International Conference on Trends and Perspectives in Linear Statistical Inference

Book of Abstracts

Tomar - Portugal

LinStat 2010

July 27 – 31, 2010
Tomar, Portugal

Edited by

K. Filipiak
Department of Mathematical and Statistical Methods,
Poznań University of Life Sciences, Poland

and

F. Carvalho
Department of Mathematics, Polytechnic Institute of Tomar, Portugal

Published by

Instituto Politécnico de Tomar

ISBN: 978-972-9473-50-0

Printed by

Instituto Nacional de Estatística

Contents

Part V. Special Sessions

Part VI. Contributed Talks

Part VII. Posters

Part VIII. Two Birthday Boys

Part IX. List of Participants

Part I

Introduction

The International Conference on Trends and Perspectives in Linear Statistical Inference, LinStat 2010, will be held in Tomar, Portugal, from 27th by 31st of July 2010. Tomar was created and building began on the castle on the 1st of March in 1160 by order of the first Grand Master of the all-powerful Order of the Knights Templar in Portugal, Gualdim Pais.

The aim of the conference is to bring together researchers sharing an interest in a variety of aspects of statistics and its applications and offer them a possibility to discuss current developments in these subjects. The format of this meeting will involve plenary talks, special sessions, contributed talks and posters.
The conference will mainly focus on a number of topics: estimation, prediction and testing in linear models, robustness of relevant statistical methods, estimation of variance components appearing in linear models, generalizations to nonlinear models, design and analysis of experiments, including optimality and comparison of linear experiments.
The work of young scientists has a special position in the LinStat 2010 to encourage and promote them. The best poster as well as the best talk will be awarded. Prizes will be given to graduate students or young scientists which have gotten their PhD in 2009 or later. Prize-winning works will be widely publicized and promoted by the conference.

It is expected that many of presented papers will be published, after refereeing, in a Special Issue of each of the journals: *Journal of Statistical Computation and Simulation* and *Communications in Statistics - Theory and Methods*, associated with this Conference. All papers submitted must meet the publication standards of mentioned journals and will be subject to normal refereeing procedure.

Committees

The Scientific Committee for this Conference comprises

- João T. Mexia (chair, New University of Lisbon, Portugal),
- Augustyn Markiewicz (vice-chair, Poznań University of Life Sciences, Poland),
- Simo Puntanen (University of Tampere, Finland),
- Götz Trenkler (Technische Universität Dortmund, Germany),
- Dietrich von Rosen (Swedish University of Agricultural Sciences, Uppsala, Sweden),
- Roman Zmyślony (University of Zielona Góra, Poland).

The Organizing Committee comprises

- Francisco Carvalho (chair, Polytechnic Institute of Tomar, Portugal),
- Katarzyna Filipiak (vice-chair, Poznań University of Life Sciences, Poland),
- Paulo C. Rodrigues (New University of Lisbon, Portugal),
- Ricardo Covas (New University of Lisbon, Portugal),
- Miguel Fonseca (New University of Lisbon, Portugal).

Call for Papers

We are pleased to announce a special issue of *Communications in Statistics – Theory and Methods* and *Journal of Statistical Computation and Simulation* (Taylor & Francis) devoted to LinStat'2010.

Communications in Statistics – Theory and Methods

It will include selected papers strongly correlated to the talks of the conference and with emphasis on advances on linear models and inference.

Coordinator-Editor: N. Balakrishnan

Guest Editors: Dietrich von Rosen and Katarzyna Filipiak

All papers submitted must meet the publication standards of Communications in Statistics (see: http://www.math.mcmaster.ca/bala/comstat/) and will be subject to normal refereeing procedure. The deadline for submission of papers is **the end of November, 2010**.

Papers should be submitted using the web site

http://mc.manuscriptcentral.com/lsta

If the author does not have account, he should create it. The contributors must choose must choose **"Special Issue – Advances on Linear Models and Inference"** as the manuscript type.

Journal of Statistical Computation and Simulation

It will include selected papers strongly correlated to the talks of the conference and with emphasis on computation and simulation.

Coordinator-Editor: S. Ejaz Ahmed

Guest Editors: João T. Mexia and Augustyn Markiewicz

All papers submitted must meet the publication standards of JSCS (see: http://jscs.statjournals.net/) and will be subject to normal refereeing procedure. The deadline for submission of papers is **the end of December, 2010**.

Papers should be sent to any of its guest editors, preferably by email in a PDF or PostScript format:

Augustyn Markiewicz: *amark@up.poznan.pl*

João T. Mexia: *jtmexia@fct.unl.pt*

Part II

Program

Program

Monday, July 26, 2010

17:00 – 19:00 **Registration and reception** (Hotel dos Templários)

Tuesday, July 27, 2010

Session I

9:00 – 10:00 Opening and Keynote Speaker:
T. Mathew: *Comparison of meta-analysis using literature and using individual patient data*

10:00 – 10:30 Coffee Break

Session II – Linear Models part I

10:30 – 11:10 J. P. Masson: *About linear models: A geometric revisitation*
11:10 – 11:45 S. Puntanen: *Comparing the BLUEs under two different linear models*

Session III – Applications part I

10:30 – 10:55 O. Bluder: *Prediction of semiconductor lifetime using bayesian linear models with mixed distributions*
10:55 – 11:20 A. Liski: *WALS estimation and an application to costs of hip fracture treatments*
11:20 – 11:45 A. Deka: *Application of statistical models for prediction of performance of students in school examination*

Session IV – Plant breeding and genetics

10:30 – 10:55 I. Mejza: *Analysing genotype by environment interaction by curvilinear regression*
10:55 – 11:20 P. C. Rodrigues: *Simulation and analysis of realistic GxE using a crop growth model with physiological parameters without GxE*
11:20 – 11:45 D. Pereira: *The environmental indexes in a Joint Regression Analysis and their meaning*

Session V – Discrimination and classification

10:30 – 10:55 N. Çalış: *Self organizing mixture network in mixture discriminant analysis: an experimental study*

10:55 – 11:20 A. S. Matos: *Dynamic PCA structure induced autocorrelation*

11:20 – 11:45 N. Shutoh: *Asymptotic expansion for the distribution of the linear discriminant function with 2-step monotone missing data*

Session VI – Financial Mathematics: Models and Statistical Methods part I

10:30 – 10:55 M. L. Esquível: *On a continuous time stock price model with regime switching, delay and threshold*

10:55 – 11:20 P. P. Mota: *Brownian motion with drift and regimes*

11:20 – 11:45 R. Gaspar: *LIBOR convexity adjustments for the Vasiček and Cox-Ingersoll-Ross models*

11:45 – 12:15 Coffee Break

Session VII

12:15 – 13:00 Invited Speaker:
S. Ejaz Ahmed: *Machine bias versus human bias: generalized linear models*

13:00 – 14:30 Lunch

Session VIII

14:30 – 15:15 Invited Speaker:
S. Gupta: *Estimation in randomized response models*

15:15 – 16:00 Invited Speaker:
C. Braumann: *Animal growth in random environments: estimation and prediction using stochastic differential equation models*

16:00 – 16:30 Coffee Break

Session IX – Linear Models part II

16:30 – 16:55 P. Druilhet: *Estimators uniformly shrinking on subspaces*

16:55 – 17:20 E. Fišerová: *Regression analysis of compositional data via linear model with type-II constraints*

17:20 – 17:45 J. Picek: *Approximations of minimum risk regression estimator*

Session X – Statistical Inference in Mixed and Multivariate Linear Models part I

16:30 – 17:10 J. Wesołowski: *Quadratic forms, Jordan algebras and the Wishart distribution*

17:10 – 17:45 D. von Rosen: *Reduced rank regression and multivariate linear models*

Session XI – Statistical Methods in Bioinformatics part I

16:30 – 16:55 D. Cuadras: *Measures of multivariate association using distances*

16:55 – 17:20 A. Sánchez-Pla: *Multivariate methods for genomic data integration and visualization*

17:20 – 17:45 M. C. Ruiz de Villa: *Searching for differential expression in microarray analysis: comparison of two nonparametric approaches*

Session XII – Financial Mathematics: Models and Statistical Methods part II

16:30 – 16:55 A. F. Prior: *Parameter estimators for a bidimensional Ornstein-Uhlenbeck process with singular diffusion*

16:55 – 17:20 B. E. Erdogan: *Predicting bankruptcy using Support Vector Machines: an application of bank bankruptcy*

17:20 – 17:45 N. Modarresi: *Flexible sampling of semi-selfsimilar Markov processes: covariance and spectrum*

17:45 – **Poster Session – Exposition**

C. Andrade: *Spectral and wavelet analysis of the Atlantic North Circulation: a case study*

L. Blanco: *Estimation of superimposed complex exponentials using covariance matching and sparsity*

N. Çalış: *A new algorithm for initial cluster centers in K-means clustering*

M. Cardoso: *Combining models in discrete discriminant analysis*

B. Ceranka: *Estimation in singular chemical balance weighing design with correlated errors*

A. Deeb: *Approximate upper confidence intervals on a ratio of sums of variances*

D. Facchinetti: *Methods for the recovery of missing data in medical research*

S. Faria: *A Poisson mixture regression model: application to financial data*

D. P. Gomes: *Resampling techniques in the optimal choice of the threshold in extremal index estimation*

M. Graczyk: *A - optimal spring balance weighing design under some condition*

L. Grilo: *A family of near-exact approximations based on truncations of the exact distribution for the generalized Wilks Lambda statistic*
Y. Hirose: *An extended least angle regression for contingency tables*
M. Hyodo: *A model selection criterion for discriminant analysis of several groups when the dimension is larger than the total sample size*
A. J. López: *Forecasting accuracy. New evidences based on the Má-competition*
T. Jurczyk: Problems caused by multicollinearity and outlier presence
K. Katulska: *D-optimal chemical balance weighing designs with* $n \equiv 0(\text{mod}4)$ *and 3 objects*
M. Kozłowska: *Block design with nested rows and columns for research on food acceptability limitation for Tetranychus urticae*
M. Maatallah: *Mendelian randomisation using instrumental variable*
K. Marczyńska: *Unreplicated experiments in early stage breeding programs*
E. Moreira: *Clustering of loglinear models using likelihood ratio tests p-values to find homogeneous regions regarding drought management*
R. Navrátil: *Rank tests of symmetry with measurement errors*
S. Nunes: *Jordan Algebras - "a first bite"*
H. Ogasawara: *Asymptotic expansions in multi-group analysis of moment structures with an application to linearised estimators*
L. N. Pereira: *Small area estimation with a longitudinal area level model under restrictions*
J. Pinto da Costa: *Boosting stumps to determine the genes involved in cell proliferation due to ascorbic acid*
J. C. M. Pires: *Comparison of several linear statistical models to predict tropospheric ozone concentrations*
P. C. Rodriuges: *Orthogonal families for one and two tier prime basis factorials*
J. Serra: *Asymptotically optimal linear bias corrections in minimum mean square error estimation*
J. Teles: *Concordance correlation coefficient: an incursion into virtual reality*
A. Tillander: *Effect of data discretization on the classification accuracy in a high-dimensional framework*
E. Vilhena: *An application of Structural Equation Modeling to test social support and physical symptoms as predictors of Quality of Life and Subjective Well-being in persons with chronic disease*
J. Xu: *Performances of the positive-rule Stein-type r-k class estimator in linear regression*

Wednesday, July 28, 2010

Session XIII

9:00 – 10:00 Keynote Speaker:
S. Datta: *Inference for accelerated failure time models for clustered time to event data*

10:00 – 10:30 Coffee Break

Session XIV – Linear Models part III

10:30 – 10:55 Y. Tian: *Some comments on estimations under a restricted linear model and its implicitly restricted linear model*
10:55 – 11:20 B. Schaffrin: *On reproducing linear estimators within the GM-model with stochastic constraints*
11:20 – 11:45 Ö. Çidar: *Estimating the principal component scores through maximum likelihood estimation under normality assumption*

Session XV – Statistical Inference in Mixed and Multivariate Linear Models part II

10:30 – 10:55 T. Yamada: *A high dimensional MANOVA test with fewer observations than the dimension*
10:55 – 11:20 T. Pavlenko: *Sparse inverse covariance estimation in the supervised classification of high-dimensional data*
11:20 – 11:45 Y. Li: *On variance estimators in PLS*

Session XVI – Robust Analysis of Linear Models part I

10:30 – 10:55 C. Amado: *Robust methods for one and two-way ANOVA*
10:55 – 11:20 J. Hinde: *Mixture extensions of linear models*

Session XVII – Nonparametrical Methods part I

10:30 – 10:55 S. Datta: *Rank aggregation and its use in bioinformatics problems*
10:55 – 11:20 J. Nevalainen: *Informative cluster size problems*
11:20 – 11:45 E. Liski: *Supervised invariant coordinate selection*

Session XVIII – Financial Mathematics: Models and Statistical Methods part III

10:30 – 10:55 J. Beleza Sousa: *Noise in machine learning Vasiček interest rate model calibration with Gaussian processes*
10:55 – 11:20 R. Carrujo: *On the evolution of a Markov open long term care population*
11:20 – 11:45 P. Vieira: *Estimation of the maximum displacement response in structures with linear behaviour*

11:45 – 12:15 Coffee Break

Session XIX

12:15 – 13:00 Invited Speaker:
Roman Zmyślony: *Likelihood ratio test with inequality constrains on parameter space*

13:00 – 14:30 Lunch

Session XX

14:30 – 15:15 Invited Speaker:
L. R. LaMotte: *On admissibility when the sample space is finite*
15:15 – 16:00 Invited Speaker:
J. Volaufová: *Model choice and testing in multivariate longitudinal models*

16:00 – 16:30 Coffee Break

Session XXI – Linear Models part IV

16:30 – 16:55 K. U. Akay: *Graphical methods for evaluating some biased estimators in mixture experiments*
16:55 – 17:20 N. Acar: *The identification of outliers in generalized linear models*
17:20 – 17:45 A. Erkoc: *A simulation study for alternative estimation technique in nonlinear models with multicollinear data*

Session XXII – Statistical Inference in Mixed and Multivariate Linear Models part III

16:30 – 16:55 C. Hao: *Influential observation in mixed linear model of repeated measures cross-over designs*
16:55 – 17:20 W. Polasek: *Sensitivity analysis of SAR estimators: a simulation study*
17:20 – 17:45 M. P. Jones: *Single-sample predictive model stability assessment via variance components estimated through resampling and cross validation*

Session XXIII – Statistical Methods in Bioinformatics part II

16:30 – 16:55 R. Oliveira: *Regression methods for multiple outcomes*
16:55 – 17:20 A. Mateus: *STATIS method applied to study diameter growth of eucalyptus stands*
17:20 – 17:45 S. Jin: *On the integer-valued mixture GARCH model*

Session XXIV – Applications part II

16:30 – 16:55 M. Babanezhad: *The effects of body mass index on pregnancy outcomes in the rural areas of north of Iran*
16:55 – 17:20 S. Panwar: *Growth rates of rice through non-linear models*

17:45 – 19:00 **Poster Session – Discussion**

19:00 – Barbecue

Thursday, July 29, 2010

Session XXV

9:00 – 10:00 Keynote Speaker:
M. S. Srivastava: *Selection of variables in multivariate regression models for large dimensions*

10:00 – 10:30 Coffee Break

Session XXVI – Linear Models part V

10:30 – 10:55 N. N. Urgan: *Logistic regression estimators comparison using Pitman's Measure of Closeness*
10:55 – 11:20 D. İnan: *A new estimator for Cox proportional hazard regression model in presence of collinearity*
11:20 – 11:45 Ö. Korucu: *An alternative approach on model selection in Generalized Linear Models*

Session XXVII – Statistical Inference in Mixed and Multivariate Linear Models part IV

10:30 – 10:55 A. Roy: *Linear models with doubly exchangeable distributed errors*
10:55 – 11:20 B. Arendacká: *Approximate interval for the between-group variance under heteroscedasticity*
11:20 – 11:45 M. Fonseca: *Extension to the models with orthogonal block structure*

Session XXVIII – Robust Analysis of Linear Models part II

10:30 – 10:55 M. Souto de Miranda: *Robust estimation of a linear Simultaneous Equations Model using GMM with limited and full information*
10:55 – 11:20 V. Leiva: *Robust statistical modeling of the failure rate using the Birnbaum-Saunderst distribution*

Session XXIX – Applications part III

10:30 – 10:55 G. Cai: *Detection and evaluation method of track bed sedimentation*
10:55 – 11:20 Y. Liang: *Safety state evaluation of urban rail vehicle in transit based on fault diagnosis and early warning*
11:20 – 11:45 X. Li: *Urban rail transit key equipment fault diagnosis method based on statistical analysis*

Session XXX – Matrix Methods

10:30 – 10:55 J. Graffelman: *Diagnostic biplots for linear models*
10:55 – 11:20 I. Kurzydło: *The Fermat's equation on the sets of matrices and the homographic functions*

11:45 – 12:15 Coffee Break

Session XXXI

12:15 – 13:00 YSA Invited Speaker:
K. Nordhausen: *Multivariate linear L_1 regression for cluster-correlated data*

13:00 – 14:30 Lunch

14:30 – Excursion

Friday, July 30, 2010

Session XXXII

9:00 – 10:00 YSA Invited Speaker:
W. Rejchel: *On rank regression, minimization of U-processes and some probabilistic inequalities*

10:00 – 10:30 Coffee Break

Session XXXIII – Statistical Inference in Mixed and Multivariate Linear Models part V

10:30 – 10:55 C. Fernandes: *Stair nesting designs*
10:55 – 11:20 M. Ohlson: *More on the Kronecker structured covariance matrix*
11:20 – 11:45 D. Klein: *On UMRU estimators in the extended growth curve model*

Session XXXIV – Financial Mathematics: Models and Statistical Methods part III

10:30 – 10:55 L. Ramos: *Sample partitioning estimation for ergodic diffusions. Applications*

10:55 – 11:20 A. Wan: *Focused information criteria, model selection and model averaging in a Tobit Model with a non-zero threshold*

11:20 – 11:45 J. Faias: *Optimal option portfolio strategies*

Session XXXV – Testing part I

10:30 – 10:55 R. Enomoto: *Multivariate normality test using Srivastava's skewness and kurtosis*

10:55 – 11:20 S. Oliveira: *Exact and near-exact distributions for likelihood ratio test statistics used to test for stationarity and circularity in multivariate models*

11.20 – 11.45 S. Takahashi: *Testing independence by step-down multiple comparison procedure*

Session XXXVI – Optimal Designs for Nonlinear Models part I

10:30 – 11:10 V. Melas: *On the methodology of optimal design for nonlinear models based on the functional approach*

11:10 – 11:35 C. Tommasi: *An adaptive sequential design for model discrimination and parameter estimation in non-linear nested models*

Session XXXVII – Experimental Designs part I

10:30 – 11:10 R. A. Bailey: *Circular designs balanced for neighbours at distances one and two*

11:10 – 11:45 P. Druilhet: *Optimal approximate repeated measurement designs and efficient exact designs*

11:45 – 12:15 Coffee Break

Session XXXVIII

12:15 – 13:00 Invited Speaker:
A. Atkinson: *Experiments for enzyme kinetic models*

13:00 – 14:30 Lunch

Session XXXIX

14:30 – 15:15 Invited Speaker:
B. Bogacka: *Adaptive experimental design in early clinical trials*

15:15 – 16:00 Invited Speaker:
S. Mejza:*Incomplete split-block designs: perspectives and challenges*

16:00 – 16:30 Coffee Break

Session XL – Statistical Inference in Mixed and Multivariate Linear Models part VI

16:30 – 16:55 Y. G. Berger: *A multivariate regression approach for estimating variances of measures of population change over time from rotating repeated surveys*
16:55 – 17:20 S. Ferreira: *Genealogical trees for segregated COBS*
17:20 – 17:45 F. Carvalho: *Coverings and light designs*

Session XLI – Distribution Theory

16:30 – 16:55 J. G. Dias: *Modeling DHS data using dynamic mixture models*
16:55 – 17:20 C. Nunes: *Control of the truncation errors for generalized F distributions*
17:20 – 17:45 K. Sultan: *Moments of generalized order statistics from some distributions*

Session XLII – Extreme values

16:30 – 16:55 M. de Carvalho: *Modelling k-sample multivariate extremes with application to extreme temperature analysis*
16:55 – 17:20 L. Henriques-Rodrigues: *Peaks over random threshold best linear unbiased estimation of the extreme value index*
17:20 – 17:45 J. Dienstibier: *Extremes of two-step regression quantiles*

Session XLIII – Experimental Designs part II

16:30 – 17:10 D. K. Ghosh: *An efficient Youden square design against the interchange of treatments*
17:10 – 17:35 F. E. S. Tan: *Robustness of uniform marginal designs for logistic mixed effects linear models with covariates*
17:35 – 18:00 B. Maus: *Optimal design for functional magnetic resonance imaging experiments based on linear models*

20:00 – Conference Dinner

Saturday, July 31, 2010

Session XLIV

9:00 – 10:00 YSA Invited Speaker:
T. Rusch: *Peeking into the black box: recursive partitioning of (generalized) linear models*

10:00 – 10:30 Coffee Break

Session XLV – Statistical Inference in Mixed and Multivariate Linear Models part VII

10:30 – 10:55 T. von Rosen: *On exact tests in unbalanced mixed linear models*
10:55 – 11:20 D. Ferreira: *Extension of maximum likelihood estimation methods to mixed linear models*
11:20 – 11:45 P. Ramos: *Interaction in mixed models*

Session XLVI – Financial Mathematics: Models and Statistical Methods part IV

10:30 – 10:55 J. Gomes: *Will it always be necessary taking into account sample selection?*
10:55 – 11:20 N. Rosati: *Estimation of a quarterly model with annual sample selection*

Session XLVII – Optimal Designs for Nonlinear Models part II

10:30 – 10:55 M. Amo-Salas: *Optimal experimental designs for models with a covariance function depending on the parameters of the model*
10:55 – 11:20 J. López-Fidalgo: *Estimating and designing for mixtures of distributions*

Session XLVIII – Testing part II

10:30 – 10:55 L. Kopylev: *On the asymptotic distribution of likelihood ratio test when parameters lie on the boundary*
10:55 – 11:20 K. Koizumi: *Confidence intervals for linear function of mean vectors in the intraclass correlation model with missing data*
11:20 – 11:45 S. Mirakhmedov: *Asymptotic efficiencies of the Greenwood's goodness-of-fit test*

Session XLIX – Experimental Designs part III

10:30 – 10:55 J. Godolphin: *Estimability and connectivity in m-way designs*
10:55 – 11:20 A. Markiewicz: *Optimality of designs under the interference model*
11:20 – 11:45 K. Filipiak: *Optimality of designs under models with interference dependence structure*

11:45 – 12:15 Coffee Break

Session L

12:15 – 13:00 Invited Speaker:
C. R. Johnson: *The logarithmic method for solving nonlinear problems: some successes*

13:00 – 14:30 Lunch

Session LI – Linear Models part VI

14:30 – 14:55 G. Tabakan: *Performance of the difference-based estimators in partially linear models*
14:55 – 15:20 Ö. G. Alma: *Performances comparison of information criterion for outlier detection in multiple regression models having multicollinearity problems using genetic algorithms*
15:20 – 15:45 J. Shabbir: *On efficient use of estimators for finite population mean*

Session LII – Financial Mathematics: Models and Statistical Methods part V

14:30 – 14:55 M. Smeets: *The application of a two-level model to the Dutch Business Survey*
14:55 – 15:20 N. Kumar: *Estimation of market capitalization and economic growth in India*
15:20 – 15:45 S. Rezakhah: *Spectral representation of multi-dimensional discrete time selfsimilar processes*

Session LIII – Testing part III

14:30 – 14:55 M. do Rosário Ramos: *Testing the significance of coefficients in the linear model. The case of the trend in a AR(1) time series*
14:55 – 15:20 P. Om do Valle: *Testing autoregressive nonnested models estimated by IV*
15:20 – 15:45 F. J. Marques: *Testing circular symmetry of a covariance matrix Ũ the exact and nearexact distributions for the likelihood ratio test statistic*

Session LIV – Nonparametrical methods part II

14:30 – 14:55 M. Amezziane: *Exact inference about kernel density estimators*

14:55 – 15:20 R. Bispo: *Statistical power of goodness of fit tests based on the empirical distribution function for Type I censored data*

15:45 – 16:15 Coffee Break

Session LV

16:15 – 17:00 Invited Speaker:
D. Pestana / J. P. Martins: *Meta-analytical issues in linear models*

17:00 – 17:45 YSA Invited Speaker:
A. Alin: *Penalizing power-divergence tests statistics for testing linear by linear association*

17:45 – Closing

Part III

Keynote Speakers

Inference for accelerated failure time models for clustered time to event data

Somnath Datta

University of Louisville, USA

Abstract

This talk deals with analysis of clustered data where the sample units are not independent. In the first part of the talk, we introduce the general phenomenon of informative or non-ignorable cluster size and demonstrate that applying standard methods such as generalized estimating equations which are not specifically designed for this issue may lead to biased inference. We show by numerous examples, both parametric and nonparametric, how to construct inferential procedures that are immured to informative cluster size. In the second half of the talk we examine the accelerated failure time regression model for survival data in detail. In keeping with the theme of this conference, we show that this issue may arise even with linear models (with transformed response) with clustered data. We also discuss how to handle right censoring in this context.

Comparison of meta-analysis using literature and using individual patient data

Thomas Mathew[1] and Kenneth Nordstrom[2]

[1] University of Maryland, Baltimore County, USA
[2] University of Oulu, Finland

Abstract

The problem of combining information from separate studies is a key consideration when performing a meta-analysis, or planning a multicenter trial. Although there is a considerable journal literature on summary versus individual patient data, recent articles in the medical literature indicate that there is still confusion and uncertainty as to the precision of an analysis based on aggregate data. In this paper we address this issue by considering the estimation of a linear function of the mean, based on linear models for individual patient data. The setup, which allows for the presence of random effects and covariates in the model, is quite general, and includes many of the commonly employed models. The one-way fixed-effects model and the two-way model without interaction and fixed or random study effects are all obtained as special cases. For this general model we derive a condition for the estimator based on summary data to coincide with the one obtained from individual patient data, extending considerably earlier work. The implications of this result for the specific models mentioned above are illustrated in detail, both theoretically and in terms of two real data sets, and the role of balance is highlighted.

Selection of variables in multivariate regression models for large dimensions

Muni S. Srivastava[1] **and Tatsuya Kubokawa**[2]

[1] University of Toronto, Canada
[2] University of Tokyo, Japan

Abstract

The Akaike information criterion, AIC, and Mallows' C_p statistic have been proposed for selecting a smaller number of regressor variables in the multivariate regression models with fully unknown covariance matrix. All these criteria are, however, based on the implicit assumption that the sample size is substantially larger than the dimension of the covariance matrix. To obtain a stable estimator of the covariance matrix, it is required that the dimension of the covariance matrix be much smaller than the sample size. When the dimension is close to the sample size, it is necessary to use ridge type of estimators for the covariance matrix. In this paper, we use a ridge type of estimators for the covariance matrix and obtain the modified AIC and modified C_p statistic under the asymptotic theory that both the sample size and the dimension go to infinity. It is numerically shown that these modified procedures perform very well in the sense of selecting the true model in large dimensional cases.

Keywords

Akaike information criterion, Mallows' C_p, Large dimension, Multivariate linear regression model, Selection of variables.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B.N., Csaki, F. (Eds.) *2nd International Symposium on Information Theory*, (pp. 267–281). Akademia Kiado, Budapest.

Akaike, H. (1974). A new look at the statistical model identification. System identification and time-series analysis. *IEEE Trans. Autom. Contr. 19*, 716–723.

Mallows, C.L. (1973). Some comments on C_p. *Technometrics 15*, 661–676.

Part IV

Invited Speakers

Machine bias versus human bias: generalized linear models

S. Ejaz Ahmed

University of Windsor, Ontario, Canada

Abstract

Penalized and shrinkage regression have been widely used in high-dimensional data analysis. Much of recent work has been done on the study of penalized least square methods in linear models. In this talk, I consider estimation in generalized linear models when there are many potential predictor variables and some of them may not have influence on the response of interest. In the context of two competing models where one model includes all predictors and the other restricts variable coefficients to a candidate linear subspace based on prior knowledge, we investigate the relative performances of absolute penalty estimator (APE), shrinkage in the direction of the subspace, and candidate subspace restricted type estimators. We develop large sample theory for the estimators including derivation of asymptotic bias and mean squared error. The asymptotics and a Monte Carlo simulation study show that the shrinkage estimator overall performs best and in particular performs better than the APE when the dimension of the restricted parameter space is large. The estimation strategies considered in this talk are also applied on a real life data set for illustrative purpose.

Penalizing power-divergence tests statistics for testing linear by linear association

Aylin Alin

Dokuz Eylül University, Izmir, Turkey

Abstract

Two different families of penalized power-divergence test statistics have been defined to find alternatives to the likelihood ratio and Pearson chi-squared statistics for testing linear by linear association in two-way contingency tables with empty cell/cells. The exact size and power properties of these statistics and the ordinary power-divergence test statistics have been studied based on extensive designed simulation study. Different penalized test statistics have been proposed depending on their exact size and power performances.

Keywords

Contingency table, Empty cells, Ordinal categorical variable, Log-linear model, Penalization.

Experiments for enzyme kinetic models

Anthony C. Atkinson

London School of Economics, UK

Abstract

The presence of an inhibitor in enzyme kinetic reactions leads to a number of potential nonlinear generalizations of the Michaelis-Menten model, depending on the particular mechanism of inhibition. The talk will present analytical expressions for the D-optimum designs for the parameters of four extended models, all nonlinear. These analytical expressions are not only important in themselves but greatly facilitate the study of design robustness: the efficiency of a proposed design can readily be established over a wide range of parameter values and the variation of the design with the parameters can be exhibited. A design robust to incorrect assumptions about the parameters can then be chosen in the light of this information.

Designs for subsets of the parameters are of interest when trying to establish the mechanism of inhibition and hence the appropriate model. It does not seem to be possible to find analytical expressions for such D_S-optimum designs. However numerical studies indicate that designs on the support points of the D-optimum designs, although with different weights, are often highly efficient.

In some cases the reduction to a simpler model occurs when two parameters have equal values. Then it is possible to rewrite the models so that parameter equality is equivalent to a new parameter being equal to zero. A D_S-optimum design can then be found for estimating this parameter as precisely as possible, leading to a powerful test of the hypothesis of parameter equality. An alterative is to use a T-optimum design for discriminating between the model in which the parameters are set equal and that in which they are not. Although the two approaches are identical for linear models, they are not so for nonlinear models. The two design approaches will be compared and the extent to which they differ elucidated.

It is intended that several of the theoretical points will be illustrated with experimental examples from the pharmaceutical industry.

Adaptive experimental design in early clinical trials

Barbara Bogacka

Queen Mary, University of London, UK

Abstract

Phase I clinical trials are the first studies to test new drug candidates in humans. The main objectives of these trials are to estimate pharmacokinetic (PK) parameters and to determine an optimal dose for further exploration in Phase II. Although both these experimental issues have attained a lot of attention in the relevant literature, majority of the contributions treat the dose-finding and the PK studies as separate experimental tasks. There are some attempts to incorporate the PK data into the dose escalation clinical trial designs, cf. Piantadosi and Liu (1996), but only in order to increase the quality of a dose-response model.
Estimation of the PK parameters is usually assumed not to depend on the dose of a drug and any clinically reasonable dose is regarded as appropriate for the PK studies. However, under such assumption it may be impossible to achieve high accuracy of the parameter estimation. Therefore, the main concept here is to treat the dose as an additional design factor and to optimize the design with respect to both the dose level selection and PK parameter estimation. This is a complex multi-criteria optimization problem.
In the proposed method we use the ethical approach for dose selection developed in Zhang et al. (2006) and based on a continuation-ratio model as in Fan and Chaloner (2004). A Biologically Optimum Dose level is searched for in an adaptive experiment with simultaneous design optimization for the PK parameter estimation.

Keywords

Biologically optimum dose, Pharmacokinetic parameters, Continuation-ratio model.

References

Fan, S. and Chaloner, K. (2004). Optimal designs and limiting optimal designs for a trinomial response. *J. Statist. Plann. Inference 126*, 347–360.

Piantadosi, S. and Liu, G. (1996). Improved designs for dose escalation studies using pharmacokinetic measurements. *Stat Med. 15*, 1605–1618.

Zhang, W., Seargent, D.J. and Mandrekar, S. (2006). An adaptive dose-finding design incorporating both toxicity and efficacy. *Stat. Med. 25*, 2365–2383.

Animal growth in random environments: estimation and prediction using stochastic differential equation models

Carlos A. Braumann, **Patrícia A. Filipe, Clara Carlos, Nuno M. Brites and Carlos J. Roquete**

Évora University, Portugal

Abstract

The purpose is to study individual animal (or plant) growth in a randomly varying environment. Many deterministic models for the growth of an individual from birth to maturity size S can be written in the form of an ordinary linear differential equation $dY(t) = b(A - Y(t))dt$, where $Y(t) = h(X(t))$, with h a strictly increasing C^1 function, $A = h(S)$ and $X(t)$ the size of the individual at age t. For example, the *Gompertz* model corresponds to $h(x) = \ln x$ and the *Bertalanffy-Richards* model to $h(x) = x^c$. Solving the differential equation, one obtains some growth curve $X(t) = g(t)$ $\left[\text{for example, in the } \textit{Gompertz} \text{ model, } g(t) = (X(0)/S)^{\exp(-bt)}\, S\right]$, which is usually adjusted to the observed values of $X(t)$ by nonlinear regression. This would be appropriate if the deviations between model and observations were due to measurement errors. However, this is totally inappropriate if the measuing instruments are quite precise and the deviations are mainly due to the effect of environmental fluctuations on the growth rate. In that case, the random variations should be introduced on the dynamics of the growth process. We propose the stochastic differential equation models $dY(t) = b\,(A - Y(t))\,dt + \sigma dW(t)$, where σ measures environmental noise intensity and $W(t)$ is a standard *Wiener* process. Properties of the model useful in livestock or forestry optimization are deduced, including studying the time required for an animal to reach a given size. We review the statistical issues of parameter estimation and prediction, both for one trajectory (*i.e.*, one animal) and several trajectories (*i.e.*, several animals) and show an application to cattle data. The generalization to the case where the average asymptotic size S varies from animal to animal according to a lognormal distribution will also be presented.

Keywords

Animal growth, Random environments, Stochastic differential equations, Estimation, Prediction.

Likelihood ratio test with inequality constrains on parameter space

Miguel Fonseca[1], **Bimal K. Sinha**[2] and **Roman Zmyślony**[3,4]

[1] New University of Lisbon, Portgual
[2] University of Maryland, Baltimore County, USA
[3] University of Opole, Poland
[4] University of Zielona Góra, Poland

Abstract

Very often constrains on the parameter space occur as consequence of the nature of the phenomena under analysis. The maximum likelihood procedures for estimation and hypothesis testing should be adapted according to these constrains. In this paper we present some results concerning estimation and hypothesis tests for linear models in this framework. Simulations are presented for the comparison of maximum likelihood estimators an hypothesis tests with and without taking account of constrains.

Estimation in randomized response models

Sat Gupta[1], **Pedro Corte Real**[2], **Javid Shabbir**[3] **and Rita Sousa**[2]

[1] University of North Carolina at Greensboro, USA
[2] New University of Lisbon, Portugal
[3] Quad-I-Azam University, Pakistan

Abstract

Randomized response models are a powerful data collection tool in social sciences. Since the original work of Warner (1965), many types of RRT models have been introduced in the literature. We take a look at several such models and introduce a new family of RRT models. Mean square error of the proposed model is compared with that of the basic RRT model using both a first order approximation and a second order approximation. We also present an extensive simulation study to further reflect on this comparison. It is observed that the MSEŠs based on first and second order approximations can be quite different for small sample size.

The logarithmic method for solving nonlinear problems: some successes

Charles R. Johnson

The College of William and Mary, Williamsburg, Virginia, USA

Abstract

By taking the proper view in exponent space, some difficult-appearing nonlinear problems may be transformed to tractable cone-theoretic problems. Often a transformation between a half-space description and a generator description of a cone is key.
Our purpose is to summarize two nice and rather different examples of this technique. Each leads to rather striking results in its own setting. We suspect that many more examples are possible.
The two examples we discuss are

1. the completion problem for TP2 matrices (all 1-by-1 and 2-by-2 minors are positive - they play a key role in the theory of total positivity), and
2. determinantal inequalities for positive definite matrices.

On admissibility when the sample space is finite

Lynn R. LaMotte

Louisiana State University Health Sciences Center, New Orleans, USA

Abstract

Any random variable Y defined on a sample space of N outcomes can be regarded as a linear function of an N-category indicator variable Z. In that case questions of inference based on Y become questions of *linear* inference based on Z. In this paper, a characterization (LaMotte 1982) of admissibility among linear estimators will be applied to investigate admissibility in general under squared-error loss in this finite-sample-space setting.

Incomplete split-block designs: perspectives and challenges

Stanisław Mejza and **Iwona Mejza**

Poznań University of Life Sciences, Poland

Abstract

Split-block designs are very often used in life science experiments and in engineering. They are specifically suited to two-factor experiments that utilize two kinds of experimental units (row plots and – crossed within them – column plots) within blocks. Levels of one of the factors (called row treatments) are assigned to the rows, and levels of the second factor (called column treatments) are assigned to the columns.

A split-block design in which all levels of a particular factor occur on the relevant kind of units (e.g. rows, columns within blocks) is called a complete (orthogonal) design. Any design that it is not complete is called incomplete (non-orthogonal) design. The paper deals with planned incomplete split-block designs. From the statistical point of view, complete split-block designs are the best. However, the incomplete designs better fit the particular (available) structures of experimental units. But in the case of incomplete designs there are many new open problems. The most important problems, which are discussed in this paper, are:

1. statistical modeling observations;
2. ANOVA;
3. general constructing methods for optimal incomplete split-block designs;
4. constructing methods for optimal incomplete designs with special reference to two kinds of treatments (test treatments and control treatments) in the experiments.

Multivariate linear L_1 regression for cluster-correlated data

Klaus Nordhausen[1], Jaakko Nevalainen[2] and Hannu Oja[1]

[1] University of Tampere, Finland
[2] University of Turku, Finland

Abstract

We consider the multivariate linear regression model $\mathbf{y_i} = \boldsymbol{\beta}'\boldsymbol{x_i} + \boldsymbol{\epsilon_i}$ where y_i is a p-variate response variable, x_i is a q-variate vector of explanatory variables, and $\boldsymbol{\epsilon}_i$ is a p-variate random error, $i = 1, ..., n$. We wish to make inference on the unknown $q \times p$ regression coefficient matrix $\boldsymbol{\beta}$. In the case of independent and identically distributed random errors, the estimates and related tests are commonly based on the use of the L_2 criterion $\sum_{i=1}^{n} ||\boldsymbol{y_i} - \boldsymbol{\beta}'\boldsymbol{x_i}||^2$ where $||\cdot||$ is the Euclidean norm. In an alternative approach, Bai et al. (1990) and Arcones (1998) considered the L_1 objective function $\sum_{i=1}^{n} ||\boldsymbol{y_i} - \boldsymbol{\beta}'\boldsymbol{x_i}||$. This approach then yields spatial median-type estimates and sign-type tests. Oja (2010) provides further references.

The assumption that $\boldsymbol{\epsilon}_1, ..., \boldsymbol{\epsilon}_n$ are independent is not true, however, if the data are clustered. The typical situation then is that instead of sampling independent and identically distributed random variables, the researcher samples observations in clusters. Observations within a cluster tend to be similar (correlated) but the clusters are independent. The clusters may, for example, be clinics with patients, schools with students, litters of rats, and so on. Also repeated measures on the same individual serves as an example.

In this talk we first shortly review the multivariate L_1 regression theory in the case of iid error variables. Then we show how the theory can be extended to the cluster-correlated case. The theory is illustrated with several examples. A small simulation study is conducted to compare the properties of different estimates.

Keywords

Clustered data, L_1 criterion, Repeated measures, Spatial median, Spatial sign.

References

Arcones, M.A. (1998). Asymptotic theory for M-estimators over a convex kernel. *Econometric Theory 14*, 387–422.

Bai, Z.D., Chen, R., Miao, B.Q. and Rao, C.R. (1990). Asymptotic theory of least distances estimate in multivariate linear models. *Statistics 4*, 503–519.
Oja, H. (2010). *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks.* New York: Springer.

Meta-analytical issues in linear models

João Paulo Martins[1,3] and Dinis Pestana[2,3]

[1] Polytechnic Institute of Leiria, Portugal
[2] Lisbon University, Portugal
[3] Center of Statistics and Applications of Lisbon University, Portugal

Abstract

Meta analysis of linear models is an active field of research, and its interaction with optimal design theory should be a main goal in the emerging area of cumulative meta analysis

In classical regression we may use one or more covariates to assess the relationship between those covariates and a dependent variable. In meta-regression a similar approach is used, with a substantial difference: the covariates are at the level of the study. Optimal design theory (Anderson, 1962; Fedorov, 1972; Kiefer and Studden, 1976) deals with the appropriate choice of observations to accomplish the estimation of the coefficients in a regression model in an optimal or quasi-optimal way (Dette and Studden, 1997; Martins et al., 2008).

When the available studies do not provide, in the light of classical meta analysis, enough "statistical evidence", the researcher may conduct a new study to add to his meta analysis, so that conclusive evidence may be reached. In this context, it is of great importance not to choose the covariates levels of this new study haphazardly, or even at random, rather they must be selected so that the extra study efficiently contributes to an enlightening cumulative meta analysis.

We develop a framework to deal with optimal or quasi-optimal choices when planning new studies whose aim is to achieve, or at least to reinforce, convincing evidence.

Optimal discriminant, optimal robust and quasi-optimal mixed designs are used to provide competitive ways of dealing with the issue.

The classical example of meta analysis of studies to assess the impact of the vaccine BCG to prevent the development of tuberculosis is used to emphasize the importance of the choice of the design in cumulative meta analysis.

Keywords

Prospective cumulative meta-analysis, Optimal designs, Discriminant designs, Mixed designs, Quasi-optimal designs.

References

Anderson, T. (1962). The choice of the degree of a polynomial regression as a multiple decision problem. *Ann. Math. Statist. 33*, 255–265.

Dette, H. and Studden, W. (1997). *The Theory of Canonical Moments with Applications in Statistics. Probability and Analysis.* Wiley-Interscience Publication, New York.

Fedorov, V. (1972). *Theory of Optimal Experiments.* Academic Press, New York.

Kiefer, J. and Studden, W. (1976). Optimal designs for large degree polynomial regression. *Ann. Statist. 4*, 1113–1123.

Martins, J.P., Mendonça, S. and Pestana, D.D. (2008). Optimal and quasi-optimal designs. *RevStat 6*, 279–307.

On rank regression, minimization of U-processes and some probabilistic inequalities

Wojciech Rejchel

Nicolaus Copernicus University, Toruń, Poland
University of Warmia and Mazury, Olsztyn, Poland

Abstract

We consider the problem of rank regression, that is basing on some features of the objects we want to predict (guess) the order between these objects. In many algorithms the $0-1$ loss function is replaced by its convex surrogate - this trick allows to find very effective procedures (support vector machines [Vapnik, 1998] or boosting [Freund, 2004]). We are interesting in statistical properties of rank regression estimators. First, we show that in the linear model one can proove the strong consistency and the asymptotical normality of estimates [Niemiro, 2009]. Moreover, we present another very popular approach to the problem, namely, we look for probabilistic inequalities concerning the risk or the excess risk (i.e. comparing to the best one in the class) of estimators [Clemencon, 2008]. Finally, we show how one can obtain significant improvements in such inequalities [Rejchel, 2009]. The methods that we use come from the theory of empirical processes and U-processes.

Keywords

Rank regression, Convex risk, Empirical process, U-process.

References

Clemenon, S., Lugosi, S. and Vayatis, N. (2008). Ranking and empirical minimization of U-statistics. *Ann. Statist. 36*, 844–874.

Freund, Y., Iyer, R., Schapire, R.E. and Singer, Y. (2004). An efficient boosting algorithm for combining preferences. *J. Mach. Learn. Res. 4*, 933–969.

Niemiro, W. and Rejchel, W. (2009). Rank correlation estimators and their limiting distributions. *Statist. Papers 50*, 887–893.

Rejchel, W. (2009). Ranking - convex risk minimization. *Proc. of WASET 56*, 172–178.

Vapnik, V.N. (1998). *Statistical learning theory.* Wiley, New York.

Peeking into the black box: recursive partitioning of (generalized) linear models

Thomas Rusch[1] **and Achim Zeileis**[2]

[1] Vienna University of Economics and Business, Austria
[2] Innsbruck University, Austria

Abstract

Recursive partitioning algorithms separate a feature space into a set of disjoint rectangles. Typically, a constant (e.g., a mean or a proportion) is fitted in every segment of the partition. While this is a simple and intuitive approach, it still lacks interpretability as to how a specific relationship between dependent and independent variables may look. Or it may be that a certain model is assumed or of interest and there is a number of candidate variables that may nonlinearly give rise to different model parameter values. We want to present an approach that offers a solution to the problem of limited interpretability of classical trees as well as providing an explorative way to assess a candidate variable's influence on a parametric model: Model-Based Recursive Partitioning (Zeileis et.al., 2008). This method conducts recursive partitioning of a parameteric model such as the generalized linear model by (1) fitting a parametric model to a data set, (2) testing for parameter instability over a set of partitioning variables, (3) splitting the model with respect to the variable associated with the highest instability. The outcome is a tree where each node is associated with a fitted parametric model. We will describe the procedure and show its versatility and suitability to gain additional insight into the relationship of dependent and independent variables by three examples, the link between professors' beauty and their teaching evaluation, the prediction of voting behaviour and a failure model for debt amortization.

Keywords

Model-based recursive partitioning, Trees, Generalized linear model, Parameter instability, Maximum likelihood.

References

Zeileis, A., Hothorn, T. and Hornik, K. (2008). Model-based recursive partitioning. *J. Comput. Graph. Statist. 17*, 492–514.

Model choice and testing in multivariate longitudinal models

Júlia Volaufová

Louisiana State University Health Sciences Center, New Orleans, USA

Abstract

Longitudinal models are widely used in many biomedical applications. The change of the mean of the response over time is modeled often by a two-stage model, which results in a mixed (linear) model. The dependences of observations over time are modeled by covariance matrix which has an assumed structure. The choice of the structure influences the behavior of test statistics for fixed effects. Here we investigate the influence of model choice criteria and behavior of approximate tests in a multivariate longitudinal setting.

Keywords

Multivariate longitudinal model, Fixed effects, Model choice criteria.

Part V

Special Sessions

Statistical methods in bioinformatics

Carles M. Cuadras

University of Barcelona, Spain

Abstract

This session is devoted to novel contributions and new insights into statistical methodology in Bioinformatics. The advent of the genomic age, where human and other genomes has been deciphered, has raised new challenges for statistics. These come from the fact that huge quantity of data that are being generated by technologies such as microarrays, or the more recent ultra deep sequencing, and from the need to integrate these data in a systems biology approach that yields a better understanding of the underlying biological processes. The appearance of these high troughput data has led to many applications of statistics as well as to development of new methods taylored to the new problems. Jointly with probabilistic modelling, or multiple testing, multivariate statistics is one of the fields where most of this activity has happened, due to the high dimensionality of these type of data.

Financial mathematics: models and statistical models

Manuel L. Esquível

New University of Lisbon, Portugal

Abstract

We welcome papers discussing models for pricing contingent claims in complete and incomplete markets as well as statistical methods useful for estimation and calibration of these models.

Functional approach and nonlinear regression models

Viatcheslav B. Melas

University of St. Petersburg, Russia

Abstract

Session is devoted to problems of constructing optimal designs for nonlinear regression models. One of the major difficulties arises from the dependence of the asymptotic covariance matrix of the parameter estimates on the vector of true, but unknown, values of the parameters. A number of well known statistical approaches are applied to overcome this difficulty Ũ locally optimal, sequential, minimax and Bayesian. Constructing the corresponding optimal designs is a very difficult problem. It is hard to find explicitly even locally optimal designs; this is possible only for the simplest models with one or two unknown parameters. An approach based on the study of support points and weights of optimal designs as implicit functions of some auxiliary parameters has been developed in the last two decades [see Melas (2006)]. One of the talks of the session will be devoted to the development of methodology based on the functional approach for constructing and studying all types of optimal designs for models of exponential and rational form. One more class of models attracting attention is models with a covariance function depending on the parameters of the model. An approach to optimal design for such class will be presented. Another problem is that of constructing an adaptive sequential design for discriminating and estimating more than two nested nonlinear models.

References

Melas, V.B. (2006). *Functional Approach to Optimal Experimental Design*. Heidelberg: Springer.

New ideas in the analysis of mixed linear models

Dietrich von Rosen

Swedish University of Agricultural Sciences, Uppsala, Sweden

Abstract

The purpose of the session is to consider linear models where both fixed and random effects are modeled simultaneously. We will also consider models where restrictions on the fixed effect are given in form of rank restrictions, so called reduced rank regression models. Topics which are of particular interest are residuals, influential observations, and special covariance structures.

Part VI

Contributed Talks

The identification of outliers in generalized linear models

Nihan Acar and **Aydin Erar**

Mimar Sinan Fine Arts University, Turkey

Abstract

Generalized linear models are used in the analysis of exponential distribution families such as Normal, Binomial, Poisson and Gamma distributions where a simple linear relationship need not be found between response variable and explanatory variables. In the analysis of data including observations that differ from main part of the data, identification of outliers is a necessary step to obtain valid results. The outlier identification problem is actually the problem of identifying those observations that lie in an area called the outlier region. There are several ways of identifying outliers. Outlier detection methods go into division depending upon number of outliers and reliability of results. According to number of outliers one step or stepwise methods can be used. To avoid masking and swamping effects robust methods are preferred rather than non-robust methods. In this study, we focus on two types of outlier identification rules and compare obtained results.

Keywords

Generalized linear models, Outliers, Outlier region.

References

Asikgil, B. and Erar, A. (2009). Examination of multiple outliers in generalized linear models. *The Announcement Book of the 6th Statistics Days Symposium*, 16–26.

Dobson, J.A. (1990). *An Introduction to Generalized Linear Models.* Chapman and Hall.

Davies, L. and Gather, U. (1993). The identification of multiple outliers. *J. Amer. Statist. Assoc. 88*, 782–792.

Kuhnt, S. and Pawlitschko J. (2005). Outlier identification rules for generalized linear models. In: Baier, D., Wernecke, D. (Eds.), *Innovations in Classification, Data Science and Information Systems* (pp. 165–172). Heidelberg: Springer.

Nelder, J.A. and Wedderburn W.M. (1972). Generalized linear models. *J. R. Stat. Soc. Ser. A. 134*, 370–384.

Graphical methods for evaluating some biased estimators in mixture experiments

Kadri Ulaş Akay

Istanbul University, Turkey

Abstract

In most practical applications, mixture data are highly collinear. In some cases, Scheffé models and other special mixture models give poor estimates of the coefficients, due to fact that constraints on the mixture components create collinearity. Effects due to collinearity can be reduced to certain extent by using alternative approaches. One of these approaches is to use biased estimators for the estimation of the coefficients. In this paper, we used a graphical method for evaluating the effect of the Liu estimator with respect to the predicted response value and the prediction variance. The obtained results are compared with the Ridge estimator.

Keywords

Collinearity, Liu estimator, Ridge estimator, Response trace, Prediction variance trace.

References

Cornell, J.A. (2002) *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data.* 3rd edition. John Wiley and Sons Inc, USA.

Cornell, J.A. and Gorman, J.W. (2003). Two new mixture models: living with collinearity but removing its influence. *J. Qual. Technol. 35(1)*, 78–88.

Jang, D.H. and Yoon, M. (1997). Graphical methods for evaluating ridge regression estimator in mixture experiments. *Comm. Statist. Simulation Comput. 26(3)*, 1049–1061.

Liu, K. (1993). A new class of biased estimate in linear regression. *Comm. Statist. Theory Methods 22(2)*, 393–402.

St. John, R.C. (1984). Experiments with mixtures, ill-conditioning, and ridge regression. *J. Qual. Technol. 16*, 149–159.

Performances comparison of information criterion for outlier detection in multiple regression models having multicollinearity problems using genetic algorithms

Özlem Gürünlü Alma

Mugla University, Turkey

Abstract

Multiple linear regression models are widely used applied statistical techniques and they are most useful devices for extracting and understanding the essential features of datasets. However, in multiple linear regression models problems arise when a serious outlier observation or multicollinearity present in the data. Multicollinearity is a linear dependency between two or more explanatory variables in the regression models which can seriously disturb the least squares estimated regression surface. The other important problem is outlier; they can strongly influence the estimated model, especially when using least squares method. Nevertheless, outlier data are often the special points of interests in many practical situations. The purpose of this study is to performances comparison of Akaike Information Criterion, Bayesian Information Criterion and Information Complexity criterion for detecting outliers using Genetic Algorithms when multiple regression models having multicollinearity problems.

Keywords

Akaike information criterion, Bayesian information criterion and information complexity criterion, Genetic algorithms, Multicollinearity, Multiple linear regression, Outlier detection.

References

Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley, USA.

Gürünlü Alma, Ö., Kurt, S. and Ugur, A. (2009). Genetic algorithms for outlier detection in multiple regression with different information criteria. *J. Stat. Comput. Simul.*, doi: 10.1080/00949650903136782.

Hoeting, J., Raftery, A.E. and Madigan, D. (1996). A method for simultaneous variable selection and outlier identification in linear regression. *Comput. Statist. Data Anal. 22*, 251–270.

Papadimitriou, S., Kitawaga, H., Gibbons, P.G. and Faloutsos, C. (2002). LOCI: Fast outlier detection using the local correlation integral. *Intel research Laboratory Technical report no. IRP-TR-02-09.*

Tolvi, J. (2004). Genetic algorithms for outlier detection and variable selection in linear regression models. *Soft Computing 8*, 527–533.

Wold, H. (1966). Estimation of Principal Components and Related Models by Iterative Least Squares. In Multivariate Analysis. In Krishnaiah, P.R. (Eds.) *Multivariate Analysis* (pp.391–420). Academic Pres, New York.

Robust methods for one and two-way ANOVA

Conceição Amado, João A. Branco and Ana M. Pires

Technical University of Lisbon, Portugal

Abstract

ANOVA is perhaps one of the most used statistical tools in applied statistics (Gelman, 2005).
All the branches of activity where experimentation is essential, do not dispense the use of this highly flexible technique that allows the users to understand the nature and to measure the magnitude of effects and interactions of a multitude of factors that naturally or artificially may contribute to the response output. Intense use of ANOVA is being recently required to analyze very large sets of data produced by computers and other technologies (data-mining, microarrays, weather prediction, credit card fraud detection, and others).
Appropriate application of the ANOVA relies on various assumptions underlying the model and if these assumptions are violated, which happens in many sets including the large data sets, the results may be misleading and do not serve to describe the reality. Unfortunately this fact is not mentioned or stressed in introductory statistics courses and the nature of the data is many times literally ignored by the user and ANOVA is applied naively, overlooking the structure of the data and the context of the problem that determine the validity of the model.
To control the effects of deviations from the model assumptions robust methods are required (Kulinskaya and Dollinge, 2007). After a short critical review of robust estimation and testing for ANOVA we concentrate on the one and two-way cases and introduce a simple robust approach. The robustness of these new tools is evaluated using Monte Carlo simulations under various distributions and in terms of different criteria. A real data set is examined to illustrate an application of the methods introduced.

Keywords

Robustness, ANOVA, Robust ANOVA.

References

Gelman, A. (2005). Analysis of variance - why is it more important than ever (with discussion). *Ann. Statist. 33*, 1–53.

Kulinskaya, E. and Dollinge, M.B. (2007). Robust weighted one-way ANOVA: Improved approximation and efficiency. *J. Statist. Plann. Inference 137*, 462–472.

Exact inference about kernel density estimators

Mohamed Amezziane

DePaul University, Chicago, Illinois, USA

Abstract

In this paper we investigate the distributional properties of kernel density estimators and discuss the hindering consequences of the normal approximation to the estimator's sampling distribution and the need to obtain an exact distribution. Such distribution is used to conduct hypothesis testing and obtain confidence intervals about this estimator.

Optimal experimental designs for models with a covariance function depending on the parameters of the model

Mariano Amo-Salas[1], Jesús López-Fidalgo[1] and Emilio Porcu[2]

[1] University of Castilla-La Mancha, Spain
[2] University of Göttingen, Germany

Abstract

Spatial design has become an important issue and increasing literature can be found on this subject. In particular, the presence of spatial autocorrelation has been noticed to be a serious element of complication for the design, as can be seen in a recent paper by Kiselak and Stehlik (2008). In this paper, we propose some scenarios induced by the use of covariance functions being represented by completely monotone functions on the real line (Berg and Forst, 1975). Some of them are well known in the geostatistical community as Matérn, Cauchy or Dagum models. In our approach we consider a stochastic process defined on the real line where the trend is represented through the sum of exponential components, whose parameters must be estimated. At the same time, we set the associated covariance function in order that its argument must depend on the same arguments of the trend, which is a very real situation in practice. This setting induces a typical reducibility problem whose solution is in general empty, unless one relaxes the conditions associated to the problem itself. We thus obtain a new class of positive definite functions depending on the trend parameters as well as a complete monotone function acting as generator. We illustrate the result through analytical examples and then compute optimal designs and compare them with previous results of the authors.

Keywords

Spatial design, Positive definite, Completely monotone.

References

Berg, C. and Forst, G. (1975). *Potential Theory on Locally Compact Abelian Groups.* Berlin: Springer.

Kiselak J. and Stehlík, M. (2008). Equidistant and D-optimal designs for parameters of Ornstein-Uhlenbeck process. *Statist. Probab. Lett. 78(12)*, 1388–1396.

STATIS method applied to study diameter growth of eucalyptus stands

Aníbal Areia[1], Ayana Mateus[2] and João T. Mexia[2]

[1] Polytechnic Institute of Setúbal, Portugal
[2] New University of Lisbon, Portugal

Abstract

Eucalyptus globulus Labill. is one of the most important economic forest species in Portugal, occupying an area of $875,000ha$ of a total forest area of $3,346,000ha$. In order to contribute to a balanced and resourceful management of Eucalyptus stand in Portugal, it is necessary to acquire models that simulate their growth under different environment conditions and treatments.

The objective of the research reported here was to analyse if diameter growth varies in a consistent manner throughout time, according to criteria that describes stand dynamics. A perspective that seemed valid to handle this problem was the STATIS methodology that uses Euclidean distances to compare configurations between statistical units observed in k different circumstances, as a measure of similarity between them. This approach was introduced by (Escoufier and L'Hermier, 1978) and developed by many authors (Lavit et al., 1994), (Areia et al., 2008).

In this method a study consists of a data matrix $X_{i'}$ with a line for objects and a column for variables and two diagonal weight matrices D_p^0 and D_n for variables and objects. To each study the following matrix is associated in order to condense the information, $A_{i'} = X_{i'}^T D_{ni'} X_{i'} D_p^0 \quad i' = 1, ..., k$

The comparison of various studies will be done through the algebraic structure associated to matrices $A_{i'}$, of type $p \times p$, $i' = 1, 2, ..., k$. This structure was developed using Hilbert-Schmidt scalar product in the space of square matrices, ie,

$$S_{i'j'} = A_{i'}|A_{j'} = tr\left(A_{i'} A_{j'}^T\right) = \sum_{v=1}^{p} \sum_{u=1}^{p} a_{v,u}\left(i'\right) a_{v,u}\left(j'\right) \quad i', j' = 1, ..., k$$

The matrix $S = [s_{i',j'}]$ with $i', j' = 1, ...,$ will be symmetric, with eigenvalues $\theta_1, ..., \theta_k$ and mutually orthogonal eigenvectors $\gamma_1, ..., \gamma_k$.

In our case the representation points are near the space spanned by γ_1, γ_2 and the following model is developed: $S = \sum_{i=1}^{2} \lambda_i \alpha_i^k \alpha_i^{k^t} + \overline{E}$
with $\alpha_i^{k^t} \alpha_j^k = \delta_{i,j} \quad i, j = 1, ..., r$, where $\overline{E} = \frac{1}{2}(E + E^t)$ and $vec(E) \sim N(0^{k^2}, \sigma^2 I_k^2)$.

Keywords

F tests, Hilbert-Schmidt product, STATIS method.

References

Areia, A., Oliveira, M. and Mexia J.T. (2008). Models for a series of studies based on geometrical representation. *Stat. Methodol. 5*, 277–288.

Escoufier, Y. and L'Hermier, H. (1978). A propos de la comparaison graphique des Matrices de Variance. *Biom. J. 20(5)*, 477–483.

Lavit, C., Escoufier, Y., Robert, S. and Pierre, T., (1994). The ACT (STATIS method). *Comput. Statist. Data Anal. 18*, 97–119.

Approximate interval for the between-group variance under heteroscedasticity*

Barbora Arendacká

Slovak Academy of Sciences, Slovakia

Abstract

We will present an approximate confidence interval for the between-group variance in one-way heteroscedastic random-effects model. Derivation of the interval is inspired by the ideas in Williams (1962) for the homoscedastic case. Simulations suggest that the proposed interval performs better than or comparably to the other available procedures as studied in Wimmer, Witkovský (2003), Hartung, Knapp (2005), Li (2007). Moreover, it is relatively easy to compute.

Keywords

Variance components, Heteroscedasticity, One-way ANOVA, Approximate confidence intervals.

References

Hartung, J. and Knapp, G. (2005). On confidence intervals for the among-group variance in the one-way random effects model with unequal error variances. *J. Statist. Plann. Inference 127*, 157–177.

Li, X. (2007). Comparison of confidence intervals on between group variance in unbalanced heteroscedastic one-way random models. *Comm. Statist. Simulation Comput. 36*, 381Ũ-390.

Williams, J.S. (1962). A confidence intervals for variance components. *Biometrika 49*, 278–281.

Wimmer, G. and Witkovský, V. (2003). Between group variance component interval estimation for the unbalanced heteroscedastic one-way random effects model. *J. Stat. Comput. Simul. 73*, 333–346.

* The work was supported by the Slovak Research and Development Agency under the contract No. LPP-0388-09.

Robust statistical modeling of the failure rate using the Birnbaum-Saunders-t distribution

Emilia Athayde[1], Cecilia Azevedo[1], Narayanaswamy Balakrishnan[2] and Víctor Leiva[3]

[1] University of Minho, Braga, Portugal
[2] McMaster University, Hamilton, Ontario, Canada
[3] University of Valparaíso, Chile

Abstract

The hazard or failure rate function is an important statistical indicator employed in lifetime analysis. The Birnbaum-Saunders (BS) model is a useful life distribution originated from a problem of material fatigue that has been largely applied to reliability and fatigue studies. The BS distribution relates the total time until the failure to some type of cumulative damage normally distributed. The generalized BS (GBS) distribution is a new class of positively skewed models with lighter and heavier tails than the classic BS distribution. One of the most important property of the GBS model is the robust estimation of its parameters, particularly when the BS-t distribution is used in the modeling. In this paper, we propose robust statistical modeling of the hazard rate by the BS-t distribution and estimate this rate and its change point using likelihood methods and the EM algorithm. Specifically, the aims of this work are (i) to produce a mathematical study of the shape of the BS-t hazard rate; (ii) to develop inference for this rate and evaluate its performance using Monte Carlo methods; (iii) to show the robustness of such a procedure; and (iv) to illustrate the obtained results by real data.

Keywords

Change point, EM algorithm, Generalized Birnbaum-Saunders distribution, Hazard rate, Lifetime data, Likelihood methods, Robustness, R language.

References

Balakrishnan, N., Leiva,V., Sanhueza, A. and Vilca, F. (2009). Estimation in the Birnbaum-Saunders distribution based on scale-mixture of normals and the EM-algorithm. *Stat. Oper. Res. Trans. 33*, 171–192.

Birnbaum, Z.W. and Saunders, S.C. (1969). A new family of life distributions. *J. Appl. Probab. 6*, 319–327.

Díaz-García, J.A. and Leiva, V. (2005). A new family of life distributions based on elliptically contoured distributions. *J. Stat. Plann. Inference 128*, 445–457

Kundu, D., Kannan, N. and Balakrishnan, N. (2008). On the hazard function of Birnbaum-Saunders distribution and associated inference. *Comput. Statist. Data Anal. 52*, 2692–2702.

The effects of body mass index on pregnancy outcomes in the rural areas of north of Iran

Manoochehr Babanezhad[1] **and Karam Nazari**[2]

[1] Golestan University, Gorgan, Golestan, Iran
[2] Sanitation Center, Amol, Mazandaran, Iran

Abstract

The effect of maternal body mass index (BMI) on the risk of maternal and neonatal outcomes were examined on a singleton term pregnancies of rural areas of Amol in North of Iran. In this study maternal height, maternal age, maternal weight in early pregnancy or in first trimester (or in first visit) were measured. The women were classified in 3 classes of BMI. We calculated the adjusted odds ratio to estimate the risk for the maternal outcomes such as, post-term delivery, pre-term delivery, rate of unplanned emergency caesarean, 3 rd-or 4th-degree lacerations, postpartum infection, transfusion requirement, and more than 3 days hospitalization. Neonatal outcomes were considered intrauterine growth retardation, birth weight, neonatal morbidity, neonatal death, stillbirth, and low Apgar score (< 7 at 5 min). We preformed recent method of instrumental variable (IV) estimator to obtain consistent estimate of the effect of BMI on considered outcomes in the presence of unmeasured confounder factors.

Keywords

Body mass index, Instrumental variable, Maternal outcomes, Neonatal outcomes, Overweight, Obesity.

References

Cedergren, M.I. (2004). Maternal morbid obesity and the risk of adverse pregnancy outcome. *Obstetrical and Gynecological Survey 59*, 489–491.

Kabiru, K. and Raynor, B.D. (2004). Obstetric outcomes associated with increase in BMI category during pregnancy. *Amer. J. Obstet. Gynecol. 191*, 928–32.

Kaiser, L.L. and Allen, L. (2002). Nutrition and lifestyle for healthy pregnancy outcome. *J. Am. Diet. Assoc. 102*, 1470–1490.

Khashan, A.S. and Kenny, L.C. (2009). The effects of maternal body mass index on pregnancy outcome. *Eur. J. Epidemiol. 24*, 697–705.

Pi-Sunyer, F.X. (1993). Medical hazards of obesity. *Ann. Intern. Med. 119*, 655–660.

Young, T.K. and Woodmansee, B. (2002). Factors that are associated with cesarean delivery in a large private practice: the importance of pre-pregnancy body mass index and weight gain. *Am. J. Obstet. Gynecol. 187*, 312–8; discussion 318–20.

Circular designs balanced for neighbours at distances one and two

Rosemary A. Bailey

Queen Mary, University of London, UK

Abstract

We consider experiments where the experimental units are arranged in a circle or in a single line in space or time. If neighbouring treatments may affect the response on an experimental unit, then we need a model which includes the effects of direct treatments, left neighbours and right neighbours. It is desirable that each ordered pair of treatments occurs just once as neighbours and just once with a single unit in between. A circular design with this property is equivalent to a special type of quasigroup.
In one variant of this, self-neighbours are forbidden. In a further variant, it is assumed that the left-neighbour effect is the same as the right-neighbour effect, so all that is needed is that each unordered pair of treatments occurs just once as neighbours and just once with a single unit in between.
I shall report progress on finding methods of constructing the three types of design.

Keywords

Neighbour-balance.

A multivariate regression approach for estimating variances of measures of population change over time from rotating repeated surveys

Yves G. Berger and **Rodolphe Priam**

University of Southampton, UK

Abstract

Measuring change over time is a central problem for many users of social, economic and demographic data and is of interest in many areas of economics and social sciences. Smith *et al.* (2003) recognised that assessing change is one of the most important challenges in survey statistics. The primary interest of many users is often in changes or trends from one time period to another. We propose a multivariate linear regression approach to estimate the variance of change.

A common problem is to compare two cross-sectional estimates for the same study variable taken on two different waves or occasions, and to judge whether the observed change is statistically significant. Assessing the significance of a change plays an important part in preliminary exploratory data analysis which helps to model a trend or a change.

Suppose, we wish to estimate the absolute change $\Delta = \tau_2 - \tau_1$ between two population totals $\tau_1 = \sum_{i\in U} y_{i1}$ and $\tau_2 = \sum_{i\in U} y_{i2}$ at wave 1 and wave 2 respectively. The quantities y_{i1} and y_{i2} are respectively the value of the variable of interest measure at wave 1 and 2. Suppose that Δ is estimated by $\hat{\Delta} = \hat{\tau}_2 - \hat{\tau}_1$, where $\hat{\tau}_1$ and $\hat{\tau}_2$ are the Horvitz-Thompson estimates. A primary interest is to test if an observed change is due to an actual change in the population or simply due to sampling errors. The variance of the change $\hat{\Delta}$ is given by

$$var(\hat{\Delta}) = var(\hat{\tau}_1) + var(\hat{\tau}_2) - 2 \times cor(\hat{\tau}_1, \hat{\tau}_2)\sqrt{var(\hat{\tau}_1)var(\hat{\tau}_2)}$$

Standard estimators can be used to estimate the variances $var(\hat{\tau}_1)$ and $var(\hat{\tau}_2)$. The correlation $cor(\hat{\tau}_1, \hat{\tau}_2)$ is the most difficult part to estimate.

The estimation of the correlation would be relatively straightforward if $\hat{\tau}_1$ and $\hat{\tau}_2$ were based upon the same sample, or if the sample remained the same from one wave to the next. Unfortunately, samples at different waves are usually not completely overlapping sets of units, because repeated surveys use rotating samples which consist in selecting new units at wave 2 to replace old units that have been in the sample for a specified number of waves.

Therefore, y_{i1} is known and y_{i2} is unknown for the units being replaced, on the other hands, y_{i1} is unknown and y_{i2} is known for the new units. For most of the units sampled at both waves, y_{i1} and y_{i2} are known.
Several methods can be used to estimate the variance (e.g. Kish 1965; Tam 1984, Holmes & Skinner 2000, Berger 2004). We propose to use a multivariate linear regression approach to estimate the correlation. Consider the multivariate model

$$\begin{pmatrix} \breve{y}_{i1} \\ \breve{y}_{i2} \end{pmatrix} = \begin{pmatrix} \beta_1^1 z_{i1} + \beta_2^1 z_{i2} + \beta_{12}^1 z_{i1} z_{i2} \\ \beta_1^2 z_{i1} + \beta_2^2 z_{i2} + \beta_{12}^2 z_{i1} z_{i2} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}$$

Where $\breve{y}_{i1} = y_{i1}/\pi_{i1}$ and $\breve{y}_{i2} = y_{i2}/\pi_{i2}$. The quantities π_{i1} and π_{i2} are respectively the inclusion probabilities for wave 1 and 2. The quantities $z_{ik} = 1$ if $i \in$ wave k sample and $z_{ij} = 0$ otherwise. The 2×1 vector residuals $(\epsilon_1, \epsilon_2)'$ can be assumed random with a bivariate normal distribution $N(0, \Sigma)$ where Σ is the 2×2 residual covariance matrix. Note that this model includes interactions between the variable z_{i1} and z_{i2}. We will show that these interactions capture the rotation of the repeated survey.
Finally, the estimator proposed for the correlation will be based upon the residuals covariance matrix Σ. Let $\hat{\Sigma}$ be the estimator of Σ. The proposed estimator for the correlation between $\hat{\tau}_1$ and $\hat{\tau}_2$ is given by

$$c\hat{o}r(\hat{\tau}_1, \hat{\tau}_2) = \frac{\hat{\Sigma}_{1,2}}{\sqrt{\hat{\Sigma}_{1,1} \hat{\Sigma}_{2,2}}}$$

where $\hat{\Sigma}_{k,\ell}$ is the component (k, ℓ) of the matrix $\hat{\Sigma}$.
We will show why the multivariate linear regression model is suitable for estimating the correlation, and how our result could be extended to include stratification.

Keywords

Change, Correlation, Horvitz-Thompson estimator, Multivariate regression, Repeated surveys, Rotation, Survey sampling.

References

Berger, Y.G. (2004). Variance estimation for measures of change in probability sampling. *Canad. J. Statist. 32*, 451–467.
Holmes, D.J. and Skinner, C.J. (2000). *Variance Estimation for Labour Force Survey Estimates of Level and Change.* GSS Methodology Series 21.
Kish, L. (1965). *Survey Sampling.* New York: John Wiley.
Smith, P., Pont, M. and Jones, T. (2003). Developments in business survey methodology in the Office for National Statistics, 1994-2000. *The Statistician 52(3)*, 257–295.
Tam, S.M. (1984). On covariances from overlapping samples. *Amer. Statist. 38*, 288–289.

Statistical power of goodness of fit tests based on the empirical distribution function for Type I censored data

Regina Bispo[1,2], Tiago A. Marques[2,3] and Dinis Pestana[4]

[1] ISPA - University Institute, Lisbon, Portugal
[2] CEAUL, University of Lisbon, Portugal
[3] CREEM, University of St Andrews, Scotland, UK
[4] University of Lisbon, Portugal

Abstract

In this study the power of common goodness of fit statistics based on the empirical distribution function (EDF) was simulated for single Type I censored data. The relative power of the Kolmogorov-Smirnov (D), Cramer-Von-Mises (W^2) and Anderson-Darling (A^2) statistics was investigated varying the null and the alternative distributions, the sample size, the degree of censoring and the significance level.
The exponential, Weibull, log-logistic and log-normal lifetime distributions are considered as they are among the most frequently used when modeling censored data.
We conclude giving some general recommendations for testing the distributional assumption of parametric survival models in homogeneous populations when using EDF goodness of fit statistics.

Keywords

Censored data, Exponential distribution, Goodness of fit, Log-logistic distribution, Log-normal distribution, Power, Simulation, Weibull distribution.

Prediction of semiconductor lifetime using bayesian linear models with mixed distributions

Olivia Bluder[1,2] **and Jürgen Pilz**[1]

[1] Alpen-Adria-University of Klagenfurt, Austria
[2] KAI-Kompetenzzentrum Automobil- und Industrie-elektronik GmbH, Austria

Abstract

Modeling and predicting lifetimes of power semiconductor devices has become more and more important during the last years. Since resources, especially time, are restricted, reliable prediction methods for lifetime of Devices under Test (DUT) are required.
For this study 10 datasets containing Cycles to Failure (CTF) of Smart Power ICs, tested with a temperature cycle stress test system, are used. Currently these tests are modeled with a log-normal distribution to predict the required parts per million (ppm) quantiles. Generally, predictions of mean lifetime are done with physical acceleration models, e.g. Arrhenius or Coffin-Manson model, but for the given data these models fail. Further difficulties arise with the given data, because the DUTs show two different failure mechanisms.
First a Bayesian linear model (LM) based on four test parameters is used. Based on previous research, for the data the normal distribution ($N(\mu, \sigma^2)$) is chosen, where the mean is modeled with a LM ($\mu = X\beta + \epsilon$). For the model parameters and σ^2 non-informative prior distributions are applied. This model shows weaknesses because it does not consider the mixed behavior of the data. To adapt the model to the DUTs behavior, a Bayesian LM with a mixture of two normal distributions is investigated. As before, non-informative priors are used, except for the intercepts, where uniform priors based on expert knowledge are applied. Since it can be demonstrated that the mixing proportion depends linearly on the peak temperature of the DUTs, this information is also included into the model. To compare the performance of the two models, leave-one-out cross validation is used. The analysis showed a significant increase in quality for the model with mixed distribution.

Keywords

Bayesian linear models, Semiconductor reliability, Cross validation, Mixed distributions.

References

Bluder, O. (2008). *Statistical analysis of smart power switch life test results.* Diploma thesis, Alpen-Adria-University of Klagenfurt, Austria.

Escobar, L.A. and Meeker, W.Q. (2006). A review of accelerated test models. *Statist. Sci. 21(4)*, 552–577.

Gill, J. (2008). *Bayesian Methods*. Boca Raton(FL): Chapman & Hall/CRC.

Glavanovics, M., Estl, H. and Bachofner, A. (2001). Reliable smart power systems ICs for automotive and industrial application - the infineon smart multichannel switch family. *Proceedings of 43. International Conference Power Electronics, Intelligent Motion, Power Quality (PCIM), Nürnberg, Germany.*

Glavanovics, M., Köck, H., Eder, H., Košel, V. and Smorodin, T. (2007). A new cycle test system emulating inductive switching waveforms. *Proceedings of 12th European Conference on Power Electronics and Applications (EPE), Aalborg, Denmark*, (1–9).

Hamada, M.S., Wilson, A.G., Reese, C.S. and Martz, H.F. (2007). *Bayesian Reliability*. New York: Springer Science + Business Media.

Shao, J. (1993). Linear model selection by cross-validation. *J. Amer. Statist. Assoc. 88*, 486–494.

Detection and evaluation method of track bed sedimentation

Guoqiang Cai and **Yu Liang**

Beijing Jiaotong University, China

Abstract

high-speed railway transportation has some incomparable advantages, such as the land high-usage, low energy consumption, little environmental pollution and high security. In the past six years, China developed high-speed railway very rapidly, and in future the momentum of development will continue. Track bed sedimentation, especially the uneven sedimentation will affect the quality of vehicle operating, or even suffer serious traffic accident and cause heavy loss to nation and passengers. A new method has been designed which is based on multi-magnetic Hall sensors. In this paper, the key issues of Automatic Measurement Method have also been discussed. In the measurement of sub-grade sedimentation, the method based on Hall sensors multi-trigger and automatic control of high-precision laser distance-phase testing is implemented; the status of track bed is correspond evaluated based on the detection. The results show that the method can be an effective measure to improve the efficiency and accuracy of automatic testing track bed sedimentation, and the evaluation system can ensure the safety of the High-speed railway transportation.

Keywords

Track bed sedimentation, Hall induction, Automatic measurement, Evaluation system.

Self organizing mixture network in mixture discriminant analysis: an experimental study

Nazif Çalış, Murat Erişoğlu and Hamza Erol

Çukurova University, Adana, Turkey

Abstract

In the recent works related with mixture discriminant analysis (MDA), expectation and maximization (EM) algorithm is used to estimate parameters of Gaussian mixtures. But, initial values of EM algorithm affect the final parameters' estimates. Also, when EM algorithm is applied two times, for the same data set, it can be give different results for the estimate of parameters and this affect the classification accuracy of MDA. Forthcoming this problem, we use Self Organizing Mixture Network (SOMN) algorithm to estimate parameters of Gaussians mixtures in MDA that SOMN is more robust when random the initial values of the parameters are used. We show effectiveness of this method on popular simulated waveform datasets and real glass data set.

Keywords

Self organizing mixture network, Mixture discriminant analysis, Waveform datasets, Glass identification, Mixture of multivariate normal distributions.

References

Dempster, P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete date via the EM algorithm. *J. R. Stat. Soc. Ser. B 39*, 1–38.

Hastie, T. and Tibshirani, R. (1996). Discriminant analysis by Gaussian mixtures. *J. R. Stat. Soc. Ser. B 58(1)*, 155–176.

Yin, H. and Allinson, N.M. (2001). Self-organizing mixture networks for probability density estimation. *IEEE Transactions on Neural Networks 12(2)*, 405–411.

On the evolution of a Markov open long term care population

Rute Carrujo[1], Gracinda Guerreiro[2], Manuel L. Esquível[2] and João T. Mexia[2]

[1] Polytechnic Institute of Tomar, Portugal
[2] New University of Lisbon, Portugal

Abstract

Long Term Care Insurance provides a combination of medical, nursing, social and community services designed to help people who have disabilities or chronic care needs. It may be provided in the person's home or in nursing homes. Our starting point is a continuous time Markov chain model for a Long Term Care population with 5 states, three of them corresponding to different degrees of dependence, one being the healthy state and the last one representing the withdrawals of population elements for which intensities are known. The Stochastic Vortices theory is used to estimate, through an open population perspective, the evolution of the population structure, which will be crucial for an accurate risk measurement and portfolio management.

Keywords

Markov chains, Stochastic Vortices, Long term care.

References

Carrujo, R. (2008). *Seguro de Dependência - Proposta de um modelo de avaliação financeiro-actuarial*. Msc Thesis, New University of Lisbon, Portugal.

Guerreiro, G.R. and Mexia, J.T. (2004). An alternative approach to bonus malus. *Discuss. Math. Probab. Stat. 24(2)*, 197–213.

Guerreiro, G.R. and Mexia, J.T. (2008). Stochastic vortices in periodically reclassified populations. *Discuss. Math. Probab. Stat. 28(2)*, 209–227.

Guerreiro, G.R. and Mexia, J.T. (2009). Stochastic vortices in periodically reclassified populations - an application to pension funds. *Proceedings of the 5th Conference in Actuarial Science and Finance, Samos 2008*, (83–97).

Habermann, S. and Pitacco, E. (1999). *Actuarial Models for Disability Insurance.* Chapman and Hall/CRC, Londres.

Resnick, S.I. (1994). *Adventures in Stochastic Processes.* Birkhäuser, Boston.

Coverings and light designs

Francisco Carvalho[1], **Ricardo Covas**[1]
and João T. Mexia[2]

[1] Polytechnic Institute of Tomar, Portugal
[2] New University of Lisbon, Portugal

Abstract

Light designs require less observations than the usual ones for experiences with the same factors for which we have the same number of levels. We now use a technique based on commutative Jordan algebras to derive light models, the coverings technique. Basically we select sub-vectors from the observations vectors of an usual model in such a way that its sets of treatments is conveniently covered, which designates the technique's name. Commutative Jordan algebras provide the tools for a proper choice of sub-vectors.
We will apply this technique to models with Commutative Orthogonal Block Structure (COBS) to derive the corresponding light models.

Keywords

Commutative Jordan algebras, COBS, Inference, Light designs.

References

Carvalho, F., Mexia, J.T. and Oliveira, M.M. (2009). Estimation in models with commutative orthogonal block structure. *J. Stat. Theory Practice 3(2)*, 523–533.

Fonseca, M., Mexia, J.T. and Zmyślony, R. (2006). Binary operation on Jordan algebras and orthogonal normal models. *Linear Algebra Appl. 417*, 75–86.

Fonseca, M., Mexia, J.T. and Zmyślony, R. (2008). Inference in normal models with commutative orthogonal block structure. *Acta et Commentationes Universitatis Tartunesis de Mathematica*, 3–16.

Jordan, P., von Neumann, J. and Wigner, E. (1934). On the algebraic generalization of the quantum mechanical formalism. *Ann. Math. 36*, 26–64.

Seely, J. (1970). Linear spaces and unbiased estimators. *Ann. Math. Statist. 41*, 1735–1745.

Zmyślony, R. (1978). A characterization of Best Linear Unbiased Estimators in the general linear model. *Lecture Notes Statistics 2*, 365–373.

Modelling k-sample multivariate extremes with application to extreme temperature analysis

Miguel de Carvalho[1] **and Anthony C. Davison**[2]

[1] New University of Lisbon, Portugal
[2] Ecole Poytechnique Fédérale de Lausanne, Switzerland

Abstract

The joint modelling of extremal events has been a subject of considerable attention both theoretically and in applications. Despite of the interest in the comovement of tail events, all approaches known in the literature consider only a spectral distribution function whereas in some applications k independent sources of information are available, each being characterized by a certain covariate x_k. Just as there is an obvious rationale for not modelling multivariate extreme values through univariate techniques, there are also strong reasons for not modelling individually the spectral density corresponding to each of the samples. Particularly, such approach would be ineffective in assessing the role that the covariate x_k would play in the interaction of extremes. This paper proposes a semiparametric formulation through which a family of k spectral densities is linked through an exponential tilt and constrained to satisfy a set of marginal moment conditions. Empirical likelihood inference and estimation for this spectral density ratio model is here obtained. An application is given wherein we contrast extreme temperatures under forest-cover versus open-site over 14 different locations in Switzerland.

Keywords

Atmospheric temperature, Empirical likelihood, Exponential tilt, Forest microclimates, Multivariate extreme values, Spectral distribution, Semiparametric modelling.

References

Boldi, M.-O. and Davison, A.C. (2007). A mixture model for multivariate extremes. *J. R. Stat. Soc. Ser. B. 69*, 217–229.

Cheng, J., Qin, J. and Zhang, B. (2009). Semiparametric estimation and inference for distributional and general treatment effects. *J. R. Stat. Soc. Ser. B. 71*, 881–904.

Exact and near-exact distributions for likelihood ratio test statistics used to test for stationarity and circularity in multivariate models

Carlos A. Coelho[1], Sandra Oliveira[2] and Filipe J. Marques[1]

[1] New University of Lisbon, Portugal
[2] Polytechnic Institute of Setúbal, Portugal

Abstract

In this paper we obtain the exact distribution for the likelihood ratio test (l.r.t.) statistics to test that in a multivariate normal model: i) the mean vector is null and the covariance matrix is circular, ii) the means in the mean vector are all equal and the covariance matrix is circular. The authors show that in the first case the exact distribution of the l.r.t. statistic may be written as an infinite mixture of Generalized Near-Integer Gamma (GNIG) distributions, while in the second case the exact distribution of the l.r.t. statistic is a Generalized Integer Gamma (GIG) distribution. For the first l.r.t. statistic, in which case the exact distribution is less manageable, it is thus desirable and useful the development of near-exact distributions. These will assume the form of finite mixtures of GNIG distributions.

Keywords

Circular covariance matrix, Sums of independent Gamma r.v.'s, Sums of independent Logbeta r.v.'s, Product of independent Beta r.v.'s.

References

Coelho, C.A. (2004). The generalized near-integer Gamma distribution, a basis for "near-exact" approximations to the distributions of statistics which are the product of an number of independent Beta random variables. *J. Multivariate Anal. 89*, 191–218.

Olkin, I. and Press, S.P. (1969). Testing and estimation for a circular stationary model. *Ann. Math. Statist. 40*, 1358–1373.

Tricomi, F.G. and Erdély, A. (1951). The asymptotic expansion of a ratio of Gamma functions. *Pacific J. Math. 1*, 133–142.

Measures of multivariate association using distances

Carles M. Cuadras[1] **and** **Daniel Cuadras**[2]

[1] University of Barcelona, Spain
[2] Catalan Oncology Institute, Spain

Abstract

Several coefficients have been proposed in measuring the multivariate association between two data sets taken on the same individuals. Ecology is a clear example, where environmental data is related to species. In genomic data we may seek relations between genotype (e.g., DNA data) and phenotypes of interest. There are also examples in biometry and psychology. Often, the data sets are represented by two quantitative matrices, where the rows are multivariate observations taken on the same individuals. Then some dependence measures (Wilks, Hotelling, Pillai, Cramer-Nicewander, etc.) based on canonical correlations can be used. However, if the two data sets are non quantitative (binary, categorical, nominal), the information can alternatively be given by a similarity or distance matrix. Then we must relate two distance matrices, and some coefficients are proposed by correlating principal coordinates. When the first data set is quantitative and univariate, these measures reduce to the squared multiple correlation coefficient.

Keywords

Multivariate statistics, Dependence measures, Canonical correlation, Distance analysis, Multidimensional scaling.

Rank aggregation and its use in bioinformatics problems

Susmita Datta

University of Louisville, USA

Abstract

We first illustrate the need to aggregate several ranked lists and produce a composite ranking through a number of practical examples. A solution of this problem is then offered through a decision theoretic optimization criterion and explained why it is better than more naive approaches. Computational methods to solve this mathematical problem are then presented. Integration of the rank aggregation methodology to three statistical problems are then discussed, namely, optimal selection of clustering algorithms and number of clusters, meta analysis of significance tests and construction of an adaptive data based classifier. Bioinformatics applications of the resulting methods are also discussed.

Application of statistical models for prediction of performance of students in school examination

Anupam Deka

Handique Girls' College, Assam, India

Abstract

Different statistical and mathematical models have been widely used for prediction indifferent situations including production, population growth, pollution etc. These procedures can also be adequately applied for prediction of students' performance in school final examination based on his performance in the test examination. Further, the overall performance of students in a state can also be predicted based on the state's earlier pass performances.
In this paper different aspects of prediction of performance of students have been examined based on his actual and hypothetical performances in the class X examination. Significantly after passing class X examination, a student decides the stream - science, arts, commerce for his future career. This has been found that these models can be used for prediction with significant efficiency.

Keywords

Normal distribution, Binomial distribution, Poisson distribution, Linear models, Goodness of fit.

Modeling DHS data using dynamic mixture models

José G. Dias

Lisbon University Institute, Portugal

Abstract

Finite mixture models have become very popular tools in modeling unobserved heterogeneity (*vide*, *e.g.*, McLachlan and Peel (2000), Dias and Willekens (2005)). This research extends the application of these models to sequential discrete data by incorporating misclassification error. The specified model is estimated by the Baum-Welch algorithm (Baum *et al.*, 1970) and the selection of the number of components in the mixture is based on the Bayesian information criterion of Schwarz (Schwarz, 1978).
Data in the illustration come from the *1996 Brazil Demographic and Health Survey* (BDHS). This data set includes a Life History Calendar of the reproductive career of the women surveyed, which identifies monthly state occupancies (*e.g.*, *being pregnant*). We selected the Northeast region of the Brazil as it is the most determinant in the evolution of the Brazilian total fertility rate (TFR). The results show three subpopulations with different dynamics.

Keywords

Finite mixture models, Latent class models, Markov chains, DHS – Demographic and Health Surveys.

References

Baum, L.E., Petrie, T., Soules, G. and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist. 41*, 164–171.

Dias, J.G. and Willekens F. (2005). Model-based clustering of sequential data with an application to contraceptive use dynamics. *Math Popul. Stud. 12(3)*, 135–157.

McLachlan, G.J. and Peel, D. (2000). *Finite Mixture Models.* New York: John Wiley & Sons.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann Statist. 6*, 461–464.

Extremes of two-step regression quantiles

Jan Dienstbier and **Jan Picek**

Technical University of Liberec, Czech Republic

Abstract

We deal with estimators of extreme value index based on two-step regression quantiles in the linear regression model. Two-step regression quantiles can be seen as a possible generalization of the quantile idea and as an alternative to regression quantiles of Koenker and Basset (1978). We derive the approximation of the tail quantile function of errors in the model on the basis of the two-step regression quantiles. Following Drees (1998) we consider a class of smooth functionals of the tail quantile function as a tool for the construction of estimators in the linear regression context. The chief examples of estimators derived in this way are versions of Pickands, maximum likelihood and probability weighted moments estimators. We illustrate the results on simulated data and real datasets such as Condroz data from Beirlant et al. (2004).

Keywords

Two-step regression quantile, Tail quantile function, Stochastical functional, Extreme value index.

References

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Application.* John Wiley & Sons, Chichester.

Jurečkova, J. and Picek, J. (2005). Two step regression quantiles. *Sankhyā 67*, 227–252.

Koenker, R. and Basset, G. (1978). Regression quantiles. *Econometrica 46*, 35–50.

Picek, J. and Dienstbier, J. (2010). Extremes of Two-step regression quantiles. *IMS Collections.* Accepted.

Optimal approximate repeated measurement designs and efficient exact designs

Pierre Druilhet

Blaise Pascal University, Clermont Ferrand, France

Abstract

Kushner (1997) and Kunert and Martin (2000) proposed new methods to obtain optimal repeated designs for direct treatment effects. Bailey and Druilhet (2004) propose another approach to obtain optimal designs for total effects. In this presentation we propose a generalization of these methods by using the extremal representation of the information matrix proposed by Gaffke (1987). Then we show how to obtain approximated repeated measurement designs in various situations and how we can derived efficient designs. At least we present an example from Druilhet and Tinsson (2009).

Keywords

Optimal designs, Approximate designs, Repeated measurement designs.

References

Bailey, R.A. and Druilhet, P. (2004). Optimality of neighbor-balanced designs for total effects. *Ann. Statist. 32*, 1650–1661.

Druilhet, P. and Tinsson, W. (2009). Optimal repeated measurement designs for a model with partial interactions. *Biometrika 96(3)*, 677–690.

Gaffke, N. (1987). Further characterizations of design optimality and admissibility for partial parameter estimation in linear regression. *Ann. Statist. 15*, 942–957.

Kunert, J. and Martin, R.J. (2000). On the determination of optimal designs for an interference model. *Ann. Statist. 28*, 1728–1742.

Kushner, H.B. (1997). Optimal repeated measurements designs: the linear optimality equations. *Ann. Statist. 25*, 2328–2344.

Estimators uniformly shrinking on subspaces

Pierre Druilhet[1] **and Alain Mom**[2]

[1] Blaise Pascal University, Clermont Ferrand, France
[2] University of Rennes 2, France

Abstract

It is well established that some classes of biased estimators are often preferable to ordinary least squares estimators in linear regression when the explanatory variables are highly correlated. Examples are Ridge regression , principal component regression (PCR), partial least-squares regression . To understand the behaviour of these estimators, many authors have investigated their shrinkage properties from two different points of view. From a global point of view, it has been established that RR, PCR and PLSR estimators have a lower Euclidean norm than the OLS estimator (see De Jong, 1995 or Goutis,1996 for PLSR). From a directional point of view, Frank and Friedman (1993) Butler and Denham (2000) and Lingjaerde and Christophersen (2000) have shown that shrinkage factors on the principal directions, i.e. those given by PCR, are between 0 and 1 for RR, PCR but not for PLSR. Druilhet and Mom (2008) have shown that shrinkage factors on the directions given by PLSR are between 0 and 1 for RR and PLSR but not for PCR.
We characterize regression on components such that the resulting estimators shrink uniformly on the subspace spanned by their weight vectors and we propose a new regression on components based on both PLSR and PCR criteria having this property.

Keywords

Biased regression, Regression on components, Shrinkage.

References

Butler, N.A. and Denham, M.C. (2000). The peculiar shrinkage properties of partial least squares regression. *J. R. Stat. Soc. Ser. B Stat. Methodol. 62*, 585–593.
De Jong, S. (1995). PLS shrinks. *J. Chemometrics 9*, 323–326.
Druilhet, P. and Mom, A. (2008). Shrinkage structure in biased regression. *J. Multivariate Anal. 99*, 232–244.
Frank, I.E. and Friedman, J.H. (1993). A statistical view of some chemometrics regression tools. *Technometrics 35*, 109–135.
Goutis, C. (1996). Partial least squares algorithm yields shrinkage estimators. *Ann. Statist. 24*, 816–824.
Lingjærde, O.C. and Christophersen, N. (2000). Shrinkage structure of partial least squares. *Scand. J. Statist 27(3)*, 459–473.

Estimation of a quarterly model with annual sample selection

Montezuma Dumangane[1] **and** **Nicoletta Rosati**[2]

[1] Portuguese Competition Authority, Lisbon, Portugal
[2] Technical University of Lisbon, Portugal

Abstract

This work aims at developing a semi-parametric methodology for the estimation of a non-linear quarterly panel data model with annual sample selection. A random annual sample is used to build a semi-parametric sample selection model that allows consistent estimation of the quarterly parameters from the non-random samples. The correction of the quarterly estimates is achieved through an extension of the model introduced by Kiriazidou (1997).
The extension is two-fold. First, the non-negative nature of the data leads to the development of a non-linear model. In fact, as pointed out in Silva and Tenreyro (2006), applying a linear model to the log-transformed data, despite being common practice, is not recommended as induces inconsistency of the estimates of at least the intercept parameter. The second extension is necessary due to the selection mechanism, since the non-random selection of the quarterly observations is done annually, therefore the frequency of the data and of the selection do not coincide.
The proposed estimator is obtained through a two-step procedure with semi-parametric weights, where the selection equation is based on a continuous random variable, allowing for unobserved heterogeneity. The methodology is applied to the estimation of a quarterly model for sales, using a Portuguese quarterly firms panel survey.

Keywords

Sample selection, Nonlinear models, Panel data, Semiparametric estimation.

References

Kiriazidou, E. (1997). Estimation of a panel data sample selection model. *Econometrica 65*, 1335–1364.

Santos Silva, J.M.C. and Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics 88*, 641–658.

Multivariate normality test using Srivastava's skewness and kurtosis

Rie Enomoto[1], **Naoya Okamoto**[2] **and Takashi Seo**[1]

[1] Tokyo University of Science, Japan
[2] Tokyo Seiei College, Japan

Abstract

We consider the multivariate normality test based on the sample measures of multivariate skewness and kurtosis defined by Srivastava (1984). Jarque and Bera (1987) proposed the test statistic by using both the univariate sample skewness and kurtosis as the univariate normality test. For the multivariate case, Koizumi, Okamoto and Seo (2009) proposed test statistics by using Srivastava's sample skewness and kurtosis, which are asymptotically distributed as χ^2-distribution. However, they did not derive variance of their test statistics. We propose a new test statistic using variance of the test statistic derived by Koizumi, Okamoto and Seo (2009). In order to evaluate accuracy of proposed test statistic, the numerical results by Monte Carlo simulation for some selected values of parameters are presented.

Keywords

Multivariate skewness, Multivariate kurtosis, Jarque-Bera test, Test for multivariate normality.

References

Jarque, C.M. and Bera, A.K. (1987). A test for normality of observations and regression residuals. *International Statist. Review 55*, 163–172.

Koizumi, K., Okamoto, N. and Seo, T. (2009). On Jarque-Bera tests for assessing multivariate normality. *J. of Statist.: Advances in Theory and Applications 1*, 207–220.

Srivastava, M. S. (1984). A measure of skewness and kurtosis and a graphical method for assessing multivariate normality. *Statist. Probab. Lett. 2*, 263–267.

Predicting bankruptcy using Support Vector Machines: an application of bank bankruptcy

Birsen Eygi Erdogan

Marmara University, Turkey

Abstract

The purpose of this study is to apply Support Vector Machines (SVM), which is a recently introduced classification method based on statistical learning theory, to bankruptcy analysis.
Although the prediction of financial distress of companies is analyzed with several statistical and machine learning techniques, the bank classification and bankruptcy prediction still needs to be investigated due to lack of adequate practice in the field of banking.
In this study SVM is implemented for analyzing financial ratios. Data sets belonging to the Turkish commercial banks are used. This work shows that Support Vector Machine's are capable of extracting useful information from financial data and can be used as a part of an early warning system.

Keywords

Bankruptcy prediction, Bank classification, Support Vector Machines.

References

Adriaanz, P. and Zantinge, D. (1996). *Data Mining.* Addison-Wesley, Harlow.

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J. Finance 23*, 589–609.

Altman, E., Marco, C. and Varetto, F. (1994). Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (the Italian experience), *J. Banking Finance 18*, 505–529.

Aziz, A., Emanuel, D.C. and Lawson, G.H. (1988). Bankruptcy prediction: an investigation of cash flow based models. *J. Management Studies*, *25(5)*, 419–437.

Barniv, R., Agarwal, A. and Leach, R. (1997). Predicting the outcome following bankruptcy filling: a three-state classification using neural networks. *Intelligent Systems in Accounting, Finance and Management 6(3)*, 177–194.

Beaver, W. (1966). Financial ratios as predictors of failures. Empirical research in accounting: selected studies. *J. Accounting Research 5*, 71–111.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). *Classification and Regression Trees.* Chapman and Hall, New York, USA.

Burges, C.J.C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery 2*, 955–974.

Boyacioglu, M.A., Kara, Y. and Baykan, O.K. (2008). Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: a comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks inTurkey. *Expert Systems with Applications 36(2)*.

Charalambous, C., Charitou, A. and Kaourou, F. (2000). Comparative analysis of artificial neural network models: application in bankruptcy prediction. *Ann. Operation Research 99*, 403–425.

Charitou, A., Neophytou, E. and Charalambous, C. (2004). Predicting corporate failure: empirical evidence for the UK. *European Accounting Review 13*, 465–497.

Coats, P.K. and Fant, L.F. (1991). A neural network approach to forecasting financial distress. *J. Business Forcasting Methods and Systems 10*, 9–12.

Demyanyk, Y.S. and Hasan, I. (2009). Financial crises and bank failures: a review of prediction methods. *Bank of Finland Research Discussion Paper No. 35/2009.* (http://ssrn.com/abstract=1518368).

Dimitras, A.I., Slowinski, R., Susmaga, R. and Zopounidis, C. (1999). Business failure prediction using rough sets. *European Journal of Operational Research 114*, 263–280.

Fethi, M.D. and Pasiouras, F. (2010). Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey. *European Journal of Operational Research 204(2)*, 189–198.

George, E. (2000). The variable selection problem. *J. Amer. Statist. Assoc. 95(452)*, 1304–1308.

Gestel, T.V., Baesens, B., Suykens, J., Espinoza, M., Baestaens, D.-E., Vanthienen, J. and Moor, B.D. (2003). Bankruptcy prediction with least squares support vector machine classifiers. *IEEE International Conference on Computational Intelligence for Financial Engineering*.

Guyon, I. and Eliseeff, A. (2003). An introduction to variable and feature selection. *J. Machine Learning Research 3*, 1157–1182.

Härdle, W., Moro, R.A. and Schäfer, D. (2005). Predicting bankruptcy with support vector machines. *SFB 649 Discussion Paper 2005-009*, (http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2005-009.pdf).

Min, J.H. and Lee, Y. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert System with Applications 28(4)*, 603–614.

Shin, K.S., Lee, T.S. and Kim, H.Y.(2005). An application of support vector machines in bankruptcy prediction model. *Expert System with Applications 28*, 127–135.

A simulation study for alternative estimation technique in nonlinear models with multicollinear data

Ali Erkoc[1], **Kadri Ulaş Akay**[2] **and Mujgan Tez**[1]

[1] Marmara University, Turkey
[2] Istanbul University, Turkey

Abstract

Nonlinear regression models have a wide use in the most of applied sciences such as chemistry, biology, agriculture. So the estimation procedure in nonlinear models is very important. A method widely used in computer algorithms for nonlinear regression is The Gauss-Newton method that uses a Taylor series expansion to approximate the nonlinear regression model with linear terms and then employs ordinary least squares to estimate the parameters. Unfortunately while this method is being employed, the multicollinearity problem is omitted. This usually causes misleading or erroneous inferences. Based on our best knowledge coming from literature survey, this problem has not been analyzed efficiently. S.H. Ngo, S. Kemény, A. Deák (2003) examined into the problem briefly by way of a simulation. The aim of this study is to introduce a new approach to estimate nonlinear regression model in case multicollinearity exists on Jacobian matrix. Finally the performance of our estimation technique will be analyzed using simulation study.

Keywords

Multicollinearity, Nonlinear regression, Iterative ridge regression, Mean squared error.

References

Hill, C.R. and Adkins, L.C. (2001). Collinearity. In: Baltagi, B.H. (Eds.) *A Companion to Theoretical Econometrics.* Wiley-Blackwell.

Belsley, D.A. (1991). *Conditioning Diagnostics Collinearity and Weak Data in Regression.* Wiley-Interscience.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: application to nonorthogonal problems. *Technometrics 12*, 69–82.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: viased estimation for nonorthogonal problems. *Technometrics 12*, 55–68.

Marquardt, D.W. and Snee, R.D. (1975). Ridge regression in practice. *Amer Statist. 29(1)*, 3–20.

Montgomery, D.C., Peck, E.A. and Vining, G.G. (2001). *Introduction to Linear Regression Analysis*. Wiley-Interscience.

Neter, J., Wasserman, W. and Kutner, M.H. (1983). *Applied Linear Regression Models*. Irwin, Homewood, Illinois.

Ngo, S.H., Kemény, S. and Deák, A. (2003). Performance of the ridge regression method as applied to complex linear and nonlinear models. *Chemometrics and Intelligent Laboratory Systems 67*, 69–78.

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New Jersey.

Zhang, J. and Ibrahim, M. (2005). A simulation study on SPSS ridge regression and ordinary least squares regression procedures for multicollinearity data. *J. Appl. Statist. 32(6)*, 571–588.

Optimal option portfolio strategies

José Faias and Pedro Santa-Clara

New University of Lisbon, Portugal

Abstract

Options should play an important role in asset allocation. They allow for kernel spanning and provide access to additional (priced) risk factors such as stochastic volatility and negative jumps. Traditional methods of asset allocation (e.g. mean-variance allocation) are not adequate for options because the distribution of returns is non-normal and the short sample makes it difficult to estimate their distribution. We propose a method to optimize option portfolios that solves these limitations. An out-of-sample exercise is performed and we show that, even when transaction costs are incorporated, our portfolio strategy delivers an annualized Sharpe ratio of 0:54 compared to 0:06 of S&P 500 index in the period between January 1996 and September 2008.

Keywords

Finance, Asset allocation, Options.

References

Eraker, B. (2007). The performance of model based option trading strategies. Working Paper.

Jones, C. (2006). A nonlinear factor analysis of S&P 500 index option returns. *J. Finance 61*, 2325–2363.

Coval, J.D. and Shumway, T. (2001). Expected option returns. *J. Finance 56*, 983–1009.

Stair nesting designs

Célia Fernandes[1], Paulo Ramos[1], Sandra Ferreira[2], Dário Ferreira[2] and João T. Mexia[3]

[1] Lisbon Superior Engineering Institute, Portugal
[2] University of Beira Interior, Covilhã, Portugal
[3] New University of Lisbon, Portugal

Abstract

The number of treatments in balanced nesting design is the product of the number of levels in each factor. In stair nesting designs the number of treatments is the sum of the factor levels. So, the stair nesting designs require fewer observations than the balanced nesting designs. Moreover it is simple to carry out inference when we use stair nesting designs. When we crossing stair nestings these advantages can also be found.

Keywords

Stair nesting, Balanced nesting, Variance components, Confidence intervals, Crossing.

References

Cox, D. and Salomon, P. (2003). *Components of Variance*. New York: Chapman and Hall.

Fernandes, C., Ramos, P. and Mexia, J.T. (2005). Optimization of nested step designs. *Biom. Lett. 42(2)*, 143–151.

Fernandes, C., Ramos, P., Saraiva,S. and Mexia, J.T. (2007). Variance components estimation in generalized orthogonal models. *Discuss. Math. Probab. Stat. 27*, 99–115.

Fonseca, M., Mexia, J.T. and Zmyślony, R. (2003). Estimators and tests for variance components in cross nested orthogonal designs. *Discuss. Math. Probab. Stat. 23(2)*, 175–201.

Fonseca, M., Mexia, J.T. and Zmyślony, R. (2006). Binary Operations on Jordan Algebras and Orthogonal Normal Models. *Linear Algebra Appl. 417*, 75–86.

Lehmann, E. (1959). *Testing statistical hypotheses*. New York: John Wiley and Sons.

Extension of maximum likelihood estimation methods to mixed linear models

Dário Ferreira[1], **Sandra Ferreira**[1], **Célia Nunes**[1] **and João T. Mexia**[2]

[1] University of Beira Interior, Covilhã, Portugal
[2] New University of Lisbon, Portugal

Abstract

We extend and compare Maximum Likelihood Estimation methods, that usually are used for estimating the parameters in unbalanced random linear models, to unbalanced mixed linear models.

Keywords

Linear mixed models, Variance components, Maximum likelihood estimation methods.

References

Ferreira, D., Ferreira, S., Nunes, C. and Mexia, J.T. (2009). Adjustment of normal mixed models through triple minimization. *InterStat*.

Fonseca, M., Carvalho, M., Oliveira, M. and Mexia, J.T. (2008). A reduction technique for conducting inference in mixed models. *LinStat'2008* - the talk.

Fonseca, M. and Mexia J.T. (2007). Inference in mixed models with stochastic search. *Proceedings of the 56th Session of the International Statistics Institute 2007*.

Hartley, H.O. and Rao, J.N.K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika 54*, 93–108.

Patterson, D. and Thompson, R. (1974). Maximum likelihood estimation of components of variance. *Proceedings of the 8th International Biometric Conference*, 197–207.

Genealogical trees for segregated COBS

Sandra Ferreira[1], **Dário Ferreira**[1], **Célia Nunes**[1],
Ricardo Covas[2] **and João T. Mexia**[3]

[1] University of Beira Interior, Covilhã, Portugal
[2] Polytechnic Institute of Tomar, Portugal
[3] New University of Lisbon, Portugal

Abstract

The crossing and nesting of segregated Commutative Orthogonal Block Structure, COBS, gives, see Ferreira et al. (2007) and Ferreira et al. (2010), segregated Commutative Orthogonal Block Structure. There are mixed models for which the Complete Sufficient Statistics, CSE, of estimable vectors are Best Linear Unbiased Estimator, BLUE, see Ferreira et al. (2010), and for which we can estimate all variance components even when the fixed effects part does not reduce to $1_n\mu$.
In Fonseca et al. (2006) we see that crossing and nesting of models are associative. Then we may use both operations to derive complex models from simpler ones. The structure and genesis of such models can, as we shall show, be presented through it's genealogical tree.

Keywords

Commutative Orthogonal Block Designs, Genealogical tree.

References

Covas, R. (2007). *Orthogonal mixed models and commutative Jordan algebras.* Ph.D Thesis – New University of Lisbon, Portugal.

Ferreira, S.S. (2006). *Inferência para modelos ortogonais com segregação.* Ph.D Thesis – University of Beira Interior, Portugal.

Ferreira, S.S., Ferreira, D. and Mexia, J.T. (2007). Cross additivity in models with cross nesting. *J. Statist. Theory Practice 1*, 377–392.

Ferreira, S.S., Ferreira, D., Nunes, C. and Mexia, J.T. (2010). Nesting segregated mixed models. *J. Statist. Theory Practice 4*. In print.

Fonseca, M., Mexia, J.T. and Zmyślony, R. (2006). Binary operations on Jordan algebras and orthogonal normal models. *Linear Algebra Appl. 417*, 75–86.

Optimality of designs under models with interference dependence structure

Katarzyna Filipiak

Poznań University of Life Sciences, Poland

Abstract

We consider optimality of complete circular block designs under two models. First we assume that observations within blocks are correlated according to circular autoregression process and uncorrelated between blocks. In the second case we consider the mixed interference model with random interference effects.
The aim of the paper is to characterize optimal designs with respect to the D- and E-optimality criterion.

Keywords

Interference model, Circular autoregression, Complete circular block design, E-optimality, D-optimality, Information matrix.

References

Filipiak, K. and Markiewicz, A. (2007). Optimal designs for a mixed interference model. *Metrika 65*, 369–386.

Filipiak, K. and Markiewicz, A. (2005). Optimality and efficiency of neighbor balanced designs for correlated observations. *Metrika 61*, 17–27.

Filipiak, K., Różański, R., Sawikowska, A. and Wojtera-Tyrakowska, D. (2008). On the E-optimality of complete designs under an interference model. *Statist. Probab. Lett. 78*, 2470–2477.

Regression analysis of compositional data via linear model with type-II constraints

Eva Fišerová and **Karel Hron**

Palacký University Olomouc, Czech Republic

Abstract

The restrictive properties of compositional data, that is multivariate data with positive parts that carry only relative information in their components (Aitchison, 1986), call for special care to be taken while performing standard statistical methods, for example, regression analysis. Among the special methods suitable for handling this problem is the total least squares procedure (TLS, orthogonal regression, regression with errors in variables, calibration problem), performed after an appropriate logratio transformation. The difficulty or even impossibility of deeper statistical analysis (confidence regions, hypotheses testing) using the standard TLS techniques can be overcome by calibration solution based on linear statistical models, namely models with type-II constraints (constraints involve in addition to the unknown model's parameters the other unobservable ones), see e.g. (Brown, 1993; Kubáček et al., 1995). This approach can be combined with standard statistical inference, for example, confidence and prediction regions and bounds, hypotheses testing, etc., suitable for interpretation of results. Here, we deal with the simplest TLS problem where we assume a linear relationship between two errorless measurements of the same object (substance, quantity). We propose an iterative algorithm for estimating the calibration line and also give confidence ellipses for the location of unknown errorless results of measurement. An illustrative example from the fields of geochemistry will be presented.

Keywords

Compositional data, Total least squares, Linear model, Calibration line.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data.* London: Chapman and Hall.

Brown, P.J. (1993). *Measurement, Regression and Calibration.* Oxford: Clarendon Press.

Kubáček, L., Kubáčková, L. and Volaufová, J. (1995). *Statistical Models with Linear Structures.* Bratislava: Veda.

Extension of models with orthogonal block structure

Miguel Fonseca[1], Paulo C. Rodrigues[1], Francisco Carvalho[2] and João T. Mexia[1]

[1] New University of Lisbon, Portugal
[2] Polytechnic Institute of Tomar, Portugal

Abstract

An approach for the analysis and inference of unbalanced mixed models is presented. Given a mixed model $\mathbf{Y}^0$ its extensions are the models $\mathbf{Y} = \mathbf{L}\mathbf{Y}^0 + \mathbf{e}$ where $\mathbf{L}$ is a matrix with linearly independent column vectors and $\mathbf{e}$ is an error vector independent from $\mathbf{Y}^0$. The study is centered on core models with orthogonal block structure (van Leeeuwen et al., 1998, 1999, Fonseca et al., 2010). Crossing and nesting of core models (Fonseca et al., 2006) is carried out as well as extensions obtained using matrices $\mathbf{L}$ such that $\mathbf{L}^+\mathbf{L} = \mathbf{I}$. These last extensions preserve balance in the core models.
This approach can be seen as an extension of the method for balanced random models with unequal frequencies in the last stage (Khuri and Ghosh, 1990), allowing, for instance, replicates to follow a linear correlation structure with exogenous covariates within each cell, and the use of mixed core models.

Keywords

Commutative Jordan algebras, Core model, Mixed models, Orthogonal block structure, Orthogonal models.

References

Fonseca, M., Mexia, J.T. and Zmyślony, R. (2010). Least squares and generalized least squares in models with orthogonal block structure. *J. Statist. Plann. Inference 140*, 71346–1352.

Fonseca, M., Mexia, J.T. and Zmyślony, R. (2006). Binary operations on Jordan Algebras and orthogonal normal models. *Linear Algebra Appl. 417*, 75–86.

Khuri, A.I. and Ghosh, M. (1990). Minimal sufficient statistics for the unbalanced two-fold nested model. *Statist. Probab. Lett. 10*, 351–353.

van Leeeuwen, D., Seely, J. and Birkes, D. (1998). Sufficient conditions for orthogonal designs in mixed linear models. *J. Statist. Plann. Inference 73*, 373–389.

van Leeeuwen, D., Birkes, D. and Seely, J. (1999). Balance and orthogonality in designs for mixed classification models. *Ann. Statist. 27*, 1927–1947.

LIBOR convexity adjustments for the Vasiček and Cox-Ingersoll-Ross models

Raquel Gaspar[1], **Bruno Gaminha**[2] and **Orlando Oliverisa**[2]

[1] Technical University of Lisbon, Portugal
[2] Coimbra University, Portugal

Abstract

In this paper we numerically implement some of the recent theoretical results concerning convexity adjustments derived within the affine term structure setup. The computation of the convexity adjustments in that setup is reduced to solving a system of ODES. Here we explore the Vasiček and Cox-Ingersoll-Ross models within LIBOR-inarrears and investigate how the convexity adjustments change with the model parameters. The two models reproduce the same behavior with the convexity adjustment showing up as an additive constant for maturity times > 5 years.

Motivation

For fixed income markets, *convexity* has emerged as an intriguing and challenging notion. Taking this effect into account correctly could provide financial institutions with a competitive advantage. The idea underlying the notion of a convexity adjustment is quite intuitive and can be easily explained in the following terms. Many fixed income products are non-standard with respect to aspects such as the timing, the currency or the rate of payment. This leads to complex pricing formulas, many of which are hard to obtain in closed-form. Examples of such products include in-arreas or in-advance products, quanto products, CMS products, or equity swaps, among others. Despite their non-standard features, these products are quite similar to plain vanilla ones whose price can either be directly obtained from the market or at least computed in closed-form. Their complexity can be understood as introducing some sort of bias into the pricing of plain vanilla instruments. That is, we may decide to use the price of plain products and adjust it somehow to account for the complexity of non-standard products. This adjustment is what is known as convexity adjustment.

Under most stochastic interest rate setups convexity adjustments cannot be computed in closed-form and market practice is to use add-hoc rules or approximations when computing them. See, for instance, Hull (2006), Pugaschesky (2001), Hart (1997), Hagan (2003), Pelsser (2000), Brigo and Mercurio (2006). Most of the times one has no clue on how big this approximation

error may be although there is the hope convexity adjustments would be of a different order of magnitude, when compared to market prices, making all errors negligible.
In this paper we focus on timming adjustments and, in particular, on what we define to be LIBOR in-arrears adjustments (LIA adjustments). In Gaspar and Murgoci (2008) it was shown that, in any affine term structure setting, LIBOR adjustments can be obtained in closed-form, up to the solution of a system of ODEs. Here and for the popular models of Vasiček (Vasiček, 1977) and Cox-Ingersoll-Ross (Cox, Ingersoll, and Cox, 1985) models we numerically solved the necessary systems of ODEs and show, for a reasonable range of parameter values, convexity adjustments may be substantial in terms of market quotes. This undermines some of the market practices. Trough numerical experiments we find out and discuss term structure shapes for LIA convexity adjustments.

Keywords

Affine term structure models, ODEs, Convexity.

References

Brigo, D. and Mercurio, F. (2006). *Interest Rate Models: Theory and Practice*, 2nd edition. Springer Finance, Heidelberg.
Cox, J., Ingersoll, J. and Ross, S. (1985). A theory of the term structure of interest rates. *Econometrica 53*, 385–407.
Gaspar, R.M. and Murgoci, A. (2008). Convexity adjustments for general affine term structure models. *ADVANCE working paper series, ISEG, Technical University of Lisbon.*
Hagan, P.S. (2003). Convexity conundrums: Pricing cms swaps, caps and floors. *Wilmott magazine*, 38–44.
Hart, Y. (1997). Unifying theory. *Risk*, 54–55.
Hull, J.C. (2006). *Options, Futures, and Other Derivative Securities*, 6th edition. Prentice Hall Internation, Inc.
Pelsser, A. (2000). *Efficient Methods for Valuing Interest Rate Derivatives.* Springer Finance, Heidelberg.
Pugaschevsky, D. (2001). Constant maturity products: Forward cms rate adjustments. *Risk 14(3)*, 125–128.
Vasiček, O. (1977). An equilibrium characterization of the term structure. *J. Financ. Econ. 5*, 177–188.

An efficient Youden square design against the interchange of treatments

Dilip Kumar Ghosh

Saurashtra University, Rajkot, Gujarat, India

Abstract

Youden(1940) introduced Youden Square Design which now we called as binary Youden Square design. In this investigation we have extended the concept of Youden square and introduced the non binary Youden square design. Further, in this article we studied the A efficiency of the Youden square design against the interchange of two treatments when (i) both the treatments belong to different rows but from the same column and (ii) both the treatments belong to different rows and different columns.
Here we found that Youden square designs are fairly robust against the interchange of two treatments for both the case (i) and (ii) with v (treatment) ≥ 5. We also concluded that as λ (a pair of treatments occur together) increases, overall A efficiency of the residual design decreases.

Keywords

Binary design, Non binary design, Youden square design, A efficiency.

Estimability and connectivity in m-way designs

Janet Godolphin

University of Surrey, UK

Abstract

The classical problem of ascertaining the connectivity status of an m-way design has received much attention, particularly in the cases where $m = 2$ and $m = 3$. In the general case, a new approach yields the connectivity status for the overall design and for each of the individual factors directly from the kernel space of the design matrix. Furthermore, the set of estimable parametric functions in each factor is derived from a segregated component of this kernel space.

The kernel space approach enables a simple derivation of some classical results. Examples are given to illustrate the main results.

Keywords

Connectivity, Contrast, Estimability, Kernel.

References

Eccleston, J.A. and Russell, K.G. (1975). Connectedness and orthogonality in multi-factor designs. *Biometrika 62*, 341–345.

Godolphin, J.D. (2006). The specification of rank reducing observation sets in experimental design. *Comput. Statist. Data Anal. 51*, 1862–1874.

Wynn, H.P. (2008). Algebraic solutions to the connectivity problem for m-way layouts: Interaction-contrast aliasing. *J. Statist. Plann. Inference 138*, 259–271.

Will it always be necessary taking into account sample selection?

João Gomes[1] **and Tiago Oliveira**[2]

[1] University of Lisbon, Portugal
[2] New University of Lisbon, Portugal

Abstract

Core topics in labour economics are "self-selection" and "sample selection" in the linear model. During the last three decades, there has been a very significant work in this area of econometrics. New methods estimation and inference have been developed, both parametric and semi-parametric. However, Ordinary Least Squares (OLS) can be a viable alternative under certain conditions. In this work, we compare OLS with Maximum Likelihood Estimation (MLE), in the linear model with some types of sample selection. We propose two new measures to quantify the lack of information of OLS.

Keywords

Sample selection, Conditional expected values, Tobit models, OLS.

Comparing the BLUEs under two different linear models

Jan Hauke[1], Augustyn Markiewicz[2] and Simo Puntanen[3]

[1] Adam Mickiewicz University, Poznań, Poland
[2] Poznań University of Life Sciences, Poland
[3] University of Tampere, Finland

Abstract

In this talk we consider two linear models, $M_1 = \{y, X\beta, V_1\}$ and $M_2 = \{y, X\beta, V_2\}$, say, which differ only in their covariance matrices. Our main focus lies on the equality of the BLUEs of $X\beta$ under these models.
The corresponding problems between the models $\{y, X\beta, I_n\}$ and $\{y, X\beta, V\}$, i.e., between OLSE and BLUE, are pretty well studied and numerous equivalent conditions for the equality of OLSE and BLUE have been introduced and several measures for the relative efficiency of the OLSE have been suggested. Our purpose is to do the corresponding considerations between the BLUEs of $X\beta$ under M_1 and M_2.

Keywords

Best linear unbiased estimator, Gauss-Markov model, Generalized least squares estimator, Ordinary least squares estimator.

References

Mitra, S.K. and Moore, B.J. (1973). Gauss–Markov estimation with an incorrect dispersion matrix. *Sankhyā Ser. A 35*, 139–152.

Mitra, S.K. and Rao, C.R. (1968). Simultaneous reduction of a pair of quadratic forms. *Sankhyā Ser. A 30*, 313–322.

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

Rao, C.R. and Mitra, S.K. (1971). *Generalized Inverse of Matrices and Its Applications*. Wiley, New York.

Peaks over random threshold best linear unbiased estimation of the extreme value index

M. Ivette Gomes[1] **and Lígia Henriques-Rodrigues**[2]

[1] University of Lisbon, Portugal
[2] Polytechnic Institute of Tomar, Portugal

Abstract

In the general theory of *Statistics*, whenever we ask the question whether the combination of information can improve the performance of an estimator, we are led to think on *Best Linear Unbiased Estimators* (BLUE), i.e., on unbiased linear combinations of an adequate set of statistics, with minimum variance among the class of such linear combinations. In *Statistics of Extremes* and regarding the estimation of the *Extreme Value Index* (EVI), the primary parameter in this area, such an approach has been considered in Gomes *et al.* (2004), where asymptotically unbiased BLUE have been studied. But these estimators, like the classical Hill estimators (Hill, 1975) are not location-invariant, contrarily to the PORT-Hill estimators, recently introduced in Araújo Santos *et al.* (2006) and further studied for finite samples in Gomes *et al.* (2008), where PORT stands for *Peaks Over Random Threshold*. In this paper we shall consider PORT-BLUE, providing an adaptive choice of the *tuning parameters* under play and an application to environmental data.

Keywords

Statistics of extremes, Semi-parametric estimation, Best linear unbiased estimators, Peaks over random threshold methodology.

References

Araújo Santos, P., Fraga Alves, M.I. and Gomes, M.I. (2006). Peaks over random threshold methodology for tail index and quantile estimation. *REVSTAT 4(3)*, 227–247.

Gomes, M.I., Figueiredo, F. and Mendonça, S. (2004). Asymptotically best linear unbiased tail estimators under a second order regular variation. *J. Statist. Plann. Inference 134(2)*, 409–433.

Gomes, M.I., Fraga Alves, M.I. and Araújo Santos, P. (2008). PORT Hill and moment estimators for heavy-tailed models. *Comm. Statist. Simulation Comput. 37(6)*, 1281–1306.

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

Diagnostic biplots for linear models

Jan Graffelman

Universitat Politècnica de Catalunya, Spain

Abstract

In observational studies multicollinearity is a common problem in regression models (Belsley et. al, 1980). Several statistics are available that help to detect collinearity such as variance inflation factors (vif's), tolerances, correlations between regression coefficients, condition indices and others. However, it often remains a complicated task to identify the offending predictors and their relationships. Graphics are extensively used in the analysis of the residuals of a regression model, but a rarely used for an analysis of multicollinearity. Recently, Friendly and Kwan (2009) have suggested the use of biplots (Gabriel, 1971; Gower and Hand, 1996) as a graphical tool for uncovering collinearity in multiple regression. They proposed biplots involving the last principal component of the matrix of predictors in particular. Biplots have mostly been used to produce pictures of data matrices (raw quantitative data matrices, contingency tables), but can also be constructed for several matrices involved in a linear model that are of diagnostic interest. Several examples of such diagnostic biplots will be discussed.

Keywords

Biplot, Multicollinearity, Variance inflation factor.

References

Belsley, D.A., Kuh, E. and Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Friendly, M. and Kwan, E. (2009). Where's Waldo? Visualizing Collinearity Diagnostics. *Amer. Statist. 63*, 56–65.

Gabriel, K.R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika 58(3)*, 453–467.

Gower, J.C. and Hand, D.J. (1996). *Biplots*. London: Chapman & Hall.

The Fermat's equation on the sets of matrices and the homographic functions

Aleksander Grytczuk and Izabela Kurzydło

University of Zielona Góra, Poland

Abstract

We consider the matrix Fermat's equation

$$X^n + Y^n = Z^n. \tag{1}$$

We investigate the solvability of the equation (1) in 2×2 rational matrices X, Y, Z.

Moreover, we consider some connections between the Fermat's equation in the set of matrices and the set of special functions, namely the set of homographic functions.

Keywords

The matrix equations, Schur's Theorem, Fermat's type Diophantine equation on matrices.

References

Cao, Z. and Grytczuk, A. (1998). Fermat's type equations in the set of 2×2 integral matrices. *Tsukuba J. Math. 22(3)*, 637–643.

Cao, Z. and Grytczuk, A. (2000). Some remarks on Fermat's equation in the set of matrices. *Acta Acad. Paedagog. Agriensis Sect. Mat. 27*, 39–46.

Grytczuk, A. (1997). Fermat's equation in the set of matrices and special functions. *Stud. Univ. Babeş-Bolyai Math. 42(4)*, 49–55.

Influential observation in mixed linear model of repeated measures cross-over designs

Chengcheng Hao[1], **Tatjana von Rosen**[1] **and Dietrich von Rosen**[2]

[1] Stockholm University, Sweden
[2] Swedish University of Agricultural Sciences, Uppsala, Sweden

Abstract

The aim of this work is to develop new methodology to detect influential observations. Linear mixed models are applied to different kinds of cross-over designs. Perturbations are performed. We consider measures of influence of the perturbations on parameter estimates and their variances. Closed-form solutions for estimators of variance components as well as fixed effect parameters in cross-over designs are utilized.

Keywords

Local influence analysis, Balaam's cross-over design, Repeated measurement, Linear mixed model.

References

Putt, M. and Chinchilli, V.M. (1999). A mixed effects model for the analysis of repeated measures cross-over studies. *Stat. Med. 18*, 3037–3058.

Mixture extensions of linear models

John Hinde

National University of Ireland, Galway, Ireland

Abstract

Mixture models provide a highly flexible framework for making extensions to the basic linear model. This talk will consider some specific regression examples that illustrate the potential of mixture models for outlier detection/accommodation, to provide more general error distributions, and to allow for heterogeneous regression relationships. Estimation can be easily implemented using the EM algorithm for a finite mixture model and extensions using a mixture of experts model will be considered.

Keywords

Mixture models, Outliers, Non-normality.

References

Aitkin, M., Francis, B., Hinde, J. and Darnell, R. (2009) *Statistical Modelling in R*. Oxford.

A new estimator for Cox proportional hazard regression model in presence of collinearity

Deniz İnan and **Müjgan Tez**

Marmara University, Turkey

Abstract

This paper considers a new approach to estimate Cox proportional hazard regression model parameters in presence of collinearity between covariates. Usually partial maximum likelihood estimator is used to estimate Cox proportional hazard regression model parameters. But when there exist collinearity partial maximum likelihood estimates can be effected seriously. Parameter estimates have large variances so they may be far from true values.
In 2007 Xue *et al.* generalized ridge regression approach to the Cox proportional hazard regression model. But especially when there exist severe collinearity this approach may not fully addresses the collinearity problem.
In this study we developed Liu-type estimator for Cox proportional hazard regression model and compared with ridge regression estimator in terms of mean squared error (MSE). Finally we evaluated its performance through simulation studies.

Keywords

Cox regression model, Liu-type estimator, Partial maximum likelihood estimator, Multicollinearity.

References

Kaçiranlar, S. and Sakallioğlu, S. (2001). Combining the Liu estimator and the principal component regression estimator. *Comm. Statist. Theory Methods 30*, 2699–2705.

Kleinbaum, D.G. and Klein, M. (2005). Parametric survival models. In: Gail, M., Samet, J., Tsiatis, A., Wong, W. (Eds.), *Survival Analysis A Self-Learning Text* (pp. 257–329). Springer, USA.

Lee, E.T. and Wang, J.W. (2003). Identification of prognostic factors related to survival time: Cox proportional hazard model. In: Balding, D.J., Bloomfield, P., Cressie, N.A.C., Fisher, N.I., Johnstone, I.M., Kadene, J.B., Ryan, L.M., Scott, D.W., Smith, A.F.M., Teugels, J.L. (Eds.), *Statistical Methods for Survival Data Analysis* (pp. 298–376). John Willey & Sons, Canada.

Liu, K. (2003). Using Liu type estimator to combat collinearity. *Comm. Statist. Theory Methods 32(5)*, 1009–1020.

Lustbader, E.D. (1986). Relative risk regression diagnosis. In: Moolgavkar, S.H., Prentice, R.L. (Eds.), *Modern Statistical Methods in Chronic Disease Epidemiology*, SIAM, Philadelphia.

Montgomery, D.C., Peck, A.E. and Vining, G.G. (2001). Multicolinearity. In: Bloomfield, P., Cressie, N.A.C., Fisher, N.I., Johnstone, I.M., Kadene, J.B., Ryan, L.M., Scott, D.W., Silverman, B.W., Smith, A.F.M., Teugels, J.L.,Barnett,V., Bradley, R.A., Hunter, J.S., Kendall, D.G. (Eds), *Introduction to the Linear Regression Analysis* (pp. 325–382). John Willey & Sons, Canada.

Xue, X., Kim, M.Y. and Shore, R.E. (2007). Cox regression analysis in presence of collinearity: an application to assessment of health risk associated with occupational radiation exposure. *Lifetime Data Anal. 13(3)*, 333–350.

On the integer-valued mixture GARCH model

Shusong Jin

Fudan University, Shanghai, China

Abstract

Integer valued time series data appear naturally in medicine, particularly in epidemiology. We introduce an integer-valued mixture GARCH model as an analogue of the mixture GARCH model to capture asymmetry and bursts in such areas. The estimation method is proposed and some large sample properties are also discussed.

Keywords

Integer-valued time series, GARCH model, Poisson process, EM algorithm.

References

Bollerslv, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Rev. Econom. Statist. 72*, 498–505.

Cox, D.R. (1955). Some statistical methods connected with series of events. *J. R. Stat. Soc. Ser. B 17*, 129–164.

Efron, B. and Tibshrani, R. (1993). *An Introduction to Bootstrap*, Chapman & Hall.

Ferland, R., Latour, A. and Oraichi, D. (2006). Integer-valued GARCH process. *J. Time Ser. Anal. 27*, 923–942.

Single-sample predictive model stability assessment via variance components estimated through re-sampling and cross-validation

Michael P. Jones

Macquarie University, Sydney, Australia

Abstract

Background: Predictive models developed on a training sample are prone to over-optimism. In small samples, perhaps due to studying a rare disease or other outcome, an independent validation sample is not always feasible. Various approaches to single-sample model assessment have been proposed including shrinkage estimators for parameter bias and leave-k-out cross-validation. Aim: To develop a measure of model predictive stability based on a training sample alone. Method: Ratio of within- and between-subject variance components (σ_w/σ_b) from leave-k-out cross-validated predicted probabilities across bootstrapped binary logistic models is derived as an index of model stability. In a simulation study σ_w/σ_b is compared with differences in model performance metrics between independent training and validation samples (Δ).
Results: The variance component index exhibits good stability across models with varying number of predictors and correlations between predictors. The index varies in both mean and variance across predictors simulated from Normal and Exponential distributions, as expected. Importantly, the variance components index correlates well with the difference in model parameters across independent samples. Conclusion: When independent sample validation is not feasible, useful insight into model stability can be gained from variance components derived from the training sample alone.

Keywords

Predictive model validation, Re-sampling, Variance components in predicted values.

References

Altman, D.G. and Andersen, P.K. (1989). Bootstrap investigation of the stability of a Cox regression model. *Stat. Med. 8*, 771–783.

Browne, M.J. (2000). Cross-validation methods. *J. Math. Psych. 44*, 108–132.

Harrell, F.E., Margolis, P.A., Gove, S., Mason, K.E., Mulholland, E.K., Lehmann, D., Muhe, L., Gatchalian, S., Eichenwald, H.F. and the who/ari young infant multicentre study group (1998). Development of a clinical prediction model for an ordinal outcome. *Stat. Med. 17*, 909–944.

Steyerberg, E.W., Bleeker, S.A., Moll, H.A., Grobbee, D.E. and Moons, K.G.M. (2000). Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J. Clin. Epidemiol. 56*, 441–447.

Steyerberg, E.W., Harrell, F.E., Borsboom, G., Eijkemans, M.J, Vergouwe, Y. and Habbema, J. (2001). Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J. Clin. Epidemiol. 54*, 774–781.

Fit generalized linear models with using of different likelihoods

Iraj Kazemi and Hoda Rashidi Nejad

Isfahan University, Iran

Abstract

We introduce generalized linear models, a statistical estimator which combines features of nonlinear and non normal regression. A GLM uses some methods such as maximum likelihood for estimating parameters in model. In these models, there is a link function that shows relations between predictors and response variables. In some cases computation of maximum likelihood in full likelihood is difficult, so we need a simple method to estimate parameters. We obtain some types of likelihoods for making easy computations. These likelihoods are conditional, profile, empirical and quasi likelihood. We also present an example with real data of productivity of bulldozer and compute estimators with using these likelihoods by SAS software.

Keywords

Linear predictor, Link function, Nuisance parameter, Quasi likelihood, Sufficient statistic.

References

Bliss, C. (1934). The method of probits. *Science 79*, 38–39.

David, H.A. (1995). First occurrence of common terms in mathematical statistics. *Amer. Statist. 29*, 21–31.

Finney, D.J. (1952). *Probit Analysis.* Cambridge University Press, Cambridge.

Fisher, R.A. (1935). Appendix to "The calculation of the dose-mortality curve" by C. Bliss. *Annals of Applied Biology 22*, 164–165.

Nelder, J.A. and Pregibon, D. (1987). An extended quasi-likelihood fnction. *Biometrika 74*, 221–132.

Pawitan, Y. (2001). *In All Likelihood: Statistical Modeling and Inference Using Likelihood.* Clarendon Press, Oxford.

Rashidi, A., Rashidi Nejad, H. and Behzadan, A.H. (2009). Multiple linear regression approach for productivity estimation of bulldozers. *Korean Conference of Construction Engineering and Management (ICCEM,ICCPM).*

Searle, S.R., Cassella, G. and McCulloch, C.E. (1992). *Variance Components.* Wiley, New York.

On UMRU estimators in the extended growth curve model

Daniel Klein and **Ivan Žežula**

P. J. Šafárik University, Košice, Slovakia

Abstract

The extended growth curve model with orthogonal spaces of the design matrices will be discussed. Rather than the nested structure the orthogonality condition seems to simplify many theoretical results. We derive necessary and sufficient existence conditions for the uniformly minimum risk unbiased estimators of the parameters with the general and some special covariance structures under the convex losses.

Keywords

Extended growth curve model, Orthogonality, UMRU estimators, Covariance structure.

References

Hu, J. (2009). Properties of the explicit estimators in the extended growth curve model. *Statistics* iFirst articles (29.10.2009), doi: 10.1080/02331880903236884.

Wu, Q.G. (1998). Existence conditions of the uniformly minimum risk unbiased estimators in extended growth curve models. *J. Statist. Plann. Inference 69*, 101–114.

Žežula, I. (2008). Remarks on unbiased estimation of the sum-of-profiles model parameters. *Tatra Mt. Math. Publ. 39*, 45–52.

Confidence intervals for linear function of mean vectors in the intraclass correlation model with missing data

Kazuyuki Koizumi

Tokyo University of Science, Japan

Abstract

In this paper, we consider the test for the equality of mean vectors in the intraclass correlation model with monotone missing data. Dunn (1961) proposed the conservative method based on Bonferroni inequality. Holland and Copenhaver (1987) considered improved Dunn's method by using Sidák inequality. We derive Bonferroni type of simultaneous confidence intervals for linear contrasts based on Bonferroni's inequality. Also we improve Bonferroni type of simultaneous confidence intervals.

Keywords

Intraclass correlation model, Monotone missing data, Simultaneous confidence intervals for linear contrasts, Bonferroni inequality, Sidák inequality.

References

Dunn, O.J. (1961). Multiple comparisons among means. *J. Amer. Statist. Assoc.* *56*, 52–64.

Holland, B.S. and Copenhaver, M.D. (1987). An improved sequentially rejective Bonferroni test procedure. *Biometrics* *43*, 417–423.

On the asymptotic distribution of likelihood ratio test when parameters lie on the boundary

Leonid Kopylev[1] **and Bimal K. Sinha**[2]

[1] US Environmental Protection Agency, USA
[2] University of Maryland, Baltimore County, USA

Abstract

This talk discusses statistical inference dealing with the asymptotic theory of likelihood ratio tests when some parameters may lie on boundary of the parameter space. Following seminal paper by Self and Liang (1987), we derive a closed form solution for the case when one parameter of interest and one nuisance parameter lie on the boundary. The asymptotic distribution is *not* always a mixture of several chi-square distributions. For the cases when one parameter of interest and two nuisance parameters or two parameters of interest and one nuisance parameter are on the boundary, we provide an explicit solution which can be easily computed by simulation.
These results can be used in many applications, e.g. one-sided confidence intervals in environmental risk assessment and testing for random effects in genetics. Contrary to the claim of some authors in the applied literature that use of chi-square distribution with degrees of freedom as in case of interior parameters will be too conservative when some parameters are on the boundary, we show that when nuisance parameters are on the boundary, that approach may often be anti-conservative.

Keywords

Nuisance parameters on the boundary, One-sided tests, Parameters of interest on the boundary.

References

Self, S.G. and Liang, K-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc. 82*, 605–610.

An alternative approach on model selection in Generalized Linear Models

Özlem Korucu and **Kadri Ulaş Akay**

Istanbul University, Turkey

Abstract

Generalized Linear Models are widely used as flexible models in which variability is described by a distribution in an exponential family. The use of Generalized Linear Models in industrial applications has become very common. The purpose of this study is to compare models in several Generalized Linear Model applications. To compare the models, some model selection criteria (AIC, SBIC and ICOMP) and the hierarchy principle are taken into consideration. The suggested approaches are illustrated on different data set in the literature.

Keywords

Generalized Linear Models, ICOMP, Hierarchy principle.

References

Clark, A.E. and Troskie, C.G. (2006). Regression and ICOMPŮA simulation study. *Comm. Statist. Simulation Comput. 35(3)*, 591–603.

Myers, R.H., Montgomery, D.C., Vining, G.G. and Robinson, T.J. (2010). *Generalized Linear Models with Applications in Engineering and the Sciences*, 2nd edition. John Wiley and Sons Inc., USA.

Peixoto, J.L. (1990). A property of well-formulated polynomial regression models. *Amer. Statist. 44(1)*, 26–30

Estimation of market capitalization and economic growth in India

Naresh Kumar

National Institute of Science Technology & Development Studies, India

Abstract

Capital formation is an integral part of economic growth and development and plays an important role in the economic theory of production and distribution. It is assumed that capital accumulation with a positive correlation and additions to the stock of capital can facilitate faster rate of growth. Traditionally growth rate depends upon growth of industrial, agriculture and service sector but stock market is also one of the major sectors for capital formation and has straight impact on the economy across the world. Hence, stock market in developing economies such as India is also growing very fast and it is estimated that Indian stock market is a trillion-dollar industry. Recently the Indian stock market is witnessing heightened activities and is increasingly gaining importance. Therefore, the present study attempts to capture the trends and patterns of market capital, domestic saving, GDP growth in Indian context using growth model. It also investigates the relationship between market capitalisation, saving and GDP growth over the last three decades or so. The results indicate further growth of market capitalisation and positive association between macro indicators.

Keywords

Economic growth, Market capitalisation, Gompertz model, Stock market.

Urban rail transit key equipment fault diagnosis method based on statistical analysis

Xi Li and **Guoqiang Cai**

Beijing Jiaotong University, China

Abstract

Along with rapid construction of urban rail transit in China, "Safe, fast and comfortable" has become development trends and needs of modern urban rail transit. However, in the development process, fault diagnosis and early warning problems to key equipment of urban rail transit vehicle remains a difficult problem. This paper first summarizes and compares the fault diagnosis technologies to different equipments, and then studies fault diagnosis and prediction methods to key equipment of rail transit vehicles with the clustering analysis, support vector machines and other statistical methods. Finally, the paper presents an example analysis to door and walking system of urban rail transit vehicle, the result shows that the presented method is effective for solving a class of key equipment failures and can be widely promoted.

Keywords

Urban rail transit vehicle, Fault diagnosis, Statistical analysis, Support vector machine, Cluster analysis.

References

Favardin, P., Lepelley, D. and Serais, J. (2002). Borda rule, Copeland method and strategic manipulation. *Rev. Econ. Des. 7(2)*, 213–228.

Hsu, D.F. and Taksa, I. (2005). Comparing rank and score combination methods for data fusion in information retrieval. *Inf. Retriev. 8(3)*, 449–480.

Hui, Y. (2009). *Efficient System-Level Fault Diagnosis of Multicomputer Systems[D].* Chongqing University. 2009.4.

Lixin, L. (2008). *Fault Diagnosis Technology Research of the Electrical Systemof SS7E Electric Locomotives[D].* Central South University. 2008.12.

Tingfang, Y. (2008). *Study on New Techniques of Online Monitoring and Fault Diagnosis for Power Transformer[D].* Huazhong University of Science and Technology. 2008.6.

On variance estimators in PLS

Ying Li

Swedish University of Agricultural Sciences, Uppsala, Sweden

Abstract

Univariate partial least squares is a method of modelling relationships between a Y variable and other explanatory variables X. It is an algorithmic approach and in the presentation we are going to use an algorithm present by I.S. Helland. According to Helland (1988), the population PLS predictor is a function of the covariance of (X, Y) and the variance of X. In this study, it is assumed that the covariance is known but the variance of X is unknown. We discuss how to estimate the variance. In particular, by maximizing the likelihood function.

Keywords

PLS, Variance estimation, ML.

Safety state evaluation of urban rail vehicle in transit based on fault diagnosis and early warning

Yu Liang and **Guoqiang Cai**

Beijing Jiaotong University, China

Abstract

urban rail traffic has become one of the most important public transit constitute part in China's (extraordinary) large cites, but the vehicle operation safety situation is not optimistic and traffic accident in transit occurs frequently. City rail vehicle operation safety is related to the life of passengers, vehicle traveling accident has seriously affected the continuous development of urban rail transit. Urban rail vehicles is mainly consisted of bogie, traction system, running control system and supplementary facilities, each part affects the vehicle safety but have different degrees. It monitors the safety state of each part by in transit fault diagnosis and early warning, applies Bayes network and the maximum likelihood estimation theory, comprehensively evaluates the urban rail transit vehicle safety status, gives the evaluation results based on the fuzzy theory, and provides quantity guidance for the safe operating of the vehicle.

Keywords

Fault diagnosis, Bayes network, Maximum likelihood estimation theory, F Safety state evaluation.

WALS estimation and an application to costs of hip fracture treatments

Antti Liski[1], Erkki P. Liski[2], Reijo Sund[3] and Merja Juntunen[3]

[1] Tampere University of Technology, Finland
[2] University of Tampere, Finland
[3] National Institute for Health and Welfare, Finland

Abstract

We present a model for hip fracture treatment costs in Finland using linear regression and estimate the average treatment costs in hospital districts. Our data of 11961 patients aged 50 or over in years 1999-2005 were obtained by combining from several national registers. The focus is to compare treatment costs between hospital districts. Seven largest districts are chosen as focus regressors that are always in the model. There are 38 auxiliary regressors– for example important comorbidities like congestive heart failure, diabetes and cancer– which are intended to reflect the mix of patients treated by a hospital.

It is known that model selection procedures can be unstable. If the inference is conditional on the chosen subset of auxiliary regressors without taking into account model uncertainty, it often means underreporting of variability. In a recent paper Magnus et al. (2010) showed that a weighted average least squares (WALS) estimator – a Bayesian combination of frequentist estimators – has major advantages over standard Bayesian model averaging (MA) estimators. Liski and Liski (2008) compared the performance of weighted average AIC and BIC (Buckland et al. 1997), Rissanen MDL criterion with exponential weights and the Mallows' type MA estimators (Hansen, 2007).

In the current paper we investigate the performance of the WALS estimator versus pretest, maximum likelihood (ML) and the MA estimators studied in Liski and Liski (2008) within a realistic set-up. We use the ML estimator in the unrestricted model as our benchmark. The context of our simulation experiments is provided by the hip fracture cost model. We are mainly interested in the distribution and estimated precision of the focus variables. It turns out that WALS estimation is a serious alternative to traditional pretest estimation, model selection and the MA techniques mentioned above.

Keywords

Model selection, Model averaging, Pretest estimation, Focus regressor, Auxiliary regressor.

References

Buckland, S.T., Burnham, K.P. and Augustin, N.H. (1999). Model selection: an integral part of inference. *Biometrics 53*, 603–618.

Hansen, B.E. (2007). Least squares model averaging. *Econometrica 75*, 1175–1189.

Liski, E.P. and Liski, A. (2008). MDL model averaging for linear regression. In Grünwald, P., Myllymäki, P., Tabus, I., Weinberger, M., Yu, B. (Eds.). *Festschrift in Honor of Jorma Rissanen on the Occasion of his 75th Birthday* (pp. 145–154). Tampere, Tampere Int. Center for Signal Processing.

Magnus, J.R., Powell, O. and Prüfer, P. (2010). A comparison of two model averaging techniques with an application to growth empirics. *J. Econometrics 154*, 139–153.

Supervised invariant coordinate selection

Eero Liski, Klaus Nordhausen and Hannu Oja

University of Tampere, Finland

Abstract

Dimension reduction plays an important role in high dimensional data analysis. Often the interest is on the dependence between a univariate random response variable y and a p-variate random vector x, where the high dimensionality of x may cause problems. One then wishes to reduce the dimension of x without losing any information about the dependence between y and x. *Principal component analysis* (PCA) is perhaps the best known dimension reduction method. It uses one scatter matrix to find the directions of maximum variance. PCA is not, however, a natural approach for a regression situation, because the scatter matrix does not take y into consideration. Therefore, PCA can be referred to as an *unsupervised* dimension reduction method.

Invariant coordinate selection (ICS) is a well known method for exploring multivariate data that uses two scatter matrices to construct an affine transformation to an invariant coordinate system. By choosing only a few directions in the invariant coordinate system based on the most extreme kurtosis values, one can use ICS in dimension reduction. ICS is also an unsupervised dimension reduction method.

Our new *supervised* dimension reduction method, supervised ICS, is based on ICS in such a way that it takes y into consideration. This is done by using a supervised scatter matrix as the second scatter matrix. A supervised scatter matrix depends on the joint distribution of y and x and is equivariant under affine transformations of x. Supervising the second scatter matrix is a major improvement in dimension reduction in a regression situation. Furthermore, the well-known supervised dimension reduction methods *sliced inverse regression* (SIR) and *principal hessian directions* (PHD) can be seen as special cases of this approach.

Keywords

Dimension reduction, Principal component analysis, Invariant coordinate selection, Sliced inverse regression, Principal hessian directions, Supervised invariant coordinate selection.

References

Liski, E., Nordhausen, K. and Oja, H. (2010). Supervised invariant co-ordinate selection. Manuscript under process.

Tyler, D.E., Critchley, F., Dümbgen, L. and Oja, H. (2009). Invariant co-ordinate selection. *J. R. Stat. Soc. Ser. B Stat. Methodol. 71*, 549–592.

Sensitivity analysis of SAR estimators: a simulation study

Shuangze Liu[1], Wolfgang Polasek[2] and Richard Sellner[2]

[1] University of Canberra, Australia
[2] Institute for Advanced Studies, Vienna, Austria

Abstract

Spatial autoregressive models come with a variety of estimators and it is interesting and useful to compare the estimators by location and covariance properties. We study the local sensitivity behavior of the main least squares estimator by using matrix derivatives results of Magnus and Neudecker (1999). Additionally we compare the estimators of the spatial autoregression (SAR) model using the covariance structure of the least squares estimators and we make efficiency comparisons using Kantorovich inequalities. Also, we calculate the Taylor approximation of the least squares estimator in the SAR model up to the second order. Finally, we demonstrate our approach by an example for GDP and employment in 239 European NUTS2 regions. We find a quite good approximation behavior of the SAR estimator in the neighborhood of $\rho = 0$, i.e. a small spatial correlation.

Keywords

Spatial autoregressive models, Least-squares estimators, Taylor approximations, Kantorovich inequality.

References

Magnus, J.R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* (revised edition) Chichester: John Wiley and Sons.

Estimating and designing for mixtures of distributions

Jesús López-Fidalgo, Raúl Martín-Martín and Maria Rodríguez-Hernández

University of Castile-La Mancha, Spain

Abstract

Maximum Likelihood Estimates (MLE) for a model with mixture of distributions is usually an unaffordable task from a computational point of view, even for simple cases when the number of distributions is known. The EM algorithm is frequently used in this context to approximate the MLE. Louis (1982) in a celebrated paper provides the information matrix for the EM ("pure") estimates. The EM algorithm provides approximate MLE, thus the information matrix to be used must be the Fisher information matrix for the marginal log-likelihood of the observations. Pure EM estimates are computed and compared to the MLE. Some comparisons of the two information matrices are also performed. Finally, optimal designs are computed for a mixture of normal distributions with modeled means throughout an explanatory variable.

References

Louis, T.A. (1982). Finding the observed information matrix when using the the EM algorithm. *J. R. Stat. Soc. Ser. B 44(2)*, 226–233.

Optimality of designs under the interference model

Katarzyna Filipiak and Augustyn Markiewicz

Poznań University of Life Sciences, Poland

Abstract

We consider optimality of complete circular block designs under the interference model. We assume that the number of blocks (b) is not proportional to the number of treatments minus one ($t-1$). It is known that for $b \neq p(t-1)$ there cannot exist circular neighbor balanced designs that are known to be universally optimal (Druilhet, 1999). The aim of the paper is to characterize D- and E-optimal designs in the class with $b = t-2$ and $b = t$ as well as universally optimal designs in the class with specific b.

Keywords

Interference model, Complete circular block design, E-optimality, D-optimality, Universal optimality, Information matrix.

References

Druilhet, P. (1999). Optimality of circular neighbor balanced designs. *J. Statist. Plann. Inference 81*, 141–152.

Filipiak, K., Różański, R., Sawikowska, A. and Wojtera-Tyrakowska, D. (2008). On the E-optimality of complete designs under an interference model. *Statist. Probab. Lett. 78*, 2470–2477.

Testing circular symmetry of a covariance matrix – the exact and near-exact distributions for the likelihood ratio test statistic

Filipe J. Marques and **Carlos A. Coelho**

New University of Lisbon, Portugal

Abstract

In this paper the exact distribution of the logarithm of the likelihood ratio test statistic for testing circular symmetry is obtained, for an odd number of variables p, under the form of a Generalized Integer Gamma distribution and for an even p, under the form of an infinite mixture of Generalized Integer Gamma distributions. For the case of an even p near-exact distributions are developed for the likelihood ratio test statistic which correspond to Generalized Near-Integer Gamma distributions or mixtures of these distributions. Numerical studies are conducted in order to assess the quality of these new approximations. Tables of exact quantiles, for odd p, and near-exact quantiles, for even p, as well as plots of the probability density functions and cumulative distributions functions for the likelihood ratio test statistic are presented.

Keywords

Near-exact distributions, Asymptotic distributions, Circular symmetry, Generalized Integer Gamma distribution, Generalized Near-Integer Gamma distribution, Mixtures.

References

Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed., New York: J. Wiley and Sons.

Nagar, D.K., Chen, J. and Gupta, A.K. (2004). Distribution and percentage points of the likelihood ratio statistic for testing circular symmetry. *Comput. Statist. Data Anal. 47*, 79–89.

Olkin, I. and Press, S.J. (1969). Testing and estimation for a circular stationary model. *Ann. Math. Statist. 40*, 1358–1373.

Coelho, C.A. (2004). The Generalized Near-Integer Gamma distribution: a basis for 'near-exact' approximations to the distributions of statistics which are the product of an odd number of independent Beta random variables. *J. Multivariate Anal. 89*, 191–218.

About linear models: A geometric re-visitation

Jean-Pierre Masson[1] **and Tadeusz Caliński**[2]

[1] Agrocampus Ouest, Rennes, France
[2] Poznań University of Life Science,Poland

Abstract

A geometric (or vector space) point of view is adopted to review some rather classic results concerning the estimation of parameters and testing the linear hypotheses in linear models. Only non-singular models are considered. An intrinsic presentation of the Best Linear Unbiased Estimators and of the Linear Hypothesis is given, this being then followed by a parameterization of the model. The relevant vector spaces are mutually related in a duality scheme. This scheme provides a profound insight into the linear nature of the model and gives an efficient way of summarizing the results. The general considerations are then applied to the Multivariate Linear Model, taking into account the tensor nature of it. The squared distance between the expectation of the observable random variables and the linear subspace specified by the linear hypothesis is clarified. It appears to be the Lawley-Hotelling trace criterion in the case of a multivariate linear hypothesis under the classic multivariate linear model.

Keywords

Gauss-Markov theorem, Loewner ordering, Lawley-Hotelling trace criterion, General linear hypothesis, Multivariate linear hypothesis.

References

Baksalary, J.K. and Puntanen, S. (1990) Characterizations of the best linear unbiased estimator in the general Gauss-Markov model with the use of matrix partial orderings. *Linear Algebra Appl. 127*, 363–370.

Caliński, T. and Lejeune, M. (1998). Choix du nombre de fonctions linéaires discriminantes. *Revue de Statistique Appliquée 46*, 31–44.

Lejeune, M. and Caliński, T. (2000). Canonical analysis applied to multivariate analysis of variance. *J. Multivariate Anal. 72*, 100–119.

Rao, C.R., Toutenburg, H., Shalabh and Heumann, C. (2008). *Linear Models and Generalizations: Least Squares and Alternatives.* Springer, Berlin.

Seber, G.A.F. (1984). *Multivariate Observations.* John Wiley Sons, Inc., New York.

Dynamic PCA structure induced autocorrelation

Ana S. Matos, Teresa S. Neves and João T. Mexia

New University of Lisbon, Portugal

Abstract

In the last years, multivariate statistical process control (MSPC) methods, and namely principal component analysis (PCA), have shown to provide a powerful approach to detection and isolation of abnormal conditions, in process industries with highly correlated variables. The use of PCA in MSPC assumes implicitly that the observations at one time are statistically independent of past observations and that the latent variables follow a Gaussian distribution. However, in chemical or biological processes, variables are driven by random noise and uncontrollable disturbances, which may cause process variables to be auto-correlated. Several statistical methods have been developed and applied to overcome the existence of autocorrelation in data. In this article, a comparative study of the performance between the well known Hotelling T^2 control chart, using residuals or one-step-ahead predictions, and the dynamic PCA (DPCA) method based on time lagged measurement vector, proposed by Ku et al. (1995), is presented. The approach developed to compare those charts is described in detail, using the average run length (ARL) as a performance indicator. Monte Carlo experiments are used to simulate three first order autoregressive models (AR(1)), with different autoregressive parameters and different variances, without cross correlation between them. ARL results for a step perturbation introduced in one, two, or all variables simultaneously show that the DPCA chart presents a steady good performance for all shift magnitudes, when compared with Hotelling T^2 chart, with significant reduction in SDRL (standard deviation run length), evidencing a better reliability in the fault detection process. Although this study, with a small dimension of variables, has shown that DPCA method achieves better results on detecting and monitoring disturbances in dynamic processes (recommended by other authors like Lee et al. (2004) and Chen and Liu (2002) as an example), the dynamic PCA cannot eliminate data autocorrelation, regardless of the time lag considered. As a consequence of this evidence, the influence of the dynamic PCA structure on data is investigated using different time series (AR, IMA, ARMA) and also random noise series. Time lag is also considered as a variable in this extension of the study. Through Monte Carlo simulation is proven that dynamic PCA structures induce dynamics into the score variables, even if all the variables correspond to random noise series. The main advantages and disadvantages of each chart

are pointed out, in a practical perspective of those who intent to use MSPC to monitoring dynamic continuous processes with a small number of variables to be controlled. Simulated and real data are used for illustration.

Keywords

Multivariate statistical process control (MSPC), Hotelling T^2 control chart, Dynamic principal component analysis (DPCA), Average Run Length (ARL).

References

Chen, J. and Liu, K.-C. (2002). On-line batch process monitoring using dynamic PCA and dynamic PLS models. *Chemical Engineering Science 57*, 63-75.

Lee, J.-M., Yoo, C. and Lee, I.-B. (2004). Statistical monitoring of dynamic processes based on dynamic independent component analysis. *Chemical Engineering Science 59*, 2995–3006.

Ku, W., Storer, R.H. and Georgakis, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems 30*, 179-196.

Optimal design for functional magnetic resonance imaging experiments based on linear models

Baerbel Maus, Gerard J. P. van Breukelen, Rainer Goebel and Martijn P. F. Berger

Maastricht University, The Netherlands

Abstract

Functional magnetic resonance imaging is a neuroimaging method which is used to study the human brain and its functional areas. In the first part of this presentation it will be shown how the general linear model is used to describe experimental functional magnetic resonance imaging (fMRI) data from one subject (Friston, Holmes, Poline, Grasby, Williams, Frackowiak and Turner, 1995). Based on the general linear model, optimal designs for one-subject fMRI experiments can be obtained by application of the D- and A-optimality criterion (Atkinson, Donev and Tobias, 2006; Maus, van Breukelen, Goebel and Berger, 2010). Because of the huge design space for fMRI experiments, a genetic algorithm (GA) is employed to find optimal designs for fMRI experiments based on a multi-objective design criterion (Kao, Mandal, Lazar and Stufken, 2009; Maus, van Breukelen, Goebel and Berger, 2010; Wager and Nichols, 2003).
The second part of the presentation will focus on the application of mixed effects models in analysis of fMRI experiments from multiple subjects (Holmes and Friston, 1998). Optimal designs for multi-subject experiments are considered and the optimal combination of number of subjects and fMRI scanner time/imaging time per subject will be studied with respect to a linear cost function.

Keywords

Optimal design of experiments, fMRI, General linear model, Mixed effects model.

References

Atkinson, A.C., Donev, A.N. and Tobias, R.D. (2007). *Optimum experimental designs, with SAS.* New York: Oxford University Press.

Friston, K.J., Holmes, A.P., Poline, J.-B., Grasby, P.J., Williams, S.C.R., Frąckowiak, R.S.J. and Turner, R. (1995). Analysis of fMRI time series - revisited. *NeuroImage 2*, 45–53.

Holmes, A.P. and Friston, K.J. (1998). Generalisability, random effects and population inference. *NeuroImage 7*, S754.

Kao, M.-H., Mandal, A., Lazar, N. and Stufken, J. (2009). Multi-objective optimal experimental designs for event-related fMRI studies. *NeuroImage 44*, 849–856.

Maus, B., van Breukelen, G.J.P., Goebel, R. and Berger, M.P.F. (2010). Robustness of optimal design of fMRI experiments with application of a genetic algorithm. *NeuroImage 49*, 2433–2443.

Wager, T.D. and Nichols, T.E (2003). Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *NeuroImage 18*, 293–309.

An adaptive sequential design for model discrimination and parameter estimation in non-linear nested models

Caterina May[1] and Chiara Tommasi[2]

[1] University of Eastern Piedmont, Italy
[2] University of Milano, Italy

Abstract

In this work, an adaptive sequential design is computed with the goal of both discriminating and estimating more than two nested non-linear models. At each step of the sequential procedure, an optimum design for this double aim is computed through the DKL-optimality criterion. The procedure is inspired to the sequential method of Biswas and Chaudhuri (2002), which can be applied only in the set up of nested linear models.

Keywords

Adaptive sequential design, Optimum designs of experiments, DKL-optimality, Consistency of selection.

References

Biswas, A. and Chaudhuri, P. (2002). An efficient design for model discrimination and parameter estimation in linear models. *Biometrika 89*, 709–718.

Tommasi, C. (2009). Optimal designs for both model discrimination and parameter estimation. *J. Statist. Plann. Inference 139*, 4123–4132.

Analysing genotype by environment interaction by curvilinear regression

Iwona Mejza[1], **Stanisław Mejza**[1], **João T. Mexia**[2], **Dulce G. Pereira**[3] **and Paulo C. Rodrigues**[2]

[1] Poznań University of Life Sciences, Poland
[2] New University of Lisbon, Portugal
[3] University of Évora, Portugal

Abstract

This paper deals with the analysis of the Genotype by Environment Interaction (GEI). The purpose of such analysis is to select genotypes that are consistently high-yielding over the range of observed or potential environments. Usually, the GEI is non-orthogonal. In the literature there are papers dealing with many statistical techniques in this area. In this paper, regression analysis is used to make inferences about GEI.
In regression analysis we should have two sets of variables, the first characterizing genotypes, and the second characterizing environments. The so-called adjusted means (or some other genotype characteristics) for genotypes usually constitute observations of the dependent variable. The problem is how to model the environmental indexes, these being the observation of independent variable. In the paper we use the environmental indexes obtained by an iterative ("zig-zag") algorithm based on the joint regression approach.
The data considered here is usually used to: (i) to predict and estimate the yield; and (ii) to provide reliable guidance for selection of the best genotypes for planting over environments. An inference concerning these problems is based on adapting two tests: (A) a test for parallelism of regression curves; and (B) a test of coincidence (regression functions are identical).
The first step should be the estimation of the regression functions (linear or curvilinear) independently for all genotypes. We recommend beginning searching for an optimal response curve in the class of nonlinear functions. Since GEI is usually present in the data, the test (A) will probably reject the respective hypothesis. Using this test repeatedly for subsets of genotypes we can find the genotypes with similar responses to environmental conditions. This group is worth considering with respect to a breeding program and with respect to use in practice. In case of failure to reject the hypothesis, it is worth applying test (B).

Keywords

Genotype × environment interaction, Regression analysis, Environmental index.

On the methodology of optimal design for nonlinear models based on the functional approach

Viatcheslav B. Melas

University of St. Petersburg, Russia

Abstract

In this paper optimal designs for nonlinear regression models are investigated on the base of the functional approach. The idea of the approach consists in considering support points and weights of optimal designs as implicit functions of some auxiliary parameters. Under certain conditions to be discussed in the paper these functions can be represented by Taylor series and the coefficients of these series can be computed using recurrent formulas. In the recent book (Melas, 2006) this approach was implemented for locally optimal and maximin efficient designs. Here we will extend the approach to L- and D- optimal Bayesian designs. It allows constructing and studying all basic types of optimal designs for nonlinear models using the same methodology. Theoretical and numerical results for several specific models of rational type will be presented.

Asymptotic efficiencies of the Greenwood's goodness-of-fit test

Sherzod Mirakhmedov and Naeem Muhammad

GIK Institute of Engineering Sciences and Technology, Topi, Pakistan

Abstract

We study the Greenwood statistic, which is sum of squares of the disjoint s-spacings. We extends the known efficiency properties of Greenwood test: (i) in the Bahadur's situation of a fixed alternative to "adjoining" domain of family of strong intermediate alternatives, and (ii) in the Pitman's situation of alternatives convergence to null hypothesis with "extremely" rate to "adjoining" domain of family of intermediate alternatives, which converges to null hypothesis with "not so fast" rate.

Keywords

Goodness-of-fit, Spacings, Asymptotic efficiency, Large deviation probabilities.

Flexible sampling of semi-selfsimilar Markov processes: covariance and spectrum

Navideh Modarresi and **Saeid Rezakhah**

Amirkabir University of Technology, Tehran, Iran

Abstract

In this paper we consider some flexible discrete sampling of a semi-selfsimilar process $\{X(t), t \in \mathbf{R}^+\}$ with scale $l > 1$. By this method we plan to have q samples at arbitrary points $\mathbf{s}_0, \mathbf{s}_1, \ldots, \mathbf{s}_{q-1}$ in interval $[1, l)$ and proceed our sampling in the intervals $[l^n, l^{n+1})$ at points $l^n\mathbf{s}_0, l^n\mathbf{s}_1, \ldots, l^n\mathbf{s}_{q-1}$, $n \in \mathbf{Z}$. Thus we have a discrete time semi selfsimilar process and introduce an embedded discrete time semi self similar process as $W(nq+k) = X(l^n\mathbf{s}_k)$, $q \in \mathbf{N}$, $k = 0, \ldots, q-1$. We also consider $V(n) = \big(V^0(n), \ldots, V^{q-1}(n)\big)$ where $V^k(n) = W(nq+k)$, as an embedded q-dimensional discrete time selfsimilar (DT-SS) process. By introducing quasi Lamperti transformation, we find spectral representation of such process and its spectral density matrix is given. Finally by imposing wide sense Markov property for $W(\cdot)$ and $V(\cdot)$, we show that the spectral density matrix of $V(\cdot)$ and spectral density function of $W(\cdot)$ can be characterized by $\{R_j(1), R_j(0), j = 0, \ldots, q-1\}$ where $R_j(k) = E[W(j+k)W(j)]$.

Keywords

Semi-selfsimilar process, Wide sense Markov, Multi-dimensional selfsimilar processes.

References

Loève, M. (1963). *Probability Theory*, 3rd edition. Van Nostrand: New York.

Modarresi, N. and Rezakhah, S. (2009). Discrete time scale invariant Markov processes. *http://arxiv.org/pdf/0905.3959v3*, 1–12.

Rozanov, Y.A. (1967). *Stationary Random Processes.* Holden-Day: San Francisco.

Brownian motion with drift and regimes

Pedro P. Mota and **Manuel L. Esquível**

New University of Lisbon, Portugal

Abstract

In the time series the study of nonlinear models is increasing and a class of nonlinear models, called threshold models can be found in Tong (1990). Our goal, as in Mota (2008), is to extend the notion of threshold processes to continuous time models and obtain estimation methods for this kind of processes. A diffusion which experiences a regime change upon crossing upper (M) and lower (m) levels will be our generic model for the stochastic process.

Keywords

Regime changes, Consistency.

References

Mota, P. (2008). *Brownian motion with drift threshold model.* Ph.D Thesis – New University of Lisbon, Portugal.

Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach.* Oxford University Press.

On a continuous time stock price model with regime switching, delay and threshold

Pedro P. Mota and Manuel L. Esquível

New University of Lisbon, Portugal

Abstract

Motivated by the need to describe bear-bull market regime switching in stock prices, we introduce and study a stochastic process in continuous time with two regimes, threshold and delay, given by a stochastic differential equation. When the difference between the regimes is simply given by different set of real valued parameters for the drift and diffusion coefficients, changes between regimes depending only on these parameters, we show that if the delay is known there are consistent estimators for the threshold as long we know how to classify a given observation of the process as belonging to one of the two regimes. When the drift and diffusion coefficients are of geometric Brownian motion type we obtain a model with parameters that can be estimated in a satisfactory way, a model that allows to differentiate regimes in some of the NYSE 21 stocks analyzed and also, that gives very satisfactory results when compared to the usual Black-Scholes model for pricing call options.

Keywords

Ergodic diffusions, Transition and invariant densities, Maximum likelihood estimators.

References

Chan, K.S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist. 2(1)*, 520–533.

Chan, K.S. and Tsay, R.S. (1998). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika 85(2)*, 413–426.

Freidlin, M. and Pfeiffer, R. (1998). A threshold estimation problem for processes with hysteresis. *Finance Stoch. 36*, 337–347.

Hansen, A.T. and Poulsen, R. (2000). A simple regime switching term structure model. *Statist. Probab. Lett. 4(4)*, 409–429.

Mota, P.P. (2008). *Brownian motion with drift threshold model.* Ph.D Thesis – New University of Lisbon, Portugal.

Petrucelli, J.D. (1986). On the consistency of least squares estimators for a threshold AR(1) model. *J. Time Ser. Anal. 7(4)*, 269–278.

Informative cluster size problems

Jaakko Nevalainen[1], **Somnath Datta**[2] **and Hannu Oja**[3]

[1] University of Turku, Finland
[2] University of Louisville, USA
[3] University of Tampere, Finland

Abstract

In spite of many recent contributions to the literature informative (or non-ignorable) cluster size settings are still not well known and understood. This is partly explained by the development of the theory to specific problems and mainly directly in the framework of modeling. The aim of this talk is to serve as a tutorial to these problems. We define the problem, describe it from different viewpoints in light of examples from the literature, and discuss different data generating mechanisms. We show how the classical statistical procedures (t-test, correlation, nonparametric location functionals,...) need to be modified in order to draw right conclusions efficiently from data with informative cluster size

Keywords

Clustered data, Informative cluster size.

Control of the truncation errors for generalized *F* distributions

Célia Nunes[1], **Dário Ferreira**[1], **Sandra Ferreira**[1] **and João T. Mexia**[2]

[1] University of Beira Interior, Covilhã, Portugal
[2] New University of Lisbon, Portugal

Abstract

F tests may not be used for all relevant hypothesis, even in rather simple models, which led to the introduction of generalized F tests, see Michalski and Zmyślony (1996, 1999).
The statistics of these tests are quotients of linear combinations of independent chi-squares, which may be non-central. Results on these distributions were obtained first for the central case, in Fonseca et al. (2002), and then for the non-central case, in Nunes and Mexia (2006). When the observations were collected under non standardized conditions the non-centrality parameters may be random. The case in which the non-centrality parameters have Gamma distributions is singled out, see Nunes et al. (2009).
The generalized F distribution are given by infinite sums. In this paper we show that there is an excellent control of the truncations errors for those sums.

Keywords

Generalized F distributions, Random non-centrality parameters, Truncation errors.

References

Fonseca, M., Mexia, J.T. and Zmyślony, R. (2002). Exact distribution for the generalized F tests. *Discuss. Math. Probab. Stat. 22*, 37–51.

Michalski, A. and Zmyślony, R. (1996). Testing hypothesis for variance components in mixed linear models. *Statistics 27*, 297–310.

Michalski, A. and Zmyślony, R. (1999). Testing hypothesis for linear functions of parameters in mixed linear models. *Tatra Mt. Math. Publ. 17*, 103–110.

Nunes, C. and Mexia, J.T. (2006). Non-central generalized F distributions. *Discuss. Math. Probab. Stat. 26(1)*, 47–61.

Nunes, C., Ferreira, S. and Ferreira, D. (2009). Generalized F tests in models with perturbations: The Gamma case. *Discuss. Math. Probab. Stat. 29(2)*, 185–197.

More on the Kronecker structured covariance matrix

Martin Ohlson[1] **and Dietrich von Rosen**[1,2]

Linköping University, Sweden
Swedish University of Agricultural Sciences, Uppsala, Sweden

Abstract

The Kronecker structured covariance matrix in multivariate normal distribution will be studied. Particularly, the mapping and parametrization which are induced by the Kronecker product are considered.
Furthermore, estimation and the uniqueness of the estimators will be discussed in the case of a covariance matrix which is a Kronecker product of several matrices.

Keywords

Kronecker product structure, Separable covariance, Maximum likelihood estimators.

References

Dutilleul, P. (1999). The MLE algorithm for the matrix normal distribution. *J. Stat. Comput. Simul. 64(2)*, 105–123.

Lu, N. and Zimmerman, D.L. (2005). The likelihood ratio test for a separable covariance matrix. *Statist. Probab. Lett. 73*, 449–457.

Naik, D.N. and Rao, S. (2001). Analysis of multivariate repeated measures data with a Kronecker product structured covariance matrix. *J. Appl. Stat. 28*, 91–105.

Roy, A. and Khattree, R. (2005). On implementation of a test for Kronecker product covariance structure for multivariate repeated measures data. *Stat. Methodol. 2*, 297–306.

Srivastava , M.S., Nahtman, T. and von Rosen, D. (2008). Models with a Kronecker product covariance structure: Estimation and testing. *Math. Meth. Statist. 17*, 357–370.

Srivastava, M.S., von Rosen, T. and von Rosen, D. (2009). Estimation in general multivariate linear models with Kronecker product covariance structure. *Sankyā*. Accepted.

Regression methods for multiple outcomes

Rosa Oliveira and **Armando Teixeira-Pinto**

University of Porto, Portugal

Abstract

In research problems, particularly in Health research, it is common to have multiple outcomes of interest in the same study. The usual approach is to analyse each outcome separately ignoring the correlation between the outcomes. However this approach does not consider the multidimensional structure of the data and may lead to inefficient estimators.

In seemingly unrelated linear regressions (SUR) context, Zellner (1962) showed that if outcomes are associated with the same set of covariates, then the maximum likelihood estimator for the regression parameters is the ordinary least squares (OLS). That is, if the outcomes are modeled with the same covariates, the multivariate linear regression and univariate regressions for the different outcomes produce exactly the same estimators. However, if the outcomes are associated with different covariates, this end result no longer applies. Namely, ignoring the correlation and fitting separate regressions originates inefficient estimates.

In this work, we study a mix setting where the outcomes share some covariates but are also associated with specific covariates. We demonstrate that for the coefficients associated with shared covariates there are efficiency gains, while for the outcome-specific covariates the efficiency gains depend on the correlation between the outcomes. Additionally, we use Monte Carlo simulations to evaluate the performance of both approaches and provide a real data example.

Keywords

Statistics, Multivariate analysis, Data analysis, Applications to biology and medical sciences.

References

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. *J. Amer. Statist. Assoc. 57*, 348–368.

Zellner, A. and Huang, D.S. (1962). Further properties of efficient estimators for seemingly unrelated regression equations. *Internat. Econom. Rev. 3(3)*, 300–313.

Searching for differential expression in microarray analysis: comparison of two nonparametric approaches

Israel Ortega[1], M.C. Ruiz de Villa[2] and Antonio Miñarro[2]

[1] Fundació Institut de Recerca del Hospital Universitario Vall d'Hebron, Barcelona, Spain
[2] University of Barcelona, Spain

Abstract

This work studies the performance of a nonparametric density estimation based method to detect differentially expressed genes in microarray experiments.
The method makes use of a density estimation by orthogonal polynomials and an stepwise algorithm to select terms based on likelihood increments. Density estimations obtained trough this methodology allow us to define a natural distance between genes.
Given the expression of a set of genes from two experimental conditions we obtain density estimates of a Student-type test statistics (f) and the corresponding density under the null hypothesis of non differential expression (f_0).
Deciding whether a gene is differentially expressed is taken based on the distance between that gene and a "representative gene" of those non differentially expressed.
In order to evaluate the performance of this method we have simulate data under several conditions and we have compared the results obtained, in term of power and false discovery rate, with those obtained using the normal mixture model of Wei Pan.
Results show that this method performs well in a wide set of conditions.

Keywords

Density estimation, Microarray analysis, Differential expression.

References

Miñarro, A. and Oller, J.M. (1992). Some remarks on the individuals-score distance and its applications to statistical inference. *Questiió 16*, 43–57.

Miñarro, A. and Oller, J.M. (1993). On a class of probability density functions and their information metric. *Sankhyá Series A 55*, 214–225.

Pan, W., Lin, J. and Le, C.T. (2003). A mixture model approach to detecting differentially expressed genes with microarray data. *Funct. Integr. Genomics 3*, 117–124.

Zhao, Y. and Pan, W. (2003). Modified nonparametric approaches to detecting differentially expressed genes in replicated microarray experiments. *Bioinformatics 19(9)*, 1046–1054.

Growth rates of rice through non-linear models

Sanjeev Panwar[1] **and Anil Kumar**[2]

[1] Indian Council of Agriculture Research, New Delhi, India
[2] Indian Agricultural Statistics Research Institute, New Delhi, India

Abstract

The god *Shiva* called rice *Vrihi*, in Sanskrit. India is one of the original centres of rice cultivation. The rice harvesting area in India is the world's largest. Indian rice cultivation is found in all states, but West Bengal, Uttar Pradesh, Madhya Pradesh, Orissa and Bihar are the major producing states. Rice is the staple food for 65% of the total population in India. The Indian population was about 1 billion people in 2000 and is still growing at a high rate (1.7% per year). Although the country exports several varieties of rice, many scientists have expressed concern that current Indian rice production techniques cannot sustain the growing domestic population. India has a large number of rice dishes and many of them are very simple to prepare. Indian pilaf rice is very flavourful and fluffy when cooked with Basmati rice.
The usual parametric approach for growth rate analysis is to assume multiplicative error in the underlying nonlinear geometric model and then fit the linearized model by "method of least squares". This paper deals with a critical study of rice yield of India and states-wise with a non-linear approach. The available data of rice during different years is taken into consideration and different statistical models are fitted for that. The time series data on annual production of Rice in India from 1962-2006 were collected from various sources. Growth rates are computed through non-linear models viz., Logistic, Gompertz and Monomolecular models and all three nonlinear models are suitable to fit.

Sparse inverse covariance estimation in the supervised classification of high-dimensional data

Tatjana Pavlenko and **Anders Björkström**

Stockholm University, Sweden

Abstract

The performance accuracy of the sample based supervised classifiers is known to be poor in a high-dimensional setting, i.e. when the data dimension p is comparable to or larger then the sample size n. To overcome this problem, we suggest to construct a sparse estimate of the class inverse covariance matrix, which in a Gaussian case can be obtained as a minimizer of its negative log-likelihood, subject to a Lasso-type penalty on its off-diagonal elements. Our procedure consists of two-stages; we first estimate the sparsity patterns in the class inverse covariance matrix and then form its block-diagonal approximation using Cuthill-McKee reordering algorithm. We then incorporate this technique in the supervised classification framework and investigate the effect of the suggested approximation on the classification accuracy in the growing dimension asymptotics, i.e. when both p and n are allowed to grow. Further, we show that our approach allows for substantial dimensionality reduction while maintaining the misclassification probability at a certain desired level.

Keywords

High-dimensional data, Sparse covariance structure, Lasso, Supervised classification, Misclassification probability.

References

Bühlmann, P. and Rütimann, P. (2009). High dimensional sparse covariance estimation via directed acyclic graphs. *Electronic J. Statist. 3*, 1133–1160.

Friedman, J., Hastie, T. and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical Lasso. *Biostatistics 9(3)*, 432–441.

Hall, P., Pittelkow, Y. and Ghosh, M. (2008). Theoretical measures of relative performance of classifiers for high dimensional data with small sample sizes. *J. R. Stat. Soc. Ser. B Stat. Methodol. 70*, 159–173.

The environmental indexes in a Joint Regression Analysis and their meaning

Dulce G. Pereira[1], **Paulo C. Rodrigues**[2,3], **Stanisław Mejza**[4] and **João T. Mexia**[2]

[1] University of Évora, Portugal
[2] New University of Lisbon, Portugal
[3] Wageningen University and Research Centre, The Netherlands
[4] Poznań University of Life Sciences, Poland

Abstract

The phenotype of an individual is determined by the genotype and the environment in which it is embedded. Plant breeders, farmers and agronomists aim to determine a superior genotype over a wide range of environmental conditions, but also over time (multi-location trials). The basic cause of differences between genotypes and stability of its production in different environments is due to these two effects that are not simply additive, i.e., data show interaction between genotype and environment (GEI) (Kang and Gauch, 1996). The data of multi-location trials have three main objectives: (i) accurately predict and estimate the production based on limited experimental data, (ii) determine the stability of productivity and the standard response of genotypes in all environments, and (iii) to provide reliable guidance for the selection of the best genotypes or treatments for planting in subsequent years, and new locations. All of those activities lead to structuring GEI. From practical point of view the GEI usually exists (it is significant in test sense) and its structuring and understanding is the most important point of inference. Many statistical techniques have been adapted and used in an attempt to describe and analyze the GEI. The Joint Regression Analysis (JRA) (Finlay and Wilkinson, 1968; Pereira and Mexia, 2008) is one of the mostly used because of its graphical capability and theoretical evidence of its usefulness. One of the biggest open problems on JRA, in which there is still no consensus, is how to obtain and interpret the environmental indexes (Pereira and Mexia, 2009).
Besides of selecting the "best" genotypes (more productive, more riches in nutrients and more resistant to diseases) for various environments, it is also in the scope of this paper to identify a physical meaning for the environmental indexes. A case study with Spring Barley (*Hordeum vulgare* L.) and Winter Wheat (*Triticum aestivum* L.) from experiments in Czech Republic will be presented.

Keywords

Joint regression analysis, Environmental indexes, Genotype by environment interaction, Spring barley, Winter wheat.

References

Finlay, K.W. and Wilkinson, G.N. (1968). The analysis of adaptation in a plant breeding programme. *Aust. J. Agr. Res. 14*, 742–754.

Kang, M.S. and Gauch, H.G. (1996). *Genotype by Environmental Interaction.* New York: CRC Press.

Pereira, D.G. and Mexia, J.T. (2008). Selection proposal of spring barley in the years from 2001 to 2004, using Joint Regression Analysis. *Plant Breeding 127*, 452–458.

Pereira, D.G. and Mexia, J.T. (2009). Comparing double minimization and zigzag algorithms in Joint Regression Analysis: the complete case. *J. Stat. Comput. Simul. 79*, 1563–5163.

Approximations of minimum risk regression estimator

Jan Picek[1] **and Jana Jurečková**[2]

[1] Technical University of Liberec, Czech Republic
[2] Charles University in Prague, Czech Republic

Abstract

The minimum risk equivariant estimator (MRE) of the regression parameter vector in the linear regression model enjoys the finite-sample optimality, but its calculation is difficult, with an exception of few special cases. We study some possible approximations of MRE, with distribution of the errors being known or unknown: A finite-sample approximation uses the Hájek–Hoeffding projection or the Hoeffding-van Zwet decomposition of an initial equivariant estimator, a large-sample approximation is based on the asymptotic representation of the same. We illustrate the finite sample behavior of the proposed approximation on simulated data.

Keywords

Asymptotic representation, Hájek–Hoeffding projection, Hoeffding–van Zwet decomposition maximal invariant, Minimum risk equivariant estimator.

References

Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist. 39*, 325–346.

Hoeffding, W. (1961). *The Strong Law of Large Numbers for U-statistics.* Inst. of Statistics Mimeo Series No 302, University of North Carolina, Chapel Hill.

Jurečková, J. and Milhaud, X. (1999). Characterization of distributions in invariant models. *J. Statist. Plann. Inference 75*, 353–361.

Jurečková, J. and Picek, J. (2009). Minimum risk equivariant estimator in linear regression model. *Statist. Decisions 27*, 37–54.

Parameter estimators for a bidimensional Ornstein-Uhlenbeck process with singular diffusion

Ana Filipa Prior[1,2], **Paula Milheiro de Oliveira**[2] **and Teresa Arede**[2]

[1] High Institute of Engineering of Lisbon, Portugal
[2] University of Porto, Portugal

Abstract

The purpose of this study is to investigate the properties of an estimator of the drift parameter of a bidimensional Ornstein-Uhlenbeck process in the special case when the diffusion matrix is singular. This particular process appears in the engineering literature as a model for mechanical systems subjected to random vibrations. Usually in the literature this estimator and its properties are obtained under some regularity conditions. In particular the diffusion matrix may not be a singular matrix.
For the stochastic differential equation $dX_t = AX_t dt + B^{\frac{1}{2}} dW_t$, where $A = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & -\frac{c}{m} \end{bmatrix}$, $B^{\frac{1}{2}} = \begin{bmatrix} 0 & 0 \\ \sigma & 0 \end{bmatrix}$ and W is a standard Wiener process, the maximum likelihood estimator of the drift parameter is obtained in this work. The stationarity and ergodicity of the process is studied and upon this properties, consistency and bias of the estimator are deduced.

Keywords

Ornstein-Uhlenbeck process, Stochastic differential equations, Maximum-likelihood estimator.

References

Arato, M.(1982). Linear stochastic systems with constant coefficients. A statistical approach. In Balakrishnan, A.V., Thoma, M. (Eds.) *Lectures Notes in Control and Information Sciences, 45*. Springer-Verlag, Berlin.

Khasminskii, R.Z., Krylov, N. and Moshchuk, N. (1999). On the estimation of parameters for linear stochastic differential equations. *Probab. Theory Related Fields 113*, 443–472.

Prakasa Rao, B.L.S. (1999). *Statistical Inference for Diffusion Type Processes.* Kendallt́s Library of Statistics 8. Wiley.

Sample partitioning estimation for ergodic diffusions. Applications

Luís Ramos, Pedro P. Mota and João T. Mexia

New University of Lisbon, Portugal

Abstract

When a diffusion is ergodic its transition density converges to its invariant density, see Durrett (1996). This convergence enabled us to introduce a sample partitioning technique that gives in each sub-sample, maximum likelihood estimators. The averages of these being a natural choice as estimators. Application of that technique for a few diffusion are given.

Keywords

Ergodic diffusions, Transition and invariant densities, Maximum likelihood estimators.

References

Bibby, B.M. and Sørensen, M. (1995). Martingale estimating. Functions for discretely observed diffusion processes. *Bernoulli 1*, 17–39.

Durrett, R. (1996). *Stochastic Calculus: A practical Introduction.* Boca Raton, CRC Press.

Küchler, U. and Sørensen, M. (1997). *Exponential Families of Stochastic Processes.* Springer-Verlag.

Iacus, S. (2008). *Simulation and Inference for Stochastic Differential Equations with R Examples.* Springer.

Øksendal, B. (1998). *Stochastic Differential Equations. An Introduction*, 5th edition. Springer-Verlag.

Sørensen, M. (1997). *Statistical Inference for Discretely Observed Diffusions.* Lecture Notes. Berlin Graduiertenkolleg.

Testing the significance of coefficients in the linear model. The case of the trend in a AR(1) time series

M. Rosário Ramos[1,2] and Marco Costa[2,3]

[1] Universidade Aberta, Lisbon, Portugal
[2] University of Lisbon, Portugal
[3] Aveiro Universtiy, Portugal

Abstract

The aim of this work is to study some tests to the regression coefficients of a model with a linear trend. It is assumed that the error term follows an AR(1), and the autoregressive parameter is unknown. The focus is on the test for one slope and on the test which compares slopes of two or more periods in the same time series. For the case of one slope tests under study are based on the Ordinary Least Squares estimators and in a nonparametric counterpart. For the second case of two slopes the study presents the matrix approach of the model and explores the behaviour of the parametric test. The autoregressive parameter is obtained through some competing estimators. The accuracy of the estimation of this parameter is also analysed. The performance of the tests is compared through a simulation study under different assumptions.
Finally, the methods are applied to a pair of series of physico-chemical variables collected within a framework of a monitoring program of water quality.

Keywords

Trend tests, Ordinary least squares, Autocorrelation, Nonparametric tests, Environmental variables.

References

Alpuim, T. and El-Shaarawi, A. (2008). On the efficiency of regression analysis with AR(p) Errors. *J. of Applied Statistics 35*, 717–737.

Costa, M. and Alpuim, T. (2010). Parameter estimation of state space models for univariate observations. *J. Statist. Plann. Inference*, doi:10.1016/j.jspi.2010.01036.

Ramos, M.R. (2006). Trend tests with application to water quality assessment: Comparison between parametric and non parametric methods. PhD Thesis (in portuguese).

Yue, S. and Wang, C.Y. (2004). The Mann-Kendall test modified by Effective Sample Size to detect trend in serially correlated hydrological series. *Water Resour. Manage. 18*, 201–218.

Interaction in mixed models

Paulo Ramos[1], **Célia Fernandes**[1] **and João T. Mexia**[2]

[1] Lisbon Superior Engineering Institute, Portugal, Portugal
[2] New University of Lisbon, Portugal

Abstract

In this work we use binary operations on commutative Jordan algebras, CJA, to study interactions between factors that nest. From two CJA we can, through these binary operations, build CJA which constitute Boolean Algebras. So when we nest the treatments from one model in each treatment of another model, we can study which sets of factors from the nested model for which we consider interactions with the factors from the previous model.

Keywords

Binary operations, Boolean algebras, Commutative Jordan algebras, Interaction, Linear models.

References

Almeida e Costa, A. (1968). *Cours d'algèbre générale, vol. 1.* Fundação Calouste Gulbenkian.

Federer, W. and King, F. (2007). *Variations on Split Plot and Split Block Experiment Designs.* John Wiley and Sons.

Fonseca, M., Mexia, J.T. and Zmyślony, R. (2006). Binary operations on Jordan algebras and orthogonal normal models. *Linear Algebra Appl. 417*, 75–86.

Jordan, P., von Neumann, J. and Wigner, E. (1934). On a algebraic generalization of the quantum mechanical formulation. *Ann. of Math. (2), 35(1)*, 29–64.

Khuri, A., Mathew, T. and Sinha, B. (1998). *Statistical Tests for Mixed Linear Models.* New York: John Wiley and Sons.

Malley, J. (2004). *Statistical Applications of Jordan Algebras.* Springer-Verlag.

Montgomery, D. (2004). *Design and Analysis of Experiments, 6th edition.* John Wiley and Sons.

Seely, J. (1971). Quadratic subspaces and completeness. *Ann. Math. Statist. 42(2)*, 710–721.

Vanleeuwen, D., Seely, J. and Birkes, D. (1998). Sufficient conditions for orthogonal designs in mixed linear models. *J. Statist. Plann. Inference 73*, 373–389.

Vanleeuwen, D., Birkes, D. and Seely, J. (1999). Balance and orthogonality in designs for mixed classification models. *Ann. Statist. 27(6)*, 1927–1947.

Testing autoregressive nonnested models estimated by IV

Efigénio Rebelo and Patrícia Oom do Valle

University of Algarve, Portugal

Abstract

Based on the Gauss-Newton regression, firstly advocated by Davidson and MacKinnon, this study proposes a methodology to test nonnested dynamic regression models, when autocorrelation is present in the corresponding disturbances of both models. The proposed methodology also considers the presence of current endogenous regressors in both models, which implies estimation by nonlinear instrumental variables. In particular, P and PA versions are now developed to autoregressive models of any order. We also show that the well-known results on nonnested spherical models are encompassed by these more general results.

Keywords

Nonnested tests, P and PA tests, Nonlinear regression function, Autocorrelation, Gauss-Newton regression, Nonlinear instrumental variables.

Multivariate methods for genomic data integration and visualization

Ferran Reverter[1], José Fernández-Real[2], Esteban Vegas[1], Francesc Carmona[1], Jacques Amar[3], Remy Burcelin[3], Eduardo García Fuentes[4], Matteo Serino[3], Francisco Tinahones[4] and Alex Sánchez-Pla[1]

[1] University of Barcelona, Spain
[2] Girona Biomedical Research Institute, Spain
[3] Institut de Medecine Moleculaire de Rangueil, Toulouse, France
[4] Hospital Interuniversitario Virgen de Victoria, Malaga, Spain

Abstract

As the developments in high throughput technologies have become more common and more accessible it is becoming usual –and affordable– to take different simultaneous approaches to study the same problem. In practice this means that different sets of data of different types (expression, proteins, metabolites...) may be available or generated for the same study, highlighting the need for methods and tools to use them in a combined way.
In recent years there have been developed many methods that integrate the analysis of different types of data. Corresponding to a certain tradition in bioinformatics many methodologies are rooted in machine learning tools such as bayesian networks, support vector machines or graph-based methods. In contrast with the high number of applications from these fields, another that seems to have contributed less to genomic data integration is multivariate statistics, which has however a long tradition in being used to combine and visualize multidimensional data. In this work we discuss the application of multivariate statistical approaches to integrate bio-molecular information by combining several multivariate statistical approaches such as principal components analysis, simple and multiple correspondence analysis and canonical correlation analysis and its variants. The techniques are applied to a real unpublished data set consisting of four different data types: DGGE bands, expression microarrays, high-throughput sequence data and clinical variables. We show how these statistical techniques can be used to perform reduction dimension and then visualize data of one type useful to explain those from other types. Whereas this is more or less straightforward when we deal with two types of data it turns to be more complicated when the goal is to visualize simultaneously more than two types. Comparison between the approaches shows that the information they provide is complementary suggesting their combined use yields more information than simply using one of them.

Keywords

Data integration, Genomic data, Visualization.

References

Hamid, J.S., Hu, P., Roslin, N.M., Ling, V., Greenwood, C.M.T. and Beyene, J. (2009). Data integration in genetics and genomics: methods and challenges. *SAGE-Hindawi Access to Research. Human Genomics and Proteomics*, doi: 10.4061/2009/869093.

Serino, M., García Fuentes, E., Quipo-Ortuño, M., Moreno, J.M., Liche, E., Amar, J., Sanchez-Pla, A., Tinahones, F., Burcelin, R. and Fernández-Real, J.M.L. (2009). A specific gut microbiota genomic profile defines insulin sensitivity. Submitted.

de Tayrac, M., Lê, S., Aubry, M., Mosser, J. and Husson, F. (2009). Simultaneous analysis of distinct Omics data sets with integration of biological knowledge: multiple factor analysis approach. *BMC Genomics*. Jan 20, 10–32.

Spectral representation of multi-dimensional discrete time self-similar processes

Saeid Rezakhah and **Navideh Modarresi**

Amirkabir University of Technology, Tehran, Iran

Abstract

We consider a discrete scale invariant (DSI) process $\{X(t), t \in \mathbf{R}^+\}$ with scale $l > 1$. We consider to have some fix number of observations in every scale, say T, and to get our samples at discrete points $\alpha^k,\ k \in \mathbf{W}$ where α is obtained by the equality $l = \alpha^T$ and $\mathbf{W} = \{0, 1, \ldots\}$. So we provide a discrete time scale invariant (DT-SI) process $X(\cdot)$ with parameter space $\{\alpha^k, k \in \mathbf{W}\}$. We find the spectral representation of the covariance function of such DT-SI process. By providing harmonic like representation of multi-dimensional self-similar processes, spectral density function of them are presented. Finally we find the spectral density matrix of such DT-SI process and its associated T-dimensional self-similar process.

Keywords

Discrete scale invariance, Spectral representation, Multi-dimensional self-similar processes.

References

Loève, M. (1963). *Probability Theory*, 3rd edition. Van Nostrand: New York.

Modarresi, N. and Rezakhah, S. (2009). Discrete time scale invariant Markov processes. *http://arxiv.org/pdf/0905.3959v3*, 1–12.

Rozanov, Y.A. (1967). *Stationary Random Processes*. Holden-Day: San Francisco.

Simulation and analysis of realistic GxE using a crop growth model with physiological parameters without GxE

Paulo C. Rodrigues[1,2], **Ep Heuvelink**[2], **Marco Bink**[2], **Leo F.M. Marcelis**[2] and **Fred van Eeuwijk**[2]

[1] New University of Lisbon, Portugal
[2] Wageningen University, The Netherlands

Abstract

A different response of genotypes across environments is frequent in multi-location trials and is known as genotype by environment interaction (GxE). The study and understanding of these interactions is a major challenge for breeders and agronomic researchers.
In this paper we use an adaptation of the LINTUL (light interception and utilization simulator) crop growth model (Spitters, 1990) with 6 physiological parameters, to simulate two-way phenotypic data tables. Each of these 6 parameters was obtained based on *a prior* knowledge regarding the characteristics of the studied crop. Considering θ the vector of the 6 parameters and $f(.)$ a nonlinear function, the phenotypic realisations (e.g. yield for genotype i and environment j) can be written as

$$Phe_{i,j} = f(\theta)_i + \varepsilon_{i,j}. \tag{1}$$

Standard tools such as AMMI (additive main effect and multiplicative interaction) (Gauch, 1992) and linear mixed models (van Eeuwijk, 1995; Basford et al. 2004) are used to analyse the GxE presented in the simulated data. A case study for yield of sweet pepper (*Capsicum annuum* L.) is presented.

Keywords

Genotype by environment interaction, Crop growth model, Linear mixed models, AMMI models, Simulation, Pepper.

References

Basford, K.E., Federer, W.T. and DeLacy, I.H. (2004). Mixed model formulations for multi-environment trials. *Agron. J. 96*, 143–147.

Gauch, H.G. (1992). *Statistical Analysis of Regional Yield Trials: AMMI Analysis of Factorial Designs.* Elsevier, Amsterdam, The Netherlands.

Spitters, C.J.T. (1990). Crop growth models: their usefulness and limitations. *Acta Horticulture 267*, 349–368.

van Eeuwijk, F. (1995). Linear and bilinear models for the analysis of multi-environment trials: I. An inventory of models. *Euphytica 84*, 1–7.

Linear models with doubly exchangeable distributed errors

Anuradha Roy

University of Texas at San Antonio, USA

Abstract

We study the general linear model (GLM) with doubly exchangeable distributed error for m random variables observed repeatedly over time and space. Doubly exchangeable linear models (DEGLMs) are suitable for three-level multivariate data, which is very common in biomedical, medical, engineering and in many other applications. The DEGLM arises when the m-dimensional error vectors are "doubly exchangeable" jointly normally distributed, which is a much weaker assumption than the independent and identically distributed error vectors in case of GLM. We estimate the parameters in the model and also find their distributions. Doubly exchangeable covariance structure assumes a block circulant covariance structure consisting of three unstructured covariance matrices for three multivariate levels. To be more precise, let t and s stand for a given point in time and a given site, respectively. Let $\boldsymbol{x}_{ts} : \Omega \to \Re^m,\ 1 \leq t \leq v,\ 1 \leq s \leq u$, be the $m-$dimensional normally distributed random vector at the t^{th} time point and at the s^{th} site. Then the random families $(\boldsymbol{x}_{1s})_{s\in\{1,\ldots,u\}}, \ldots, (\boldsymbol{x}_{vs})_{s\in\{1,\ldots,u\}}$ are assumed to be exchangeable. Furthermore, for fixed t, the family of random variables $(\boldsymbol{x}_{ts})_{s\in\{1,\ldots,u\}}$ is exchangeable. That is, doubly exchangeable covariance structure has a blocked constant covariance matrix over time, like the compound symmetric covariance structure has constant variance over time for the univariate repeated measures data or just one-level multivariate data.

Keywords

Linear models, Doubly exchangeable covariance structure.

References

Roy, A. and Leiva, R. (2007). Discrimination with jointly equicorrelated multi-level multivariate data. *Advances in Data Analysis and Classification 1(3)*, 175–199.

Small sample estimation in dynamic panel data models

Lorelie Santos and **Erniel Barrios**

University of the Philippines Diliman, Philippines

Abstract

This paper uses simulated data to investigate both the small and large sample properties of the within-groups (WG) and first difference generalized method of moments (FD-GMM) estimators of a dynamic panel data (DPD) model. The magnitude of WG and FD-GMM estimates are almost the same for square panels. WG estimator performs best for long panels such as those with time dimensions as large as 50. The advantage of FD-GMM estimator however, is observed on panels that are long and wide, say with time dimension at least 25 and cross-section dimension size of at least 30. For small-sized panels, we developed parametric bootstrap versions of WG and FD-GMM estimators. Simulations indicate the advantages of the bootstrap methods under small sample cases wherein the variances of the individual effects and the disturbances are of similar magnitude. Thus, while WG and FD-GMM estimators are asymptotically optimal, small samples can still exhibit such optimality through the integration of the bootstrap method.

Keywords

Dynamic panel data model, Within-groups estimator, First-difference generalized method of moments estimator, Parametric bootstrap.

References

Alvarez, J. and Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica 71*, 1121–1159.

Anderson, T. and Hsiao, C. (1981). Estimation of dynamic models with error components. *J. Amer. Statist. Assoc. 76(375)*, 598–606.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: monte carlo evidence and an application to employment equations. *Rev. Econom. Stud. 58*, 277–297.

Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-component models. *J. Econometrics 68*, 29–45.

Baltagi, B. (2005). *Econometric Analysis of Panel Data*, 3rd edition. New York: John Wiley and Sons.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *J. Econometrics 87*, 115–143.

Bond, S. (2002). Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic J. 1(2)*, 141–162.

Buhlman, P. (2002). Bootstrap for time series. *Statist. Sci 1(1)*, 52–72.

Carlstein, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist. 14*, 1171–1179.

Chigira, H. and Yamamoto, T. (2006). A bias-corrected estimation for dynamic panel models in small samples. In *Hi-Stat Discussion Paper Series No.177*. Institute of Economic Research, Hitotsubashi University.

Efron B. (1979). Bootstrap method: another look at the Jacknife. *Ann. Statist. 7(1)*, 1–26.

Everaert, G. and Pozzi, L. (2007). Bootstrap based bias correction for dynamic panels. *J. Econom Dynam. Control 31*, 1160–1184.

Hayakawa, K.(2007). Small sample bias properties of the system GMM estimator in dynamic panel data models. *Economics Lett. 95*, 32–38.

Hsiao, C. (2003). *Analysis of Panel Data*, 2nd edition. Cambridge: Cambridge University Press.

Hsiao, C. and Tahmiscioglu, A. (2008). Estimation of dynamic panel data models models with both individual and time-specific effects. *J. Statist. Plann. Inference 138*, 2698–2721.

Islam, N. (2000). Small sample performance of dynamic panel data estimators: a monte carlo study on the basis of growth data. In: Baltagi, B. (Eds.) *Advances in Econometrics: Nonstationary Panels, Panel Cointegration and Dynamic Panels* (Vol. 15). New York: Elsevier Science.

Judson, R. and Owen, A. (1999). Estimating dynamic panel data models: a practical guide for macroeconomist. *Economics Lett. 65(1)*, 9–15.

Kiviet, J. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *J. Econometrics 68*, 53–78.

Kiviet, J. and Bun, M. (2001). The accuracy of inference in small samples of dynamic panel data models: simulation evidence and empirical results. *Tinbergen Institute Discussion Paper TI 2001-006/4*, University of Amsterdam and Tinbergen Institute.

Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist. 17*, 1217–1241.

MacKinnon, J. (2002). Bootstrap inference in econometrics. *Canad. J. Economics 35*, 615–645.

MacKinnon, J. (2006). Bootstrap methods in econometrics. *Economic Record 82* (Special Issue), S2–S18.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*, 1417–1425.

On reproducing linear estimators within the GM-model with stochastic constraints

Burkhard Schaffrin

Ohio State University, Columbus/OH, USA

Abstract

In a Gauss-Markov Model (GMM) with fixed constraints, all the relevant estimators do perfectly satisfy these constraints. As soon as they become stochastic, most estimators are allowed to satisfy them only approximately, thereby leaving room for non-vanishing residuals to describe the deviation from the prior information.
Sometimes, however, linear estimators may be preferred that are able to perfectly reproduce the prior information in form of stochastic constraints, including their variances and covariances. As typical example may be considered the case where a geodetic network ought to be densified without changing the higher-order point coordinates that are usually introduced together with their variances and (some) covariances. Traditional estimators are based on the "Helmert" or "S-transformation", respectively an adaptation of the fixed-constraints Least-Squares estimator.
Here, it will be shown that neither approach generates the optimal reproducing estimator which will be presented in detail and compared with the other reproducing estimators in terms of their MSE-risks.

Keywords

Gauss-Markov Model, Stochastic constraints, Variance/covariance preservation, Reproducing estimators.

References

Schaffrin, B. (1997). On suboptimal geodetic network fusion. *Poster paper, presented at the IAG General Meeting, Rio de Janeiro, Brazil, Sept. 1997.*

Schaffrin, B. (2003). Reproducing estimates via least-squares: An optimal alternative to the Helmert transformation. In Grafarend, E.W., Krumm, F.W., Schwarze, V.S (Eds.), *Geodesy-The Challenge of the 3rd Millennium* (pp.387–392), Springer: Berlin.

Fok, H.S., Iz, H.B. and Schaffrin, B. (2009). Comparison of four geodetic network densification solutions. *Survey Review 41*, 44–56.

On efficient use of estimators for finite population mean

Javid Shabbir[1] **and Sat Gupta**[2]

[1] Quaid-i-Azam University, Islamabad, Pakistan
[2] University of North Carolina, Greensboro, USA

Abstract

Koyuncu and Kadilar (2009) introduced a family of estimators and found that their estimators are more efficient than Khoshnevisan et al. (2007) family of estimators under certain conditions. Shabbir and Gupta (2010) introduced the exponential ratio type estimator for population mean. Recently Koyuncu and Kadilar (2010) suggested an improved ratio type estimator for estimating the population mean. In this paper we propose a difference type estimator and exact bias and mean square error (MSE) are computed for any sample size. The proposed estimator is more efficient than the mean per unit estimator, traditional ratio and regression estimators as well as Shabbir and Gupta (2010), Koyuncu and Kadilar (2009, 2010) estimators. Some real data sets are used to observe the performances of the estimators.

Keywords

Ratio estimator, Regression estimator, MSE, Efficiency.

Asymptotic expansion for the distribution of the linear discriminant function with 2-step monotone missing data

Nobumichi Shutoh and **Takashi Seo**

Tokyo University of Science, Japan

Abstract

We discuss the linear discriminant function with 2-step monotone missing data and the asymptotic expansion for the distribution of the function. In discriminant analysis, asymptotic expansions for the distributions of the discriminant functions play important roles for obtaining the asymptotic approximations of the probabilities of misclassification and so on. Okamoto (1963) derived an asymptotic expansion for the linear discriminant function with complete data. Similarly to Shutoh and Seo (2010), Shutoh (2010) derived an asymptotic expansion for the distribution of the linear discriminant function with 2-step monotone missing data, i.e., an extension for the result of Okamoto (1963) up to the terms of the first order. We present the outline for the derivation of the result of Shutoh (2010). Finally we give the numerical evaluations of our result by Monte Carlo simulation for some selected parameters.

Keywords

Linear discriminant analysis, Asymptotic expansion, Probability of misclassification, Missing data.

References

Okamoto, M. (1963). An asymptotic expansion for the distribution of the linear discriminant function. *Ann. Math. Statist. 34*, 1286–1301.

Shutoh, N. (2010). Asymptotic expansions relating to discrimination based on 2-step monotone missing samples. Submitted.

Shutoh, N. and Seo, T. (2010). Asymptotic expansion of the distribution of the Studentized linear discriminant function based on 2-step monotone missing samples. *Comm. Statist. Simulation Comput.* Accepted.

The application of a two-level model to the Dutch Business Survey

Marc Smeets, Virginie Blaess and Sabine Krieg

Statistics Netherlands, The Netherlands

Abstract

The Dutch Structural Business Survey (SBS) is an annual survey that measures the total production of the Dutch enterprises and describes the cost-benefit structure of the different economic sectors. The primary publication cells are given by the so-called industries that are based on the economic classification of the enterprises. The official statistics of national statistical offices are preferably based on empirical evidence and as little as possible on model assumptions. Design-based or model-assisted methods like the generalized regression estimator (GREG) are therefore preferred in general. The sample sizes of the industries, however, are too small to produce reliable figures based on the GREG. This is especially due to the influence of outliers. For this reason, a model-based approach is considered based on the theory of small area estimation (Rao, 2003). By applying a multilevel model information can be borrowed from other industries to improve the accuracy of the estimates. For the estimation of the total turnover there are some covariates available like tax-turnover, number of persons employed and age of the enterprise. Based on this information we have constructed a two-level model. By means of a case study the multilevel approach is compared with the GREG. In this talk we present the two-level model and show the results of the case study.

Keywords

Small area estimation, Multilevel models, Outliers.

References

Rao, J.N.K. (2003). *Small Area Estimation*. New York: John Wiley.

Noise in machine learning Vasicek interest rate model calibration with Gaussian processes

João Beleza Sousa

Superior Institute of Engineering of Lisbon, Portugal

Abstract

Gaussian processes for machine learning can be used to calibrate the Vasicek interest rate model under the risk neutral measure.
Calibrating the model for one particular zero coupon bond does not require any modification of the Vasicek model.
However, to calibrate the model for a set of zero coupon bonds, noise must be assumed in the observed zero coupon bond prices.
In this paper we introduce noise in the Vasicek interest rate model calibration with Gaussian processes for machine learning.

Keywords

Vasicek interest rate model, Arbitrage free risk neutral measure, Interest rate model calibration, Gaussian processes for machine learning, Zero coupon bond prices.

References

Beleza Sousa, J., Esquível, M.L. and Gaspar, R.M. (2010). Machine learning Vasicek model calibration with Gaussian processes. *J. Statist. Plann. Inference.* Accepted.

Robust estimation of a linear Simultaneous Equations Model using *GMM* with limited and full information

Manuela Souto de Miranda[1], João A. Branco[2] and Anabela Rocha[1]

[1] University of Aveiro, Portugal
[2] Technical University of Lisbon, Portugal

Abstract

Common procedures for the estimation of the coefficients of the linear Simultaneous Equations Model are based on the least squares principle and on the generalized method of moments principle.
Whatever may be the estimation principle, it is important to consider the choice between the limited information approach (*LIM*) (sequential estimation of each individual equation) or the full information approach (*FIM*), (simultaneous estimation of the whole system). Specialists (see for instance Srivastava and Tiwary (1978), Judge *et al.* (1998) and Greene (2003))are not unanimous about the best choice since both approaches show some advantage, depending on the estimation principle.
The present study focus on the generalized method of moments estimator (*GMM*). The estimator has received an increasing attention in recent literature, since it demands less restrictive assumptions than the least squares estimators and it can include, in a natural way, some optimality conditions which are often present in Economic problems. This motivated a comparative study of *LIM versus FIM* under *GMM* estimation (Souto de Miranda *et al.* (2007)). A robust version of the *GMM* was presented in Rocha (2010) for the *FIM* approach. In the present work we give the *LIM* corresponding robust version, which requires a much simplified computation process. A simulation study based on several distributions of the errors of the model and on different sample sizes, makes it possible to evaluate the performance of the new estimator.

Keywords

Linear simultaneous equations model, Limited information methods, Full information methods, Robust estimation, Generalized method of moments.

References

Greene, W. (2003). *Econometric Analysis*, 5th edition. Prentice-Hall: New Jersey.

Judge, G., Hill, R., Griffiths, W., Lutkepohl, H. and Lee, T. (1988). *Introduction to the Theory and Practice of Econometrics*, 2nd edition. John Wiley & Sons: New York.

Rocha, A. (2010). *Estimação robusta em Modelos Lineares de Equações Simultâneas.* Ph.D Thesis - University of Aveiro, Portugal.

Souto de Miranda, M., Rocha, A. and Branco, J. (2007). Método dos momentos generalizado: a ambígua supremacia da informação completa sobre a informação limitada. In Ferrão, M.E., Nunes, C., Braumann, C. (Eds.) *Actas do XIV Congresso Anual da Sociedade Portuguesa de Estatística.* Lisbon, Edições SPE.

Srivastava, V. and Tiwary, R. (1978). Efficiency of two-stage and three-stage least squares estimators. *Econometrica 46*, 1495–1498.

Moments of generalized order statistics from some distributions

Khalaf S. Sultan and **Tagreed S. Al-Malki**

King Saud University, Riyadh, Saudi Arabia

Abstract

In this paper, we derive explicit forms of the moments of the generalized order statistics (GOSs) from the power function and log-logistics distributions. Then, we deduce the moments of the ordinary order statistics (OOSs) and record values (RVs) as special cases. Also, we use the moments of the record values from the logistics distribution to develop the best linear unbiased estimate (BLUE) of the scale parameter. Finally, we show the usefulness of the BLUE through Monte Carlo Simulations.

Keywords

Order statistics, Record values, Single moments, Product moments, Scale parameter, Coefficients of the BLUE and Monte Carlo simulations.

References

Ahsanullah, M. (1997). Generalized order statistics from power function distribution. *J. Appl. Statist. Sci. 5*, 283–290.

Balakrishnan, N. and Sultan, K.S. (1998). ecurrence relations and identities for moments of order statistics. In Balakrishnan, N., Rao, C.R. (Eds.) *Handbook of Statistics, 16, Order Statistics: Theory and Methods* (pp. 149–228). North-Holland, Amsterdam.

Kamps, U. (1995). *A Concept of Generalized Order Statistics.* Teubner, Stuttgart.

Performance of the difference-based estimators in partially linear models

Gülin Tabakan

Aksaray University, Turkey

Abstract

In this study, we consider a commonly used partially linear model. We proposed a restricted difference-based ridge estimator for the vector of parameters in a partially linear model with one smoothing term when additional linear restrictions on the parameter vector are assumed to hold. Afterwards, the performance of difference-based estimators in partially linear models are evaluated with a Monte Carlo simulation study.

Keywords

Lagrangian function, Difference-based ridge regression estimator, Partially linear model.

References

Engle, R.F., Granger, C.W.J., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc. 81*, 310–320.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics 12*, 55–67.

Hoerl, A.E., Kennard, R.W. and Baldwin, K.F. (1975). Ridge regression: some simulation. *Comm. Statist. Simulation Comput. 4*, 105–123.

Kibria, B.M.G. (1996). On preliminary test ridge regression estimators for linear restriction in a regression model with non-normal disturbances. *Communications in Statistics A 25*, 2349–2369.

Lawless, J.F. (1978). Ridge and related estimation procedure. *Communications in Statistics A. 7*, 39–164.

McDonald, G.C. and Galerneau, D.I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *J. Amer. Statist. Assoc. 70*, 407–416.

Tabakan, G. and Akdeniz, F.(2010). Difference-based ridge estimator of parameters in partial linear model. *Stat. Papers*, doi: 10.1007/s00362-008-0192-6.

Yatchew, A. (1997). An elemantary estimator of the partial linear model. *Economics Lett. 57*, 135–143.

Testing independence by step-down multiple comparison procedure

Sho Takahashi[1], **Takahiro Nishiyama**[1], **Takashi Seo**[1] **and Tsunehisa Imada**[2]

[1] Tokyo University of Science, Japan
[2] Tokai University, Japan

Abstract

We consider testing independence among components of the random vector in multivariate normal populations. The likelihood ratio test statistic and the chi-square distribution which is the asymptotic distribution are used for this test. In addition, the modified likelihood ratio test statistic which makes the chi-square approximation better is known (see, e.g., Muirhead (1982)). In this study, we consider the case to test which components there is a correlation among of the random vector. Also, we propose a step-down multiple comparison procedure based on the closed testing procedure (Marcus, Peritz and Gabriel (1976)) to perform simultaneous test for independence among components of the random vector.

Keywords

Closed testing procedure, Modified likelihood ratio, Step-down multiple comparison procedure, Testing for independence.

References

Marcus, R., Peritz, E. and Gabriel, K.R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika 63*, 655–660.
Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory.* New York: Wiley.

Robustness of uniform marginal designs for logistic mixed-effects linear models with covariates

Frans E. S. Tan[1] **and Fetene B. Tekle**[2]

[1] Maastricht University, The Netherlands
[2] Tilburg University, The Netherlands

Abstract

Optimal design theory deals with the assessment of the optimal joint distribution of all independent variables prior to data collection. In many practical situations, however, covariates are involved for which the distribution is not previously determined. These variables will be called Śprior-uncontrolled variablesŠ. The optimal design problem may then be reformulated in terms of finding the optimal marginal distribution for a specific set of variables. In general, the optimal solution depends on the unknown (conditional) distribution of the prior-uncontrolled variables. In this presentation sufficient conditions will be given under which the uniform design of a subset of independent discrete variables is DA Ũ maximin. The sufficient conditions are formulated for Generalized Linear Mixed Models with an arbitrary number of quantitative and qualitative independent variables and random effects. Further, we studied robustness of DA-maximin marginal designs when some of the conditions for uniformity are invalid. In particular, the performance of the uniform marginal design is of interest when the range of regression parameters is not symmetric around zero. The performance of uniform designs when neglecting these variables will also be discussed.

Keywords

D_A - maximin, Marginal designs, Prior-uncontrolled variables, Uniform designs.

References

Cook, R.D. and Thibodeau, L.A. (1980). Marginal restricted D-optimal designs. *J. Amer. Statist. Assoc. 75(370)*, 366–371.

Ouwens, M.J.N.M., Tan, F.E.S. and Berger, M.P.F. (2006). A maximin criterion for the logistic random intercept model with covariates. *J. Statist. Plann. Inference 136(3)*, 962–981.

Tan, F.E.S. (2010). Conditions for DA Ũ maximin marginal designs for generalized linear models to be uniform. *Comm. Statist. Theory Methods.* To appear.

Tekle, F.B., Tan, F.E.S. and Berger, M.P.F. (2008). D-optimal design for binary longitudinal responses. *Comput. Statist. Data Anal. 52(12)*, 5253–5262.

Estimating the principal component scores through maximum likelihood estimation under normality assumption

Yücel Tandoğdu and Övgü Çıdar

Eastern Mediterranean University, North Cyprus

Abstract

Functional principal component analysis(FPCA) methodology is used in trajectory estimation. Principal component scores ξ_k play an important role in FPCA. ξ_k is a random variable with zero expectation and variance $E(\xi_k^2) = \lambda_k$, λ_k being the eigenvalues of the data covariance matrix. In the estimation of a trajectory, the estimates for ξ_k has to be determined. Since the distribution of ξ_k is unknown, it has to be estimated from available data. In the case of difficulty in determining the distribution of ξ_k, a transformation to normality will facilitate the robust estimation of the principal component scores. Further the maximum likelihood estimators of the statistics for transformed ξ_k will have some important asymptotic properties. Transformed scores are then used in the estimation of trajectories. Different approaches are available for employing ξ_k in FPCA. The methodology developed in this study is compared with the principal component analysis through conditional expectation (PACE) which is an accepted method in estimation problems through FPCA. Estimated trajectories using transformed scores have produced better results compared to those obtained from PACE method.

Keywords

Smoothing, Principal component analysis, Maximum-likelihood, Functional principal component scores, Transformation to normal.

References

Cheng, T. and Feser, V.M.P. (2002). High breakdown estimation of multivariate mean and covariance with missing observaitons. *British J. Math. Statist. Psychology 55*, 317–335.

Li, P. (2005). *Box-Cox Transformations: an Overview.* Department of Statistics, University of Connecticut.

Manly, B.F.J. (1976). Exponential data transformations. *Statistician 25*, 37–42.

Müller, H.G. (2005). Functional modelling and classification of longitudinal data. *Scand. J. Statist. 32*, 223–240.

Park, H.M. (2008). *Univariate Analysis and Normality Test Using SAS, Stata, and SPSS, working paper.* The UITS center for Statistical and Mathematical Computing, Indiana University.

Rubin, D.B. and Dempster, A.P. (1977). Maximum-likelihood from incomplete data via EM algorithm. *J. R. Stat. Soc. Ser. B Stat. Methodol. 39*, 1–38.

Yao, F., Müller, H.G. and Wang, J.L. (2005). Functional data analysis for sparse longitudinal data. *J. Amar. Statist. Assoc. 100*, 577–590.

Yeo, I.K. and Johnson, R. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika 87*, 954–959.

Some comments on estimations under a restricted linear model and its implicitly restricted linear model

Yongge Tian

China Economics and Management Academy,
Central University of Finance and Economics, Peking, China

Abstract

In the investigation of the restricted linear model $\{\mathbf{y},\ \mathbf{X}\boldsymbol{\beta} \mid\ \mathbf{A}\boldsymbol{\beta} = \mathbf{b},\ \sigma^2\boldsymbol{\Sigma}\}$, the parameter constraints $\mathbf{A}\boldsymbol{\beta} = \mathbf{b}$ are often handled by transforming into an implicitly restricted linear model. The two models are not necessarily equivalent when estimating the unknown parameters under different optimal criteria. In this talk, I present necessary and sufficient conditions for the ordinary least-squares estimations under the two models to be equal. In addition, we consider relations between the best linear unbiased estimations under the two models.

Keywords

Restricted linear model, Implicitly restricted model, OLSE, BLUE, Equality of estimators, Moore-Penrose inverse of matrix, Matrix rank method.

Logistic regression estimators comparison using Pitman's Measure of Closeness

Nurkut Nuray Urgan

Namik Kemal University, Turkey

Abstract

The multicollinearity may occur in many applications of the logistic regression models, in this case Ridge estimator is used to combat multicollinearity in logistic regression models. Liu logistic regression estimator, which combines the advantages of Ridge and Stein estimates, is proposed as an alternative to the Ridge Logistic estimator in the sense of mean square estimation criterion. This paper provides theoretical results about the performance of Liu and Ridge logistic regression estimator under the criterion of Pitman's Measure of closeness as an alternative criterion to the mean square estimation criterion.

Keywords

Logistic regression, Ridge logistic estimator, Liu logistic estimator, Pitman's Measure of Closeness.

References

Hoerl, A.E. and Kennard, R.W. (1970a). Ridge regression: biased estimation for non-orthogonal problems. *Technometrics 12*, 55–67.

Hoerl, A.E. and Kennard, R.W. (1970b). Ridge regression: applications to non-orthogonal problems. *Technometrics 12*, 69–82.

Liu, K. (1993). A new class of biased estimate in linear regression. *Comm. Statist. Theory Methods 22(2)*, 393–402.

Mason, R.L. and Blaylock, N.W. (1991). Ridge regression estimator comparisons using Pitman's Measure of Closeness. *Comm. Statist. Theory Methods 20*, 3629–3641.

Mason, R.L., Keating, J.P., Sen, P.K. and Blaylock, N.W. (1990). Comparison of linear estimators using Pitman's Measure of Closeness. *J. Amer. Statist. Assoc. 85*, 579–581.

Ozkale, M.R. and Kaciranlar, S. (2007). Comparisons of the unbiased ridge estimation to the other estimations. *Comm. Statist. Theory Methods 36(4)*, 707–723.

Estimation of the maximum displacement response in structures with linear behaviour

Pedro Vieira[1], **Paula Milheiro-Oliveira**[2] **and Álvaro Cunha**[2]

[1] University of Trás-os-Montes e Alto Douro, Portugal
[2] University of Porto, Portugal

Abstract

The response of a structure characterized by a single degree of freedom and excited by gaussian white noise is investigated in terms of its maximum absolute displacement. The mean and mean square of the maximum absolute displacement are studied based upon different approaches. Statistics and mean values are computed by simulating the process obtained by discretizing the exact solution of the motion equation and are compared with analytical techniques: the Poisson method, the Vanmarcke method and the ACC method. The Poisson and Vanmarcke methods constitute classical approximations to this problem. The ACC (Advanced Censored Closure) method is more recent (Muscolino, 2007) and is based on the assumption of a Gumbel model for the maximum of i.i.d. Gaussian variables. It is known to be suitable for stationary responses but it produces bad results when the response is non-stationary. In this study, we show that the Generalized Extreme Value (GEV) distribution is more suitable to solve this problem than the Gumbel model. The parameters of the GEV distribution were estimated based on the L-moments method. As a consequence the ACC method was modified to include a GEV distribution of type II. We also concluded that the results obtained using the GEV distribution in the ACC are more acurate than those obtained by the other two methods.

Keywords

Stochastic differential equations, Maximum absolute response, ACC method, GEV distribution.

References

Muscolino, G and Palmeri, A. (2003). Largest absolute value statistics for the response of linear structures under random excitations. *Proc. CSM-4, Comput. Stochastic Mechanics*, 431–440.

Muscolino, G and Palmeri, A. (2007). Maximum response statistics of MDOF linear structures excited by non-stationary random processes. *Comput. Methods Appl. Mech. Engrg. 194*, 1711–1737.

Senthilnathan, A. and Lutes, L.D. (1991). Nonstationary maximum response statistics for linear structures. *J. Engrg. Mech. 117(2)*, 305–327.

Reduced rank regression and multivariate linear models

Dietrich von Rosen[1], Tatjana von Rosen[2] and Yonghui Liu[3]

[1] Swedish University of Agricultural Sciences, Uppsala, Sweden
[2] Stockholm University, Sweden
[3] Shanghai Finance University, China

Abstract

We present a number of new results concerning rank restrictions in general multivariate linear models. Both restrictions on the mean and dispersion matrix will be considered. The focus will be on the mathematical treatment and the main purpose is to show several directions of extensions of existing models as well as their treatment.

On exact tests in unbalanced mixed linear models

Tatjana von Rosen[1] **and Dietrich von Rosen**[2]

[1] Stockholm University, Sweden
[2] Swedish University of Agricultural Sciences, Uppsala, Sweden

Abstract

Exact tests for variance components in unbalanced mixed linear models with patterned covariance structures are presented. The derivation utilizes orthogonal transformations and a subsequent resampling. The work is related to Gallo & Khuri (1990) and Öfversten (1993).

Keywords

Exact tests, Hypothesis testing, Unbalanced mixed models, Variance components.

References

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.

Öfversten, J. (1993). Exact tests for variance components in unbalanced mixed linear models. (English, French summary). *Biometrics 49*, 45–57.

Focused information criteria, model selection and model averaging in a Tobit Model with a non-zero threshold

Alan Wan[1], **Zhang Xinyu**[2] **and Zhou Sherry**[1]

[1] City University of Hong Kong, Hong Kong
[2] Chinese Academy of Sciences, China

Abstract

In a recent paper, Claeskens and Hjort (2003) developed a Focused Information Criterion (FIC) for model selection. Unlike common model selection methods that lead to a single model for all purposes, the FIC selects different models for different purposes. In another paper, Hjort and Claeskens (2003) presented model averaging as an alternative to model selection, and suggested a local mis-specification framework for studying the limiting distributions and asymptotic risk properties of post model selection and model average estimators in parametric models. Despite the bourgeoning literature on Tobit models, little work has been undertaken with respect to model selection explicitly in the Tobit context. In this paper, we propose FICs for variable selection allowing for such measures as the MAD, MSE, and expected LINEX errors in a Type I Tobit model with an unknown threshold. We also develop a model average Tobit estimator using values of a smoothed version of the FIC as weights. The finite sample performance of model selection and model average estimators resulting from various FICs is studied via a Monte Carlo experiment, where the possibility of using a model screening procedure prior to combining the models is also demonstrated. Finally, we present an example from a well-known study on married womenŠs working hours to illustrate the estimation methods discussed.

Keywords

Backward elimination, Censored regression, Local mis-specification, Model screening.

References

Claeskens, G. and Hjort, N.L. (2003). The focused information criterion [with discussion]. *J. Amer. Statist. Assoc. 98*, 900–916.

Quadratic forms, Jordan algebras and the Wishart distribution

Jacek Wesołowski

Warsaw University of Technology, Poland

Abstract

Celebrated Lukacs (1955) characterization of the gamma distribution was extended to Wishart distribution in Olkin and Rubin (1962), and more recently in Casalis and Letac (1996), Bobecka and Wesołowski (2002) and Boutoria, Hassairi, Massam (2010). In all these results different versions of original condition of independence of $X+Y$ and $X/(X+Y)$ for independent random variables X and Y were considered, as for example independence of $X+Y$ and $(X+Y)^{-1/2}X(X+Y)^{-1/2}$ for independent positive definite random matrices.

It seems that the very essence of the Wishart distribution is hidden not in independence properties mentioned above, but rather in an invariance property of regression of quadratic forms in the following sense: Let the space $\mathcal{Q}$ of quadratic forms on $\mathbb{R}^n$ be splitted in a direct sum $\mathcal{Q}_1 \oplus \ldots \oplus \mathcal{Q}_k$, le X and Y be independent random vectors in $\mathbb{R}^n$, let there exist a real number a such that $\mathbb{E}(X|X+Y) = a(X+Y)$ and real distinct numbers $b_1, ..., b_k$ such that $\mathbb{E}(q(X)|X+Y) = b_i q(X+Y)$ for any q in $\mathcal{Q}_i$. Then $\mathcal{Q} = \mathcal{Q}_1 \oplus \mathcal{Q}_2$, $\mathbb{R}^n$ can be structured in a Euclidean Jordan algebra and X and Y have Wishart distributions on symmetric cones related to the Jordan structures. Moreover, the subspaces $\mathcal{Q}_1$ and $\mathcal{Q}_2$ can be characterized as eigenspaces of an intriguing linear operator Ψ acting on the space of symmetric endomorphisms of the Euclidean Jordan algebra. Theses results will appear soon in Letac and Wesołowski (2010).

Keywords

Backward elimination, Censored regression, Local mis-specification, Model screening.

References

Bobecka, K. and Wesołowski, J. (2002). The Lukacs-Olkin-Rubin theorem without invariance of the "quotient". *Studia Math. 152*, 147–160.

Boutoria, I., Hassairi, A. and Massam, H. (2010). Extension of the Olkin and Rubin characterization of the Wishart dsitribution on homogeneous cones. *arXiv 1002.1451v1*, 1–19.

Casalis, M. and Letac, G. (1996). The Lukacs-Olkin-Rubin characterization of the Wishart distribution on symmetric cones. *Ann. Statist. 24*, 763–786.

Letac, G. and Wesoł owski, J. (2010). Why Jordan algebras are natural in statistics: regression of quadratic forms and Wishart distribution. *Bull. Soc. Math. France*. To appear.

Lukacs, E. (1955). A characterization of the gamma distribution. *Ann. Math. Statist. 26*, 319–324.

Olkin, I. and Rubin, H. (1962). A characterization of the Wishart distribution. *Ann. Math. Statist. 33*, 1272–1280.

A high dimensional MANOVA test with fewer observations than the dimension

Takayuki Yamada[1] **and Muni S. Srivastava**[2]

[1] Kitasato University, Minato, Tokyo, Japan
[2] University of Toronto, Canada

Abstract

We consider the problem of testing a linear hypothesis in a multivariate linear model when the $N \times p$ observation matrix is normally distributed with unknown covariance matrix, and $N \leq p$, This includes the case of testing the equality of several mean vectors. A test is proposed which is a generalized version of the two-sample test proposed by Srivastava and Du (2008). The asymptotic null and non-null distributions are obtained. The performance of this test is compared, theoretically as well as numerically, with the coresponding generalized version of the two-sample Dempster's (1958) test, or equivalently Bai-Sarandasa test (1996) who gave its asymptotic version.

Keywords

MANOVA, Asymptotic null and non-null distributions, High-dimension, Power comparison.

References

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statist. Sinica 6*, 311–329.

Dempster, A.P. (1968). A high dimensional two sample significance test. *Ann. Math. Statist. 29*, 995–1010.

Fujikoshi, Y., Himeno, T. and Wakaki, H. (2004). Asymptotic results of a high dimensional MANOVA test and power comparison when the dimension is large compared to the sample size. *J. Japan Statist. Soc. 34*, 19–26.

Srivastava, M.S. and Fujikoshi, Y. (2006). Multivariate analysis of variance with fewer observations than the dimension. *J. Multivariate Anal. 97*, 1927–1940.

Srivastava, M.S. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *J. Multivariate Anal. 99*, 386–402.

Part VII

Posters

Spectral and wavelet analysis of the Atlantic North Circulation: a case study

Cristina Andrade[1,3], João A. Santos[2] and João Corte-Real[3]

[1] Polytechnic Institute of Tomar, Portugal
[2] University of Trás-os-Montes and AltoDouro, Vila Real, Portugal
[3] University of Évora, Portugal

Abstract

Wavelet analysis has been used for numerous studies with atmospheric data and is becoming a common tool for analysing one-dimensional time series, as it is useful in determining the dominant modes within a time series, especially in studying the displacement characteristics of a moving oscillating structure. This property allows the analysis of its temporal evolution and the detection of short duration events, even in large time series.

Due to the significant influence of the North Atlantic ridge (usually associated to the Azores high) in Iberian Peninsula winter climate, a better understanding of the physical mechanisms that contribute to its development and maintenance is highly relevant. Aiming to isolate common periodicities, three target areas were initially chosen to this case study (two over North America at both tropospheric and stratospheric levels and one over the North Atlantic at tropospheric levels). These areas were chosen taking into account previous results, where some dynamical precursors of strong and persistent North Atlantic ridges were already identified at specific locations over North America and the North Atlantic. The analyses is focused in daily mean atmospheric fields at different isobaric levels and for selected winters, with highly contrasting dynamical conditions over the North Atlantic. Therefore, two different approaches are undertaken. Firstly, a power spectral analysis allows a preliminary insight in order to identify common significant oscillations, which might be considered a manifestation of a dynamical connection at the different target areas. Secondly, a wavelet analysis significantly enhances this case study.

Keywords

North Atlantic ridge, Wavelet analysis, Power spectral analysis.

References

Grinsted, A., Jevrejeva, S. and Moore, J. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Proc. Geophys. 11*, 561–566.

Torrence, C. and Compo, G.P. (1998). A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc. 79*, 61–78.

Estimation of superimposed complex exponentials using covariance matching and sparsity

Luis Blanco[1], Montse Nájar[2] and Francisco Rubio[1]

[1] Technological Center for Telecommunications of Catalonia, Barcelona, Spain
[2] Polytechnic University of Catalonia, Barcelona, Spain

Abstract

The inferencial problem of estimating superimposed complex exponentials in general linear models is an old problem that plays a key role in multiple fields such as seismics, medical diagnosis, array signal processing, etc. Recent works based on sparse representations show promising results that outperform traditional subspace methods.
A new algorithm based on a covariance-matching approach that preserves sparsity is derived. The method considers a covariance matrix model based on an overcomplete basis representation that tries to fit the variance of the unknown parameters to the sample covariance estimate. Sparsity is enforced by means of an l_1-norm penalty term. The problem is reduced to an objective function that can be solved efficiently using the LARS/homotopy algorithm subject to a positive constraint. The algorithm proceeds in an iterative fashion solving small linear systems of equations until a stopping criterion is fulfilled.

Keywords

Sparse signal representation, Least absolute shrinkage and selection operator, Overcomplete basis, Least angle regression and selection.

Estimation in singular chemical balance weighing design with correlated errors

Bronisław Ceranka and Małgorzata Graczyk

Poznań University of Life Sciences, Poland

Abstract

The problem of estimation of the total weight of objects in a chemical balance weighing design when the errors have the same variances and are negative correlated has been studied. A lower bound for the variance of the estimated total weight is obtained and a necessary and sufficient condition for this lower bound to be attained is given. Some designs for which the lower bound is attained are constructed.

Keywords

Chemical balance weighing design, Incomplete block design, Total weight.

References

Banerjee, K.S. (1975). *Weighing Designs for Chemistry, Medicine, Economics, Operations Research, Statistics.* New York: Marcel Dekker Inc.

Ceranka, B. and Katulska, K. (1996). On the estimation of total weight in chemical balance weighing designs under the covariance matrix of errors $\sigma^2\mathbf{G}$. *Advances in Statistical Software 5, SoftStat '95, Stuttgart: Lucius* & *Lucius*, 461–468.

Chacko, A. and Dey, A. (1978). On the estimation of total weight in chemical balance weighing designs. *Austral. J. Statist. 20*, 83–86.

Dey, A. and Gupta, S.C. (1977). Singular weighing designs and the estimation of total weight. *Comm. Statist. Theory Methods 6*, 289–295.

Kageyama, S. (1988). Optimum chemical balance weighing designs for estimating the total weight. *Comm. Statist. Theory Methods 17*, 2697–2704.

Pukelsheim, F. (1993). *Optimal Design of Experiment.* New York: John Wiley and Sons.

Raghavarao, D. (1971). *Constructions and Combinatorial Problems in designs of Experiments.* New York: John Wiley Inc.

Shah, K.R. and Sinha, B.K. (1989). *Theory of Optimal Designs.* Berlin, Heidelberg: Springer-Verlag.

Approximate upper confidence intervals on a ratio of sums of variances

Ali Deeb[1] and Adel Elgafghuf[2]

[1] El-fateh University, Libya
[2] 7th of October University, Libya

Abstract

Confidence intervals are very useful statistical inference tool, they are usually constructed from completely specified distribution, however, in some cases the exact distribution of the statistic of interest is unknown, then an approximate distribution is derived and used to form an approximate confidence interval for the parameter under consideration. Satterthwaite (1946) gave an approximation of the distribution of a linear combination of chi-squared variates as another chi-square with appropriate degrees of freedom. These degrees of freedom are generally functions of the unknown variance components and hence must be estimated. Satterthwaite (1946) and Birch (1990) approximations are popular methods to construct confidence intervals on $\lambda = (k_1\theta_1 + k_2\theta_2)/(k_3\theta_3 + k_4\theta 4)$ where θ_i represents a variance of normal population and $k_i > 0$ for $i = 1, 2, 3, 4$. This article proposes other methods for constructing approximate upper confidence intervals on $\lambda = (\theta_1+\theta_2)/(\theta_3+\theta_4)$ and comparing them with those constructed under Satterthwaite and Birch methods

References

Satterthwaite , F.E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin 2*, 110–114.

Birch, N.J., Burdick, R.K. and Ting, N. (1990). Confidence intervals and bounds for a ratio of summed expected mean squares. *Technometrics 32*, 437–444.

A new algorithm for initial cluster centers in K-means clustering

Murat Erişoğlu, Nazif Çalış and Sadullah Sakallýoğlu

Çukurova University, Turkey

Abstract

Clustering is a very well known technique in data mining. One of the most widely used clustering techniques is the k-means clustering. Solutions obtained from k-means clustering are dependent on the initialization of cluster centers. In this paper, we propose an algorithm to compute initial cluster centers for K-means clustering. Firstly two principal variables are selected by the proposed methods. The data set is partitioned one at a time until the number of cluster equals to the predefined number of clusters. The algorithm has been applied to different data sets and good results are obtained.

Keywords

K-means algorithm, Initial cluster centers, Clustering algorithm, Rand index, Classification accuracy.

References

Blake, C.L. and Merz, C.J. (1998). UCI Repository of machine learning databases. *University of California, Irvine, Department of Information and Computer Science.*

Bradley, P.S. and Fayyad, U.M. (1998). Refinning initial points for k-means clustering. In: *Proceeding of The 15. International Conference on Machine Learning.*

Deelers, S. and Auwatanamongkol, S. (2007). Enhancing k-means algorithm with initial cluster centers derived from data partitioning along the data axis with the highest variance. *Int. J. Computer Science 2(4)*, 247–252.

Jain, A.K. and Dubes, R.C. (1988). *Algorithms for Clustering Data.* Prentice Hall, Englewood Cliffs, NJ.

Khan, S.S. and Ahmad, A. (2004). Cluster center initilization algorithm for k-means clustering. *Pattern Recognition Letters 25*, 1293–1302.

Likas, A., Vlassis, N. and Jakob, J.V. (2003). The global k-means clustering algorithm. *Pattern Recognition 36*, 451–461.

Macqueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. In Cam, L.M.L., Neyman, J. (Eds.) *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA: University of California. Pres., 281–297.

Rand, W.M. (1971). Objective criteria for the evaluation of clustering methods, *J. Amer. Statist. Assoc. 66*, 846–850.

Ruspine, E.H. (1970). Numerical methods for fuzzy clustering. *Inform. Sci. 2*, 319–350.

Methods for the recovery of missing data in medical research

Danya Facchinetti[1], Federico Tavola[1], Marie Claire Cantone[1] and Augusto Giussani[2]

[1] University of Milan, Italy
[2] Helmholtz Zentrum München - German Research Center for Environmental Health (GmbH), Germany

Abstract

The data collected in medical analysis are useful to describe physiological processes. Due to the clinical environment, in which logistics and experimental protocols have to follow first the patient needs, data collection can't usually be the same within a population study as for the time of sample collection. Therefore the data collected of the single patients of a population can't be directly comparable to each other. This aspect limits the capability to define a statistical model able to describe the process studied.
In order to bypass this problem, different statistical approaches have been studied to reconstruct the missing data in every data set related to the patients, so that at the end every data set would have the same number of data at the same time point, which make them directly comparable as a synthesis of real and simulated data. The different approaches studied in this work are the smoothing exponential, the moving average, regressive and autoregressive models.
Once the different data sets are directly comparable and then analyzed, a model is defined and then validated, applying it to the initial data sets without the simulated data.
In this work it has been considered, as an example, the typical case of a PET (Proton Emission Tomography) study in order to define the biokinetic of the radiopharmaceutical, due to dosimetric purposes, in which starting from real data sets collected from plasma, urine and organ images, missing time point data have been simulated.

Keywords

Missing data, Time series analysis, PET, SPECT.

References

Box, G.E.P., Jenkins, G.M. and Reinsel G.C. (2008). *Time Series Analysis.* New Jersey: Wiley.

Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis.* New Jersey: Wiley.

Molenberghs, G. and Kenward, M.G. (2007). *Missing Data in Clinical Studies.* West Sussex: Wiley.

Santamaria, L. (1981). *Analisi delle serie storiche.* Bologna: Il Mulino.

A Poisson mixture regression model: application to financial data

Susana Faria and Fátima Gonçalves

University of Minho, Guimarães, Portugal

Abstract

Poisson regression has been recognized as an important tool for analyzing the effects of covariates on count data. However, conditional on covariates, the Poisson assumption of mean-variance equality may not be valid when data are potentially overdispersed. Alternative methods of analysis have been proposed to deal with the overdispersion problem. Poisson mixture regression models are appropriate when the extra variability comes from the unobserved heterogeneity of the population, which composes of two or more subgroups mixed in various proportions, in the presence of covariate information.
These models are applied to a real data set for credit-scoring purposes. We try to predict the number of defaulted payments of a client. The number of components in the mixture model is deemed to be unknown and estimated from the data. Using covariates in all components we aim to reveal the impact of demographic and financial variables in creating different groups of clients and to predict the group to which each client belongs, as well as his expected number of defaulted payments.

Keywords

Mixture Poisson regression models, EM algorithm, Overdispersion, Count data.

References

Fruhwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models.* Heidelberg: Springer.
Karlis, D. and Rahmouni, M. (2007). Analysis of defaulters' behaviour using the Poisson-mixture approach. *IMA J. Manag. Math. 18(3)*, 297–311.
Wang, P.M., Puterman, M.L., Cockburn, I. and Le, N. (1996). Mixed Poisson regression models with covariate dependent rates. *Biometrics 52*, 381–400.
Wedel, M., DeSarbo, W.S., Bult, J.R. and Ramaswamy, V. (1993). A latent class poisson regression model for heterogeneous count data. *J. Appl. Econometrics 8*, 397–411.

Resampling techniques in the optimal choice of the threshold in extremal index estimation

Dora Prata Gomes[1] and Manuela Neves[2]

[1] New University of Lisbon, Portugal
[2] Technical University of Lisbon, Portugal

Abstract

Classical extreme-value theory for stationary sequences of random variables has to deal with one of the parameters governing the behaviour of the extremes, the extremal index. This parameter plays a special role in the description of the dependence between exceedances over a high threshold.
The estimation of the extremal index is then performed on the basis of the k order statistics in the sample or on the exceedances over a high level u_n. The estimators considered in the literature, despite of having good asymptotic properties, present high variance for high levels and a high bias when the level decreases, showing then a strong dependence on the high threshold u_n, for finite samples. A compromise between these two measures is obtained by considering the mean squared error, MSE. A question that has been often addressed is the choice of u_n, that minimizes the mean squared error. The selection of that threshold is then required.
By using resampling techniques (i.e. bootstrap and/or jackknife) an heuristic approach for estimating the level u_n, that asymptotically gives the minimum MSE of an estimator of the extremal index is suggested, based on Lahiri *et. al.* (2007) method. The key idea of the method considers the bootstrap estimation of the variance and the bias of the block bootstrap estimator. The proposed rule is based on the Jackknife-after-Bootstrap (JAB) that yields a nonparametric estimator of the variance of a block bootstrap estimator.
A simulation study as well as a real case study are shown.

Keywords

Bootstrap, Estimation, Extremal index, Jackknife.

References

Lahiri, S.N., Furukawa, K. and Lee, Y.-D. (2007). A nonparametric plug-in rule for selecting the optimal block length for the block bootstrap methods. *Stat. Methodol. 4*, 292–321.

A - optimal spring balance weighing design under some condition

Małgorzata Graczyk

Poznań University of Life Sciences, Poland

Abstract

In the paper, some investigation of the A-optimal spring balance weighing design under the assumption that the errors have different variances are dealt with. Its topic is concerned with the determining of the lower bound of the $\mathrm{tr}\left(\mathbf{X}'\mathbf{G}^{-1}\mathbf{X}\right)^{-1}$. Moreover, the new construction method of the optimal design is presented.

Keywords

A-optimal design, Balanced incomplete block design, Spring balance weighing design.

References

Jacroux, M. and Notz, W. (1983). On the optimality of spring balance weighing designs. *Ann. Statist. 11*, 970–978.

Pukelsheim, F. (1993). *Optimal Design of Experiment.* New York: John Wiley and Sons.

Raghavarao, D. (1971). *Constructions and Combinatorial Problems in Designs of Experiments.* New York: John Wiley Inc.

Shah, K.R. and Sinha, B.K. (1989). *Theory of Optimal Designs.* Berlin, Heidelberg: Springer-Verlag.

A family of near-exact approximations based on truncations of the exact distribution for the generalized Wilks Lambda statistic*

Luís M. Grilo[1] and Carlos A. Coelho[2]

[1] Polytechnic Institute of Tomar, Portugal
[2] New University of Lisbon, Portugal

Abstract

In multivariate analysis the generalized Wilks Lambda statistic is used to test the independence among m sets of random variables, under the normality assumption. For the case where at least two sets, among the m sets, have an odd number of variables, we do not have the exact distribution in a manageable form, adequate for further manipulation. Thus, we expressed the exact characteristic function of this well known statistic under the form of the characteristic function of an infinite mixture of Generalized Integer Gamma distributions and, based on truncations of this exact characteristic function, we obtained a family of near-exact approximations, as finite mixtures of Generalized Integer Gamma distributions and Generalized Near-Integer Gamma distributions, which by construction match the first two exact moments. The members of the family of near-exact approximations developed this way display an asymptotic behaviour for increasing number of variables. The corresponding near-exact cumulative distribution functions are obtained in a concise and manageable form, relatively easy to implement computationally, allowing for the computation of virtually exact moments and quantiles. We undertake a comparative study for small sample sizes, using two proximity measures based on the Berry-Esseen bounds, to assess the performance of the near-exact approximations, for different numbers of sets and different numbers of variables in each set, and also to compare them with a near-exact approximation based on factorizations of the exact characteristic function.

Keywords

Independent random variables, Characteristic function, Sum of Gamma, Mixtures, Proximity measures, Small sample sizes.

* This research received the financial support of the Portuguese Foundation for Science and Technology through grant 2010 ISFL-1-297 from FCT/MCTES/PT

An extended least angle regression for contingency tables

Yoshihiro Hirose and Fumiyasu Komaki

University of Tokyo, Japan

Abstract

In Hirose and Komaki (2010), we extended the least angle regression algorithm (Efron, Hastie, Johnstone, and Tibshirani, 2004). The extended least angle regression is used for estimating parameters and selecting models in the generalized linear regression problem while the original least angle regression is used in the linear regression problem. In this presentation, we treat contingency tables. We consider the dually flat space of distributions corresponding to contingency tables. The extension is based on the information geometry of dually flat spaces. Information geometry is an generalization of Euclidean geometry. We illustrate the extended least angle regression algorithm for contingency tables and show results for some examples.

Keywords

Contingency table, Dually flat space, Information geometry.

References

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression (with discussion). *Ann. Statist. 32*, 407–499.

Hirose, Y. and Komaki, F. (2010). An extension of least angle regression based on the information geometry of dually flat spaces. *J. Comput. Graph. Statist.* Accepted.

A model selection criterion for discriminant analysis of several groups when the dimension is larger than the total sample size

Masashi Hyodo[1], Takayuki Yamada[2] and Takashi Seo[1]

[1] Tokyo University of Sciences, Japan
[2] Kitasato University, Tokyo, Japan

Abstract

The program of the Conference is concerned with the problem of selecting variables in multiple group discriminant analysis for high-dimensional data with fewer observations than the dimension. We consider a selection criterion based on approximately unbiased for AIC type of risk. When the dimension is large compared to the sample size, AIC type of risk cannot be defined. We propose AIC by replacing maximum likelihood estimator with ridge-type estimator. This idea follows Srivastava and Kubokawa (2008). It has been further extended by Yamamura et al. (2010). Simulation revealed that the proposed AIC performs well.

Keywords

Akaike information criterion, Discriminant analysis, Ridge-type estimator, High dimensional data.

References

Srivastava, M.S. and Kubokawa, T. (2008). Akaike information criterion for selecting components of the mean vector in high dimensional data with fewer observations. *J. Japan Statist. Soc. 38*, 259–283.

Yamamura, M., Yanagihara, H. and Srivastava, M.S. (2010). Variable selection in multivariate linear regression models with fewer observations than the dimension. *Japanese J. Appl. Statist. 39(1)*, 1–19.

Problems caused by multicollinearity and outlier presence

Tomáš Jurczyk

Charles University, Prague, Czech Republic

Abstract

This poster is an illustration of problems caused by multicollinearity and outlier presence in the data. Through simple examples we can see behaviour of classical least squares method but also behaviour of robust least trimmed squares (LTS) method in different situations. The most interesting is investigation of functionality of just mentioned LTS in revealing outliers in the situation where majority of the data suffers from multicollinearity. This is closely knit with the possibility of using LTS (or its generalization LWS) as a tool for multicollinearity detection. We are going to present also a proposal of a new method which can be potential candidate for solving both problems (multicollinearity as well as outlier presence) together. Our proposal is logical combination of techniques used for dealing with multicollinearity and techniques used for dealing with outliers.

Keywords

Multicollinearity, Ridge regression, Least trimmed squares.

D-optimal chemical balance weighing designs with $n \equiv 0 \pmod 4$ and 3 objects

Krystyna Katulska and Łukasz Smaga

Adam Mickiewicz University, Poznań, Poland

Abstract

In this paper the problem of estimation of the individual weights of three objects using a chemical balance weighing design is considered. We use the criterion of D-optimality. We assume that the variance matrix of errors is the matrix of first-order autoregressive process. Such problems were discussed in Li and Yang (2005) and also in Yeh and Lo Huang (2005). We present new results of D-optimal designs in certain class of designs with the design matrix $\boldsymbol{X} \in M_{n\times 3}(\pm 1)$ such that each column of matrix $\boldsymbol{X}$ has at least one 1 and one -1.

Keywords

Chemical balance weighing design, D-optimality, First-order autoregressive process.

References

Galil, Z. and Kiefer, J. (1980). D-optimum weighing designs. *Ann. Statist. 8*, 1293–1306.

Horn, R.A. and Johnson, C.R. (1985). *Matrix Analysis*. Cambridge University Press, Cambridge.

Hotelling, H. (1944). Some improvements in weighing and other experimental techniques. *Ann. Math. Statist. 15*, 297–306.

Li, C.H. and Yang, S.Y. (2005). On a conjecture in D-optimal designs with $n \equiv 0 \pmod 4$. *Linear Algebra Appl. 400*, 279–290.

Yeh, H.G. and Lo Huang, M.N. (2005). On exact D-optimal designs with 2 two-level factors and n autocorrelated observations. *Metrika 61*, 261–275.

Block design with nested rows and columns for research on food acceptability limitation for *Tetranychus urticae*

Maria Kozłowska[1], Agnieszka Łacka[1] and Anna Skorupska[2]

[1] Poznań University of Life Sciences, Poland
[2] Institute of Plant Protection, Poznań, Poland

Abstract

We presented properties of some partially balanced block designs with nested rows and columns. The designs are considered for near-factorial experiments, when there are a levels of experimental factor A and b levels of experimental factor B and there is one control treatment added. We carried out our consideration by the derived mixed linear model resulting from randomization of blocks, rows and columns. For this model of observations, some properties of estimation of treatment contrasts are discussed. We calculated the efficiency factors of estimation of treatment contrasts.
We formulated theorems for some partially balanced block designs with nested rows and columns to be designs possessing special properties. Plant protection experiment on limitation of food acceptability for *T. urticae* is given to show how the obtained results can be applied.

Keywords

Block design with nested rows and columns, Mixed model, Factorial experiments, Near-factorial experiments, *Tetranychus urticae*.

References

Kozłowska, M., Łacka, A., Krawczyk, R. and Kozłowski, R.J. (2010). Some block designs with nested rows and columns for research on pesticide dose limitation. *Environmetrics*. Accepted.

Łacka, A. and Kozłowska, M. (2009). Planning of factorial experiments in a block design with nested rows and columns for environmental research. *Environmetrics 20(6)*, 730–742.

Łacka, A., Kozłowska, M. and Bogacka, B. (2009). Estimation and testing hypothesis in a block design with nested rows and columns. *Biom. Lett. 46(2)*, 113–128.

Łacka, A., Kozłowska, M. and Kozłowski, J. (2009). Some optimal block designs with nested rows and columns for research on alternative methods of limiting slug damage. *Statist. Papers 50(4)*, 837–846.

Forecasting accuracy. New evidences based on the Má-competition

Ana Jesus López, Blanca Moreno and Rigoberto Pérez

University of Oviedo, Spain

Abstract

Forecasting availability has widely increased, suggesting the need of analyzing the adequacy of different alternative methods. One of the main empirical researches in this field is the M-Competition developed by Makridakis and Hibon, whose last edition (M3) is referred to 2000.
The M3 Competition is based in 3003 time series and 24 forecasting techniques, including naive procedures, explicit trend models, decomposition methods, ARIMA model, expert systems and neural networks, and leading to some interesting results (which are quite similar to those obtained in previous editions). Furthermore, this investigation confirms two interesting facts: first, more sophisticated procedures do not necessarily improve the quality of the obtained results and second, the accuracy of the combined forecasts increases with regard to the individual procedures.

Since the M3-Competition only includes five different accuracy measures, in this paper we propose the use of new indicators, including the Theil index (U) and some other measures based on the Information Theory.

More specifically, given a variable Y we focus on the quadratic unquietness of Y defined by R. Pérez (1985) as follows:

$$H(Y) = \frac{2}{T}\sum_t \left(\frac{E(Y)}{Y} - 1\right)$$

and we propose the quadratic information related to the forecasts as a measure of the forecasting adequacy defined as follows:

$$IC(Y,\hat{Y}) = H(Y) - H\left(Y/[\hat{Y}]\right)\left(1 - r_{Y,\hat{Y}}\right)$$

where $r_{Y,\hat{Y}}$ is the linear correlation coefficient between actual and forecasted values and $H\left(Y/[\hat{Y}]\right)$ is the quadratic unquietness conditioned to the forecasting intervals, given by the expression:

$$H\left(Y/[\hat{Y}]\right) = \sum_j p\left([\hat{Y}_j]\right)\frac{E(Y)}{E[\hat{Y}_j]}H\left(Y/[\hat{Y}_j]\right)$$

The application of these measures to the M3 Competition database leads to some interesting results about the forecasting accuracy of the considered procedures.

Keywords

Forecasting, M-competition, Theil index, Quadratic information.

References

López, A.J., Moreno, B. and Pérez, R. (2003). Forecast evaluation based on information measures. *Proceedings of the International Statistical Institute 54th Session*. Berlin.

Makridakis, S. and Hibon, M. (2000). The M3-competition: results, conclusions and implications. *Int. J. Forecasting 16*, 451–476.

Pérez, R., Gil, M.A. and Gil, P. (1986). Estimating the uncertainty associated with a variable in a finite population. *Kybernetes 15*, 251–256.

Theil, H. (1966). *Applied Economic Forecasting*. North Holland Publishing.

Combining models in discrete discriminant analysis

Anabela Marques[1], Ana Sousa Ferreira[2] and Margarida Cardoso[3]

[1] Setúbal Polytechnic, Portugal
[2] University of Lisbon, Portugal
[3] ISCTE - University Institute, Lisbon, Portugal

Abstract

Diverse Discrete Discriminant Analysis (DDA) models perform differently on different sample observations (Brito et al. (2006)). This fact has encouraged research in combined models for DDA. This research seems to be specially promising when the *a priori* classes are not well separated or when small or moderate sized samples are considered, which often occurs in practice. In this work we evaluate the performance of a linear combination of two DDA models (Marques et al. (2008)): the First-Order Independence Model (FOIM) and the Dependence Trees Model (DTM) (Celeux and Nakache (1994). The proposed methodology also uses a Hierarchical Coupling Model (HIERM) when addressing multiclass classification problems, decomposing the multi-class problems into several bi-class problems, using a binary tree structure (Sousa Ferreira (2000)). The analysis is based both on simulated and real datasets. Results include measures of precision regarding a training set, a test set and cross-validation. The R software is used for the algorithm's implementation.

Keywords

Combining model, Discrete discriminant analysis, First-order independence model, Dependence trees model.

References

Brito, I., Celeux, C. and Sousa Ferreira, A. (2006). Combining methods in supervised classification: A comparative study on discrete and continuous problems. *REVSTAT 4(3)*, 201–225.

Celeux, G. and Nakache, J.P. (1994). *Analyse Discriminante sur Variables Qualitatives*. Polytechnica.

Marques, A., Sousa Ferreira, A. and Cardoso, M. (2008). Uma proposta de combinação de modelos em Análise Discriminante Discreta. In Oliveira, I. et al. (Eds.) *Estatística - Arte de Explicar o Acaso* (pp. 393-403). Edições SPE.

Sousa Ferreira, A. (2000). *Combinação de Modelos em Análise Discriminante sobre Variáveis Qualitativas.* Ph.D Thesis - New University of Lisbon, Portugal.

Mendelian randomisation using instrumental variable

Magid Maatallah

University of London, UK

Abstract

In epidemiological research, the causal effect of a modi able phenotype or exposure on a disease is often of public health interest. Randomised controlled trials to investigate this effect are not always possible and inferences based on observational data can be confounded. However, if we know of a gene closely linked to the phenotype without direct effect on the disease, it can often be reasonably assumed that the gene is not itself associated with any confounding factors – a phenomenon called Mendelian randomisation. These properties de ne an instrumental variable and allow estimation of the causal effect, despite the confounding, under certain model restrictions. In this paper, we present a formal framework for causal inference based on Mendelian randomisation and suggest using directed acyclic graphs to check model assumptions by visual inspection. This framework allows us to address limitations of the Mendelian randomisation technique that have often been overlooked in the medical literature.

Unreplicated experiments in early stage breeding programs

Katarzyna Marczyńska and Stanisław Mejza

Poznań University of Life Sciences, Poland

Abstract

In plant breeding trials, during the early stages of the improvement process, it is not possible to use an experimental design that satisfies the requirement of replicating all the treatments because of the large number of genotypes involved, the small amount of seed and the low availability of resources. Hence, the unreplicated designs are used for early generation testing when hundreds or even thousands new genotypes need evaluation in the same trial using a limited amount of seed that is enough for one replicated only. To control the real or potential heterogeneity of experimental units, control (check) plots are arranged in the trial.
There are many methods of using information resulting from check plots. In the paper the main tool of exploring this information will be based on a response surface methodology (RSM). At the beginning we will try to identify response surface characterizing experimental environments. The obtained response surface we will be then used to adjust the observations for genotypes. Finally, so adjusted data will be used for inference concerning the next steps of breeding program. The theoretical considerations will be illustrated with the example dealing with spring barley.

Clustering of loglinear models using likelihood ratio tests p-values to find homogeneous regions regarding drought management*

Elsa Moreira and João T. Mexia

New University of Lisbon, Portugal

Abstract

The Alentejo in the southern Portugal a is a region that often is subject to severe and extreme droughts. The main concern of the present study was to know if Alentejo should or not be considered a region aiming at drought management. So, montly time series of Standardized Precipitation Index (SPI) with 67 years correspondent to 40 meteorological stations in the region were selected in order to be analyzed. These series were used to compute contingency tables for the transitions between drought classes, considering 4 levels of drought classes. Then, to this contingency tables appropriate loglinear models were fitted in order to obtain estimates of the drought class transitions probabilities. Finished the model fitting phase, the starting point for a second phase, was in fact the existence of 40 objects, described by the correspondent loglinear models, that we wish to classify and group according to some common characteristics. So, the similarity of the objects reduces to the similarity of their loglinear models. Since the models are members of the same family, the models similarity is defined as the closeness of their vectors of model parameters. The technique used to finding similarity between loglinear model parameters was the Likelihood Ratio Tests (LRT) where the similarity measure is the asymptotic p-value associated with the model linking test. If the p-value is large, the null hypotheses of equality between the vectors of parameters is likely to be true. A p-value similarity matrix was then computed in order to detect possible groups of similar objects. Two model linking tests were performed, one considering that the loglinear model parameters are all of interest and another considering that just some of the parameters are of interest. After several attempts of identify clusters that could delimit regions, none was found. Therefore, according to the present study, the Alentejo could be considered as an homogeneous region aiming at drought management.

Keywords: Loglinear models, Drought class transitions, Likelihood ratio test, Model linking, p-value.

* This work was partially supported by Financiamento Base 2010 ISFL-1-297 from FCT/MCTES/PT.

References

Agresti, A. (1990). *Categorical Data Analysis.* J. Wiley & Sons: New York.

Lehmann, E.L. (1997). *Testing Statistical Hypotheses.* Reprint of the 2nd ed. publ. by Wiley 1986. Springer: New York.

Moreira, E.E., Paulo, A.A., Pereira, L.S. and Mexia, J.T. (2006). Analysis of SPI drought class transitions using loglinear models. *Journal of Hydrology 331*, 349–359.

Nelder, J.A. (1974). Loglinear models for contingency tables: a generalization of classical least squares. *Appl Stat. 23*, 323–329.

Zhu, R. and El-Shaarawi, A.H. (2009). Model clustering and its application to water quality monitoring. *Envirometrics 20*, 190–205.

Rank tests of symmetry with measurement errors

Radim Navrátil

Charles University, Prague, Czech Republic

Abstract

In many practical applications we often need to test the hypothesis that the new treatment is better than the current, or that older twin has different properties than younger, or for examle the left eye can see sharper than the right one. In all these situations we use one-sample test of symmetry. We will consider rank tests for their simplicity, robustness and other profitable properties. In many cases when the values of the random variable of our interest are obtained by measurement can happen that we do not get the accurate value of the random variable, but we get the value affected by measurement error. Application of parametric methods in this case in not very convenient, because we do not know the exact distribution of the errors and their estimation can make the situation more difficult. We will show how easy the rank tests can deal with this situation.

Keywords

Ranks, Tests of symmetry, Measurement errors.

References

Jurečková, J. and Picek, J. (2009). Rank tests in partially linear and measurement errors models. Submitted.

Jurečková, J., Picek, J. and Saleh, A.K.Md.E. (2009). Rank tests and regression rank score tests in measurement error models. *Comput. Statist. Data Anal.*, doi: 10.1016/j.csda.2009.08.020.

Jurečková, J., Kalina, J., Picek, J. and Saleh, A.K.Md.E. (2009). Rank tests of linear hypothesis with measurement errors both in regressors and responses. *KPMS Preprint 66*, MFF UK, Praha.

Jordan Algebras - "a first bite"

Sandra Nunes[1], Sandra Monteiro[1], Sandra Oliveira[1], Dina Salvador[1] and João T. Mexia[2]

[1] Polytechnic Institute of Setúbal, Portugal
[2] New University of Lisbon, Portugal

Abstract

The main goal of this presentation is to give a general perspective on Jordan algebras, their structure and their role in ares such as Quantum Mechanics, Physics, Analysis, Geometry, Statistics and more recently Optimization and Combinatory. It is not our purpose to describe the Jordan theory in an exhaustive way but only to introduce this so important subject, highlighting those who have contributed most to that development.

Keywords

Jordan algebras, The Jordan River, The classical formulation, Zel'manov's Exceptional Methods.

References

Jacobson, N. (1968). Structure and Representations of Jordan Algebras. *Amer. Math. Soc. Colloq. Publ. 39*.

Jordan, P., Von Neumann, J. and Wigner, E. (1934). On an algebraic generalization of the quantum mechanical formalism. *Ann. Math. 36*, 29–64.

Koecher, M. (1999). *The Minnesota Notes on Jordan Algebras and their Applications*. Lecture Notes in Mathematics, Springer.

McCrimmon, K. (2004). *A Taste of Jordan Algebras*. Springer.

McCrimmon, K. (1978). Jordan algebras and their applications. *Bull. Amer. Math. Soc. 84*, 612–627.

Asymptotic expansions in multi-group analysis of moment structures with an application to linearised estimators

Haruhiko Ogasawara

Otaru University of Commerce, Japan

Abstract

Asymptotic expansions of the joint distributions of functions of sample means and central moments up to an arbitrary order in multiple populations are given by Edgeworth expansions. The asymptotic distributions of the parameter estimators in moment structures under null/fixed alternative hypotheses and the chi-square statistics based on asymptotically distribution-free theory under fixed alternatives are given as applications of the above results. Asymptotic expansions of the null distributions of the chi-square statistics are also derived. For parameter estimators with the chi-square statistic, the linearized estimators are dealt with as well as fully iterated estimators.

Keywords

Moment structure analysis, Asymptotic expansions, Multiple populations, ADF theory, Implicit functions.

References

Ogasawara, H. (2010). Asymptotic expansions in multi-group analysis of moment structures with an application to linearised estimators. *Comm. Statist. Theory Methods.* To appear.

Small area estimation with a longitudinal area level model under restrictions

Luis N. Pereira[1] and Pedro S. Coelho[2]

[1] University of Algarve, Portugal
[2] New University of Lisbon, Portugal

Abstract

Large scale sample surveys are usually designed to produce reliable design-based estimates of various characteristics of interest for large geographical regions or subgroups of a population. However, for effective planning in a wide variety of fields, there is a growing demand to produce similar estimates for smaller geographical areas and sub-populations for which adequate samples are not available. In fact, sample sizes are often very small or even zero in many small areas of interest, which results in unreliable direct design-based small area estimates. This makes it necessary to employ indirect estimators that Şborrow informationŤ from related small areas through linking models, using recent census and current administrative data, in order to increase the effective sample size and thus precision. Such indirect estimators are often based on explicit linear mixed models that provide a link to a related small area through the use of supplementary data. The empirical best linear unbiased prediction (EBLUP) approach is the most popular method for the estimation of small area parameters of interest. However, in order to guarantee the calibration in publication and to protect against possible model misspecification, the mean of the small area estimates is often forced to equal the design-based estimate for a larger region for which the design-based estimate is sufficiently accurate. Thus, calibrated or benchmarked small area estimators are needed. In this paper, it is proposed a benchmark small area estimator based on a longitudinal area level model under restrictions. It is showed that such estimator, which guarantees the concordance between the model-based small area estimates and the design-based estimate for a larger region, is the best linear unbiased predictor (BLUP) accordingly to the overall criterion defined by Wang *et al.* (2008). It is considered a direct and a synthetic estimator for a larger region. The mean squared error (MSE) of the benchmarked predictor is also discussed. Finally, it is presented an empirical study using data from the Prices of the Habitation Transaction Survey conducted by the Portuguese Statistical Office.

Keywords

Calibration, Linear mixed models, MSPE of the benchmarked EBLUP, Small area estimation.

References

Pfeffermann, D. and Tiler, R.B. (2006). Small area estimation with state space models subject to benchmark constraints. *J. Amer. Statist. Assoc. 101(476)*, 1387–397.

Rao, J.N.K. and Yu, M. (1994). Small-area estimation by combining time-series and cross-sectional data. *Canad. J. Statist. 22(4)*, 511–528.

Ugarte, M.D., Militino, A.F. and Goicoa, T. (2009). Benchmarked estimates in small areas using linear mixed models with restrictions. *Test 18(2)*, 342–364.

Wang, J., Fuller, W.A. and Qu, Y. (2008). Small area estimation under a restriction. *Surv. Methodol. 34(1)*, 29–36.

Boosting stumps to determine the genes involved in cell proliferation due to ascorbic acid

Joaquim Pinto da Costa and Filipe Sousa

University of Porto, Portugal

Abstract

Boosting is an iterative algorithm that performs a linear combination of multiple predictions (function estimates) and can be applied both to regression and classification. Usually the individual predictors have mediocre performance in terms of misclassification error rates, like for instance stumps which are decision or regression trees with just one split. Nevertheless, the final linear combination found by boosting improves the performance of base classifiers, both theoretically and empirically, and produces a highly accurate classification rule. In our application we have a very small number of observations (9) and thousands of variables (genes) and so it will be very difficult to find a single accurate classifier. For that reason we decided to use stumps, as the data is not enough to build larger decision trees. However, by building many stumps, which will model different aspects of the data, and then combine them linearly with boosting, we hope to find a good classifier.

Our individual classifiers, stumps, are expected to present a large bias, due to their simplicity. Boosting is a bias reduction technique and typically improves the performance of a single (simple) tree model. The procedure starts by building in the first iteration a stump to predict if a certain example belongs to class ω_1 (in our application ascorbic acid) or not. Those observations which were not correctly classified by this stump will have a larger weight in the next iteration; the opposite happens with the ones correctly classified. In the end we have a set of M models (stumps in our case) and a weighted linear combination of them is found as the final model.

The ada package (Culp et al. 2006) implements the original AdaBoost algorithm (Freund and Schapire 1996, 1997) with other extensions. Some important features incorporated in this package include the use of both regression and classification trees for boosting and also various useful plots that aid in assessing variable importance and relationships between subsets of variables. This last feature is particularly interesting for us because, as we have thousands of variables (genes), we want not only an accurate final prediction model but mainly a way of choosing amongst those thousands of genes, the ones that matter.

The motivation for this work is to identify the genes involved in cell proliferation due to ascorbic acid (AA) and its stable form ascorbic acid 2-phosphate

(AA2P). Our skin, besides providing cover for the underlying soft tissues, it also performs many additional functions, including protection against injury, bacterial invasion and desiccation. Obviously any lesion must be rapidly and efficiently repaired in order to keep homeostasis. Skin consists of two layers: an outer epidermis and a deeper connective tissue layer, the dermis. Fibroblasts, the most abundant cell type in the connective tissue, are responsible for the synthesis of almost the entire extracellular matrix and thus contribute to skin regeneration. There is evidence that fibroblasts found at sites of lesions proliferate more and more actively secrete extracellular matrix.
Ascorbic acid (AA), also known as vitamin C, is a sugar acid with antioxidant properties. ROS (reactive oxygen species that appear due to lesions and cytotoxic molecules) which contain unpaired electrons may interact with nucleic acids, proteins or lipids destroying them. Ascorbic acid can terminate these chained radical reactions by being a stable electron donor in interactions with free radicals.
Ascorbic acid 2-phosphate (AA2P) is a stable form of vitamin C and, as ascorbic acid, it is involved in enhanced cell proliferation and relative rate of extracellular matrix synthesis.
Our aim is to determine the genes involved in cell proliferation due to AA and AA2P treatments and we use the same dataset that was used in (Duarte & al. 2009). The dataset is a microarray matrix where the whole human genome is scanned to extract the expression profiles. Both treatments and controls (scorbutic cells) were analyzed in triplicate.
Important genes are those where expression profiles of AA and AA2P treated cells are significantly higher than scorbutic cells. This means a particular gene over expression being transcriptional products more abundant and available to cell proliferation.

Keywords

Boosting, Decision trees, Supervised classification, Microarrays.

References

Culp, M., Johnson, K. and Michailidis, G. (2006). ada: An R package for stochastic boosting. *J. Statist. Software 17*, 1–27.

Freund, Y. and Schapire, R. (1996). Experiments with a new boosting algorithm. In *International Conference on Machine Learning*, 148–156.

Freund, Y. and Schapire, R. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. System Sci. 55(1)*, 119–139.

Duarte, T.L., Markus, S.C. and George, D.D. (2009). Gene expression profiling reveals new protective roles for vitamin C in human skin cells. *Free Radical Biology & Medicine 46*, 78–87.

Comparison of several linear statistical models to predict tropospheric ozone concentrations

José C. M. Pires, Fernando G. Martins, Maria C. M. Alvim-Ferraz and Maria C. Pereira

University of Porto, Portugal

Abstract

This study aims to evaluate the performance of five linear statistical models in the prediction of the next day hourly average ozone (O_3) concentrations. The selected models are: (i) multiple linear regression (MLR); (ii) principal component regression (PCR); (iii) independent component regression (ICR); (iv) quantile regression (QR); and (v) partial least squares regression (PLSR). As far as it is known, no study was presented comparing the performance of these five linear models for predicting tropospheric ozone concentrations. Moreover, it is the first time that ICR is applied with this aim. The considered ozone predictors are meteorological data (hourly averages of temperature, relative humidity and wind speed) and environmental data (hourly average concentrations of sulphur dioxide, carbon monoxide, nitrogen oxide, nitrogen dioxide and ozone) of the previous day collected in an urban site with traffic influences. The analysed period was May and June 2003.

The QR model, which tries to model the entire distribution of the O_3 concentrations, presents better performance in the training step, because it tries to model the entire distribution of the O_3 concentrations. However, it presents worst predictions in the test step. This means that a new procedure should be found better than the one applied (k nearest neighbours algorithm) that can be able to estimate the percentiles of the output variable in the test dataset with more precision. From the five statistical models tested in this study, the PLSR model presents the best predictions of tropospheric ozone concentrations.

Keywords

Air pollution, Tropospheric ozone, Statistical models, Concentration level prediction.

Orthogonal families for one and two tier prime basis factorials

Paulo C. Rodrigues[1,2], Vera Jesus[1] and João T. Mexia[1]

[1] New University of Lisbon, Portugal
[2] Wageningen University and Research Centre, The Netherlands

Abstract

When the experiments in a family have the same structure and correspond to the treatments of an orthogonal design, we have an orthogonal family of experiments.
The aim of this paper is to show how Jordan algebras (Seely, 1970a, b, 1971; Seely and Zyskind, 1971; Fonseca et al., 2006; Jesus et al., 2009) can be used to derive the appropriate analysis of variance for certain classes of linear regression model, namely to carry out joint analysis for orthogonal families of experiments. The case in which the individual experiments are one and two strata prime basis factorials is singled out.

Keywords

Commutative Jordan algebras, Double tier, Factorial designs, Families of models, Orthogonal models.

References

Fonseca, M., Mexia, J.T. and Zmyślony, R. (2006). Binary operations on Jordan algebras and orthogonal normal models. *Linear Algebra Appl. 417*, 75–86.

Jesus, V., Mexia, J.T., Fonseca, M. and Zmyślony, R. (2009). Binary operations and canonical forms for factorial and related models. *Linear Algebra Appl. 430*, 2781–2797.

Seely, J. (1970). Linear spaces and unbiased estimation. *Ann. Math. Stat. 41*, 1725–1734.

Seely, J. (1970). Linear spaces and unbiased estimation - application to the mixed linear model. *Ann. Math. Stat. 41*, 1735–1748.

Seely, J. (1971). Quadratic subspaces and completeness. *Ann. Math. Stat. 42*, 710–721.

Seely, J. and Zyskind, G. (1971). Linear spaces and minimum variance unbiased estimation. *Ann. Math. Stat. 42*, 691–703.

Asymptotically optimal linear bias corrections in minimum mean square error estimation

Jordi Serra and Francisco Rubio

Technological Centre for Telecommunications of Catalonia, Spain

Abstract

We investigate the class of bias corrections given by affine functions of parameter estimators minimizing the mean square error (MSE) in a general linear model. We consider both linear minimum MSE estimators in a Bayesian setting as well as linear unbiased minimum variance estimators that are given as a function of the population covariance matrix of the observation.
In principle, naive methods based on simple scaling and leading to an optimal correction factor that depends upon the population covariance matrix are not realizable if the latter is unknown. We concentrate on the set of solutions relying on directly replacing the covariance matrix with its sample covariance matrix estimator. By resorting to some recent results from random matrix theory, we derive a family of realizable estimators that are asymptotically optimal under a general asymptotic regime considering high dimensional observations and relatively small sample-sizes.

Keywords

Linear estimation, Bias correction, Sample covariance matrix, Random matrix theory.

Concordance correlation coefficient: an incursion into virtual reality

Júlia Teles[1], Emília Duarte[2], Luís Teixeira[1] and Francisco Rebelo[1]

[1] Technical University of Lisbon, Portugal
[2] Superior School of Design, Lisbon, Portugal

Abstract

Lin (1989) defined a new reproducibility index, named concordance correlation coefficient, to quantify the agreement between two measurements. This coefficient has two components: precision, which evaluates how far the observations deviate from the adjusted linear regression line, and accuracy, that takes account of the deviation between the regression line and the concordance line. To quantify the agreement between more than two measurements several generalizations have been proposed (e.g. Barnhart et al., 2002). If the objective is to give emphasis to the agreement of the lower or higher observations, a weighted concordance coefficient should be used, e.g., the top-down concordance coefficient (Iman and Conover, 1987).
Virtual Reality (VR) allows the generation of realistic scenarios that have been used to study people behavior in several scopes, namely to assess behavioral compliance with safety warnings (Duarte et al., 2010). The type of warnings and the level of environmental clutter can result in different space exploration strategies. The concordance correlation coefficient was used to evaluate the agreement of space exploration maps among five experimental conditions, defined by the type of warning and the level of clutter. Top-down concordance coefficient was also used and the results of these two coefficients were compared. The potentialities of these measures of agreement in the analysis of space exploration maps in VR-based studies are discussed.

Keywords

Agreement, Concordance correlation coefficient, Virtual reality.

References

Barnhart, H.X., Haber, M. and Song, J.L. (2002). Overall concordance correlation coefficient for evaluating agreement among multiple observers. *Biometrics 58*, 1020–1027.

Duarte, M.E.C., Rebelo, F. and Wogalter, M. (2010). The potential of virtual reality (VR) for evaluating warning compliance. *Hum. Factors Ergon. Manuf. 20(6).* In print.

Iman, R.L. and Conover, W.J. (1987). A measure of top-down correlation. *Technometrics 29(3)*, 351–357.

Lin, L.I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics 45*, 255–268.

Effect of data discretization on the classification accuracy in a high-dimensional framework

Annika Tillander and Tatjana Pavlenko

Stockholm University, Sweden

Abstract

Computionally efficient methods for classification are known to preform better when they are based on the discrete data, and this is especially important in a high dimensional setting, i.e. when the number of feature variables, p, is comparable to or greater than the number of observations, n. The goal of this on-going study is to empirically evaluate discretization of the continuous features and explore the effect of this procedure on the performance properties of a high dimensional classifier.

In order to represent a of variety discretization measures we focus on the hierarchical structure in Peng et al. (2009) and embed it into a high dimensional framework. We then explore six different discretization methods that can be seen in terms of *supervised* or *unsupervised*. Supervised discretization methods, also known as class-driven discretization technique, utilize the class membership, whereas unsupervised methods are known as thresholding as they mainly discretize the feature variables with regard to interval width or interval frequency.

As the goal of classification is to correctly predict a class membership of an observation we optimize the discretization procedure using the misclassification probability as a measure of the classification accuracy. To compare the classification performance between continuous feature and discretized ones we consider three supervised methods, k-nearest neighbor, Naive Bayes and a type of C4.5. To capture the effect of high dimensionality we investigate a variety of p for a fixed n, which makes it possible to evaluate the combined effect of the discretization and growing dimensionality.

Since the discretization is a data transformation procedure another aspect of this step is to investigate how the dependence structure is affected by the discretization. Accounting for such structures can improve accuracy and lead to models that are more interpretable (see Deng and Yuan 2009). For comparison of the dependence structure for discretized features with the original ones, two different performance measures were used; Frobenius norm and Kullback-Leibler loss. Various types of covariance were considered in the data generating process to explore the effect of discretization on the data dependence structure.

Current results reveal that a supervised entropy-based measure gives the over all lowest misclassification probabilities, whereas it alters the dependence

structure mostly. The discretization method that best retain the dependence structure is an unsupervised binning measure.

Keywords

Supervised classification, Discretization, High dimensionality, Misclassification probability.

References

Peng, L., Qing, W. and Yuija, G. (2009). Study on comparison of discretization methods. *4th International Conference on Artificial Intelligence and Computional Intelligence*, 380–384.

Deng, X. and Yuan, M. (2009). Large Gaussian covariance matrix estimation with Markov structures. *J. Comput. Graph. Statist. 18(3)*, 640–657.

An application of Structural Equation Modeling to test social support and physical symptoms as predictors of Quality of Life and Subjective Well-being in persons with chronic disease

Estela Vilhena[1,2], José Luís Pais Ribeiro[2,6], José Maia[2], Isabel Silva[3], Luísa Pedro[4,6], Rute Meneses[3], Helena Cardoso[2,5], Madalena Abreu[2], Mariana Henriques[2], Vera Melo[2], Ana Martins[5], António Martins da Silva[2,5] and Denisa Mendonça[2]

[1] Polytechnic Institute of Cávado and Ave, Portugal
[2] University of Porto, Portugal
[3] University of Fernando Pessoa, Portugal
[4] Polytechnic Institute of Lisbon, Portugal
[5] Hospital Center of Porto, Portugal
[6] UIPES Portugal

Abstract

The aim of the present study was to test a hypothetical model to examine whether social support and physical symptoms are relevant predictors of Quality of Life (QoL) domains (general, physical and mental well-being) and Subjective Well-being (SWB), in Portuguese patients with chronic diseases. Structural Equation Models (SEM) were used to test the quality of the hypothesized model, estimating the simultaneous effects of predictors on the outcome variables. A sample of 572 patients was recruited from central hospitals in various districts of Portugal. All completed self-reported questionnaires assessing socio-demographic and clinical variables, social support (MOS Social Support Survey) and physical symptoms (Portuguese version of Psychosomatic Symptom Checklist). The hypothesized model fitted the data reasonably well. It was found that social support had an independent positive impact on both QoL and SWB, after controlling for physical symptoms, whereas physical symptoms had a negative impact on the QoL and SWB. These findings suggest that, the greater the social support patients perceive the more facilitated will be their adjustment, which in turn will affect their quality of life.

Keywords

Chronic disease, Quality of life, Structural equation modeling.

Performances of the positive-rule Stein-type r-k class estimator in linear regression

Jianwen Xu and Hu Yang

Chongqing University, China

Abstract

In this paper, the Stein-type r-k class estimator and positive-rule Stein-type r-k class estimator are introduced for the parameter vector in a linear regression model when it is suspected that the parameter vector may be restricted to a linear manifold. Bias and quadratic risks of the proposed estimators are derived and some sufficient conditions on the ridge parameter k and the departure parameter Δ are derived for the superiority of the positive-rule Stein-type r-k class estimator over the r-k class estimator by Baye and Parker (1984), the restricted r-k class estimator by Xu and Yang (2010) and the Stein-type r-k class estimator, respectively.

Keywords

Uncertain prior information, r-k class estimator, Restricted r-k class estimator, Positive-rule Stein-type r-k class estimator, Quadratic risk.

References

Baye, M.R. and Parker, D.F. (1984). Combining ridge and principal component regression: A money demand illustration. *Comm. Statist. Theory Methods 13(2)*, 197–205.

Xu, J. and Yang, H. (2010). On the restricted r-k class estimator and restricted r-d class estimator in linear regression. *J. Stat. Comput. Simul.* To appear.

Part VIII

Two Birthday Boys

Many happy returns, Simo

Erkki P. Liski

University of Tampere, Finland

Fig. 1. Erkki Liski and Simo Puntanen, Montreal 1976.

Simo Puntanen was born on the 20th July 1945 at Peräseinäjoki, the eldest son of Viljami Puntanen and his wife Kyllikki, before the family moved to Tampere, where he spent his childhood. He married Soile in the late 1960s and their first daughter Niina was born at the appointed time.

Simo started his studies in 1965 at a new faculty of the University of Tampere – the Faculty of Economics and Administration which was founded just in 1965. I started my studies one year later at the same faculty. In those days our university went through a period of rapid growth and development. The professorship in computer science – the first one in the Nordic countries – was founded in 1965, the full professorship in statistics in the same year and

Fig. 2. Erkki and Leena Liski, Anna, Simo, Niina and Soile Puntanen, Finland 1976.

the mathematics program started at the beginning of 1967. Just statistics, mathematics and computer science were also Simo's main subjects.

Soon after his basic studies in the late 60's Simo acted as part time assistant in Mathematics. As far as I remember, I attended his mathematics exercise group in the spring term 1967. In 1970 he started acting as lecturer is Statistics. In these times we got to know each other, as well as Pentti Huuhtanen and Tarmo Pukkila. Our professor was Eino Haikala – a very inspiring teacher and a great personality. He was not a usual professor, but also a farmer, gardener and a divine storyteller, among others.

This was the beginning time, and there was no research tradition at our department. So we had no patterns of research or supervision of postgraduate studies. Everybody had to be self-guided. We hardly had any clear understanding of the demands of an academic career. However, Simo got a research training which was better than the training according to our usual standards. He attended a special research training program that was a pilot experiment for a new type of research training. The idea was to give in the beginning of studies in short time intensive training especially in mathematics, statistics and computer science, i.e., in subjects which were considered important in

Fig. 3. Simo Puntanen (ISR Short Book Reviews Editor), Ali S. Hadi (ISR Editor-in-Chief) and Erkki Liski at the statistics library, Tampere 1991.

Fig. 4. Jan Hauke and Simo Puntanen, Krakow, Poland 2009.

Fig. 5. Augustyn Markiewicz, Simo Puntanen, and Götz Trenkler, Tampere 2007.

research. This pilot lasted two years and these ideas were due to put into action in a new faculty of methodological sciences. The faculty did not become true and the pilot training was put into practice only once. The soul of this new methodological research training was young professor Seppo Mustonen, now Professor Emeritus from the University of Helsinki. Mustonen was an advisor of the strong and charismatic rector Paavo Koli and Mustonen had an important role in the emergence of the faculty of Economics and Administration. But the new faculty of methodological sciences did not become true, and disappointed with this turn of events Mustonen left Tampere.

Throughout his career Simo has been very active in developing the use of computer in teaching of statistics. Especially the Survo system – a creation of Seppo Mustonen – is his favourite. Before coming to Tampere in the mid-1960s Mustonen worked at the electronics department of the Finnish Cable Factory, that was one of the first places in Finland where the research of automatic data processing started. Mustonen's task was to take care of statistical programming and there he started developing a statistical program Survo. Various forms of Survo have existed during the last 50 years and it has developed to a general computing environment. The use of Survo system in teaching basic statistics courses in Tampere University started since the year 1975 and Simo was heavily involved in this development work. The role of computer in teaching of statistics has also been one central theme in his

Fig. 6. Harri Hietikko, Simo Puntanen, Tarmo Pukkila, and Pentti Huuhtanen in COMPSTAT 84, Prague 1984.

Fig. 7. Simo Puntanen, C.R. Rao (Opponent), Eino Haikala (Custodian). Simo's doctoral dissertation in Tampere 1987.

research. Simo has taken care of building and sustaining the Survo culture in our University for decades to date.

Professor Haikala retired in 1976 and Tarmo Pukkila became the professor of statistics. Tarmo started to emphasize the importance of research and the necessity of international co-operation. In practice this meant attending and organizing conferences, inviting foreign scientists etc. However, before the start of this primary era of international co-operation we did with Simo one major trip to United States. Simo has relatives in New Hampshire and he had decided to visit them in 1976. Presumably his plan was to travel with his wife but they had two small children and Soile was pregnant with her third child when the trip was due to start in August 1976. Their son Ville was happily born in 1977. So I became the substitute of Soile. We landed at the JFK International Airport, New York, and went then by bus to Keene in New Hampshire. George Styan had visited our University first time in January 1976, and hence we got the idea to drop in on George and McGill University, although George was not personally present in Montreal. I do not remember if we knew that in advance. In Vermont, on our way to Montreal, we run across in a motel a charming lady who was traveling by car to Montreal. She promised to take us with her to Montreal next day. Next morning the lady had vanished and we took a taxi to Montreal. At previous night the musicians of the band playing in the motel had tipped us off about interesting targets in Montreal and given us some addresses. In Montreal we dropped in one of the recommended targets. It took a while to realize that we were in a gay bar, and then we rushed out – we were severely homophobic. Montreal Olympics 1976 were already over but we visited olympic stadium, of course. These were the Lasse Viren Olympics and the long distance running was our sport.

In the 70s Simo organized many Nokia Summer Olympics. Nokia is the locality where Simo's summer cottage was situated and where he nowadays lives. The 1500 m, 5000 m or 10000 m run was the serious business, but the fabulous party after the games at the summer cottage was the real experience. The cottage was in a marvelous situation near a lake, and besides the cottage has an interesting history. Simo's farther Viljami was a journalist, like Simo's second daughter Anna nowadays. Viljami knew a lot of people, among others Arvo "Poika" Tuominen who was also a journalist and much more. Viljami bought the cottage from Arvo "Poika". His nickname Poika was given because his boyish looks – poika means "boy" in Finnish.

In 1977 Simo organized a very special and immemorial meeting with Arvo "Poika" Tuominen (1894–1981) who was a Finnish communist revolutionary and later a social democratic journalist and politician. We had an opportunity to discuss with Tuominen. Finally, it was not a discussion but more like Tuominen's presentation and performance. He was a very charismatic personality. In 1922 and subsequently he was imprisoned for maintaining contacts with the Soviet Union and the banned Communist Party. In 1932 Tuominen was paroled and then he secretly went to the Soviet Union. He was given a

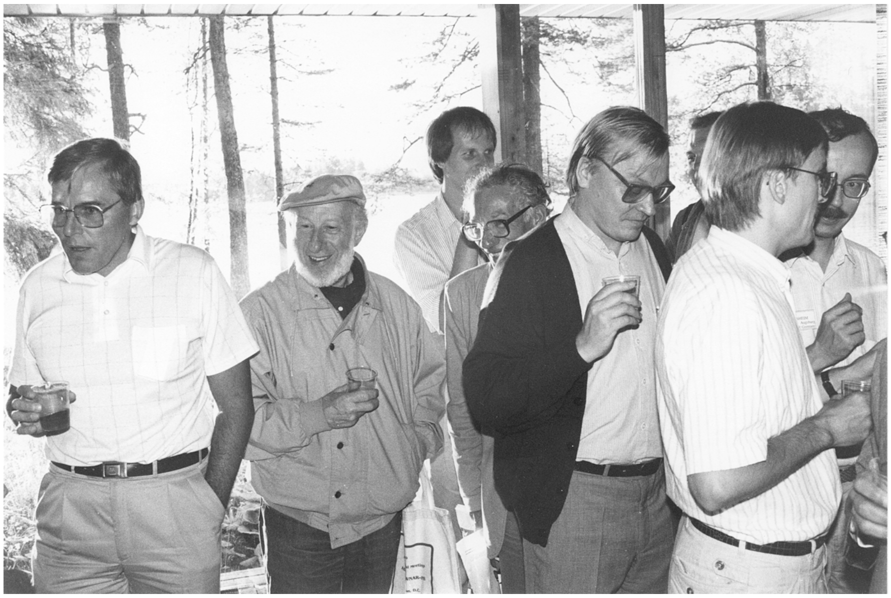


Fig. 8. Tarmo Pukkila, Ingram Olkin, Dietrich von Rosen, Heinz Neudecker, Erkki Liski, Simo Puntanen, Friedrich Pukelsheim. Sauna party in Tampere 1990.

Fig. 9. George Styan, Simo and Soile Puntanen, Halifax 2003.

crash course at the Lenin Party School and was appointed General Secretary of the Finnish Communist Party, also becoming a member of the Comintern Executive Committee Presidium. He had direct personal contacts with Stalin and other big communist leaders. Tuominen witnessed the Great Purge first-hand until he was able to leave Moscow for Stockholm in early 1938. On November 13, 1939, Stalin ordered him to return to Moscow in order to become the head of the Communist government of the Finnish Democratic Republic which Stalin planned to install in Finland. Tuominen decided to stay in Stockholm. He has written many books on his life. Soon after this Tuominen meeting we started with Simo a joint project Stalin, Statistics in Linguistics, which inspired some joint research and led us to the 7th World Congress of Applied Linguistics in Brussels, 1984.

The First International Tampere Seminar on Linear Statistical Models and their Applications in 1983 was an important and a far-reaching event in the development of our statistics unit, and in 1987 followed The 2nd International Tampere Conference in Statistics, which was a much larger event. Simo was heavily involved in taking the initiatives to organize the conferences and putting them into practice. The International Workshop on Linear Models, Experimental Design and Related Matrix Theory organized in 1990 started a series of annual events, of which the workshop organized this year in Shanghai was the 19th. Simo has had, as we know, a central role in organizing these conferences, not forgetting his role in organizing many other conferences too. LinStat1984 in Poznań was an impressive experience for Simo and me. Then we had first time an opportunity to meet famous Polish statisticians like Caliński, Baksalary, Kala etc. The party organized by Jan Hauke was immemorial and Jerzy Baksalary was the leading figure there. Thereafter Polish–Tampere exchange in Statistics started. Now Jerzy's students like Augustyn Markiewicz are at forefront. Simo has been a leading light in cultivating our friendship and collaboration with the Polish School of Statisticians. So, it is certainly very appropriate to propose a toast to Simo for his 65th birthday in LinStat2010. Many happy returns, Simo!

Acknowledgements

Special thanks go to Jarmo Niemelä for his assistance with the final editing of this writing, and in particular, for photo processing. Thanks go also to Pentti Huuhtanen for giving two photos. Many of the photos have been stolen without permission from Simo's collections.

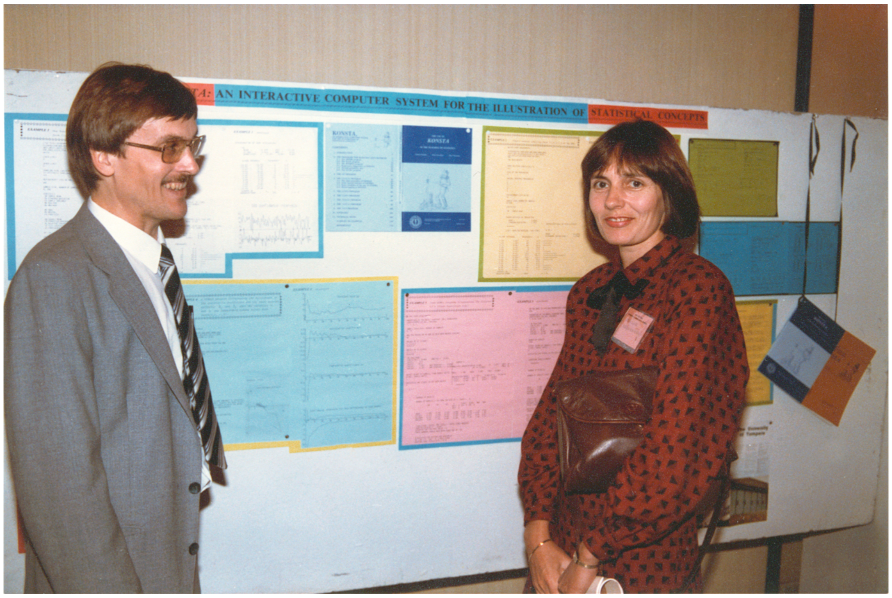


Fig. 10. Simo Puntanen and Greta Ljung, Frontiers in Computational Statistics, Boston 1984.

Fig. 11. Simo and Soile (right) Puntanen, the Sword-Whetting in Tampere 2000.

A photo album for Simo Puntanen in celebration of his 65th birthday on 20 July 2010

George P. H. Styan

McGill University, Montreal, Canada

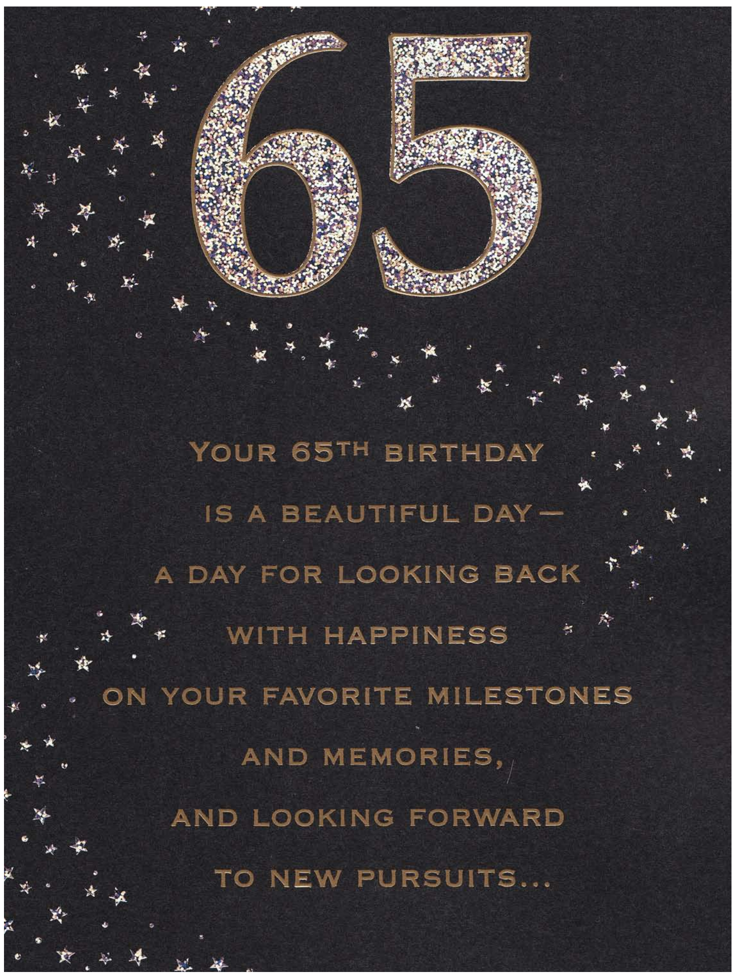
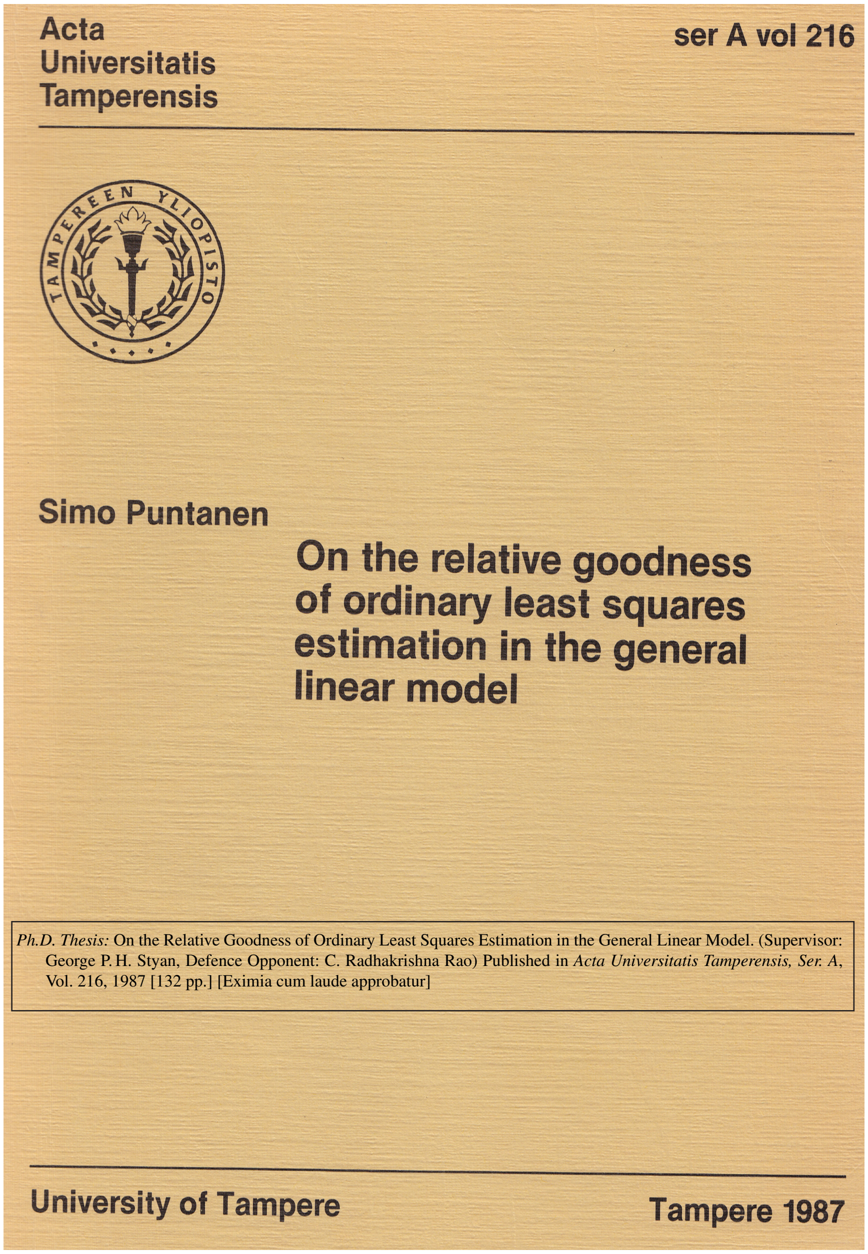


Fig. 1. Many Happy Returns, Simo: 20 July 2010.



Ph.D. Thesis: On the Relative Goodness of Ordinary Least Squares Estimation in the General Linear Model. (Supervisor: George P. H. Styan, Defence Opponent: C. Radhakrishna Rao) Published in *Acta Universitatis Tamperensis, Ser. A*, Vol. 216, 1987 [132 pp.] [Eximia cum laude approbatur]

Fig. 2. Simo Puntanen, PhD: *Eximia cum laude approbatur*.

Fig. 3. Four generations: Jarkko Isotalo, Simo Puntanen, George P. H. Styan, T. W. Anderson; Uppsala, Sweden 2006. (Photo by Soile Puntanen).

Fig. 4. The 8th International Workshop on Matrices and Statistics: Tampere, Finland, 7–8 August 1999. Local and International Organizing Committee Chair: Simo Puntanen.

The Friend and Professor João Tiago Mexia

Francisco Carvalho[1], Paulo C. Rodrigues[2], Elsa Moreira[2] and Miguel Fonseca[2]

[1] Polytechnic Institute of Tomar, Portugal
[2] New University of Lisbon, Portugal

Fig. 1. João Tiago Mexia, Będlewo, Poland, 2009.

João Tiago Praça Nunes Mexia was born in the 2nd of June of 1939, the third child of a traditional catholic family.

In 1962 he got his degree in Forestry Engineering from the Technical University of Lisbon. Simultaneously, in 1961 he also got a degree in Pulp and Paper Technology, a Technical degree for Forestry Workers Formation (sponsored by FAO in Sweden) and in 1962 a degree in Tropical Forestry.

Between 1970 and 1971 he got approved in a high number of courses of Mathematics at Faculty of Sciences of the University of Lisbon, all of them with excellent grades.

From 1961 to 1962 he worked at the Missão de Estudos Agronómicos do Ultramar [Agronomic Studies Overseas Commission], followed by a commission in the Portuguese army.

From 1964 to 1965, and while in the army, he was requested by the Instituto de Investigação Agronómica de Angola [Agronomic Research Institute of Angola], where he organized the statistical procedures.

In 1964, while in the army, he was the head of the Secção de Estatística do Instituto de Investigação Médica de Angola [Statistics Section of the Medical Research Institute of Angola].

Between 1967 and 1968, he got a scholarship of the Gulbenkian Foundation for a project in *Statistical Theory of Extreme Values*.

In 1968, he returned to the Missão de Estudos Agronómicos do Ultramar [Agronomic Studies Overseas Commission], becoming Senior Research Officer of Statistical Mathematics and Experimental Design section.

In 1976, he was promoted to Researcher of the Junta de Investigação Científica do Ultramar [Overseas Scientific Research Committee], being appointed the delegate of that committee at the Ministry of Education for the Statistics Consulting Committee.

In 1981, he was appointed member of the Gastroenterology Centre of Lisbon Universities, and in 1982 appointed as researcher of the Agriculture Production Centre of the Technical University of Lisbon. In the same year he was promoted to Principal Researcher at Junta de Investigação Científica do Ultramar [Overseas Scientific Research Committee] and appointed head of the Missão de Estudos Agronómicos do Ultramar [Agronomic Studies Overseas Commission].

In 1986, he was approved, by unanimity in absolute merit, for the position of Coordinator Researcher of the Instituto de Investigação Científica Tropical [Tropical Scientific Research Institute]. This period extends to 1989, during which he had to analyze a large number of experiments, not all properly designed. This lead him to try to generalize the assumption underlying the ANOVA method. These ideas were incorporated in his PhD Thesis "Heterocedasticidade Controlada, Espaços Vectoriais Quociente e Testes F para Hipóteses sobre Vectores Médios" ["Controlled Heterocedasticity, Quotient Vector Spaces and F Tests for Hypothesis on Mean Vectors"], which he presented, without a supervisor, at the Faculty of Sciences and Technology of the New University of Lisbon (FCT/UNL) in 1988. His PhD was unanimous approved with *summa cum laude*.

He started working in Universities in 1971. From 1971 to 1973 he was responsible for the course of Infinitesimal Calculus and Elements of Probability Theory at the Technical University of Lisbon. Later on, from 1979 to 1986, he was Invited Teaching Assistant (half time) of the Biomathematics Department of the Medical Sciences Faculty of the New University of Lisbon. From 1986 to 1988, he was Invited Assistant Professor at Mathematics Department of FCT/UNL where taught Probability and Statistics, Economical Statistics and Econometrics. After his PhD, in 1998, he became a full time Invited Associated Professor. In 1989, he got the position of Associated Professor and in 1992 was appointed as Full Professor at FCT/UNL. During this time, he supervised the teaching of Statistics at FCT/UNL.

Fig. 2. Fernando Santana and João Tiago Mexia, IWMS, Tomar, Portugal, 2008.

His parallel work connected with Risk Management and Actuarial Sciences made him: the elected president of the Committee of the Portuguese Actuaries Institute Bulletin; Member of the Consulting Committee of Portuguese Assurance Institute, appointed by the Minister of Finances; and Member of the Exam Committee for Actuaries.

His research activities continue to the present time, despite his status of Emeritus Full Professor since June 2009. He is currently associated with the PhD Program in Statistics and Risk Management at FCT/UNL. Moreover, from 1999 to 2009 he directed the research Center of Mathematics and Applications (CMA) of FCT/UNL. The Center was evaluated four times during this period by international panels, starting with “Fair” in 1999 and ending with “Very Good” in 2009. He has also organized several scientific meetings being part of Organizing and Scientific Committees.

Fig. 3. João T. Mexia, Roman Zmyślony and Miguel Fonseca, Będlewo, Poland, 2005.

His experience in Experimental Designs and Linear Statistical Inference lead him to collaborate with many institutions. Namely, (i) Coffee Institute of Angola; (ii) Technical University of Lisbon (ISA); (iii) Lisbon University (Medical School); (iv) Évora University; (v) Coffee Rusts Research Centre; (vi) Overseas Agriculture Garden Museum; (vii) Trás-os-Montes e Alto Douro University; (viii) Portuguese Plant Breeding Station; and (ix) Portuguese Agronomic Station.

Fig. 4. João T. Mexia, Manuela Oliveira, Miguel Fonseca and Carlos A. Coelho, Texas, USA, 2005.

Great part of his work was devoted to the supervision and guidance of PhD students. So far he has supervised 28 PhD thesis (15 as main supervisor and 13 as co-supervisor). This work was conducted at New University of Lisbon, University of Évora, University of Beira Interior and Open University. His ongoing research activity still guides and influences the work of 10 current PhD candidates.

Fig. 5. Some of João's PhD students (former and current) during his retirement ceremony, 2009.

His main research topic concerns Linear Statistical Inference, with over 70 papers published in international journals with referee. His broad view of statistics allowed him to participate in several multidisciplinary research projects in such areas as Environmental Sciences, Agronomy, Forestry, Computer Sciences and Medicine.

For the last ten years he has actively cooperated with Polish statisticians, namely Professors Tadeusz Caliński, Stanisław Mejza, Roman Zmyślony and Augustyn Markiewicz.

Fig. 6. Roman Zmyślony, Tadeusz Caliński and João T. Mexia, Wielkopolski National Park, Poznań, Poland, 2004.

Fig. 7. Tadeusz Caliński, Miguel Fonseca, João T. Mexia and Roman Zmyślony, Wielkopolski National Park, Poznań, Poland, 2004.

One of the highlights of his academic career was being the promoter of the *Honoris Causa* Degree awarded by the New University of Lisbon to Professor C. R. Rao in 2006.

Fig. 8. Stanisław Mejza, Tadeusz Caliński, João T. Mexia and Augustyn Markiewicz, C. R. Rao *Honoris Causa* Degree ceremony, Tomar, Portugal, 2006.

Fig. 9. C. R. Rao and João T. Mexia, C. R. Rao *Honoris Causa* Degree ceremony, Tomar, Portugal, 2006.

Acknowledgements

Special thanks go to Carlos A. Coelho and Dinis Pestana for allowing us to use some information from a former biography of Professor João T. Mexia.

Part IX

List of Participants

Participants

1. **Nihan Acar**
   Mimar Sinan Fine Arts University, Istanbul, Turkey,
   nihan.acar@msgsu.edu.tr
2. **S. Ejaz Ahmed**
   Department of Mathematics and Statistics, University of Windsor,
   Ontario, Canada, *seahmed@uwindsor.ca*
3. **Kadri Ulaş Akay**
   Arts and Sciences Faculty, Marmara University, Istanbul, Turkey,
   kulas@istanbul.edu.tr
4. **Aylin Alin**
   Department of Statistics, Faculty of Arts and Sciences, Dokuz Eylül
   University, Izmir, Turkey, *aylin.alin@deu.edu.tr*
5. **Özlem Gürünlü Alma**
   Mugla University, Kötekli/Muğla, Turkey, *ozlem.gurunlu@hotmail.com*
6. **Conceição Amado**
   Department of Mathematics, Technical Superior Institute, Lisbon,
   Portugal, *camado@math.ist.utl.pt*
7. **Mohamed Amezziane**
   Department of Mathematics, DePaul University, Chicago, Illinois, USA,
   mamezzia@depaul.edu
8. **Mariano Amo-Salas**
   Department of Mathematics, University of Castilla-La Mancha, Ciudad
   Real, Spain, *mariano.amo@uclm.es*
9. **Cristina Andrade**
   Department of Mathematics, Polytechnic Institute of Tomar, Tomar,
   Portugal, *c.andrade@ipt.pt*
10. **Barbora Arendacká**
    Institute of Measurement Science, Slovak Academy of Sciences,
    Bratislava, Slovakia, *barendacka@gmail.com*
11. **Emilia Athayde**
    Department of Mathematics, Minho University, Braga, Portugal,
    mefqa@math.uminho.pt
12. **Anthony Atkinson**
    Department of Statistics, The London School of Economics and
    Political Science, London, UK, *a.c.atkinson@lse.ac.uk*
13. **Cecília Azevedo**
    Department of Mathematics, Minho University, Braga, Portugal,
    cecilia@math.uminho.pt

14. **Manoochehr Babanezhad**
Department of Applied Mathematics ad Computer Science, Ghent University, Ghent, Belgium, *manoochehr.babanezhad@ugent.be*

15. **Rosemary A. Bailey**
School of Mathematical Sciences, Queen Mary, University of London, London, UK, *r.a.bailey@qmul.ac.uk*

16. **Yves Berger**
School of Social Sciences, University of Southampton, Southampton, UK, *y.g.berger@soton.ac.uk*

17. **Regina Bispo**
Applied Psychology High Institute, Lisbon, Portugal, *rbispo@ispa.pt*

18. **Anders Björkström**
Institute of Mathematics, University of Stockholm, Stockholm, Sweden, *bjorks@math.su.se*

19. **Luis Blanco**
Technological Center for Telecommunications of Catalonia, Barcelona, Spain, *lblanco@cttc.es*

20. **Olivia Bluder**
Alpen-Adria University of Klagenfurt, Klagenfurt, Austria, *olivia.bluder@k-ai.at*

21. **Barbara Bogacka**
School of Mathematical Sciences, Queen Mary, University of London, London, UK, *b.bogacka@qmul.ac.uk*

22. **João Branco**
Department of Mathematics, Technical Superior Institute, Lisbon, Portugal, *jbranco@math.ist.utl.pt*

23. **Carlos A. Braumann**
Department of Mathematics, University of Évora, Évora, Portugal, *braumann@uevora.pt*

24. **Nuno Brites**
Department of Mathematics, University of Évora, Évora, Portugal, *m44661@alunos.uevora.pt*

25. **Guoqiang Cai**
State Key Laboratory for Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China, *gqcai@bjtu.edu.cn*

26. **Nazif Çalış**
Department of Statistics, çukurova University, Adana, Turkey, *ncalis@cu.edu.tr*

27. **Derya Çalişkan**, Department of Statistics, Hacettepe University, Ankara, Turkey, *deryacal@hacettepe.edu.tr*

28. **Margarida Cardoso**
ISCTE, Lisbon University Institute, Lisbon, Portugal, *mgsc@iscte.pt*

29. **Francesco Carmona**
Department of Statistics, University of Barcelona, Barcelona, Spain, *fcarmona@ub.edu*

30. **Rute Carrujo**
Department of Mathematics, Polytechnic Institute of Tomar, Tomar, Portugal, *rute.carrujo@fct.unl.pt*

31. **Francisco Carvalho**
Department of Mathematics, Polytechnic Institute of Tomar, Tomar, Portugal, *fpcarvalho@ipt.pt*

32. **Miguel de Carvalho**
Center of Mathematics and Applications, Faculty of Sciences and Technology, New University of Lisbon, Caparica, Portugal, *mbcarvalho@fct.unl.pt*

33. **Bronisław Ceranka**
Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Poznań, Poland, *bronicer@up.poznan.pl*

34. **Övgü Çıdar**
Department of Mathematics, Eastern Mediterranean University, Cyprus, *ovgu.cidar@emu.edu.tr*

35. **Carlos Agra Coelho**
Mathematics Department, Faculty of Sciences and Technology, New University of Lisbon, Caparica, Portugal, *cmac@fct.unl.pt*

36. **Joaquim Costa**
Department of Mathematics, Faculty of Sciences, University of Porto, Porto, Portugal, *jpcosta@fc.up.pt*

37. **Ricardo Covas**
Department of Mathematics, Polytechnic Institute of Tomar, Tomar, Portugal, *ricardocovas@gmail.com*

38. **Carles Cuadras**
Department of Statistics, Faculty of Biology, Barcelona University, Barcelona, Spain, *ccuadras@ub.edu*

39. **Daniel Cuadras**
Catalan Oncology Institute, Barcelona, Spain, *dcuadras@iconcologia.net*

40. **Somnath Datta**
Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, Louisville, USA, *somnath.datta@louisville.edu*

41. **Susmita Datta**
Department of Bioinformatics and Biostatistics, School of Public Health and Information Sciences, University of Louisville, Louisville, USA, *susmita.datta@louisville.edu*

42. **Ali Deeb**
Department of Statistics, Faculty of Sciences, Al Fateh University, Tripoli, Libya, *deebali@hotmail.com*

43. **Anupam Deka**
Department of Statistics, Handique Girls' College, Guwahati, India, *anupamdeka4@rediffmail.com*

44. **José G. Dias**
ISCTE, Lisbon University Institute, Lisbon, Portugal, *jose.dias@iscte.pt*

45. **Jan Dienstbier**
Faculty of Natural Sciences, Humanities and Teaching, Technical University of Liberec, Liberec, Czech Republic, *dienstbi@karlin.mff.cuni.cz*

46. **Pierre Druilhet**
Mathematics Laboratory, Blaise Pascal University, Aubière Cedex, France, *pierre.druilhet@math.univ-bpclermont.fr*

47. **Hilmar Drygas**
Kassel University, Kassel, Germany, *hilmar@drygas.de*

48. **Rie Enomoto**
Tokyo University of Science, Tokyo, Japan, *j1410701@ed.kagu.tus.ac.jp*

49. **Birsen Eygi Erdogan**
Arts and Sciences Faculty, Marmara University, Istanbul, Turkey, *birsene@marmara.edu.tr*

50. **Ali Erkoc**
Arts and Sciences Faculty, Marmara University, Istanbul, Turkey, *ali66_math@hotmail.com*

51. **Manuel L. Esquível**
Department of Mathematics, Faculty of Sciences and Technology, New University of Lisbon, Caparica, Portugal, *mle@fct.unl.pt*

52. **Danya Facchinetti**
University of Milan, Milan, Italy, *danya.facchinetti@unimi.it*

53. **José Faias**
Economics Faculty, New University of Lisbon, Lisbon, Portugal *jfaias@fe.unl.pt*

54. **Susana Faria**
Department of Mathematics and Applications, University of Minho, Braga, Portugal, *sfaria@math.uminho.pt*

55. **Célia Ferandes**
Superior Institute of Engineering of Lisbon, Lisbon, Portugal, *cfernandes@deetc.isel.ipl.pt*

56. **Dário Ferreira**
Department of Mathematics, University of Beira Interior, Covilhã, Portugal, *dario@mat.ubi.pt*

57. **Sandra Saraiva Ferreira**
Department of Mathematics, University of Beira Interior, Covilhã, Portugal, *sandra@mat.ubi.pt*

58. **Katarzyna Filipiak**
Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Poznań, Poland, *kasfil@up.poznan.pl*

59. **Eva Fišerová**
Department of Mathematical Analysis and Applications of Mathematics, Palacký University of Olomouc, Olomouc, Czech Republic, *fiserova@inf.upol.cz*

60. **Miguel Fonseca**
Center of Mathematics and Applications, Faculty of Sciences and Technology, New University of Lisbon Caparica, Portugal, *fmig@fct.unl.pt*

61. **Raquel Gaspar**
Management Department, High Institute of Economy and Management, Technical University of Lisbon, Lisbon, Portugal, *rmgaspar@iseg.utl.pt*

62. **Dilip Kumar Ghosh**
Departmet of Statistics, Saurashtra University Rajkot, Rajkot, Gujarat, India, *ghosh_ dkg@yahoo.com*

63. **Janet Godolphin**
Department of Mathematics, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, UK, *j.godolphin@surrey.ac.uk*

64. **Dora Prata Gomes**
Faculty of Sciences and Technology, New University of Lisbon, Caparica, Portugal, *dsrp@fct.unl.pt*

65. **João Gomes**
Department of Statistics and Operations Research, Faculty of Engineering Sciences, University of Lisbon Lisbon, Portugal, *jjgomes@fc.ul.pt*

66. **Sónia Gouveia**
Department of Mathematics, Faculty of Sciences, University of Porto, Porto, Portugal, *sonia.gouveia@fc.up.pt*

67. **Małgorzata Graczyk**
Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Poznań, Poland, *magra@up.poznan.pl*

68. **Jan Graffelman**
Department of Statistics and Operations Research, Polytechnic University of Catalonia, Barcelona, Spain, *jan.graffelman@upc.edu*

69. **Luis Grilo**
Department of Mathematics, Polytechnic Institute of Tomar, Tomar, Portugal, *lgrilo@ipt.pt*

70. **Sat Gupta**
Department of Mathematics and Statistics, University of North Carolina, Greensboro, USA, *sngupta@uncg.edu*

71. **Chengcheng Hao**
Department of Statistics, Stockholm University, Stockholm, Sweden, *chengcheng.hao@stat.su.se*

72. **Lígia Henriques-Rodrigues**
Department of Mathematics, Polytechnic Institute of Tomar, Tomar, Portugal, *lhenriques@ipt.pt*

73. **John Hinde**
Department of Mathematics, National University of Ireland, Galway, Ireland, *john.hinde@nuigalway.ie*

74. **Yoshihiro Hirose**
Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo, Tokyo, Japan, *hirose@stat.t.u-tokyo.ac.jp*

75. **Masashi Hyodo**
Tokyo University of Science, Tokyo, Japan, *hyodoh_h@yahoo.co.jp*

76. **Deniz İnan**
Department of Methodology and Statistics, Arts and Sciences Faculty, Marmara University Istanbul, Turkey, *denizlukuslu@marmara.edu.tr*

77. **Shusong Jin**
Fudan University, Shanghai, China, *jinss@fudan.edu.cn*

78. **Charles R. Johnson**
Department of Mathematics, College of William and Mary, Williamsburg, Virginia, USA, *crjohnso@math.wm.edu*

79. **Michael Jones**
Department of Psychology, Faculty of Human Sciences, Macquarie University, Sydney, Australia, *mike.jones@psy.mq.edu.au*

80. **Tomáš Jurczyk**
Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Prague, Czech Republic, *jurczyk@karlin.mff.cuni.cz*

81. **Krystyna Katulska**
Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poznań, Poland, *krakat@amu.edu.pl*

82. **Soleiman Khazaei**
University of Dauphine, Paris, France, *khazaei@ceremade.dauphine.fr*

83. **Daniel Klein**
Faculty of Sciences, Pavol Josef Šafárik University in Košice, Košice, Slovakia, *daniel.klein@upjs.sk*

84. **Kazuyuki Koizumi**
Tokyo University of Science, Tokyo, Japan, *zumisj@rs.kagu.tus.ac.jp*
85. **Leonid Kopylev**
US Environmental Protection Agency, Washington, D.C., USA, *kopylev.leonid@epa.gov*
86. **Özlem Korucu**
Istanbul University, Istanbul, Turkey, *kozlem@istanbul.edu.tr*
87. **Maria Kozłowska**
Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Poznań, Poland, *markoz@up.poznan.pl*
88. **Jitender Kumar**
Kurukshetra University, Kurukshetra, India, *kulwantsinghus@yahoo.com*
89. **Naresh Kumar**
National Institute of Science Technology & Development Studies, New Delhi, India, *nareshqumar@yahoo.com*
90. **Izabela Kurzydło**
Faculty of Mathematics, Computer Science and Econometrics, University of Zielona Góra, Zielona Góra, Poland, *I.Kurzydlo@wmie.uz.zgora.pl*
91. **Lynn Roy LaMotte**
Biostatistics Program, School of Public Health, Louisiana State University, New Orleans, USA, *llamot@lsuhsc.edu*
92. **Ying Li**
Department of Energy and Technology, Biometrics, and Systems Analysis, Swedish University of Agriculture Sciences, Uppsala, Sweden, *Ying.Li@et.slu.se*
93. **Xi Li**
State Key Laboratory of Rail Traffic Safety and Control, Beijing Jiaotong University, Beijing, China, *bjtulx@yahoo.com*
94. **Yu Liang**
State Key Laboratory for Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China, *09114224@bjtu.edu.cn*
95. **Antti Liski**
Tampere University of Technology, Tampere, Finland, *antti.liski@tut.fi*
96. **Eero Liski**
Department of Mathematics and Statistics, University of Tampere, Tampere, Finland, *eero.liski@tut.fi*
97. **Erkki P. Liski**
Department of Mathematics, Statistics and Philosophy, University of Tampere, Tampere, Finland, *epl@uta.fi*
98. **Ana Jesús López**
Department of Applied Economics, University of Oviedo, Oviedo, Spain, *anaj@uniovi.es*

99. **Jesús López-Fidalgo**
Department of Mathematics, University of Castilla-La Mancha, Ciudad Real, Spain, *jesus.lopezfidalgo@uclm.es*

100. **Magid Maatallah**
King's College London, University of London, London, UK, *msmatala@yahoo.com*

101. **Joana Malta**
National Institute of Statistics, Lisbon, Portugal, *joana.malta@ine.pt*

102. **Katarzyna Marczyńska**
Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Poznań, Poland, *kasiapaw@up.poznan.pl*

103. **Augustyn Markiewicz**
Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Poznań, Poland, *amark@up.poznan.pl*

104. **Anabela Marques**
Department of Mathematics, High School of Technology of Barreiro, Polytechnic Institute of Setúbal, Setúbal, Portugal, *Anabela.Marques@estbarreiro.ips.pt*

105. **Filipe J. Marques**
Center of Mathematics and Applications, Faculty of Sciences and Technology, New University of Lisbon, Caparica, Portugal, *fjm@fct.unl.pt*

106. **João Paulo Martins**
High School of Technology and Management, Polytechnic Institute of Leiria, Leiria, Portugal, *jpmartins@estg.ipleiria.pt*

107. **Jean-Pierre Masson**
Agrocampus Rennes, Rennes, France, *masson.giard@wanadoo.fr*

108. **Ayana Mateus**
Department of Mathematics, Faculty of Sciences and Technology, New University of Lisbon, Caparica, Portugal, *amf@fct.unl.pt*

109. **Thomas Mathew**
Department of Mathematics and Statistics, University of Maryland, Baltimore, USA, *mathew@math.umbc.edu*

110. **Ana Matos**
Faculty of Sciences and Technology, New University of Lisbon, Caparica, Portugal, *asvm@fct.unl.pt*

111. **Baerbel Maus**
Department of Methodology and Statistics, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands, *baerbel.maus@stat.unimaas.nl*

112. **Iwona Mejza**
Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Poznań, Poland, *imejza@up.poznan.pl*

113. **Stanisław Mejza**
Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Poznań, Poland, *smejza@up.poznan.pl*

114. **Viatcheslav Melas**
Department of Statistical Modelling, St. Petersburg University, St.Petersburg, Russia, *v.melas@pochta.tvoe.tv*

115. **João Tiago Mexia**
Department of Mathematics, Faculty of Sciences and Technology, New University of Lisbon, Caparica, Portugal, *jtm@maul.fct.unl.pt*

116. **Maria Souto de Miranda**
Department of Mathematics, University of Aveiro, Aveiro, Portugal, *manuela.souto@ua.pt*

117. **Navideh Modarresi**
Amirkabir University of Technology, Tehran, Iran, *namomath@aut.ac.ir*

118. **Sandra Inês da Cunha Monteiro**
Superior School of Business Sciences, Polytechnic Institute of Setúbal, Setúbal, Portugal, *smonteiro@esce.ips.pt*

119. **Elsa Moreira**
Center of Mathematics and Applications, Faculty of Sciences and Technology, New University of Lisbon, Caparica, Portugal, *efnm@fct.unl.pt*

120. **Pedro Mota**
Department of Mathematics, Faculty of Sciences and Technology, New University of Lisbon, Caparica, Portugal, *pjpm@fct.unl.pt*

121. **Jaakko Nevalainen**
Department of Statistics, Faculty of Social Sciences, University of Turku, Turku, Finland, *jaakko.nevalainen@utu.fi*

122. **Radim Navratil**
Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Prague, Czech Republic, *navratil@karlin.mff.cuni.cz*

123. **Hoda Rashidi Nejad**
Isfahan university, Isfahan, Iran, *hoda_mehr@yahoo.com*

124. **Takahiro Nishiyama**
Tokyo University of Science, Tokyo, Japan, *nishiyam@rs.kagu.tus.ac.jp*

125. **Klaus Nordhausen**
Department of Mathematics, Tampere School of Public Health, Tampere University, Tampere, Finland, *klaus.nordhausen@uta.fi*

126. **Célia Nunes**
Departement of Mathematics, University of Beira Interior, Covilhã, Portugal, *celia@mat.ubi.pt*

127. **Sandra Nunes**
Department of Economics and Management, Superior School of Business Sciences, Polytechnic Institute of Setúbal, Setúbal, Portugal, *sandra.nunes@esce.ips.pt*

128. **Haruhiko Ogasawara**
Department of Information and Management Science, Otaru University of Commerce, Otaru, Japan, *hogasa@res.otaru-uc.ac.jp*

129. **Martin Ohlson**
Department of Mathematics, Linköping University, Linköping, Sweden, *martin.ohlson@liu.se*

130. **Hannu Oja**
Tampere School of Public Health, University of Tampere, Tampere, Finland, *hannu.oja@uta.fi*

131. **Paula Milheiro de Oliveira**
Mathematics Center, Faculty of Engineering, University of Porto, Porto, Portugal, *poliv@fe.up.pt*

132. **Rosa Oliveira**
Faculty of Medicine, University of Porto, Porto, Portugal, *rcoliveira@med.up.pt*

133. **Sandra Oliveira**
Department of Economics and Management, Superior School of Business Sciences, Polytechnic Institute of Setúbal, Setúbal, Portugal, *sandra.oliveira@esce.ips.pt*

134. **Patrícia Oom do Valle**
Faculty of Economics, University of Algarve, Faro, Portugal, *pvalle@ualg.pt*

135. **Sanjeev Panwar**
Indian Council of Agricultural Research, New Delhi, India, *scientist1775@yahoo.co.in*

136. **Tatjana Pavlenko**
Department of Statistisc, Stockholm University, Stockholm, Sweden, *tatjana.pavlenko@stat.su.se*

137. **Dulce Gamito Pereira**
Investigation Center of Applied Mathematics, Colégio Luis António Verney, University of Évora, Évora, Portugal, *dgsp@uevora.pt*

138. **Luis Nobre Pereira**
School of Management, Hospitality and Tourism, University of Algarve, Faro, Portugal, *Lmper@ualg.pt*

139. **Dinis Pestana**
Department of Statistics and Operational Research, Faculty of Sciences, University of Lisbon, Lisbon, Portugal, *dinis.pestana@fc.ul.pt*

140. **Jan Picek**
Faculty of Natural Sciences, Humanities and Teaching, Technical University of Liberec, Liberec, Czech Republic, *jan.picek@tul.cz*

141. **Iola Pinto**
High Institute of Engineering of Lisbon, Lisbon, Portugal, *ipinto@eetc.isel.ipl.pt*

142. **Ana M. Pires**
Department of Statistics, Technical University of Lisbon, Lisbon, Portugal, *apires@math.ist.utl.pt*

143. **José Pires**
Faculty of Engineering, University of Porto, Porto, Portugal, *jcpires@fe.up.pt*

144. **Isabel Pitacas**
Department of Mathematics, Polytechnic Institute of Tomar, Tomar, Portugal, *ipitacas@ipt.pt*

145. **Wolfgang Polasek**
Department of Economics and Finance, Institute of Advanced Studies, Vienna, Austria, *polasek@ihs.ac.at*

146. **Bozidar Popovic**
Statistical Office of the Republic of Serbia, Belgrade, 5 Milana Rakica St., Serbia, *bozidarpopovic@gmail.com*

147. **Ana Filipa Prior**
Department of Mathematics, High Institute of Engineering of Lisbon, Lisboa, Portugal, *anafpontes@gmail.com*

148. **Simo Puntanen**
Department of Mathematics, Statistics and Philosophy, University of Tampere, Tampere, Finland, *simo.puntanen@uta.fi*

149. **Luis Ramos**
Department of Mathematics, Faculty of Sciences and Technology, New University of Lisbon, Caparica, Portugal, *lpcr@fct.unl.pt*

150. **Maria do Rosário Ramos**
Department of Sciences and Technology, Aberta University, Lisbon, Portugal, *ramos@cii.fc.ul.pt*

151. **Paulo Ramos**
Superior Institute of Engineering of Lisbon, Lisbon, Portugal, *pramos@deetc.isel.ipl.pt*

152. **Wojciech Rejchel**
Faculty of Mathematics and Computer Sciences, Nicolaus Copernicus University, Toruń, Poland, *wrejchel@gmail.com*

153. **Pedro Réu**
Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden, *pedro.reu@ki.se*

154. **Saeid Rezakhah**
Amirkabir University of Technology, Tehran, Iran, *rezakhah@aut.ac.ir*

155. **Maria da Conceiçao de Oliveira Nunes Rocha**
Department of Mathematics, Faculty of Sciences, University of Porto, Porto, Portugal, *mnrocha@fc.up.pt*

156. **Paulo Canas Rodrigues**
Center of Mathematics and Applications, Faculty of Sciences and Technology, New University of Lisbon, Caparica, Portugal, *paulocanas@fct.unl.pt*

157. **Nicoletta Rosati**
Department of Mathematics, High Institute of Economy and Management, Technical University of Lisbon, Lisbon, Portugal, *nicoletta@iseg.utl.pt*

158. **Dietrich von Rosen**
Department of Biometry and Engineering, Swedish University of Agricultural Sciences, Uppsala, Sweden, *dietrich.von.rosen@et.slu.se*

159. **Tatjana von Rosen**
Department of Statistics, Stockholm University, Stockholm, Sweden, *tatjana.vonrosen@stat.su.se*

160. **Anuradha Roy**
Department of Management Science and Statistics, University of Texas at Santo Antonio, Santo Antonio, USA, *anuradha.roy@utsa.edu*

161. **Maria del Carme Ruiz de Villa**
Department of Statistics, Biology Faculty, University of Barcelona, Barcelona, Spain, *mruiz_de_villa@ub.edu*

162. **Thomas Rusch**
Department of Statistics and Mathematics, Vienna University of Economics and Business, Vienna, Austria, *thomas.rusch@wu.ac.at*

163. **Dina Salvador**
Department of Statistics, Superior School of Technology, Polytechnic Institute of Setúbal, Setúbal, Portugal, *dina.salvador@estsetubal.ips.pt*

164. **Alex Sánchez-Pla**
Department of Statistics, Biology Faculty, University of Barcelona, Barcelona, Spain, *asanchez@ub.edu*

165. **Burkhard Schaffrin**
School of Earth Sciences, The Ohio State University, Columbus, Ohio, USA, *schaffrin.1@osu.edu*

166. **Takashi Seo**
Tokyo University of Science, Tokyo, Japan, *seo@rs.kagu.tus.ac.jp*

167. **Jordi Serra**
Technological Center for Telecommunications of Catalonia, Barcelona, Spain, *jordi.serra@cttc.es*

168. **Javid Shabbir**
Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan, *jsqau@yahoo.com*

169. **Nobumichi Shutoh**
Tokyo University of Science, Tokyo, Japan, *j1409702@ed.kagu.tus.ac.jp*

170. **Carlos Sousa**
University of Algarve, Faro, Portugal, *cfsousa@ualg.pt*

171. **João Beleza Sousa**
Superior Institute of Engineering of Lisbon, Lisbon, Portugal, *jsousa@deetc.isel.ipl*

172. **Marc Smeets**
Statistics Netherlands, The Netherlands, *mset@cbs.nl*

173. **Muni S. Srivastava**
Department of Statistics, University of Toronto, Toronto, Ontario, Canada, *srivasta@utstat.toronto.edu*

174. **Khalaf Sultan**
Statistics & Operations Research Department, College of Sciences, King Saud University, Riyadh, Saudi Arabia, *ksultan@ksu.edu.sa*

175. **Gülin Tabakan**
Aksaray University, Aksaray, Turkey, *gtabakan@aksaray.edu.tr*

176. **Sho Takahashi**
Tokyo University of Science, Tokyo, Japan, *j1410704@ed.kagu.tus.ac.jp*

177. **Frans Tan**
Department of Methodology and Statistics, Maastricht University, Maastricht, The Netherlands
frans.tan@stat.unimaas.nl

178. **Júlia Teles**
Interdisciplinary Centre for the Study of Human Performance, Faculty of Human Kinetics, Technical University of Lisbon, Lisbon, Portugal, *jteles@fmh.utl.pt*

179. **Müjgan Tez**
Department of Methodology and Statistics, Arts and Sciences Faculty, Marmara University, Istanbul, Turkey, *mtez@marmara.edu.tr*

180. **Yongge Tian**
China Economics and Management Academy, Central University of Finance and Economics, Beijing, China, *yongge.tian@gmail.com*

181. **Annika Tillander**
Institute of Statistics, Stockholm University, Stockholm, Sweden, *annika.tillander@stat.su.se*

182. **Chiara Tommasi**
Department of Economic Sciences and Statistics, University of Milano, Milano, Italy, *chiara.tommasi@unimi.it*

183. **Nurkut Nuray Urgan**
Department of Mathematics, Namık Kemal University, Tekirdağ, Turkey, *nurgan@nku.edu.tr*

184. **Corália Vicente**
Institute of Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal, *cvicente@icbas.up.pt*

185. **Pedro Vieira** Department of Mathematics, University of Trás-os-Montes e Alto Douro, Vila Real, Portugal, *pmfvieira@hotmail.com*

186. **Estela Vilhena**
Polytechnic Institute of Cávado and Ave, Barcelos, Portugal, *evilhena@ipca.pt*

187. **Júlia Volaufová**
School of Public Health, Louisiana State University Health Sciences Center, New Orleans, USA, *jvolau@lsuhsc.edu*

188. **Alan Wan**
City University of Hong Kong, Hong Kong, *msawan@cityu.edu.hk*

189. **Jianxin Wei**
Department of Staistics, Uppsala University, Uppsala, Sweden, *jianxin.wei@statistik.uu.se*

190. **Jacek Wesołowski**
Department of Probability and Mathematical Statistics, Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland, *wesolo@mini.pw.edu.pl*

191. **Jianwen Xu**
Department of Statistics and Actuarial Science, Chongqing University, Chongqing, China, *xjw@cqu.edu.cn*

192. **Takayuki Yamada**
Kitasato Institute for Life Sciences, Kitasato University, Tokyo, Japan, *yamadat@pharm.kitasato-u.ac.jp*

193. **Ivan Zezula**
Faculty of Sciences, Pavol Josef Šafárik University in Košice, Košice, Slovakia, *ivan.zezula@upjs.sk*

194. **Andrzej Zieliński**
Department of Applied Mathematics, Warsaw University of Life Sciences, Warsaw, Poland, *andrzejz@chello.pl*

195. **Roman Zmyślony**
Faculty of Mathematics, Computer Sciences and Econometrics, University of Zielona Góra, Zielona Góra, Poland, *r.zmyslony@wmie.uz.zgora.pl*

Index