PROBASTAT 2011

The Sixth International Conference on Mathematical Statistics
honoring the 80th birthday of Professor Lubomír Kubáček

ABSTRACTS

July 4-8, 2011, Smolenice Castle, Slovak Republic

PROBASTAT 2011
The Sixth International Conference on Mathematical Statistics
honoring the 80^{th} birthday of Professor Lubomír Kubáček
July 4 – 8, 2011
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

SPONSORED BY
National Bank of Slovakia
SAS Institute GmbH, Slovakia

PROBASTAT 2011
Abstracts

EDITORS
Radoslav Harman, Ján Somorčík & Viktor Witkovský
PUBLISHER
Institute of Measurement Science
Slovak Academy of Sciences, Bratislava, Slovak Republic
Bratislava, June 2011

Contents

Invited Talks

OPTIMAL DESIGN OF EXPERIMENTS WITH VERY LOW REPLICATION

ROSEMARY A. BAILEY
Queen Mary, University of London, UK

Trials of new crop varieties usually have very low average replication. Thus one possibility is to have a single plot for each new variety and several plots for a control variety, with the latter well spread out over the field. A more recent proposal is to ignore the control, and instead have two plots for each of a small proportion of the new varieties.

Variation in the field may be accounted for by a polynomial trend, by spatial correlation, or by blocking. However, if the experiment has a second phase, such as making bread from flour milled from the grain produced in the first phase, then that second phase usually has blocks. The optimality criterion used is usually the A criterion: the average variance of the pairwise differences between the new varieties. I shall compare designs under the A criterion when the average replication is much less than two.

EPIDEMICS ON RANDOM NETWORKS WITH HOUSEHOLD STRUCTURE

FRANK BALL
University of Nottingham, UK

There has been a growing interest in models for epidemics among structured populations, which incorporate realistic departures from homogeneous mixing whilst maintaining mathematical tractability. Two classes of structured population epidemic models that have attracted considerable recent attention are network models (in which there is a random graph describing possible infectious contacts) and household models (in which the population is partitioned into households with different contact rates for within- and between-household infection). In this talk I describe and analyse a model for the spread of an SIR (susceptible $\rightarrow$ infective $\rightarrow$ removed) epidemic that includes both of these features. The analysis includes deriving a threshold parameter which determines whether or not an epidemic with few initial infectives can become established and lead to a major outbreak, and determining the probability and expected relative final size of a major outbreak. The model is compared and contrasted with standard household and network models. Vaccination strategies are briefly considered, as are limitations and extensions of the model.

Acknowledgements

Based on work done jointly with David Sirl (Loughborough University) and Pieter Trapman (Stockholm University).

ADAPTIVE DESIGNS FOR INFERENCE AND ETHICS

ALESSANDRA GIOVAGNOLI
University of Bologna, Italy

In experiments aimed at comparing two or more treatments, the target treatment allocation may be known, e.g. balance, or depend on the unknown parameters of the statistical model; the latter is known in the statistical literature as "local optimality". A possible solution consists in sequential experimentation, so hat assignments can be redressed towards the unknown target. Moreover, randomization, namely assigning statistical units to the treatments randomly, is widely regarded as the most scientifically sound approach in order to mitigate several types of bias. However, in clinical trials randomization is not always ethical or practical, and the attention has recently turned to "adaptive" designs, aimed at minimizing the number of subjects exposed to the inferior treatments, which are also sequential.

The first part of the talk will review the theoretical issues involved in adapting the experiment to past allocations and/or observations. Some recent proposals of adaptive designs for clinical research that trade off ethical and inferential demands will be presented in the second part.

References

[1] BALDI ANTOGNINI, A. and GIOVAGNOLI, A. (2010). Compound Optimal Allocation for Individual and Collective Ethics in Binary Clinical Trials. *Biometrika* **97** 935–946.
[2] BALDI ANTOGNINI, A. and GIOVAGNOLI, A. (2006). On the asymptotic inference for response-adaptive experiments. *Metron* **64** 29–45.
[3] BALDI ANTOGNINI, A. and GIOVAGNOLI, A. (2005). On the large sample optimality of sequential designs for comparing two or more treatments. *Sequential Analysis* **24** 205–217.

MARKOV CHAINS WITH A SPECIFIED DIRECTED GRAPH

STEPHEN KIRKLAND
National University of Ireland, Maynooth, Ireland

We consider discrete time, time homogeneous ergodic Markov chains on a finite state space. Many such Markov chains are equipped with a combinatorial structure that arises by considering the zero-nonzero pattern of the corresponding transition matrix. This combinatorial information can be encoded in a directed graph, and it is natural to wonder how the structure of that directed graph is reflected in various quantities of interest associated with the Markov chain. In this talk we will discuss several results along these lines, focusing on how the directed graph influences the eigenvalues of the transition matrix, the entries in the stationary vector, and the Kemeny constant.

ESTIMATING INCIDENCE FROM CROSS-SECTIONAL BIOMARKER DATA

MICHAL KULICH*, ARNOŠT KOMÁREK, MAREK OMELKA,
DANIEL HLUBINKA and ZDENĚK HLÁVKA
Charles University, Prague, Czech Republic

Disease incidence is usually estimated by following a cohort of disease-free individuals over a specified period, subjecting them to repeated disease tests and registering the observed disease occurrences. This procedure may be expensive, time consuming, and inappropriate when the disease test is invasive or affects future incidence. In the context of HIV incidence assessment, several biological markers related to the duration of the infection have been developed. It was suggested that the number of subjects infected within some time period be estimated from a single measurement of such a marker on each of the infected individuals. We review the procedures for estimating incidence from such data and argue that they are generally inconsistent. We present an alternative approach based on mixture analysis.

OPTIMAL DESIGN FOR NONLINEAR MIXED EFFECTS MODELS: OPTIMIZATION OF SAMPLING SCHEMES FOR POPULATION PHARMACOKINETIC/PHARMACODYNAMIC STUDIES

SERGEI LEONOV
GlaxoSmithKline, Collegeville, USA

We concentrate on design of experiments for nonlinear mixed effects models which typically arise in various population pharmacokinetic/pharmacodynamic (PK/PD) studies. We briefly discuss various population optimal design tools (PFIM, PopDes, PopED, WinPOPT) and then focus on PkStaMp library which is intended for constructing locally D-optimal designs for compartmental PK and PK/PD models. Examples of optimal sampling schemes for several PK and combined PK/PD models are provided.

ADAPTIVE AND INTERACTING MARKOV CHAIN MONTE CARLO METHODS

ERIC MOULINES
Telecom Paris Tech, France

Adaptive and interacting Markov chain Monte Carlo algorithms (MCMC) have been recently introduced in the literature. These novel simulation algorithms are designed to increase the simulation efficiency to sample complex distributions. Motivated by some novel algorithms

introduced recently (such as the Adaptive Metropolis (AM) algorithm and the Equi-Energy (EE) sampler), we develop a general methodological and theoretical framework, covering a large number of adaptation algorithms. Within this theoretical framework, we are able to cope with both *internal* and *external* adaptation schemes. In the external scheme, the adaptation is carried out using a certain number of auxiliary processes which can be interact. We also consider the case where the parameter to adapt is infinite dimensional. This framework covers, among other, the AM and the EE samplers, but much more general variants (coupling the two adaptations techniques) can be considered as well.

EQUALITIES OF THE BLUES AND/OR BLUPS IN TWO LINEAR MODELS

SIMO PUNTANEN

University of Tampere, Finland

In this talk we consider two linear models, M_1 and M_2, which differ in their covariance matrices. We review conditions under which the best linear unbiased estimator (BLUE) of an estimable parametric function under M_1 continues to be BLUE also under M_2. Similarly, we consider linear models, L_1 and L_2, with new unobserved future observations, and give conditions that the best linear unbiased predictor (BLUP) of the new observation under the model L_1 continues to be BLUP also under the model L_2. These results can be applied for studying the BLUPs of random effects in two mixed models.

References

[1] HASLETT, S. J. and PUNTANEN, S. (2010a). Equality of BLUEs or BLUPs under two linear models using stochastic restrictions. *Statistical Papers* **51** 465–475.

[2] HASLETT, S. J. and PUNTANEN, S. (2010b). A note on the equality of the BLUPs for new observations under two linear models. *Acta et Commentationes Universitatis Tartuensis de Mathematica* **14** 27–33.

[3] ISOTALO, J., MÖLS, M. and PUNTANEN, S. (2006). Invariance of the BLUE under the linear fixed and mixed effects models. *Acta et Commentationes Universitatis Tartuensis de Mathematica* **10** 69–76.

[4] MARKIEWICZ, A., PUNTANEN, S. and STYAN, G. P. H. (2010). A note on on the interpretation of the equality of OLSE and BLUE. *Pakistan Journal of Statistics* **26** 127–134.

[5] MITRA, S. K. and MOORE, B. J. (1973). Gauss–Markov estimation with an incorrect dispersion matrix. *Sankhyā, Ser. A* **35** 139–152.

[6] RAO, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Berkeley, California, 1965/1966* **1** (Le Cam, L. M. and J. Neyman, J., eds.) 355–372. University of California Press, Berkeley.

A COMPACT FUNCTIONAL ESTIMATE OF A FUNCTIONAL VARIANCE-COVARIANCE OR CORRELATION KERNEL

JAMES O. RAMSAY

McGill University, Ottawa, Canada

In functional data analysis, as in its multivariate counterpart, estimates of the bivariate covariance kernel $\sigma(s,t)$ and its inverse are useful for many things. However, the dimensionality of functional observations often exceeds the sample size available to estimate $\sigma(s,t)$. Then the analogue of the multivariate sample estimate is singular and non-invertible. Even when this is not the case, the high dimensionality of the usual estimate often implies unacceptable sample variability and loss of degrees of freedom for model fitting. The common practice of employing low-dimensional principal component approximations to $\sigma(s,t)$ to achieve invertibility also raises serious issues.

This talk describes a functional and nonsingular estimate of $\sigma(s,t)$ defined by an expansion in terms of finite element basis functions that permits the user to control the resolution of the estimate as well as the time lag over which covariance may be nonzero. This estimate also permits the estimation of covariances and correlations at observed pairs of sampling points, and therefore has applications to many classical statistical problems, such as discrete but unequally spaced time and spatial series.

VOLATILITY ESTIMATION FROM NOISY OBSERVATIONS AND LE CAM THEORY

MARKUS REIß

Humboldt-Universität zu Berlin, Germany

We consider discrete-time observations of a continuous martingale under measurement error. This serves as a fundamental model for high-frequency data in finance, where an efficient price process is observed under microstructure noise. It is shown that this nonparametric model is in Le Cam's sense asymptotically equivalent to a Gaussian shift experiment in terms of the square root of the volatility function σ and a nonstandard noise level. As an application, new rate-optimal estimators of the volatility function and simple efficient estimators of the integrated volatility are constructed.

EXPLICIT ESTIMATORS IN UNBALANCED MIXED LINEAR MODELS

DIETRICH VON ROSEN*

Swedish University of Agricultural Sciences, Uppsala, Sweden

TATJANA VON ROSEN

Stockholm University, Sweden

and

JÚLIA VOLAUFOVÁ

Louisiana State University Health Sciences Center, New Orleans, USA

Consider the following mixed linear model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma} + \epsilon,$$

where $\boldsymbol{\beta}$ is to be estimated and $\boldsymbol{y}$: $n \times 1$, the random vector, $\boldsymbol{X}$: $n \times k$, $\boldsymbol{Z}$: $n \times l$ are known matrices both supposed to be of full column rank, $\mathcal{C}(\boldsymbol{X}) \subseteq \mathcal{C}(\boldsymbol{Z})$, where $\mathcal{C}(\cdot)$ denotes the column vector space, $\boldsymbol{\gamma} \sim N_l(\boldsymbol{0}, \sigma_\gamma^2 \boldsymbol{I}_l)$, $\boldsymbol{\epsilon} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$, where where $\boldsymbol{\gamma}$ and $\boldsymbol{\epsilon}$ are supposed to be independent and the variance parameters σ_γ^2 and σ^2 are unknown.

The aim of the research project is to derive explicit estimators for $\boldsymbol{\beta}$ and the variance parameters σ_γ^2 and σ^2. Moreover, some explicit tests concerning the parameters will also be derived.

The model may be unbalanced and then it becomes complicated to find explicit estimators with reasonable uncertainty as well as to construct exact tests of appropriate power.

Testing of both variance components and fixed effects in an unbalanced mixed model relies usually on approximations, in particular, Satterthwaite's approximation of the test statistics. The derived tests have unknown distributions, both under the null and alternative hypotheses, due to the lack of independence and chi-squaredness of the mean squares involved. In this presentation we will adopt ideas from, among others, Gallo & Khuri (1990) and Öfversten (1993) where a resampling approach has been presented.

References

[1] GALLO, J. and KHURI, A. I. (1990). Exact tests for the random and fixed effects in an unbalanced mixed two-way cross-classification model. *Biometrics* **46** 1087–1095.

[2] ÖFVERSTEN, J. (1993). Exact tests for variance components in unbalanced mixed linear models. *Biometrics* **49** 45–57.

ON R-ESTIMATORS IN ERRORS-IN-VARIABLES MODELS

SILVELYN ZWANZIG
Uppsala University, Sweden

Errors-in-variables models are regressions models where the independent variables are observed with an additional error. That makes the inference much more complicated. The naive use of regression estimators by disregarding the errors in the variables delivers biased, inconsistent estimators. Under normal error distributions the maximum likelihood estimator in the errors-in-variables model coincides with the orthogonal regression estimator and is optimal and consistent.

In this talk rank methods are considered. The naive use of the regression rank estimators minimizing the Jaeckel's dispersion is not appropriate. These estimators are based on vertical residuals and are inconsistent. In errors-in-variables models rank estimators should be defined by using the ranks of the orthogonal residuals.

Another approach is presented for the structural model, that is an errors-in-variables model where the unobserved variables are i.i.d. In this model we can find a transformation depending on the true parameter, such that the transformed observations are uncorrelated. A rank estimation is now defined as a parameter value in the transformation for which the rank correlation measure of the transformed observations is zero. By the way the Pearson correlation delivers the orthogonal regression estimator.

Short Contributions & Posters

BIVARIATE WEIBULL DISTRIBUTIONS DERIVED FROM COPULA FUNCTIONS IN THE PRESENCE OF CURE FRACTION AND CENSORED DATA

JORGE ALBERTO ACHCAR* and EDSON ZANGIACOMI MARTINEZ

Universidade de Sao Paulo, Brazil

In this paper, we introduce bivariate Weibull distributions derived from copula functions in presence of cure fraction, censored data and covariates. Two copula functions are explored in this paper: the FGM (Farlie - Gumbel - Morgenstern) copula and the Gumbel copula. Inferences for the proposed models are obtained under the Bayesian approach, using standard MCMC (Markov Chain Monte Carlo) methods. An illustration of the proposed methodology is given considering a medical data set.

GENERALIZED FIDUCIAL PIVOTS IN RANDOM EFFECTS MODELS

BARBORA ARENDACKÁ

Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia

We will consider generalized fiducial pivots for variance components in 2-way random effects models (partially balanced and unbalanced), alternative to those considered by Burch (2007). Similar possibilities for fiducial generalized inference for variance components exist in both the homoscedastic and heteroscedastic 1-way random effects models and we will try to point out the general concept behind. That will lead us to a discussion of a possible inference in a general 2-way random effects model (without interactions), not covered by Burch.

Acknowledgements

This work was supported by the Slovak Research and Development Agency under the contract No. LPP-0388-09 and further by the grants VEGA 1/0077/09 and VEGA 2/0019/10.

References

[1] BURCH, B. D. (2007). Generalized confidence intervals for proportions of total variance in mixed linear models. *Journal of Statistical Planning and Inference* **137** 2394–2404.

CARBON DATING OF THE SHROUD OF TURIN: PARTIALLY LABELLED REGRESSORS AND THE DESIGN OF EXPERIMENTS

ANTHONY C. ATKINSON*

London School of Economics, UK

and

MARCO RIANI

University of Parma, Italy

Tradition claims that the Turin Shroud (TS) is the cloth Christ was enveloped in when placed in a tomb in Palestine about 2000 years ago. However, results from radio carbon dating in 1988 led to dates between 1260 and 1390 AD, which agrees with the undisputed historical records of the existence of the TS, which go back to AD 1357.

The results for the TS show more inter-laboratory variability than those of control samples also dated in 1988. It is possible to explain this heterogeneity by using the partial information on the spatial location of the samples sent to the four laboratories. 387,072 bivariate regressions are required to allow for all spatial allocations of subsamples. The major problem is that the samples were taken from one corner of the TS. A question is how the samples might have been taken to provide clear evidence of the age of the complete fabric.

PREDICTION OF FUTURE FIELD FAILURES BASED ON CLAIM DATA

JAIWOOK BAIK

Korea National Open University, Seoul, Korea

Manufactures must predict the number of future field failures frequently based on claim data, especially when an unanticipated failure mode is detected in the field. These predictions are needed to quantify future warranty costs and ensure a sufficient amount of spare parts available. In this talk various types of claim data are considered and proper approaches to the estimation of parameters are discussed.

CLASSIFICATION OF BREATH GAS COMPOUNDS FOR DETECTION OF SEVERE DISEASES

KATARÍNA BARTOŠOVÁ
Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia

In the present time, by development of analytical methods for measuring small quantities of Volatile Organic Compounds (VOCs), the Breath Analysis has become an interesting method because of the possibility of early noninvasive diagnosis of severe diseases such as lung cancer. We describe appropriate binary classification methods to the Breath Analysis.

The first described method is the well-known Fisher linear discriminant analysis (FLDA) method using the Youden index. In our situation the observed data include variability of repetitive measurements, noisy data. A solution to this problem is a robust formulation that stems from the Support Vector Machine (SVM) method. For classification of data when some of the entries are missing (NaN, Not a Number), e.g. error device, we use Decision Tree (DT) classification method.

OPTIMUM DESIGN OF EXPERIMENTS FOR ENZYME INHIBITION KINETIC MODELS

BARBARA BOGACKA*
Queen Mary, University of London, UK
MACIEJ PATAN
University of Zielona Gora, Poland
PATRICK J. JOHNSON and KURESH A. YOUDIM
Pfizer Ltd, Sandwich, UK
and
ANTHONY C. ATKINSON
London School of Economics, UK

We find closed-form expressions for the D-optimum designs for three- and four-parameter nonlinear models arising in kinetic models for anzyme inhibition. We calculate the efficiency of designs over a range of parameter values and make recommendations for design when the parameter values are not well known. Experiments performed in Pfizer laboratories confirm the theoretical results and show that huge savings can be made by use of D-optimum designs for such models.

ON COMPUTATIONAL COMPLEXITY OF CONSTRUCTION OF C-OPTIMAL LINEAR REGRESSION MODELS OVER FINITE EXPERIMENTAL DOMAINS

MICHAL ČERNÝ*
University of Economics, Prague, Czech Republic
and
JAROMÍR ANTOCH
Charles University, Prague, Czech Republic

Assume the linear regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. A lot of work has been already devoted to the construction of the $\boldsymbol{c}$-optimal design in such a setting. In our lecture we will discuss selected complexity-theoretic results on finding $\boldsymbol{c}$-optimal designs over finite experimental domains. We show implications of the complexity-theoretic results for the analysis of existing algorithms and for construction of new algorithms for the design problem. We point out a close relation between the design of $\boldsymbol{c}$-optimal experiments and linear programming. We show that the design problem does not have efficient parallel algorithms (assuming some complexity-theoretic conjectures generally believed to be true). We study the question whether the design problem has strongly polynomial algorithms. We distinguish between the approximate and the exact version of the design problem and show how they differ from complexity-theoretic point of view. We introduce restricted versions of the design problem. We point out several open questions in the field.

Acknowledgements

Supported by Project No. F4/18/2011 of Internal Grant Agency of University of Economics, Prague.

References

[1] ARORA, S. and BARAK, B. (2009). *Computational complexity. A modern approach.* Cambridge University Press.
[2] ČERNÝ, M. and HLADÍK, M.. Two complexity results on c-optimality in experimental design. To appear in: *Computational Optimization and Applications*. DOI: 10.1007/s10589-010-9377-8.
[3] ČERNÝ, M., HLADÍK, M. and SKOČDOPOLOVÁ, V. (2010). On computationally complex instances of the c-optimal experimental design problem: Breaking RSA-based cryptography via c-optimal designs. In: *Proceedings of CompStat 2010* (Lechevallier, Y. and Saporta, G., eds.) 879–886. Springer, Proceedings in Computational Statistics.
[4] GREENLAW, R., HOOVER, H. and RUZZO, W. (1995). *Limits to parallel computation. P-completeness theory.* Oxford University Press.
[5] HARMAN, R., and JURÍK, T. (2008). Computing c-optimal experimental designs using the simplex method of linear programming. *Computational Statistics and Data Analysis* **53** 247–254.
[6] PÁZMAN, A. (1986). *Foundations of optimum experimental design.* Reidel Publishing Company, Dordrecht.
[7] PUKELSHEIM, F. and RIEDER, S. (1992). Efficient rounding in approximate designs. *Biometrika* **79** 763–770.

A COMPARISON OF THE SIMULTANEOUS TWO-SIDED TOLERANCE INTERVALS FOR A LINEAR REGRESSION MODEL

MARTINA CHVOSTEKOVÁ
Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia

Simultaneous tolerance intervals for the linear regression model are constructed using a vector of observations, so that with confidence level $1-\alpha$, at least a specified proportion γ of the distribution of the response variable is to be contained in the corresponding tolerance interval simultaneously for all possible predictor values. The most common application for the simultaneous tolerance intervals is in multiple-use calibration problem. All known methods for constructing the intervals are conservative, they exceed the nominal level of confidence. We introduce the method based on the confidence-set approach, i.e. the form of the confidence set for the unknown parameters of the model is specified. The exact $(1-\alpha)$-confidence set is constructed by using likelihood ratio test for testing the null hypothesis about all parameters of the model simultaneously. The method is numerical compared with the known method for constructing simultaneous tolerance intervals for a linear regression based on the confidence set approach. In particular, we compare the Wilson method (1967), the Limam-Thomas method (1988), the modified Wilson method (1988), and the introduced method based on the estimated confidence in the specified simple linear regression model.

Acknowledgements

The paper was supported by the Slovak Research and Development Agency (APVV), grant SK-AT-0003-08, 0096-10/STATGUM/ and by the Scientific Grant Agency of the Slovak Republic (VEGA), grant 1/0077/09 and 2/0019/10.

References

[1] LIMAM, M. M. T. and THOMAS, R. (1988). Simultaneous Tolerance Intervals for the Linear Regression Model. *Journal of the American Statistical Association* **83** 801–804.
[2] WILSON, A., L. (1967). An Approach to Simultaneous Tolerance Intervals in Regression. *The Annals of Mathematical Statistics* **38** 1536–1540.

REGRESSION MODELS AND THE COORDINATE-FREE APPROACH

HILMAR DRYGAS
University of Kassel, Germany

We consider the linear model $y = \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$, $E(\varepsilon) = 0$, $E(\varepsilon\varepsilon') = \sigma^2 I_n$. Let $\hat{\beta}_k$ be the Gauss-Markov estimator of β_k, which is assumed to be estimable. $y - \hat{\beta}_k x_k$ is a random

vector with expectation $\beta_1 x_1 + \cdots + \beta_{k-1} x_{k-1}$ and singular covariance matrix. It is, however, more convenient to determine the covariance-operator rather than the covariance matrix.

CRITERION-ROBUST DESIGNS FOR THE MODEL OF SPRING BALANCE WEIGHING WITH A CONSTANT BIAS

LENKA FILOVÁ* and RADOSLAV HARMAN
Comenius University, Bratislava, Slovakia
and
THOMAS KLEIN
Technische Universität München, Germany

We analyze the linear regression model with a nonzero intercept term on the vertices of a d-dimensional unit cube. This setting may be interpreted as a model of weighing d objects on a spring balance with a constant bias.

Using criteria based on partial sums of eigenvalues of the information matrix and the methods of semidefinite programming, we compute designs that are robust with respect to orthogonally invariant information functions, the so-called $\mathbb{O}$-maximin efficient designs. We also compare the performance of these designs to commonly used D-, A-, and E-optimal designs.

ONE KIND OF REDUCTION OF MULTIVARIATE STATISTICAL MODELS

EVA FIŠEROVÁ* and LUBOMÍR KUBÁČEK
Palacký University Olomouc, Czech Republic

Multivariate statistical models, or the so-called multivariate multiple regression models, are utilized widely for studying relationships between a set of multiresponse data and a set of regressors. Multivariate view in a multiresponse multiple regression situations is important since multiresponse data should be modelled jointly in general.

A multivariate model for a p-dimensional multiresponse data and a k-dimensional set of regressors (including an intercept) includes a $(k \times p)$-dimensional matrix $\mathbf{B}$ of unknown regression coefficients. In some situations, the multivariate model can be reduced into a system of simpler models. For example, if each multiresponse is not intercorrelated, the multivariate model can be reduced into p-tuple independent univariate models. Similarly, if multiresponses can be decomposed into two not intercorrelated parts, then the multivariate model can be reduced into two independent multivariate models. A decomposition of multiresponses leads to a corresponding decomposition of the parameter matrix $\mathbf{B}$ with zero blocks on a minor diagonal.

In the contribution we will study when the multivariate model can be approximated by simpler models in general. Naturally, a part of model's parameters can be sometimes neglected if their

magnitude is sufficiently small. The aim of the contribution is to present some conditions under which the neglecting of some model's parameters leads to the reduction of the multivariate model into two simpler multivariate models with more accurate estimators than in the original one.

PARAMETER ESTIMATION OF TIME SERIES MODELS BASED ON ROUNDED DATA

MEI-HUI GUO* and GEN-LIANG LI
National Sun Yat-sen University, Kaohsiung, Taiwan

Most recorded data are rounded to the nearest decimal place due to the precision of the recording mechanism. This rounding entails errors in estimation and measurement. In this paper, we compare the performances of three types of estimators based on rounded data of time series models, namely A-K corrected estimator (Stam and Cogger, 1993), approximate maximum likelihood estimator (Bai, Zheng, Zhang and Hu, 2009) and the short, overlapping series (SOS) estimator (Zhang, Liu and Bai, 2010). In order to perform the comparison, the A-K corrected estimators for the MA(1) model are derived theoretically. To improve the efficiency of the estimation, two types of variance-reduction estimators based on linear combinations of aforementioned three estimators are proposed. Simulation results show the proposed variance reduction estimators significantly improve the estimation efficiency.

References

[1] Bai, Z. D., Zheng, S. R., Zhang, B. X. and Hu, G. R. (2009). Statistical analysis for rounded data. *Journal of Statistical Planning and Inference* **139** 2526–2542.
[2] Stam, A. and Cogger, K. O. (1993). Rounding errors in autoregressive processes. *International Journal of Forecasting* **9** 487–508.
[3] Zhang, B. X., Liu, T. Q. and Bai, Z. D. (2010). Analysis of rounded data from dependent sequences. *Annals of the Institute of Statistical Mathematics* **62** 1143–1173.

EMPIRICAL PREDICTORS IN FINITE DISCRETE SPECTRUM LINEAR REGRESSION MODELS

MARTINA HANČOVÁ
Pavol Jozef Šafárik University Košice, Slovakia

In a linear regression model of time series the best linear unbiased predictor (BLUP) and its mean squared error generally depends on variance-covariance components of its finite observation. In practice these components are unknown and can be only estimated. We present some results

from our study on properties of an empirical predictor in a finite discrete spectrum linear regression model which is originated by plugging the so-called modified natural parameter estimators into the theoretical BLUP of the model.

A CONDITIONAL DISTRIBUTION APPROACH TO UNIFORM SAMPLING ON SPHERES AND BALLS IN LEBESGUE SPACES

RADOSLAV HARMAN* and VLADIMÍR LACKO
Comenius University, Bratislava, Slovakia

Liang and Ng (2008) proposed a componentwise conditional distribution method for L_p-uniform sampling on L_p-norm n-spheres. On the basis of properties of a special family of L_p-norm spherical distributions we suggest a wide class of algorithms for sampling uniformly distributed points on n-spheres and n-balls in L_p spaces, generalizing the approach of Harman and Lacko (2010), and including the method of Liang and Ng as a special case. We also present results of a numerical study proving that the choice of the best algorithm from the class significantly depends on the value of p. The presentation is based on a paper by Lacko and Harman (to appear in Metrika).

References

[1] LIANG, J. and NG, K. W. (2008). A method for generating uniformly scattered points on the L_p-norm unit sphere and its applications. *Metrika* **68** 83–98.
[2] HARMAN, R. and LACKO, V. (2010). On Decompositional Algorithms for Uniform Sampling from n-Spheres and n-Balls. *Journal of Multivariate Analysis* **101** 2297–2304.
[3] LACKO, V. and HARMAN, R. A conditional distribution approach to uniform sampling on spheres and balls in L_p spaces. *Metrika*, to appear.

OPTIMAL DESIGNS FOR NONPARAMETRIC ESTIMATION OF ZEROS OF REGRESSION FUNCTIONS

ZDENĚK HLÁVKA
Charles University, Prague, Czech Republic

An estimator of location of some important point is often more interesting than an estimator of the regression curve itself. Such interesting points are often defined as zeros (roots) of a regression function or its derivatives. Note, for example, that the location of maximum may be found by looking at the zero of the first derivative or that the inflection point of a growth curve corresponds to a zero of the second derivative. In this contribution, we review asymptotic properties of

nonparametric estimators of zeros of regression functions and, using a fixed-design nonparametric regression estimator and a classical calculus of variations, we derive the asymptotically optimal distribution of the fixed design points. Results are obtained both for independent and mixing random errors.

SEQUENTIAL MONITORING IN AUTOREGRESSION

ZDENĚK HLÁVKA and MARIE HUŠKOVÁ*

Charles University, Prague, Czech Republic

CLAUDIA KIRCH

Karlsruhe Institute of Technology, Germany

and

SIMOS MEINTANIS

National and Kapodistrian University of Athens, Greece

Procedures for monitoring changes autoregressive time series while controlling the overall size of the sequential test will be presented. The changes both in the error distribution and autoregressive parameters will be discussed. The proposed procedures utilize the empirical characteristic function of properly estimated residuals. The limit behavior of the test statistic is investigated under the null hypothesis as well as under alternatives. Since the asymptotic null distribution contains unknown parameters, a bootstrap procedure is proposed in order to actually perform the test, and corresponding results on the finite-sample performance of the new methods are presented.

DATA DEPTH IN CLASSIFICATION

DANIEL HLUBINKA* and STANISLAV NAGY

Charles University, Prague, Czech Republic

The data depth is a nonparametric tool for ordering multivariate (or functional) data. The nonparametric and robust nature of the method was recently used for classification. In our contribution we shall discuss weak and strong points of the depth approach to discrimination. In particular we will discuss definitions of depth functions and their use for discrimination of functional data.

PROPERTIES OF DATA DEPTH CALIBRATION SET

KLÁRA HORNIŠOVÁ
Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia

Properties of single-use calibration set based on simplicial data depth proposed by Zappa are examined in computationally simple case when both predictor and response are 1-dimensional.

LINEAR REGRESSION WITH COMPOSITIONAL EXPLANATORY VARIABLES USING LOGRATIO APPROACH

KAREL HRON*
Palacký University, Olomouc, Czech Republic
and
PETER FILZMOSER
Vienna University of Technology, Austria

Compositional data are defined as multivariate observations, describing quantitatively the parts of some whole, which carry exclusively relative information between the parts; usually, they are represented as proportions or percentages. However, compositional explanatory variables should not be directly used in a linear regression model because any inference statistic can become misleading. While various approaches for this problem were proposed, here an approach based on the isometric logratio transformation is used. It turns out that the resulting model is easy to handle, and that parameter estimation can be done like in usual linear regression. Moreover, it is possible to use the isometric logratio variables for inference statistics in order to obtain an appropriate interpretation of the model.

SECURITY ANALYSIS OF RFID MUTUAL AUTHENTICATION PROTOCOL

YU-JUNG HUANG
I-Shou University, Kaohsiung City, Taiwan

The widespread deployment of RFID technologies may generate new threats to security and user privacy. One of the main drawbacks of RFID technology is the weak authentication systems between a reader and a tag. In general, ”weak” authentication systems that either leak the password

directly over the network or leak sufficient information while performing authentication allow intruders to deduce or guess the password. There has been much focus on the development of more secure RFID identification and authentication protocols in the research community. Due to the lightweight nature of RFID tags, it is an extremely challenging task for designing a secure authentication protocols by relying only on the minimal features available on an RFID tag. This paper presents vulnerability analysis related to RFID (Radio Frequency Identification) tag,reader mutual authentication by using the XOR operation. We compare the level of security offered by examining whether the security properties of the access password and kill password under conditional probability models on permutations. The vulnerability analysis proves not only the enhancement of the security level, but also its performance in the mutual authentication between a RFID tag and a reader.

MARKOV CHAIN PROPERTIES IN TERMS OF COLUMN SUMS OF THE TRANSITION MATRIX

JEFFREY J. HUNTER

Auckland University of Technology, New Zealand

Questions are posed regarding the influence that the column sums of the transition probabilities of a stochastic matrix (with row sums all one) have on the stationary distribution, the mean first passage times and the Kemeny constant of the associated irreducible discrete time Markov chain. Some new relationships, including some inequalities, and partial answers to the questions, are given using a special generalized matrix inverse that has not previously been considered in the literature on Markov chains.

PAIRWISE COMPARISONS FOR PARALLEL PROFILE MODELS WITH MIXED EFFECTS

HIROTO HYAKUTAKE

Kyushu University, Fukuoka, Japan

There are several linear and nonlinear models for analyzing repeated measurements. The mean response for an individual depends on the regression parameters specific to that individual. One of the simple form is the sum of vectors of fixed parameters and random effects. When the models with mixed effects for several groups are parallel, pairwise comparisons of level differences are considered. For the comparisons, simultaneous confidence intervals are given.

CHAOS AND STABILITY FOR DISCRETE RANDOM DYNAMICAL SYSTEMS

KATARÍNA JANKOVÁ
Comenius University, Bratislava, Slovakia

Several ways of introduction of randomness into discrete dynamical systems generated by continuous interval maps are considered. Results obtained for chaos and stability of discrete dynamical systems with small random perturbations are applied to certain types of random dynamical systems. Sufficient conditions for nonchaotic behavior of these systems are given.

GENERALIZED SHAPE SPACE PCA, SYMMETRIC TWO-BLOCK PLS, AND LDA WITH RESPECT TO BENDING ENERGY

STANISLAV KATINA
University of Glasgow, UK
and
Comenius University, Bratislava, Slovakia

We present three multivariate models which decompose biological signal into low-dimensional linear combinations of high-dimensional measurements. We generalize shape-space Principal Component Analysis (PCA), symmetric two-block Partial Least Squares (PLS, of two shape blocks, and one shape block and one block of external variables), and Linear Discriminant Analysis (LDA) with respect to bending energy. In this setting, bending energy represents curvature and reflects spatial relationship between studied variables, here centered Procrustes Shape Coordinates (PSC) calculated by a Generalized Procrustes Analysis (Bookstein 1991) using suitable alignment and transformation of the data. The covariance matrix of PSC is decomposed to within-module and between-module (cross-block) sub-matrices and generalized, i.e. pre- and post-multiplied by bending energy matrix. All methods are discussed in three subspaces: full shape space, non-affine subspace of global and local bending patterns.

The methods mentioned above are applied to Cartesian coordinates of 72 (semi)landmarks (on eyebrow, eyes, nose, lips, and chin) and skin patches sized 150×150 pixels from the cheek taken from 40 standardized 2D facial photographs. The visualisation is performed by suitably extrapolated Thin Plate Spline deformation grids and fields of vectors superimposed on the surface of Procrustes distances in 2D. The programs written in R language (R Development Core Team 2011) are used in examples and figures throughout the talk.

New suggested methods are useful in the decomposition of biological signal into non-affine components which are usually not visible if the data are analysed by classical multivariate methods. Our proposals suggest a possibly novel and interesting fusion of spatial bending patters

with the standard modern data-analytic toolkit, with direct practical implications in applied craniometrics and anthropometrics more generally.

Acknowledgements

The research was supported by Wellcome Trust grant WT086901MA and VEGA grant 1/0077/09. For data acquisition and pre-processing the author thanks Lisa Oberzaucher.

References

[1] BOOKSTEIN, F. L. (1991). *Morphometric Tools for Landmark Data: Geometry and Biology.* Cambridge University Press.

[2] R DEVELOPMENT CORE TEAM (2011). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

PROBABILITY LIMIT IDENTIFICATION FUNCTIONS ON SEPARABLE METRIZABLE SPACES

PAVEL KŘÍŽ
Charles University, Prague, Czech Republic

Simons in 1971 introduced the concept of the probability limit identification function (PLIF). This function identifies almost surely the value of the probability limit of a sequence of random variables on the basis of one realization of the sequence. Štěpán in 1973 proved the existence of the PLIF for real-valued random variables under the continuum hypothesis and Blackwell showed in 1980 that such function can not be Borel measurable. The present poster shows the existence of the PLIF on any separable metrizable topological space under the continuum hypothesis and application of such PLIF to functional representations in stochastic analysis.

PLANNING OF EXPERIMENTS FOR A NONAUTONOMOUS ORNSTEIN-UHLENBECK PROCESS

VLADIMÍR LACKO
Comenius University, Bratislava, Slovakia

We study optimal sampling designs for processes governed by the equation

$$\mathrm{dX_t} = \kappa(\bar{\mathrm{X}} - \mathrm{X_t})\mathrm{dt} + \sigma(\mathrm{t})\mathrm{dW_t},$$

where X_0 and $\bar{X}$ are unknown parameters. We will show that there always exists a nondegenerate optimal design for parameters estimation. Moreover, it turns out that low-volatility points are

more informative. To demonstrate the results we will compute D-optimal designs for the model with exponential volatility.

Acknowledgements

The work has been supported by the VEGA-Grant No. 1/0077/09 and UK-Grant No. 383/2011.

References

[1] LACKO, V. Optimal Sampling Designs for Nonautonomous Ornstein-Uhlenbeck Process, submitted.

[2] HARMAN, R. and ŠTULAJTER, F. (2009). Optimality of Equidistant Sampling Designs for a Nonstationary Ornstein-Uhlenbeck process. *Proceedings of the 6th St. Petersburg Workshop on Simulation* 1097–1101.

L_2 CONSISTENCY

PETR LACHOUT
Charles University, Prague, Czech Republic

The contribution will discuss L_2 consistency of estimators. We will discuss calculation of the mean and the variance of weighted sums of random variables. This simple tool gives us results for a bright family of estimators possessing approximation by a ratio of weighted sums of observations. Also dependency structure can be incorporated using this instrument. We intend to present an application to parameter estimators in linear regression models.

ESTIMABILITY AND RESTRICTIONS IN LINEAR MODELS

LYNN ROY LAMOTTE
Louisiana State University Health Sciences Center, New Orleans, USA

In a linear model $X\beta$ for the mean vector of a multivariate random variable Y, estimable functions of β are customarily defined as those functions $P'\beta$ for which unbiased linear estimators exist. Testable hypotheses about β, on the other hand, are linear relations that restrict the model.

Statistical computing packages, like SAS and ANOVA programs in R, refuse to deal with non-estimable functions. That an estimator is not unbiased doesn't seem to be so fatal a flaw that we should be forbidden even to look at it. Then why is non-estimability so bad?

By examining the relation between estimability and restrictions, this paper will show that the mean vector carries no information at all about non-estimable functions, and that therefore any statement about a non-estimable function based on the mean vector is false. As part of this development, useful representations of linear restrictions on affine sets are shown.

USING INFORMATION MATRIX FOR EVALUATION OF THE EFFECTIVENESS IN PARAMETER ESTIMATION FOR DIFFERENT STATISTICAL MODELS

ANDRZEJ MICHALSKI
Wroclaw University of Environmental and Life Sciences, Wroclaw, Poland

In the paper different forms of information matrix by analogy with classical Fisher's information in statistical models with varying degree of generality will be presented. The information matrices have been determined for the different methods of estimation: FIML (full information maximum likelihood), LIML (limited information maximum likelihood) in the multivariate linear regression models, and also for nonlinear regression models (Magnus et al., 1990) and next for generalized mixed linear models (McCullagh and Nelder, 1990; Wand, 2007). To evaluate the effectiveness of ML- and REML- estimators of variance components in the mixed linear models with two variance components has been used the asymptotic information matrix (see Gnot et al., 2004) as a result of applying generalized Cramér-Rao inequality (cf. Lehman, 1983).

References

[1] GNOT, S., MICHALSKI, A. and URBANSKA-MOTYKA, A. (2004). On some properties of ML and REML estimators in mixed normal models with two variance components. *Discussiones Mathematicae - Probability and Statistics* **24** 109–126.
[2] LEHMAN, E.L. (1983). Theory of Point Estimation. Wiley, New York.
[3] MAGNUS, J. R. and NEUDECKER, H. (1990). Matrix Differential Calculus with Applications in Statistics and Econometrics, revised ed., J. Wiley & Sons, Chichester-N.York-Brisbane-Singapore-Toronto.
[4] MCCULLAGH, P. and NELDER, J. A. (1990). Generalized Linear Models, second ed., Chapman & Hall, London.
[5] WAND, M. P. (2007). Fisher information for generalized linear mixed models. *Journal of Multivariate Analysis* **98** 1412–1416.

STATISTICS FOR EXPERIMENTERS IN R

IVAN MIZERA* and TING DONG
University of Alberta, Edmonton, Canada

An R package entitled BHH2+ is discussed, aiming at providing the open source software support of certain methods for analyzing experiments, which, somewhat surprisingly, appeared to be missing in the volume of publicly available contributed R code. In particular, the focus will be on some specific methods championed by the book of Box, Hunter, and Hunter, as well as on the method of moments for mixed models.

ONE-DIMENSIONAL REPLICATED CALIBRATION MODEL WITH QUADRATIC CALIBRATION FUNCTION

KATEŘINA MYŠKOVÁ
Mendel University, Brno, Czech Republic

The article deals with estimation of unknown parameters in one-dimensional replicated calibration model with a quadratic calibration function. The knowledge of the incomplete indirect measurement model with the second type conditions for the first order parameters is used here for estimation of parameters of mean values and parameters of the calibration function. In addition, it is possible to estimate unknown parameters of covariance matrix in the replicated model. For this purpose we use MINQUE method. Kenward-Roger approach is applied for construction of confidence intervals.

AN ANALYSIS OF THE SURVIVAL TIMES OF THE PATIENTS WITH ACUTE MYELOID LEUKEMIA BY USING A CHANGING FORM OF THE HAZARD FUNCTION

DIMITRIOS PANAGIOTIS NIKOLAKIS* and JOHN E. GOULIONIS
University of Piraeus, Greece

Survival pattern can be conveniently inspected by using the hazard function. It is informative to examine the hazard function, when time to response data is being analyzed. In this paper Cox Proportional Hazard Model is fitted to patient data of Acute Myeloid Leukemia AML. The natural log of the hazard function is the response and two of the explanatory variables age and the cellularity level of the patient. Three different forms of the hazard rates are incorporated in the hazard function and are tried out as response.

The fitted model based on three different hazard function revealed, that estimates of regression coefficients and their relevant standard error are quit close, when h2(t) and h3(t) are used as hazard rate. Both reflect positive but insignificant effect for the two explanatory variables, while h1(t) is giving misleading results.

A CASE STATISTICAL STUDY FOR RISK FACTORS RELATIONSHIP TO EPILEPSY

DIMITRIOS PANAGIOTIS NIKOLAKIS* and JOHN E. GOULIONIS
University of Piraeus, Greece

Epilepsy is a chronic neurological condition and the world⁄s common serious brain disorder.

Objectives: The goal of this case-control study is to identify the significance of certain risk factors for epilepsy. The risk factors examined are age, head trauma, central nervous system

infections e.g. meningitis, abnormal perinatal history, family history, socioeconomic status and parental consanguinity.

Methodology: We designed a case-control study for 700 patients (512 males and 188 females) attending the outpatient neurology clinic of Athens Teaching Hospital during a two years (2007-2009) period. In this study the patients were examined and their personal and medical data were taken. For each patient, the phenomenon of epilepsy was studied in relation to different risk factors. The odd ratios have been computed to see the significant factors associated with epilepsy.

Conclusions: Positive family history for epilepsy, head trauma, sex, febrile convulsions and abnormal perinatal history, low education, were shown to have a statistically significant association with epilepsy in patients attending the Athens Hospital. Although consanguinity is widely practiced in Athens, it appears that it does not increase the risk of epilepsy probably due to the small contribution of monogenetic recessive epilepsies to the population with epilepsy.

AMS 62K15, 62K99

NOVEL APPROACHES OF DATA-MINING IN EXPERIMENTAL PHYSICS

GENNADY ALEXEEVICH OSOSKOV
Joint Institute for Nuclear Research, Dubna, Russia

Contemporary experiment in high energy and nuclear physics are characterized by very high rate of produced data and their extremely sophisticated structure, while the wanted physical effects are hidden in huge background exceeding useful information by many orders of magnitude. Therefore data mining for processing experimental data led to many multiparametric problems, two of them are considered: (i) hypothesis testing and classification approaches based on artificial neural networks and boosted decision trees (ii) clustering of large amounts of data by so-called growing neural gas. Some examples from the practice of the Joint Institute for Nuclear research are given to show how to prepare data to deal with those approaches.

THE DISTRIBUTION OF FRACTIONAL DOMINATION NUMBER OF A RANDOM DIGRAPH FAMILY BASED ON ONE-DIMENSIONAL UNIFORM DATA

ENES ÖZEL* and ELVAN CEYHAN
Koç University, Istanbul, Turkey

In this work, we generalize the concept of minimum dominating sets and domination number such that, for a value of $p \in [0, 1]$, a minimum fractional dominating set has the minimum number of vertices that dominate at least $(100 \times p)\%$ of the vertices. The cardinality of a minimum fractional dominating set is called the fractional domination number. We illustrate the fractional domination number on a particular family of interval catch digraphs, called proximity catch digraphs (PCDs) based on one-dimensional uniform data, where the PCDs have an expansion parameter. We derive the exact and asymptotic distribution of the fractional domination number of this digraph family for the entire ranges of expansion and domination parameters.

References

[1] PRIEBE, C. E., DEVINNEY, J. G. and MARCHETTE, D. J. (2001). On the distribution of the domination number for random class cover catch digraphs. *Statistics & Probability Letters* **55** 239–246.

[2] CEYHAN, E. (2008). The distribution of the domination number of class cover catch digraphs for nonuniform one-dimensional data. *Discrete Mathematics* **308** 5376–5393.

[3] HAYNES, T. W., HEDETNIEMI, S. T. and SLATER, P. J. (1998). Fundamentals of Domination in Graphs, Marcel Dekker, New York.

COMPARATIVE EXPRESSIONS FOR DESIGN IN MODELS WITH CONSTRAINTS

ANDREJ PÁZMAN
Comenius University, Bratislava, Slovakia

We summarize and compare briefly expressions (partly known), which are needed for (local) optimal experimental design in regression models with and without constraints. Compared are: conditions for estimability, the expressions for the variance matrix of estimators, the formulas for the D-, A-, E-, c-optimality criteria and for their gradients, the expressions needed in the Elfving theorem and in the equivalence theorems and the role of the information matrix. We indicate the common geometric background of these results.

ON FAVORABLE INFERENCE FOR HEAVY TAILED DATA

RASTISLAV POTOCKÝ
Comenius University, Bratislava, Slovakia
and
MILAN STEHLÍK*
Johannes Kepler University, Linz, Austria

If the model F; underlying the data, is a regularly varying function with index $-1/\alpha$, $\alpha > 0$ it is usually supposed that the top scaled order statistics are Pareto distributed. Hill (1975) derived a procedure of Pareto tail estimation by the MLE. Later on, many authors tried to robustify the Hill estimator, but they still rely on maximum likelihood (e.g. Fraga Alves, 2001, Gomes and Oliveira, 2003, and Li et al., 2010). However, the influence function of Hill estimator is slowly increasing, but unbounded. Hill procedure is thus no robust and many authors tried to make the original Hill robust (see Beran and Schell, 2010, Vandewalle et al., 2007).

During the talk we discuss the recent developments on estimation and testing under presence of heavy tailed data (see Stehlík et al., 2010, Potocký et al., 2011). In Stehlík et al. (2011)

we introduced the general form of the robust Jarque-Bera (JB) test which is optimal for testing against the individual or contaminated Pareto alternative. As a reference for such a contamination we consider different Pareto distributions. We concentrate mainly on simulation results for moderate and small samples. However, we also prove consistency and asymptotic distribution for introduced tests. We show that we may take a t-Hill estimator introduced in the paper as the suitable measure of nominal level of Pareto tail parameter. To guarantee the consistency of the whole procedure, we also prove the consistency of t-Hill estimator. The introduced general class of robust tests of the normality is illustrated on the selected datasets of financial time series.

References

[1] BERAN, J., SCHELL, D. (2010). On robust tail index estimation. *Computational Statistics and Data Analysis,* doi:10.1016/j.csda.2010.05.028.
[2] GOMES, M. I. and OLIVEIRA, O. (2003). Censoring estimators of a positive tail index. *Statistics & Probability Letters* **3** 147–159.
[3] FRAGA ALVES, M. I. (2001). A location invariant Hill-type estimator. *Extremes* **4** 199–217.
[4] HILL, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3** 1163–1174.
[5] LI, J., PENG, Z. and NADARAJAH, S. (2010). Asymptotic normality of location invariant heavy tail index estimator. *Extremes* **13** 269–290.
[6] STEHLÍK, M., POTOCKÝ, R., WALDL, H. and FABIÁN, Z. (2010). On the favourable estimation of fitting heavy tailed data *Computational Statistics* **25** 485–503.
[7] POTOCKÝ, R., STEHLÍK, M. and WALDL H. (2011). Catastrophic claims: new challenge for insurance companies, Technical report.
[8] STEHLÍK, M., FABIÁN, Z. and STŘELEC, L. (2010). *Journal of Statistical Planning and Inference,* to appear.
[9] VANDEWALLE, B., BEIRLANT, J., CHRISTMANN, A. and HUBERT, M. (2007). A robust estimator for the tail index of Pareto-type distributions. *Computational Statistics & Data Analysis* **51** 6252–6268.

ON ESTIMATIONS IN A MULTIVARIATE RCA(1) MODEL

ZUZANA PRÁŠKOVÁ
Charles University, Prague, Czech Republic

Random coefficient autoregressive models belong to a broad class of conditional heteroscedastic time series models because of their varying conditional variance and as such may be used in various applications.

We say that a process of random vectors $X_t \in R^m, t \in Z$, follows the multivariate first-order random coefficient autoregressive model (RCA(1)), if X_t for each $t \in Z$ satisfies

$$X_t = (\beta + B_t)X_{t-1} + Y_t,$$

where β is an $m \times m$ matrix of (unknown) parameters, $\{B_t, t \in Z\}$ is a sequence of $m \times m$ random matrices and $\{Y_t, t \in Z\}$ is an $m \times 1$ random error process. Matrices of the second moments of the random elements B_t and Y_t represent other parameters of the model.

In contrast to the univariate RCA models, where the quasi-maximum likelihood estimators of the parameters are commonly studied in the literature, multivariate models are rarely investigated, and due to large number of parameters and computational difficulties, least-squares estimators are considered, only.

Here we will discuss a class of weighted estimators that preserve the computational simplicity of the least-squares estimators but seems to be more stable, and study their asymptotic properties under milder moment assumptions that need the least-squares estimators. This advantage can be utilized in a construction of on-line monitoring schemes for the stability of the parameters of RCA models.

AN APPROACH TO CONSTRUCTION OF RESPONSE SURFACE DESIGNS WITH QUALITATIVE AND QUANTITATIVE FACTORS

PETER Z. G. QIAN
University of Wisconsin-Madison, USA
and
YUFEN HUANG*
National Chung Cheng University, Chiayi, Taiwan

Construction of response surface designs with qualitative and quantitative factors was posed in Cox (1984) as one of eleven open problems in design of experiments. Unfortunately, this problem has not been fully resolved to date, especially for situations with two or more qualitative factors. We propose an approach to tackling this long-standing problem by exploiting a slicing structure of orthogonal arrays with four levels. This approach constructs a new type of design, called a sliceable response surface design, to accommodate both qualitative and quantitative factors. Such a design has two desirable properties. First, when collapsed over the qualitative factors, it is a second-order design with good estimation efficiency for the quantitative factors. Second, it can be sliced into sub-designs, corresponding to different qualitative levels, each of which is guaranteed to be a first-order design. This method is easy to implement, entails no substantial computing and can accommodate a number of qualitative factors. Examples for illustrating the effectiveness of the proposed method are provided.

ON TESTING PARALLELISM OF REGRESSION LINES

FRANTIŠEK RUBLÍK
Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia

Suppose that $k > 1$ be a fixed integer. Let

$$Y_{ij} = a_i + b_i X_{ij} + \varepsilon_{ij}, \quad i = 1, \ldots, k, \;\; j = 1, \ldots, n_i,$$

where i denotes the index of regression line, X_{ij} are known constants, a_i, b_i are unknown regression constants, $\{\varepsilon_{ij};\ j = 1, \ldots, n_i\}$ is random sample from the distribution with unknown continuous distribution function F_i and these random samples are independent. A multiple comparison rule for testing the hypothesis $b_1 = \ldots = b_k$ is presented and its asymptotic null distribution is derived under the Sen regularity conditions, employed in the paper of Sen in AMS 1969.

THE COMPARISON OF ESTIMATORS OF VARIANCE PARAMETERS IN THE GROWTH CURVE MODEL

RASTISLAV RUSNAČKO
Pavol Jozef Šafárik University, Košice, Slovakia

The standard Growth Curve Model was introduced by Potthoff and Roy in 1964. This model is also known as the generalized multivariate analysis of variance model or Potthoff and Roy model and it has provided a useful statistical model for various areas of study including economics, psychology, medicine or biology. The estimation of unknown parameters in this model and compare their statistical properties has long been an interesting topic of research among statisticians. We will compare estimators of variance parameters in this model.

FRÉCHET CLASSES OF COPULAS

MARTIN RYBÁR
Pavol Jozef Šafárik University, Košice, Slovakia

Given fixed margins of multivariate distribution function, one might be interested in the class of such functions with prescribed margins, the so called Fréchet class. However such classes are not necessarily nonempty. Some conditions for nonemptiness have already been derived. Approach from copula theory is being used and some special Fréchet classes are studied.

NETWORK-RELATED PROBLEMS IN OPTIMAL EXPERIMENTAL DESIGN AND SECOND ORDER CONE PROGRAMMING

GUILLAUME SAGNOL
Zuse Institut Berlin, Germany

In the past few years several applications of optimal experimental designs have emerged to optimize the measurements of certain quantities in communication networks. The optimal design

problems arising from this kind of applications share three interesting properties: (i) measurements are only available at a small number of locations of the network; (ii) each monitor can simultaneously measure several quantities, which can be modelled by "multiresponse experiments"; (iii) the observation matrices depend on the topology of the network. In this talk, I will give an overview of these experimental design problems and present recent results for the computation of optimal designs by second order cone programming (SOCP). I will also present some new results which have emerged from an application to the optimization of toll enforcement on German highways: there we formulate the problem of optimally distributing the experimental effort over space and time for the estimation of the mean of a multivariate AR(1) process, and show a simple case in which the A- and the E-optimal design problems can be solved by a tractable second-order cone program. Some experimental results will also be presented.

ASYMPTOTICS OF THE SPATIAL MEDIAN FOR CLUSTERED DATA

JÁN SOMORČÍK

Comenius University, Bratislava, Slovakia

Recently, some papers dealing with multivariate clustered data have been published. Under multivariate clustered data we understand any collection of identically distributed random vectors grouped in clusters: any two vectors from different clusters are independent, while two vectors from the same cluster do not have to be.

In this setting, asymptotic properties of the well-known spatial median as estimator of location have been studied by some authors under certain assumptions. Our talk would like to provide some weakenings of these assumptions. Also the methodology of our proofs could be interesting since it has not been used in the clustered data setting till now.

BEST LINEAR UNBIASED PREDICTORS IN SPECTRALLY ORTHOGONAL FDSLRMS

FRANTIŠEK ŠTULAJTER and RADOSLAV HARMAN*

Comenius University, Bratislava, Slovakia

We consider the problem of computing the mean square error and its limit of the best linear unbiased predictor in finite discrete spectrum linear regression models. This is done under the assumption that spectral vectors in the models for finite observation of time series are orthogonal, but are not orthogonal to the mean value vectors.

DESIGN OF EXPERIMENT FOR SIGMOIDAL FUNCTIONS: MEASUREMENT OF HYSTERESIS LOOPS

MICHAELA TUČKOVÁ*

Palacký University Olomouc, Czech Republic

RADOSLAV HARMAN

Comenius University, Bratislava, Slovakia

and

JIŘÍ TUČEK, PAVEL TUČEK and LUBOMÍR KUBÁČEK

Palacký University Olomouc, Czech Republic

The aim of this article is a specific application of experimental design theory to the measurement of magnetization in nanomaterial research. The output of the measuring procedure is a hysteresis loop which can be approximated by sigmoidal functions with unknown parameters. The measurement of magnetization is very expensive and time consuming, making the design of experiments an important part of the process. We present the optimal experimental designs for sigmoidal functions corresponding to two physical models—Langevin and Brillouin.

TESTING HYPOTHESES IN LINEAR MODELS WITH TWO VARIANCE-COVARIANCE COMPONENTS

JÚLIA VOLAUFOVÁ

Louisiana State University Health Sciences Center, New Orleans, USA

In linear models with two variance-covariance components, there is an extensive literature on testing linear hypothesis about the variance component, especially testing against zero. Here we shall investigate exact tests of the simultaneous hypothesis about the mean and variance, a special case of which is the hypothesis about the variance only. Statistical properties such as accuracy of p-value and power will be investigated.

Acknowledgements

Joint work with Lynn R. LaMotte and Viktor Witkovský.

PARTIAL-SUMS DISTRIBUTIONS FROM A NEW POINT OF VIEW

GEJZA WIMMER*

Masaryk University Brno, Czech Republic

Mathematical Institute, Slovak Academy of Sciences, Bratislava, Slovakia

Matej Bel University, Banská Bystrica, Slovakia

and

JÁN MAČUTEK

Comenius University, Bratislava, Slovakia

A partial-sums discrete probability distribution $\{P_x\}_{x\in\mathbb{N}_0}$ is given by an equation

$$P_x = \sum_{j\geq x} f(x,j)P_j^*, \qquad x \in \mathbb{N}_0, \tag{1}$$

where $\{P_j^*\}_{j\in\mathbb{N}_0}$ is a probability mass function of the parent distribution and $f(x,j)$ is an appropriate function (such that $\{P_x\}_{x\in\mathbb{N}_0}$ sums to 1).

Distributions arising from the equation (1) were used as models e.g. in linguistics, musicology (Wimmer and Altmann, 2001b) and economy (Willmot, 1986). Theoretical properties of partial-sums distributions were investigated e.g. in Wimmer and Altmann (2001a) and Mačutek (2003).

Our contribution follows a twofold aim:

1. So far, the partial summations (1) were seen either as a tool for creating new discrete distributions (for a given parent distribution $\{P_j^*\}_{j\in\mathbb{N}_0}$ some functions $f(x,j)$ were chosen and the resulting partial-sums distributions $\{P_x\}_{x\in\mathbb{N}_0}$ was obtained), or as a link among distributions or distribution families (see Wimmer and Altmann (2000)). We demonstrate that every two discrete probability distributions are connected by a partial summation. Due to this fact, some applications of partial-sums distributions to mathematical modelling (especially) in quantitative linguistics must be revised.
2. We show that if in (1) the Poisson distribution is the parent distribution and

$$f(x,j) = Ce^{\theta}\frac{(a_1)^{(x)}\ldots(a_k)^{(x)}}{(b_1)^{(x)}\ldots(b_r)^{(x)}}\binom{j}{x}\left\{\sum_{i=0}^{j-x}(-1)^i\binom{j-x}{i}\frac{(a_1+x)^{(i)}\ldots(a_k+x)^{(i)}}{(b_1+x)^{(i)}\ldots(b_r+x)^{(i)}}\right\},$$

one obtains a very general distribution family which contains both Kemp hypergeometric and Kemp-Dacey hypergeometric families (see Wimmer and Altmann, 1999).

References

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

[2] WILLMOT, G. (1986). Mixed compound Poisson distributions. *ASTIN Bulletin Supplement* **16** 59–79.

[3] WIMMER, G. and ALTMANN, G. (2000). On the generalization of the STER distribution applied to generalized hypergeometric parents. *Acta Universistatis Palackiensis Olomouciensis Facultas Rerum Naturalium Mathematica* **39** 215–247.

[4] WIMMER, G. and ALTMANN, G. (1999). Thesaurus of univariate discrete probability distributions, Stamm-Verlag, Essen

[5] WIMMER, G. and ALTMANN, G. (2001a). A new type of partial-sums distributions. *Statistics & Probability Letters* **52** 359–364.

[6] WIMMER, G. and ALTMANN, G. (2001b). Models of rank-frequency distributions in language and music, Text as a linguistic paradigm: levels, constituents, constructs. *Festschrift in honour of Luděk Hřebíček* (Uhlířová, L., Wimmer, G., Altmann, G., Köhler, R., eds.) 283–294. WV, Trier.

ON SOME EXACT SIMULTANEOUS LIKELIHOOD-BASED CONFIDENCE REGIONS FOR TWO VARIANCE COMPONENTS

VIKTOR WITKOVSKÝ

Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia

The goal of the paper is to present methods to construct the exact simultaneous confidence regions for the variance components and/or their functions in normal linear regression model with two variance components. We briefly present the approach based on inverting the exact (restricted) likelihood ratio test (RLRT), as suggested by Crainiceanu and Ruppert (Statistical Methodology, 2004), as well as the method based on inverting the RLRT in model with canonical parametrization, as suggested by Witkovský (2010).

References

[1] CRAINICEANU, C. M. and RUPPERT, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **66** 165–185.

[2] WITKOVSKÝ, V. (2010). On some exact simultaneous likelihood-based confidence regions for the parameters of normal LRM. (In Slovak). In: *Sborník XIX. letní školy biometriky, Lednice 6.–10. 9. 2010.*

List of Participants

Jorge	ACHCAR	achcar@fmrp.usp.br
Jaromír	ANTOCH	antoch@karlin.mff.cuni.cz
Barbora	ARENDACKÁ	barendacka@gmail.com
Anthony C.	ATKINSON	a.c.atkinson@lse.ac.uk
Jaiwook	BAIK	jbaik@knou.ac.kr
Rosemary A.	BAILEY	r.a.bailey@qmul.ac.uk
Frank	BALL	Frank.Ball@nottingham.ac.uk
Katarína	BARTOŠOVÁ	katarina.cimermanova@gmail.com
Ondřej	BLÁHA	blaha.ondrej1@gmail.com
Barbara	BOGACKA	B.Bogacka@qmul.ac.uk
Michal	ČERNÝ	cernym@vse.cz
Martina	CHVOSTEKOVÁ	chvosta@gmail.com
Hilmar	DRYGAS	hilmar@drygas.de
Lenka	FILOVÁ	filova@fmph.uniba.sk
Eva	FIŠEROVÁ	eva.fiserova@upol.cz
Alessandra	GIOVAGNOLI	alessandra.giovagnoli@unibo.it
John E.	GOULIONIS	goulionis25@gmail.com
Marian	GRENDÁR	umergren@savba.sk
Mei-Hui	GUO	guomh@math.nsysu.edu.tw
Martina	HANČOVÁ	martina.hancova@upjs.sk
Radoslav	HARMAN	harman@fmph.uniba.sk
Zdeněk	HLÁVKA	hlavka@karlin.mff.cuni.cz
Daniel	HLUBINKA	hlubinka@karlin.mff.cuni.cz
Klára	HORNIŠOVÁ	Klara.Hornisova@savba.sk
Karel	HRON	hronk@seznam.cz
Yu-Jung	HUANG	yjhuang@isu.edu.tw
Yufen	HUANG	yfhuang@peony.math.ccu.edu.tw
Jeffrey	HUNTER	jeffrey.hunter@aut.ac.nz
Marie	HUŠKOVÁ	huskova@karlin.mff.cuni.cz
Hiroto	HYAKUTAKE	hyakutak@math.kyushu-u.ac.jp
Katarína	JANKOVÁ	jankova@fmph.uniba.sk
Gabriela	KAKUSHKIN	gkakushkin@vcc.ca
Stanislav	KATINA	katina@fmph.uniba.sk
Stephen	KIRKLAND	stephen.kirkland@nuim.ie
Pavel	KŘÍŽ	pavel-kriz@post.cz
Lubomír	KUBÁČEK	kubacekl@inf.upol.cz
Michal	KULICH	kulich@karlin.mff.cuni.cz
Petr	LACHOUT	lachout@karlin.mff.cuni.cz
Vladimír	LACKO	lackovladimir@gmail.com
Lynn R.	LAMOTTE	llamot@lsuhsc.edu

Sergei	LEONOV	Sergei.2.Leonov@gsk.com
Ján	MAČUTEK	macutek@fmph.uniba.sk
Andrzej	MICHALSKI	andrzej.michalski@up.wroc.pl
Ivan	MIZERA	mizera@stat.ualberta.ca
Eric	MOULINES	eric.moulines@telecom-paristech.fr
Kateřina	MYŠKOVÁ	myskova@mendelu.cz
Dimitrios P.	NIKOLAKIS	dim_pan_nik@hotmail.com
Gennady	OSOSKOV	ososkov@jinr.ru
Enes	ÖZEL	enozel@ku.edu.tr
Andrej	PÁZMAN	pazman@fmph.uniba.sk
Rastislav	POTOCKÝ	potocky@fmph.uniba.sk
Zuzana	PRÁŠKOVÁ	praskova@karlin.mff.cuni.cz
Simo	PUNTANEN	Simo.Puntanen@uta.fi
James	RAMSAY	ramsay@psych.mcgill.ca
Marcus	REIß	mreiss@math.hu-berlin.de
František	RUBLÍK	umerrubl@savba.sk
Rastislav	RUSNAČKO	rastislav.rusnacko@student.upjs.sk
Martin	RYBÁR	martin.rybar@student.upjs.sk
Guillaume	SAGNOL	sagnol@zib.de
Ján	SOMORČÍK	somorcik@fmph.uniba.sk
Milan	STEHLÍK	Milan.Stehlik@jku.at
Danica	STRAPCOVÁ	Danica.Strapcova@fmph.uniba.sk
František	ŠTULAJTER	stulajter@fmph.uniba.sk
Michaela	TUČKOVÁ	tuckova.michaela@email.cz
Júlia	VOLAUFOVÁ	jvolau@lsuhsc.edu
Dietrich	VON ROSEN	Dietrich.von.Rosen@et.slu.se
Gejza	WIMMER	wimmer@mat.savba.sk
Viktor	WITKOVSKÝ	witkovsky@savba.sk
Silvelyn	ZWANZIG	zwanzig@math.uu.se

Index of Authors