHARACTERISTIC FUNCTION

LEONARD SANTANA and JAMES SAMUEL ALLISON
North-West University, Potchefstroom, South Africa
and
SIMOS MEINTANIS
University of Athens, Greece

A new test for symmetry based on the novel idea of the probability weighted empirical characteristic function is introduced. The performance of the test is demonstrated by considering a test for symmetry of the error distribution in regression models. Along with the new tests, procedures based on the ordinary empirical characteristic function as well as other more well-known procedures are implemented and compared.

A NEW HYBRID METHOD FOR TIME SERIES FORECASTING

BUSENUR SARICA and EROL EĞRİOĞLU
Marmara University, Istanbul, Turkey
and
BARIŞ AŞIKGIL
Mimar Sinan Fine Arts University, Istanbul, Turkey

In this study, a new hybrid forecasting method is proposed. The proposed method is called autoregressive adaptive network fuzzy inference system (AR-ANFIS). The AR-ANFIS has been shown in a network structure. The architecture of the network has two parts, the first part is an ANFIS structure and the second part is a linear autoregressive model structure. In literature AR models and ANFIS are widely used in time series forecasting. AR models are linear models and they are used according to model based strategy. A nonlinear model is employed by using ANFIS. Moreover ANFIS is a kind of data based modelling system like artificial neural network models. In this study, a linear and nonlinear forecasting model is proposed by creating a hybrid method of AR and ANFIS. The new model has advantages of databased and model based approaches. AR-ANFIS is trained by using particle swarm optimization and fuzzification is done by using fuzzy c-means method. AR-ANFIS method has been tested on some real world time series data and it is compared with other time series forecasting methods. As a consequence of applications, it is shown that the proposed method can produce accurate forecasts.

NEW ADVANCES IN DIRECTIONAL MULTIPLE-OUTPUT QUANTILE REGRESSION

MIROSLAV ŠIMAN
Czech Academy of Sciences, Prague, Czech Republic

The poster displays what has recently been achieved in the field of directional multiple-output quantile regression (DMOQR) by the author and his collaborators. The presented results include
(1) an Octave toolbox for DMOQR,
(2) an R package for DMOQR,
(3) its local polynomial extensions,
(4) its application to the construction of growth charts,
(5) its consequences for the (exact) computation of some projection-based statistics, and
(6) its use for defining a multivariate precision capability index.

OPTIMAL SEQUENTIAL PROCEDURES FOR BUYING–SELLING PROBLEMS

GEORGY SOFRONOV
Macquarie University, Sydney, Australia

We consider a buying–selling problem when at least two stops of a sequence of random variables are required. We observe these random variables sequentially and have to decide when we must stop, given that there is no recall allowed, that is, a random variable once rejected cannot be chosen later on. Our decision to stop depends on the observations already made, but does not depend on the future which is not yet known. Our objective is to maximize an expected reward. We derive an optimal stopping rule and the value of the game.

BEST PRACTICE IN METAMODELING FOR DATA DERIVED FROM CIVIL ENGINEERING APPLICATIONS

MARIA KERSTIN STEINER and TOM LAHMER
Bauhaus University, Weimar, Germany

High computation time has arisen as an outstanding obstacle for researchers and engineers of civil engineering problems, especially with the development of computers and new numerical methods. One approach of trying to overcome this problem is the application of metamodels by significantly reducing the computation cost of models, without a substantial loss in model quality. In this contribution, several metamodeling approaches such as polynomial regression, moving

least squares approximation, and the Kriging method are observed to enable a comparison of prediction quality and model complexity. As one of many engineering applications where the use of these methods has proved to be convenient, a variance-based sensitivity analysis of the finite elements model of a triaxial test is performed.

BIAS REDUCTION STUDIES IN NON-PARAMETRIC REGRESSION WITH APPLICATIONS

CORNELIA JOHANNA SWANEPOEL and MARIKE COCKERAN
North-West University, Potchefstroom, South Africa

The effect of three improvement methods on recently developed and well-established non-parametric kernel regression estimators is investigated. The improvement methods are applied to the Nadaraya-Watson estimator with cross-validation bandwidth selection as well as with plug-in bandwidth selection, a local linear estimator with plug-in bandwidth selection and a bias-corrected nonparametric estimator, using cross-validation bandwidth selection. The resulting regression estimators are evaluated by minimizing the mean integrated squared error. Two improvement methods are bootstrapped-based. Bagging is an acronym for bootstrap aggregating and is primarily a variance reduction tool. Bragging stands for bootstrap robust aggregating. A robust estimator is obtained by using the sample median over bootstrap estimates instead of the sample mean as in bagging. Boosting is the third improvement method and aims to reduce the bias component of the estimator iteratively. A selection of algorithms, involving the various improvement methods, estimators, models to be estimated, parametric set-ups and bandwidth selection methods (also involving various bootstrap strategies) are presented, displaying clear steps in terms of handling the data to obtain the goals of the study. Results are briefly summarized and illustrated, followed by conclusions and remarks regarding computational problems such as grid issues and bandwidth selection issues.

BERNSTEIN ESTIMATION FOR A COPULA DERIVATIVE WITH APPLICATION TO CONDITIONAL DISTRIBUTION AND REGRESSION FUNCTIONALS

JAN WILLEM HENDRIK SWANEPOEL
North-West University, Potchefstroom, South Africa
and
PAUL JANSSEN and NOËL VERAVERBEKE
University of Hasselt, Belgium

Bernstein estimators attracted considerable attention as smooth nonparametric estimators for distribution functions, densities, copulas and copula densities. In this talk we present a parallel result for the first order derivative of a copula function. We discuss how this result leads to

Bernstein estimators for a conditional distribution function and its important functionals, such as the regression and quantile functions. Results of independent interest, such as an almost sure oscillation behavior of the empirical copula process and a Bahadur type almost sure asymptotic representation for the Bernstein estimator of a regression quantile function, are also presented. The outcome of a simulation study demonstrates the good performance of the proposed estimators.

NON-RANDOMIZED RESPONSE MODELS FOR SENSITIVE SURVEYS WITH NONCOMPLIANCE

MAN LAI TANG
Hang Seng Management College, Hong Kong, China

Collecting representative data on sensitive issues has long been problematic and challenging in social issue studies (e.g., history of child abuse). Alternative data collection techniques that can be adopted to study sensitive questions validly become more important and necessary. As an alternative to the famous Warner randomized response model, non-randomized response triangular model has recently been developed to encourage participants to provide truthful responses in surveys involved sensitive questions. Unfortunately, both randomized and non-randomized response models could underestimate the proportion of subjects with the sensitive characteristic as some respondents do not believe that these techniques can protect their anonymity. As a result, some authors hypothesized that lack of trust and noncompliance should be highest among those who have the most to lose and the least to use for the anonymity provided by using these techniques. Some researchers noticed the existence of noncompliance and proposed new models to measure noncompliance in order to get reliable information. However, all proposed methods are based on randomized response models which require randomizing devices, restrict the survey to only face-to-face interview and are lack of re-productivity. Taking the noncompliance into consideration, we introduce new non-randomized response techniques in which no covariate is required. Asymptotic properties of the proposed estimates for sensitive characteristic and noncompliance probabilities are presented. Some simulation results will be shown to support our proposed method. A real example about premarital sex among university students is use to demonstrate our methodologies.

AN À LA TUKEY CONFIDENCE INTERVAL FOR REGRESSION SLOPE

RÓBERT TÓTH and JÁN SOMORČÍK
Comenius University, Bratislava, Slovakia

We investigated the application of a Tukey's methodology in Theil's regression to obtain a confidence interval for the true slope in the straight line regression model. For already ten years, the resulting confidence interval is implemented in an R package that is widely used

and/or recommended in the literature. However, we illustrate by Monte Carlo simulations, that this approach, unlike the classical Theil's approach based on Kendall's tau, deflates the true confidence level of the resulting interval. We are also able to provide rigorous proofs in case of $n = 4$ data points (in general) and in case of $n = 5$ data points (under uniform distribution of errors).

ARTIFICIAL INTELLIGENCE OPTIMIZATION METHODS APPLICATION IN COMPARTMENT MODELS

ÖZLEM TÜRKŞEN
Ankara University, Turkey
and
MÜJGAN TEZ
Marmara University, Istanbul, Turkey

Mathematical modeling of pharmacokinetics has a crucial role in drug development. In pharmacokinetic studies, compartment models are commonly used for modeling. In order to estimate the pharmacokinetic parameters, a derivative-based method, called stripping, has been commonly used in drug studies until now. In this study, artificial intelligence optimization methods, which are Genetic Algorithm (GA), hybrid of GA with Nelder-Mead Simplex (NMS) method (GAHNMS) and Particle Swarm Optimization (PSO), are used for parameter estimation procedure. By using these intelligence methods, calculations become easier because these methods are all population based and do not need any assumptions on compartment models. A data set with two compartment model is preferred as application from the literature. It is seen from the results that the suggested intelligence method GAHNMS is more preferable among the other methods with consistence parameter estimates and small error function values.

CUSUM PROCEDURES BASED ON SIGNED SEQUENTIAL RANKS

CORLI VAN ZYL and FREEK LOMBARD
North–West University, Potchefstroom, South Africa

The statistical literature abounds in results about, and refinements of, cumulative sum (CUSUM) procedures and their application in diverse scientific fields. By far the majority of the literature is concerned with procedures based on specific parametric assumptions regarding the form of the underlying data generating mechanism. These procedures are typically rather sensitive to deviations from the underlying distributional assumptions. We propose non-parametric CUSUM procedures to detect changes from known initial values of the location or scale parameters of a symmetric distribution. Firstly, we motivate the necessity of a non-parametric CUSUM procedure by illustrating the non-robustness to deviations from the normality assumption of the

standard normal CUSUM procedure. Secondly, we propose a CUSUM procedure based on signed sequential ranks to detect a change in location and we relate the appropriate offset to the target change size in the location parameter. Thirdly, we also propose a non-parametric CUSUM procedure to detect a change in the scale parameter. The performance of the CUSUMs is evaluated by theoretical calculations supplemented by Monte Carlo simulation. The use of the proposed procedures is illustrated by application to some data from the mining industry.

A SIMULATION STUDY OF THE PERFORMANCE OF CONCOMITANT VARIABLES IN FINITE MIXTURE REGRESSION MODELS

KRISTÝNA VAŇKÁTOVÁ and EVA FIŠEROVÁ
Palacký University, Olomouc, Czech Republic

The classical linear regression is based on an assumption of homogeneous population. Finite mixture regression models are a popular technique for modelling unobserved heterogeneity in the population that a sample represents. Such a sample can be clustered into groups by modelling the conditional distribution of the response given the explanatory variable as a mixture. In most applications, the parameters of a mixture of linear regression models are estimated by maximum likelihood using the expectation maximization (EM) algorithm.

In order to characterize the different components and improve regression parameter estimates and predictions the use of concomitant variables has been proposed. Our goal is to compare the performance of standard mixture of regressions and mixture of regressions with concomitant variable. A simulation study is designed to investigate this problem and it is focused on the accuracy of parameter estimations and following predictions. R package flexmix is used for modelling of mixtures using the EM algorithm.

MODEL IDENTIFICATION AND FOLLOW-UP INFERENCE IN FIXED EFFECTS LINEAR MODELS

JÚLIA VOLAUFOVÁ, LYNN ROY LAMOTTE and ONDREJ BLAHA
Louisiana State University Health Sciences Center – School of Public Health, New Orleans, USA

The R^2 statistic in fixed-effects regression settings, for a given set of responses, is a monotone increasing function in the number of explanatory variables. It is used as a basis for selection algorithms, such as the "all-submodels" method. In a variety of statistical applications it is not unusual to establish first a parsimonious model from among a large number of explanatory variables and after the "best" model is identified to pursue inference about the parameter of interest. Here we investigate the adjusted R^2, as well as other criteria, such as AIC, BIC, and SBIC, from a slightly different point of view. Assuming that there is an underlying true model,

we try to address the question, how well or how likely a given criterion identifies the true model. In the second part, we investigate a setting in which the effect of interest is part of the linear fixed effects model; however, additional explanatory variables are included as additive effects by use of a model identification process using different information criteria. The resulting model hence may be miss-specified. The accuracy of p-values of usual test statistics in the "best" selected model are investigated. We illustrate our results in a simulation study.

STATISTICAL INFERENCE IN COX MODELS

KORAKOT WICHITSA-NGUAN
University of Potsdam, Germany

The estimation of parameters or testing of hypotheses about parameters is formulated in given models, especially in a Cox model

$$\lambda(t|\mathbf{x}) = \lambda_0(t)\exp(\boldsymbol{\beta}_0^T\mathbf{x}). \tag{1}$$

Here the estimation of $\boldsymbol{\beta}_0$ will be discussed. We start with observations, which were realizations of i.i.d. copies $(T_i, \delta_i, \mathbf{X}_i)$, $i = 1, \ldots, n$ of $(T, \delta, \mathbf{X})$. Let $T_i = \min\{T_i^*, C_i\}$ where T_i^* is the individual i'th survival time and $\delta_i = \mathbf{1}(T_i^* \leq C_i)$ is the censoring indicator function ($\delta_i = 1$ if event has occurred, 0 if the lifetime is censored), where C_i is the individual i'th censoring time. The $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^T$ is the individual i'th random covariate. The observed information matrix is given by $\mathbf{I}_n(\boldsymbol{\beta})$ and we have

$$n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{\mathrm{D}} \mathsf{N}(0, \Sigma^{-1}(\boldsymbol{\beta}_0, \lambda_0))$$

for the maximum partial likelihood estimate $\hat{\boldsymbol{\beta}}_n$ and

$$\Sigma(\boldsymbol{\beta}_0, \lambda_0) = \mathrm{plim}_{n\to\infty} n^{-1}\mathbf{I}_n(\boldsymbol{\beta}_0)$$

and

$$\Sigma(\boldsymbol{\beta}_0, \lambda_0) = \mathrm{plim}_{n\to\infty} n^{-1}\mathbf{I}_n(\hat{\boldsymbol{\beta}}_n).$$

If one likes to estimate $\boldsymbol{\beta}_0$ then the basic property of the estimate is to have a well defined asymptotic variance matrix $\Sigma^{-1}(\boldsymbol{\beta}_0, \lambda_0)$.

The observed information matrix $\mathbf{I}_n(\boldsymbol{\beta})$ depends on the $t_1, \ldots, t_n$, $\delta_1, \ldots, \delta_n$ and $\mathbf{x}_1, \ldots, \mathbf{x}_n$. This we express in the notation of the form

$$\mathbf{I}_n(\boldsymbol{\beta}) = \mathbf{I}_n(\boldsymbol{\beta}; t_1, \ldots, t_n; \delta_1, \ldots, \delta_n; \mathbf{x}_1, \ldots, \mathbf{x}_n). \tag{2}$$

but this we will write only in the cases when we need this dependence explicitly. In general we use the shorter notation $\mathbf{I}_n(\boldsymbol{\beta})$.

The $\mathbf{X}_i$, $i = 1, 2, ...$ are covariates which characterize conditions of the considered process described by the model. We assume here that these conditions can be controlled as it is given in many technological or medical problems. If we will consider e.g. the failure times of car tyres for different countries then we use the countries as covariates and for estimating $\boldsymbol{\beta}_0$ we can choose

the countries where we have to measure for estimating β_0 in an optimal way. This means that the model (1) holds, the parameter β_0 is to be estimated and the covariates can be chosen. This can be considered as a problem of experimental design, e.g. one chooses the places where one observes the survival times. As usually in experimental design and estimation problems we formulate the assumptions for estimability and derive the criteria for optimal choices of covariates.

Our procedure of estimating is the maximum partial likelihood estimate which is the appropriate estimate in the Cox model (1) with a general censoring distribution C, covariates $\mathbf{X}$ and an unknown baseline hazard rate $\lambda_0(t)$.

Optimal designs under a survival framework have been considered in López–Fidalgo et al. (2009), Garcet–Rodríguez et al.(2008), Balakrishnan and Han (2007), Schmidt and Schwabe (2015). In these papers the maximum-likelihood estimate was considered. Our approach bases on the maximum partial likelihood estimate(MPLE) and hence we have a completely other asymptotic variance which has to be minimized. One has to find conditions for estimability of the unknown parameters, for asymptotic estimability and one has to determine appropriate representations of the observed information matrix of the MPLE or of the asymptotic variance of the MPLE. This will be represented in this paper. We find that for the Cox model the asymptotic estimability depends only on the support points of the covariate $\mathbf{X}$. Of course an optimal design or an optimal covariate depends moreover on the whole distribution of $\mathbf{X}_i$, T_i^* and the censoring C_i of the observation. We discuss several questions. One sees that the model is nonlinear in the parameter β_0 and the linearity of the logarithm of the hazard rate does not help. This will be referred to the fact that the number of support point of an optimal covariate depends on the value of β_0. For $\beta_0 = 1$ the optimal covariate contains two support points, for $\beta_0 \geq 3$ an optimal covariate contains more than 3 support points. Some properties of the optimal covariates are discussed, partly in cases with graphical plots.

References

[1] BALAKRISHNAN, N. and HAN, D. (2007). Optimal progressive type-II censoring schemes for nonparametric confidence intervals of quantiles. *Communications in Statistics – Simulation and Computation* **36** 1247–1262.

[2] GARCET–RODRÍGUEZ, S., LÓPEZ–FIDALGO, J. and MARTÍN–MARTÍN, R. (2008). Some complexities in optimal experimental designs introduced by real life problems. *Tatra Mountains Mathematical Publications* **39** 135–143.

[3] LÓPEZ–FIDALGO, J., RIVAS–LÓPEZ, M. J. and DEL CAMPO, R. (2009). Optimal designs for Cox regression. *Statistica Neerlandica* **63** 135–148.

[4] SCHMIDT, D. and SCHWABE, R. (2015) On optimal designs for censored data. *Metrika* **78** 237–257.

ON COMPUTING DISTRIBUTION OF THE ANDERSON-DARLING AND THE CRAMÉR-VON MISES STATISTICS BY NUMERICAL INVERSION OF THEIR CHARACTERISTIC FUNCTIONS

VIKTOR WITKOVSKÝ

Slovak Academy of Sciences, Bratislava, Slovakia

The exact small sample distribution of the goodness-of-fit test based on the Anderson-Darling statistic, and/or the Cramér-von Mises statistic, still remains to be unknown or difficult to evaluate. Although the asymptotic distribution is known, see e.g. Anderson and Darling (1952) and Csörgő and Faraway (1996), the analytical inversion of the Fourier (Laplace) transform of such distributions frequently leads to complicated and computationally rather strange expressions. Recently, an efficient method for numerical evaluation of the asymptotic Anderson-Darling distribution, based on a sophisticated recurrence relation, was proposed by Marsaglia and Marsaglia (2004). The algorithm is currently available also in the R-package goftest.

As an alternative, here we suggest to consider a method to compute the distribution (PDF/CDF) based on a direct numerical inversion of the characteristic function (either by using the FFT algorithm, see Hürlimann (2013), and/or other implementations of the numerical inversion formulae, see e.g. Gil-Pelaez (1951)). We shall discuss also other related computational issues of the numerical inversion of the characteristic functions, in particular focused on the applications related to the stochastic (Gaussian) processes.

Acknowledgements

This work was supported by the Slovak Scientic Grant Agency VEGA, grants VEGA 2/0047/15 and VEGA 2/0043/13.

References

[1] ANDERSON, T. W. and DARLING, D. A. (1952). Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. *Annals of Mathematical Statistics* **23** 193–212.

[2] CSÖRGŐ, S. and FARAWAY, J. J. (1996). The exact and asymptotic distributions of Cramér-von Mises statistics. *Journal of the Royal Statistical Society, Series B* **58** 221–234.

[3] MARSAGLIA, G. and MARSAGLIA, J. (2004). Evaluating the Anderson-Darling Distribution. *Journal of Statistical Software* **9**, 1–5.

[4] HÜRLIMANN, W. (2013). Improved FFT approximations of probability functions based on modified quadrature rules. *International Mathematical Forum* **8** 829–840.

[5] GIL–PELAEZ, J. (1951). Note on the inversion theorem. *Biometrika* **38** 481–482.

A SEMIPARAMETRIC REGRESSION MODEL WITH ERRORS IN ALL VARIABLES

SEÇIL YALAZ TOPRAK
Dicle University, Diyarbakır, Turkey
MÜJGAN TEZ
Faculty of Science, Marmara University, Istanbul, Turkey
and
HASAN İLHAN TUTALAR
Dicle University, Diyarbakır, Turkey

This paper develops a method for semiparametric partially linear regression model with measurement errors in all variables. The method is likened to kernel deconvolution method which relies on the assumption that measurement errors densities are known. The availability of two error-contaminated measurements of the independent variables is used to achieve the identification. In the application, the performances of $\hat{\beta}$ and $\hat{g}_n(t)$ are investigated through Monte Carlo experiments.

List of Participants

James S.	Allison	james.allison@nwu.ac.za
Mariano	Amo–Salas	mariano.amo@uclm.es
Jaromír	Antoch	antoch@karlin.mff.cuni.cz
Barbora	Arendacká	barendacka@gmail.com
Oykum E.	Aşkin	oykumesra@gmail.com
Alena	Bachratá	alena.bachrata@fmph.uniba.sk
Katarína	Bartošová	katarina.cimermanova@gmail.com
Eva	Benková	eva.benkova@jku.at
Katarína	Burclová	katarina.burclova@gmail.com
Wai	Chan	wchan@psy.cuhk.edu.hk
Martina	Chvosteková	martina.chvostekova@upol.cz
Somnath	Datta	somndatta@gmail.com
Hilmar G.	Drygas	hilmar@drygas.de
Kent M.	Eskridge	keskridge1@unl.edu
Lenka	Filová	filova@fmph.uniba.sk
Eva	Fišerová	eva.fiserova@upol.cz
Nancy	Flournoy	flournoyn@missouri.edu
Andrej	Gajdoš	andrejg44@gmail.com
Mousa	Golalizadeh	golalizadeh@modares.ac.ir
Nesrin	Güler	nesring@sakarya.edu.tr
Martina	Hančová	martina.hancova@upjs.sk
Radoslav	Harman	harman@fmph.uniba.sk
Zdeněk	Hlávka	hlavka@seznam.cz
Daniel	Hlubinka	hlubinka@karlin.mff.cuni.cz
Marie	Hušková	huskova@karlin.mff.cuni.cz
Deniz	İnan	denizlukuslu@marmara.edu.tr
Jozef	Jakubík	jozef.jakubik.jefo@gmail.com
Stanislav	Katina	katina@math.muni.cz
Roger	Koenker	rkoenker@uiuc.edu
Veronika	Kopčová	veronika.kopcova@student.upjs.sk
Michaela	Koščová	michaela.koscova@fmph.uniba.sk
Fİkriye	Kurtoğlu	fkurtoglu@cu.edu.tr
Petr	Lachout	petr.lachout@mff.cuni.cz
Lynn R.	LaMotte	llamot@lsuhsc.edu
Jesús	López–Fidalgo	jesus.lopezfidalgo@uclm.es
Matus	Maciak	maciak@karlin.mff.cuni.cz
Ján	Mačutek	jmacutek@yahoo.com
Jaroslav	Marek	jaroslav.marek@upce.cz
Simos G.	Meintanis	simosmei@econ.uoa.gr

Bojana	MILOŠEVIĆ	bojana@matf.bg.ac.rs
Ivan	MIZERA	imizera@yahoo.com
Märt	MÖLS	martm@ut.ee
Christine H.	MÜLLER	cmueller@statistik.tu-dortmund.de
Werner G.	MÜLLER	werner.mueller@jku.at
Stanislav	NAGY	stanislav.nagy@wis.kuleuven.be
Marko	OBRADOVIĆ	marcone@matf.bg.ac.rs
M. R.	ÖZKALE	mrevan@cu.edu.tr
Victor	PATRANGENARU	vic@stat.fsu.edu
Zbyněk	PAWLAS	pawlas@karlin.mff.cuni.cz
Andrej	PÁZMAN	pazman@fmph.uniba.sk
Michal	PEŠTA	michal.pesta@mff.cuni.cz
Jan	PICEK	jan.picek@tul.cz
Charl	PRETORIUS	charl.pretorius@nwu.ac.za
Luc	PRONZATO	pronzato@i3s.unice.fr
Simo	PUNTANEN	simo.puntanen@uta.fi
Lizanne	RAUBENHEIMER	l.raubenheimer@ru.ac.za
Ina	REICHERT	ina.reichert@uni-weimar.de
Juan M.	RODRIGUEZ–DIAZ	juanmrod@usal.es
Samuel	ROSA	samuel.rosa@fmph.uniba.sk
Zuzana	ROŠŤÁKOVÁ	zuzana.rostakova@gmail.com
Leonard	SANTANA	leonard.santana@nwu.ac.za
Pascal	SARDA	sarda@univ-tlse2.fr
Busenur	SARICA	busenur.sarica@marmara.edu.tr
Miroslav	ŠIMAN	siman@utia.cas.cz
Georgy	SOFRONOV	georgy.sofronov@mq.edu.au
Ján	SOMORČÍK	somorcik@fmph.uniba.sk
Maria K.	STEINER	maria.steiner@uni-weimar.de
Cornelia J.	SWANEPOEL	cornelia.swanepoel@nwu.ac.za
Jan W. H.	SWANEPOEL	jan.swanepoel@nwu.ac.za
Gábor	SZŰCS	szucs@fmph.uniba.sk
Man L.	TANG	mltang@hsmc.edu.hk
Müjgan	TEZ	mtez@marmara.edu.tr
Özlem	TÜRKŞEN	turksen@ankara.edu.tr
Florin	VAIDA	fvaida@ucsd.edu
Corli	VAN ZYL	vanzylcorli@gmail.com
Kristýna	VAŇKÁTOVÁ	kristyna.vankatova@centrum.cz
Júlia	VOLAUFOVÁ	jvolau@lsuhsc.edu
Korakot	WICHITSA-NGUAN	baibuabok9@hotmail.com
Gejza	WIMMER	wimmer@mat.savba.sk
Viktor	WITKOVSKÝ	witkovsky@savba.sk
Seçil	YALAZ TOPRAK	secilyalaz@gmail.com

Index of Authors