International Conference
on Trends and Perspectives
in Linear Statistical Inference

Book of Abstracts

September 2 – 6, 2024
Poprad, Slovakia

Edited by

Daniel Klein
Institute of Mathematics
P. J. Šafárik University, Košice, Slovakia

and

Monika Mokrzycka
Institute of Plant Genetics
Polish Academy of Sciences, Poznań, Poland

Contents

Part V. Abstracts

Part I

Introduction

International Conference on Trends and Perspectives in Linear Statistical Inference, LinStat'2024, will be held from 2 till 6 September, 2024, in Poprad, Slovakia, at Hotel Satel. This is the follow-up of the LinStat series held in Będlewo, Poland (2008, 2012, 2018, 2021), in Tomar, Portugal (2010, 2022), in Linköping, Sweden (2014), and in Istanbul, Turkey (2016).

The purpose of the meeting is to bring together researchers sharing an interest in a variety of aspects of statistics and its applications as well as matrix analysis and its applications to statistics, and offer them a possibility to discuss current developments in these subjects. The conference will mainly focus on research of linear modelling methods, estimation, testing, and predictions in economic, medical, and social science data, analysis of multivariate and high-dimensional data with special regard to complex data structures, 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, and applications of matrix methods in statistics.

Special session topic:

- Projection Pursuit

The conference will include invited talks given by

- Yoav Benjamini (Israel)
- Lynn Roy LaMotte (USA)
- Nicola Loperfido (Italy)
- Augustyn Markiewicz (Poland)
- Dietrich von Rosen (Sweden)
- Douglas Wiens (Canada)
- Viktor Witkovský (Slovakia)
- Emelyne Umunoza Gasana (Rwanda)
- Mateusz John (Poland)

Organizers

- Institute of Mathematics, Faculty of Science, P. J. Šafárik University, Košice, Slovakia
- Institute of Mathematics, Poznań University of Technology, Poland
- Institute of Plant Genetics, Polish Academy of Sciences, Poznań

Committees

The Scientific Committee for this Conference comprises

- Ivan Žežula (Slovakia) - Chair
- Anthony C. Atkinson (UK)
- Augustyn Markiewicz (Poland)
- João Tiago Mexia (Portugal)
- Simo Puntanen (Finland)
- Dietrich von Rosen (Sweden)
- Müjgan Tez (Turkey)
- Götz Trenkler (Germany)
- Roman Zmyślony (Poland)

The Organizing Committee comprises

- Daniel Klein (Slovakia) - Chair
- Katarzyna Filipiak (Poland)
- Jozef Hanč (Slovakia)
- Martina Hančová (Slovakia)
- Monika Mokrzycka (Poland)

Part II

Program

International Conference on Trends and Perspectives
in Linear Statistical Inference

Scientific Programme

Sunday, 1. 9. 2024

13:00 – 18:00	Registration
19:00 – 21:00	Dinner

Monday, 2. 9. 2024

7:30 – 9:00	Breakfast	
9:00 – 9:45	Registration	
9:45 – 10:00	Opening	
Plenary Session		*Chair: Ivan Žežula*
10:00 – 11:00	**Yoav Benjamini** *Hierarchical analysis of experiments involving complex ANOVA*	
11:00 – 11:30	Coffee break	
Session I		*Chair: Douglas Wiens*
11:30 – 12:00	**Lynn Roy LaMotte** *A resolution for unbalanced ANOVA*	
12:00 – 12:30	**Júlia Volaufová** *Estimation and testing in simple linear random coefficient model with some covariance observations missing*	
13:00 – 14:30	Lunch	
Plenary Session		*Chair: Daniel Klein*
15:00 – 16:00	**Mateusz John** *Estimation and testing in hierarchical model*	
16:00 – 16:30	Coffee break	
Session II		*Chair: Monika Mokrzycka*
16:30 – 17:00	**Katarzyna Filipiak** *On the relationship between the estimators under the multivariate and matrix-variate t distribution*	
17:00 – 17:30	**Malwina Mrowińska** *Likelihood ratio test for covariance matrix under the matric-variate t distribution*	
18:00 – 19:00	Dinner	
19:30 – 21:00	Welcome reception	

Scientific Programme

Tuesday, 3. 9. 2024

7:30 – 9:00	Breakfast	
Plenary Session		*Chair: Shuangzhe Liu*
9:00 – 10:00	**Nicola Loperfido** *Projection pursuit: theory, applications and challenges*	
10:00 – 10:30	Coffee break	
Invited Session I - Projection Pursuit		*Chair: Nicola Loperfido*
10:30 – 11:00	**Lutz Duembgen** *Projection pursuit via kernel mean embeddings*	
11:00 – 11:30	**H. Sherry Zhang** *Studying the performance of the jellyfish search optimiser for the application of projection pursuit*	
11:30 – 11:45	Health break	
Invited Session II - Projection Pursuit		*Chair: Cinzia Franceschini*
11:45 – 12:15	**Xiaomeng Ju** *Projection-based Bayesian regression for matrix-valued predictors*	
12:15 – 12:45	**Martin Eppert** *Recovering imbalanced clusters via gradient-based projection pursuit*	
13:00 – 14:30	Lunch	
Plenary Session		*Chair: Lynn Roy LaMotte*
15:00 – 16:00	**Douglas Wiens** *To ignore dependencies is perhaps not a sin*	
16:00 – 16:30	Coffee break	
Invited Session III - Projection Pursuit		*Chair: Nicola Loperfido*
16:30 – 17:00	**Cinzia Franceschini** *Investigating food attitudes of Italian children using Projection Pursuit*	
17:00 – 17:30	**Aurore Archimbaud** *Invariant Coordinate Selection and Fisher discriminant subspace beyond the case of two groups*	
18:00 – 19:00	Dinner	

Scientific Programme

Wednesday, 4. 9. 2024

7:30 – 9:00	Breakfast
Plenary Session	*Chair: Dietrich von Rosen*
9:00 – 10:00	**Emelyne Umunoza Gasana** *Edgeworth-type expansion of the density of the classifier when growth curves are classified via likelihood*
10:00 – 10:30	Coffee break
Session III	*Chair: Augustyn Markiewicz*
10:30 – 11:00	**Stephen Haslett** *How is retention of BLUEs in linear model under changes of error covariance relevant to Official Statistics, experimental design, and genetics?*
11:00 – 11:30	**Simo Puntanen** *The fundamental BLUE equation in linear models revisited*
11:30 – 11:45	Health break
Session IV	*Chair: Simo Puntanen*
11:45 – 12:30	**Shuangzhe Liu** *On heavy-tailed matrix variate regression models and statistical diagnostics*
12:30 – 13:00	**Emanuel Ben-David** *Gaussian DAG models with symmetries*
13:00 – 14:30	Lunch
Plenary Session	*Chair: Júlia Volaufová*
15:00 – 16:00	**Viktor Witkovský** *Effortless calculation of complex probability distributions using CharFunTool: Computational methods, tools and applications*
16:00 – 16:30	Coffee break
Session V	*Chair: Viktor Witkovský*
16:30 – 17:00	**Jozef Hanč** *Double exponential quadrature in calculating the distribution of products and quotients of random variables*
17:00 – 17:45	**Dietrich von Rosen** *What can we learn from the classification of growth curves*
19:30 – 22:00	Conference Dinner

International Conference on Trends and Perspectives in Linear Statistical Inference

Scientific Programme

Thursday, 5. 9. 2024

7:30 – 9:00	Breakfast
9:00 – 17:00	Excurssion (including lunch)
18:00 – 19:00	Dinner

Friday, 6. 9. 2024

7:30 – 9:00	Breakfast	
Plenary Session		*Chair: Stephen Haslett*
9:00 – 10:00	**Augustyn Markiewicz** *Characterization of matrix quadratic spaces with application to the estimation of covariance matrices*	
10:00 – 10:30	Coffee break	
Session VI		*Chair: Mateusz John*
10:30 – 11:00	**Shubhabrata Das** *On constraints and interrelations between nonparametric and polychoric correlations for variables measured on the Likert scale*	
11:00 – 11:30	**Adam Mieldzioc** *Regularization and estimation of linearly structured covariance matrix*	
11:30 – 11:45	Health break	
Session VII		*Chair: Katarzyna Filipiak*
11:45 – 12:15	**Monika Mokrzycka** *Block covariance matrix estimation with structured off-diagonal blocks*	
12:15 – 12:45	**Daniel Klein** *Matrix mean testing in models with a special class of variance matrices*	
12:45 – 13:00	Closing	
13:00 – 14:30	Lunch	
15:00 – 16:00	Open discussions	
16:00 – 16:30	Coffee break	
16:30 – 18:00	Open discussions	
18:00 – 19:00	Dinner	

Part III

Invited Speakers

Hierarchical analysis of experiments involving complex ANOVA

Yoav Zeevi, Liad Mudrik and Yoav Benjamini

Tel Aviv University, Israel

Abstract

In experimental psychology researchers often resort to 3 factors (and more) ANOVA, and the reported research may include more than just one ANOVA. The abstract and discussion often highlight the very limited set of results selected from all levels. We explain the use of the hierarchical FDR testing to address such problems. We then demonstrate the impact of using FDR control to reduce the replicability problem reported in Psychology Reproducibility Project.

A resolution for unbalanced ANOVA

Lynn Roy LaMotte

LSU Health New Orleans, USA

Abstract

Multiple regression models include ANOVA effects with dummy variables, which causes containment issues. Contrast coding has long been used to avoid those problems. Effects can be tested with extra SSE due to deleting corresponding sets of regressors, but it is not always clear exactly what is tested. In models with empty cells, determining what part of an effect is estimable requires extra steps. It is established here that those problems are avoided with contrast coding. Then the model formed by deleting an effect's regressors is the correct restricted model for the target ANOVA effect; the non- centrality parameter is zero if and only if all estimable linear functions of the effect are zero; and no other numerator sum of squares that tests the effect has a greater non-centrality parameter or lesser degrees of freedom.

Keywords

Restricted model - full model SS; Main effects SS; Contrast coding

Projection pursuit: theory, applications and challenges

Nicola Loperfido

Università degli Studi di Urbino Carlo Bo, Italy

Abstract

Projection pursuit is a multivariate statistical technique aimed at finding interesting low-dimensional data projections. It looks for the data projection which maximizes the projection pursuit index, that is a measure of its interestingness. After an interesting projection is found, it is removed to facilitate the search for other interesting features. Projection pursuit addresses three major challenges of multivariate analysis: the curse of dimensionality, the presence of irrelevant features and the limitations of visual perception. Its applications have been hampered by several difficulties of computational, interpretative and inferential nature. Additional problems arise when data are high-dimensional, that is when there are more variables than units. This talk outlines the main features of projection pursuit and its connections with other multivariate techniques. The theory is illustrated with both real and simulated datasets.

Characterization of matrix quadratic spaces with application to the estimation of covariance matrices

Katarzyna Filipiak[1], **Augustyn Markiewicz**[2] **and Malwina Mrowińska**[1]

[1] Poznań University of Technology, Poland
[2] Poznań University of Life Sciences, Poland

Abstract

In this paper we give a general characterization of quadratic spaces of symmetric matrices which enables the characterization of the quadratic spaces of block matrices with separately linearly structured blocks, possibly with non-zero off-diagonal blocks. This characterization is applied to the problem of estimating covariance matrices with a linear block structure. For a quadratic structure, the appropriate estimation method is the least squares method. Otherwise, the shrinkage method is recommended. Since the properties of the estimator depend on the choice of the target space – a quadratic subspace of the structure space – it is necessary to characterize all such subspaces as possible.

What can we learn from the classification of growth curves

Dietrich von Rosen

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

Abstract

Discrimination between two populations following the growth curve model is considered. A likelihood-based classification procedure is established, in the sense that we compare the two likelihoods given that the new observation belongs to respective population. The possibility to classify the new observation as belonging to an unknown population is discussed, which is shown to be natural when considering growth curves.

To ignore dependencies is perhaps not a sin

Douglas Wiens

University of Alberta, Edmonton, Canada

Abstract

We present a result according to which certain functions of covariance matrices are maximized at scalar multiples of the identity matrix. In a statistical context, in which such functions measure loss, this has surprising consequences. In particular it implies the robustness against dependence of experimental designs constructed assuming independence. Another consequence is that the ordinary least squares (ols) estimate of a correctly specified regression function can be robust, among generalized least squares (gls) estimates, against dependence. An implication is that it can be not only safe, but optimal to ignore such departures from the usual assumption of i.i.d. errors. We then consider regression models in which the response function is possibly misspecified, and show that ols is minimax if the design is uniform on its support, but that this often fails otherwise. We go on to investigate the interplay between minimax gls procedures and minimax designs, leading us to extend, to robustness against dependencies, an existing observation - that robustness against model misspecifications is increased by splitting replicates into clusters of observations at nearby locations.

Effortless calculation of complex probability distributions using CharFunTool: Computational methods, tools and applications

Viktor Witkovský

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

Abstract

This presentation is focused on the calculation of complex probability distributions using CharFunTool, a powerful toolbox for characteristic functions. We explore various applications in statistics, insurance, measurement, and metrology, demonstrating how CharFunTool simplifies and enhances both analytical and computational processes. Through selected examples, we showcase its utility in addressing real-world problems and improving accuracy across diverse fields. Attendees will gain practical insights into leveraging CharFunTool for efficient distribution analysis, making it an invaluable resource for researchers and practitioners alike.

Statistical inference often requires the computation of non-standard probability distributions for estimators or test statistics, which can be fully specified by their moments or characteristic functions. Standard asymptotic approximations and small sample approximations based on first moments and cumulants are commonly used but can be inadequate for specific situations. In such cases, more sophisticated approximations, such as near-exact distributions, can be employed. However, for a wide range of applications, numerical inversion of characteristic functions is a highly efficient and sufficient method.

In this talk, we illustrate the applicability of numerical inversion of characteristic functions through several examples, including the exact null and non-null distributions of selected tests in multivariate statistical analysis, the application of empirical characteristic functions, and the exact distribution of the bootstrap mean. We will also discuss practical experiences and potential issues associated with calculating the required characteristic functions and related exact distributions.

Keywords

Characteristic function, Numerical inversion, CharFunTool.

Acknowledgements

The work was supported by the Slovak Research and Development Agency, project APVV-21-0216, and by the projects VEGA 2/0023/22 and VEGA 2/0120/24.

Edgeworth-type expansion of the density of the classifier when growth curves are classified via likelihood

Emelyne Umunoza Gasana[1], Dietrich von Rosen[2] and Martin Singull[3]

[1] University of Rwanda, Kigali, Rwanda
[2] Swedish University of Agricultural Sciences, Uppsala, Sweden
[3] Linköping University, Linköping, Sweden

Abstract

When classifying repeated measurements using the Growth Curve model, also known as bilinear regression model, it can happen that the observations to classify might not belong to any of the two predetermined populations. [1] derived a two-step classification rule taking into account this perspective. Probabilities of misclassification of the two-step likelihood-based discriminant rule are established for the classification of growth curves where the distribution for the classifier is approximated using an Edgeworth-type expansion.

References

[1] von Rosen, D. and Singull, M. (2022). Classification of repeated measurements using growth curves, *Linköping University Electronic Press.*

Estimation and testing in hierarchical model

Mateusz John

Poznań University of Technology, Poland

Abstract

The aim of the presentation is to propose tests for covariance structures in doubly multivariate models. Due to the hierarchical nature of the considered experiments, block matrices are appropriate structures. We are considering block structures belonging to the quadratic subspaces. The proposed tests, among others the likelihood ratio test, the Rao score test and the Wald test, are compared with each other in terms of the speed of convergence to the limiting chi-square distribution and the power. For the comparison simulation methods are used. Moreover, since the maximum likelihood estimators of unknown parameters have an important role in each considered test, the presentation shows that these estimators can be obtained by projecting onto an appropriate quadratic subspace. Presented results are illustrated using real data example.

Part IV

Special Session

Projection Pursuit

Nicola Loperfido

Università degli Studi di Urbino Carlo Bo, Italy

Projection pursuit I

The first talk in this session, delivered by Lutz Duembgen, is entitled "Projection pursuit via kernel mean embeddings". It presents a method for finding interesting data projections by combining global search and local optimization. The latter maximizes the mean discrepancy between the empirical distribution of the projected data and a data-driven Gaussian mixture distribution. The second talk in this session, delivered by Sherry Zhang, is entitled "Studying the performance of the jellyfish search optimiser for the application of projection pursuit". Projection pursuit is more problematic in the presence of non-smooth projection indices and small squint angles. The talk quantifies these concepts and addresses the problems they pose. The latter task is accomplished via the jellyfish search optimiser, a swarm-based algorithm.

Projection pursuit II

The first talk in this session, delivered by Xiaomeng Ju, is entitled "Projection-based Bayesian regression for matrix-valued predictors" and is based on a joint work with Hyung G. Park and Thaddeus Tarpey. It presents a novel Bayesian, projection-pursuit-type estimator of the regression model with connectivity predictors encoded as symmetric matrices, summarized via dimension reduction projections. The second talk in this session, delivered by Martin Eppert, is entitled "Recovering imbalanced clusters via gradient-based projection pursuit" and is based on a joint work with Satyaki Mukherjee and Debarghya Ghoshdastidar. It proposes a method for identifying projections containing two imbalanced clusters using a gradient-based technique to optimize the projection index, while assessing sample complexity.

Projection pursuit III

The first talk in this session, delivered by Cinzia Franceschini, is entitled "Investigating food attitudes of Italian children using Projection Pursuit" and is based on a joint work with Nicola Loperfido. It investigates the attitudes of Italian children towards food and its consumption in school canteens. The data at hand make a case for projection pursuit when variable selection for clustering is sought. The second talk in this session, delivered by Aurore Archimbaud, is entitled "Invariant Coordinate Selection and Fisher discriminant subspace beyond the case of two groups" and is based on a joint work with Colombe Becquart, Anne Ruiz-Gazen, Luka Prilc and Klaus Nordhausen. The talk investigates some properties of invariant coordinate selection, a strong competitor of projection pursuit, related to clustering in more than two groups.

Invited speakers:

- Aurore Archimbaud (France)
- Lutz Duembgen (Switzerland)
- Martin Eppert (Germany)
- Cinzia Franceschini (Italy)
- Xiaomeng Ju (USA)
- H. Sherry Zhang (USA)

Part V

Abstracts

Invariant Coordinate Selection and Fisher discriminant subspace beyond the case of two groups

Aurore Archimbaud[1], Colombe Becquart[2], Anne Ruiz-Gazen[2], Luka Prilc[2], Klaus Nordhausen[3]

[1] TBS Business School, Toulouse, France
[2] Toulouse School of Economics, Toulouse, France
[3] University of Jyvaskyla, Finland

Abstract

Invariant Coordinate Selection (ICS) is a powerful unsupervised multivariate method designed to identify the structure of multivariate datasets on a subspace. It relies on the joint diagonalization of two scatter matrices and is particularly relevant as a dimension reduction tool prior to clustering or outlier detection. It goes beyond the well-known Principal Components Analysis (PCA) method by not relying on maximizing the inertia but on optimizing a generalized kurtosis and is not only invariant by an orthogonal transformation of the data but by any affine transformation. Unlike PCA, ICS has a theoretical foundation that explains why the identified subspace should contain relevant information and under what conditions it might fail to reveal such subspace. More precisely, some theoretical results proved that under some elliptical mixture models, the subspace spanned by the first and/or last components carries the information regarding the multivariate structure and recovers the Fisher discriminant subspace, whatever the choice of scatter matrices. These general results have traditionally been examined in detail primarily for specific scatter combinations within a two-cluster framework. In this study, we expand these investigations to include more clusters and scatter combinations. Based on these expanded theoretical insights and supported by numerical studies, we conclude that ICS is indeed suitable for dimension reduction with some guidelines on which cases it might fail.

Gaussian DAG models with symmetries

Emanuel Ben-David

US Census Bureau, USA

Abstract

A Gaussian DAG model is a family of multivariate normal distributions where the zero elements of the modified Cholesky matrix, the triangular matrix obtained by the Cholesky decomposition of the covariance matrix, are determined by a directed acyclic graph (DAG). In this talk, we introduce a new subclass of Gaussian DAG models by imposing further restrictions on the elements of the Cholesky matrix, resulting in more parsimonious models with much fewer parameters. This research is motivated by the work of Lauritzen and Højsgaard, which introduced undirected graphical Gaussian models with edge and vertex symmetries. We will explain how color graphs naturally represent these restricted DAG models and discuss the estimation problem. We will focus on the maximum likelihood estimation of the parameters. Estimating parameters for these models presents several challenges due to the additional constraints imposed by the color graphs. One significant difficulty is ensuring that the restrictions are appropriately incorporated into the estimation process, which can complicate the optimization algorithms used.

On constraints and interrelations between nonparametric and polychoric correlations for variables measured on the Likert scale

Shubhabrata Das

Indian Institute of Management, Bangalore, India

Abstract

Nonparametric correlations, such as Spearman's rank correlation and Kendall's tau, have been widely utilized for nearly a century to analyze ordinal data. Chatterjee correlation [1] is a more recent development on that front which captures associations within a broader structure. This study examines these three measures, particularly in the context of ordinal variables measured using the Likert scale. All these measures have adaptations for handling ties. However, for Likert scale variables, ties occur frequently as a norm rather than exception. We specifically investigate whether the Chatterjee correlation can be modified to effectively handle Likert scale data. Since all the above correlations are based on rank, they inherently stem from the perspective of sampled data. The corresponding population parameters have been partially addressed: Nešlehová [2] has explored this for Kendall's tau and Spearman's correlation, while the strong consistency of the Chatterjee correlation also provides some insights. We reflect on these population parameters, with a particular focus on their potential role in factor analysis with Likert scale data, aiming to improve the interpretation of underlying factors. Finally, we explore the possible connections between the different nonparametric correlations and polychoric correlation.

Keywords

Chatterjee correlation, Factor analysis, Ordinal, Kendal's tau, Spearman's rank correlation.

References

[1] Chatterjee, S. (2020). A New Coefficient of Correlation. *Journal of the American Statistical Association, 116(536)*, 2009–2022.

[2] Nešlehová, J. (2007). On rank correlation measures for noncontinuous random variables. *Journal of Multivariate Analysis, 98(3)*, 544–567.

Projection pursuit via kernel mean embeddings

Lutz Duembgen

University of Bern, Switzerland

Abstract

Detecting and visualizing interesting structures in high-dimensional data is a ubiquitous challenge. If one aims for linear projections onto low-dimensional spaces, a well-known problematic phenomenon is the Diaconis–Freedman effect: under mild conditions, most projections do not reveal interesting structures but look like scale mixtures of spherically symmetric Gaussian distributions. We present a method which combines global search strategies and local projection pursuit via maximizing the maximum mean discrepancy (MMD) between the empirical distribution of the projected data and a data-driven Gaussian mixture distribution. Here, MMD is based on kernel mean embeddings with Gaussian kernels.

Recovering imbalanced clusters via gradient-based projection pursuit

Martin Eppert[1], Satyaki Mukherjee[2] and Debarghya Ghoshdastidar[1]

[1] Technical University of Munich, Germany
[2] National University of Singapore, Singapore

Abstract

Projection Pursuit is a classic exploratory technique for finding ”interesting” projections of a dataset. We propose a method for identifying projections containing two imbalanced clusters using a gradient-based technique to optimize the projection index. As sample complexity is a major limiting factor in Projection Pursuit, we analyze our algorithm's sample complexity within a Planted Vector setting where we can observe that imbalanced clusters can be recovered more easily than balanced ones. This differs from previous literature that primarily focuses on the recovery of projections containing symmetric distributions such as balanced clusters [1] and Bernoulli-Rademacher [4]. While some work has addressed recovering skewed directions [3] the sample complexity required for such recovery has not been previously analyzed. Additionally, we give a generalized result that allows for the analysis of the sample complexity for a variety of data distributions and projection indices. We compare these results to computational lower bounds in the Low-Degree-Polynomial framework [2]. While there still is a small gap between the upper and lower bounds, we give empirical evidence that by numerically whitening the data before gradient ascent, the lower bound of the sample complexity can be achieved for large imbalances. Finally, we experimentally evaluate our method's applicability to multi-class classification problems on FashionMNIST.

Keywords

Gradient-Based Methods, Projection Pursuit, Optimization.

Acknowledgements

This work has been supported by the German Research Foundation (DFG) through DFG-ANR PRCI ”ASCAI” (GH 257/3-1).

References

[1] Davis, D., Diaz, M. and Wang, K. (2021). Clustering a mixture of gaussians with unknown covariance. *ArXiv Preprint ArXiv:2110.01602*.
[2] Kunisky, D., Wein, A. and Bandeira, A. (2019). Notes on Computational Hardness of Hypothesis Testing: Predictions using the Low-Degree Likelihood Ratio.
[3] Loperfido, N. (2018). Skewness-based projection pursuit: A computational approach. *Computational Statistics And Data Analysis. 120*, 42–57.
[4] Mao, C. and Wein, A. (2022). Optimal Spectral Recovery of a Planted Vector in a Subspace.

On the relationship between the estimators under the multivariate and matrix-variate t distribution

Katarzyna Filipiak[1], **Daniel Klein**[2], **Stepan Mazur**[3] **and Malwina Mrowińska**[1]

[1] Poznań University of Technology, Poland
[2] P. J. Šafárik University, Košice, Slovakia
[3] Örebro University, Örebro, Sweden

Abstract

One of the most common distributional assumptions of many statistical models is the normal distribution. However, in many applied fields such as, for example, macro- and financial economics, heavy tailed distribution are more appropriate to consider. In this talk one of the most popular heavy-tailed distributions, the multivariate t distribution, is considered. Additionally, in this talk we relax the assumption about independence to uncorrelated observations, which is indistinguishable under normality assumption.
The aim of this talk is to present the maximum likelihood estimators of location and scale parameters of multivariate t distribution, however, since the observations are uncorrelated (not independent), for this purpose the matrix t distribution will be used. Observe, that the definition of matrix t distribution depends on the stochastic representation of matrix-variate t distribution random variable and thus two alternative forms of distribution will be considered.
Finally, basic statistical properties of maximum likelihood estimators of location and scale parameters, such as biasedness and sufficiency, will be shown.

Keywords

Multivariate t distribution, Matrix-variate t distribution, Maximum likelihood estimators.

Acknowledgements
The author was supported by the Poznan University of Technology under Grant no. 0213/SBAD/0119.

Investigating food attitudes of Italian children using Projection Pursuit

Cinzia Franceschini[1] **and Nicola Loperfido**[2]

[1] Sassari University, Sassari, Italy
[2] Urbino University, Urbino, Italy

Abstract

The University of Gastronomic Sciences (Pollenzo, Italy) investigated the attitude of Italian children towards food and its consumption in school canteens. Data were collected from questionnaires administered to 1108 children in 9 primary Italian schools. We first clustered original data by means of model-based clustering and k-means clustering. Then we used principal component analysis to reduce the number of variables before clustering. We obtained the best clustering using k-means on the data projected onto the directions found using projection pursuit, a multivariate statistical technique aimed at finding interesting low-dimensional data projections. By means of projection pursuit we also obtained the most satisfactory variable selection, compared to other statistical methods implemented in R packages. We conclude that the data at hand make a case for projection pursuit when variable selection for clustering is sought.

Keywords

Cluster analysis, Dimension reduction, Multivariate data, Variable selection.

References

[1] Arabie, P. and Hubert, L. (1994). Cluster analysis in marketing research. In: Bagozzi, R. P.(Eds.), *Advanced methods of marketing research*, (pp.160–189). Blackwell, Oxford.
[2] Fop, M. and Murphy, T.B. (2018). Variable selection methods for model-based clustering. *Statistics Surveys. 12*, 18–65.
[3] Neal, M.R. and McNicholas, P.D. (2024). Flexible Variable Selection for Clustering and Classification. *arXiv:2305.16464v2 [stat.ME]*
[4] Sun, J. (2006). *Projection Pursuit. Encyclopedia of Statistical Sciences* (10).
[5] Yoshikazu Terada (2014). Strong Consistency of Reduced K-means Clustering. *Scandinavian Journal of Statistics. 41*, 913 –931

Double exponential quadrature in calculating the distribution of products and quotients of random variables

Jozef Hanč and **Martina Hančová**

P. J. Šafárik University, Košice, Slovakia

Abstract

We explore the application of double exponential (DE) quadrature [1, 2] for fast and exact calculations of the distribution of products and quotients of independent random variables, which are important not only in linear statistical inference and uncertainty estimation in metrology but also in other fields. The DE quadrature, which uses the trapezoidal rule with a DE transformation, is employed for the numerical integration of the required probability density expressed as the Mellin convolution integral [3]. In our numerical explorations, we also compare and cross-check our results with well-known analytic formulas [3], recent numerical approaches [4], and Monte Carlo and bootstrap simulations. Since the DE quadrature is well-suited for CPU and GPU parallelization [5], HPC can also be effectively utilized and explored.

Keywords

Mellin convolution integral, exact probability calculations, fast numerical integration, high-performance computing (HPC), numerical simulations

Acknowledgements

This work was supported by the Slovak Research and Development Agency under the Contract No. APVV-21-0369, No. APVV-21-0216 and by the Slovak Scientific Grant Agency VEGA under grant VEGA 1/0585/24.

References

[1] Mori, M. (2005). Discovery of the Double Exponential Transformation and Its Developments. *Res. Inst. Math. Sci. 41(4)*, 897–935.

[2] Hančová, M., et al. (2022). A practical, effective calculation of gamma difference distributions. *J. Stat. Comput. Simul. 92(11)*, 2205–2232.

[3] Springer, M. D. (1979). *The Algebra of Random Variables*. Wiley.

[4] Jaroszewicz, S., and Korzeń, M. (2012). Arithmetic Operations on Independent Random Variables. *SIAM J. Sci. Comput. 34(3)*, A1241–A1265.

[5] Bailey, D. H., and J. M. Borwein. (2011). High-Precision Numerical Integration: Progress and Challenges. *J. Symb. Comput. 46(7)*, 741–54.

How is retention of BLUEs in linear model under changes of error covariance relevant to Official Statistics, experimental design, and genetics?

Stephen Haslett

Massey University, New Zealand; Australian National University, Australia; University of Wollongong, Australia

Abstract

The necessary and sufficient conditions for BLUEs to be retained in a linear model under a change of error covariance go back to [1]. More recent research shows it is also possible to retain the BLUE covariance or the sum of squares of errors. The requirements for retention of BLUEs and BLUE covariance can be extended so that every possible submodel also retains its BLUEs and its BLUE covariance. These properties have applications in Official Statistics, experimental design, and genetics. Examples of these applications will form the focus of the presentation.

Keywords

Best linear unbiased estimator (BLUE), BLUE covariance matrix, Confidentialised Unit Record File (CURF), data cloning, encryption, genetics, linear model misspecification, Official Statistics, strong singularity, submodels, sum of squares of errors, weak singularity.

Acknowledgements

This presentation is part of joint research with Jarkko Isotalo, Augustyn Markiewicz and Simo Puntanen.

References

[1] Rao, C.R. (1971). Unified theory of linear estimation. *Sankhyā Ser. A*, 33, 371–394. [Corrigenda (1972): 34, p. 194 and p. 477.]

Projection-based Bayesian regression for matrix-valued predictors

Xiaomeng Ju, Hyung G. Park and Thaddeus Tarpey

New York University School of Medicine, United States

Abstract

We present a novel Bayesian approach for nonlinear regression with connectivity predictors encoded as symmetric matrices or symmetric positive definite matrices. Unlike methods that vectorize matrices as predictors, resulting in a large number of parameters, our approach extracts informative features through dimension reduction projections, yielding a projection-pursuit-type estimator of the regression model. We establish the identifiability conditions of the proposed model and impose sparsity priors on the projection directions to prevent overfitting and enhance interpretability of the parameter estimates. The performance of our approach is evaluated through simulation studies and a case study investigating the relationship between brain connectivity features and behavioral outcomes.

Keywords

Bayesian nonlinear regression, Projection pursuit, Matrix predictors

Matrix mean testing in models with a special class of variance matrices

Daniel Klein and **Ivan Žežula**

P. J. Šafárik University, Košice, Slovakia

Abstract

In recent years, multivariate models with matrix-valued observations attract more and more attention, see e.g. [1], [2] or [3]. One of the main problems of their practical use is usually the problem of small sample size, which causes numerical instability or even rank deficiency of the sample variance matrix. As a result, mean testing can be problematic or even impossible. Various special variance structures are used, when possible, to tackle the problem. However, this brings the need to derive the distribution of the test statistic for every special variance matrix structure. We will present the solution for a general class of test procedures, when the variance matrix can be decomposed into parts belonging to some commutative quadratic subspace. Three classical test procedures – Likelihood ratio, Rao score and Wald test – are studied. Exact null distribution as well as the distribution under large- and high-dimensional regime are derived. This covers many previous results, which can be viewed as special cases of the general model introduced here.

Keywords

Multivariate linear model, Special variance structure, Quadratic subspace, Mean testing.

Acknowledgements

This work was supported by the Slovak Research and Development Agency under the Contract No. APVV-21-0369, and grant VEGA No. 1/0585/24.

References

[1] Roy, A., Zmyślony, R., Fonseca, M., Leiva, R. (2016). Optimal estimation for doubly multivariate data in blocked compound symmetric covariance structure. *Journal of Multivariate Analysis* 144, 81–90.

[2] Žežula, I., Klein, D., Roy, A. (2018). Testing of multivariate repeated measures data with block exchangeable covariance structure. *Test* 27(2), 360–378.

[3] Dai, D., Hao, C., Jin, S., Liang Y. (2023). Regularized estimation of Kronecker structured covariance matrix using modified Cholesky decomposition. *Journal of Statistical Computation and Simulation*, DOI: 10.1080/00949655.2023.2291536.

On heavy-tailed matrix variate regression models and statistical diagnostics

Shuangzhe Liu[1] **and Conan Liu**[2]

[1] University of Canberra, Canberra, Australia
[2] University of New South Wales, Sydney, Australia

Abstract

Matrix variate distributions and matrix regression models are powerful tools for analysing multivariate data with inherent matrix structure. These methods extend traditional univariate and multivariate techniques to handle more complex data structures, such as those found in genomics, neuroscience, and image analysis. In this talk, we introduce a framework for regression models under heavy-tailed matrix variate distributions. We begin by discussing several matrix variate distributions and then explore the general linear model under the matrix variate normal distribution, along with its relevant variations based on heavy-tailed distributions. Additionally, we cover important sensitivity analysis and statistical diagnostics for these models, highlighting potential future research problems and applications. Our goal is to provide insights into the theoretical and practical aspects of heavy-tailed matrix variate regression models, paving the way for further advancements in this area.

Keywords

Complex data structure, Heavy-tailed distribution, Matrix variate regression model, Sensitivity analysis, Statistical diagnostics.

Block covariance matrix estimation with structured off-diagonal blocks

Monika Mokrzycka[1] and Malwina Mrowińska[2]

[1] Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland
[2] Poznań University of Technology, Poland

Abstract

This talk deals with the estimation of a block covariance matrix with off-diagonal blocks corresponds to the part of autoregression of order one structure, AR(1). Commonly used maximum likelihood estimation is challenging and time-consuming, thus we propose also another approach based on the least squares method. Some estimates are not always well-conditioned and may not even be definite. Thus, the improvement based on a shrinkage method and an additional algebraic approach is applied. The considered structure can be also expressed as a sum of two matrices: block diagonal matrix and AR(1) matrix. New approach based on estimation of whole AR(1) structure is presented and several estimation method are proposed. All considered estimates are compared with respect to some statistical properties and time needed to determine them.

Keywords

Block covariance matrix, Autoregression of order one structure, Maximum likelihood estimation, Least squares method, Shrinkage method.

Likelihood Ratio Test for Covariance Matrix under the matric-variate t distribution

Katarzyna Filipiak[1], Daniel Klein[2], Stepan Mazur[3] and Malwina Mrowińska[1]

[1] Poznań University of Technology, Poland
[2] P. J. Šafárik University, Košice, Slovakia
[3] Örebro University, Sweden

Abstract

In many applied fields such as, for example, macro and financial economics the distribution of variables is skewed and heavy tailed. Multivariate t-distribution assumption provides a better solutions than multivariate normality when variables are heavy tailed. In the multivariate normal case uncorrelated observations are independent which is not the case for any other distribution. In this talk matric-variate t distribution under the assumption of uncorrelated observations (instead of independent) is considered.
The likelihood ratio test is proposed and its distributional properties are verified under the null hypothesis with fully precised covariance matrix or precised up to the constant.

Keywords

Matric-variate t distribution, Hypothesis testing, Likelihood Ratio Test

Acknowledgements
The author was supported by the Poznan University of Technology under Grant no. 0213/SBAD/0119.

The fundamental BLUE equation in linear models revisited

Simo Puntanen

Tampere University, Tampere, Finland

Abstract

In the world of linear statistical models there is a particular matrix equation, $\mathbf{G}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{X} : \mathbf{0})$, which is sufficiently important that it is sometimes called the fundamental BLUE equation. In this equation, $\mathbf{X}$ is a model matrix, $\mathbf{V}$ is the covariance matrix of $\mathbf{y}$ in the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, and we are interested in finding the best linear estimator, BLUE, of $\mathbf{X}\boldsymbol{\beta}$. Any solution $\mathbf{G}$ for this equation has the property that $\mathbf{Gy}$ provides a representation for the BLUE of $\mathbf{X}\boldsymbol{\beta}$: this is the message of the fundamental BLUE equation, whose main developer was the late Professor C. R. Rao in early 1970s.
This talk is part of joint work with Stephen J. Haslett, Jarkko Isotalo and Augustyn Markiewicz.

Keywords

BLUE, BLUP, covariance matrix, equality of the BLUEs, linear sufficiency, misspecified model.

References

[1] Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob., Vol. 1.* (L.M. Le Cam, J. Neyman, eds.) Univ. of Calif. Press, Berkeley, pp. 355–372.

[2] Rao, C. R. (1971). Unified theory of linear estimation. *Sankhyā Ser. A*, 33, 371–394. [Corrigenda (1972): 34, p. 194 and p. 477]

[3] Rao, C. R. (1973). Representations of best linear estimators in the Gauss–Markoff model with a singular dispersion matrix. *Journal of Multivariate Analysis*, 3, 276–292.

Estimation and testing in simple linear random coefficient model with some covariance observations missing

Júlia Volaufová

LSUHSC, New Orleans, USA

Abstract

In this talk, we deal with a simple linear random coefficients model (LRCM), in which we assume that some of the observations of the covariates may be missing. For several decades, estimation and prediction have been extensively studied in a variety of rather complex statistical models under the assumption of missing observations. However, much fewer results are available in such settings about statistical inference (hypotheses tests). Under rather specific simplifying assumptions, we extend and apply the distributional model for the missing covariates introduced by [1] in a linear fixed effects model. Our primary interest is to investigate properties of commonly used tests in the classic LRC models about the mean parameters, using maximum likelihood estimates obtained by the EM algorithm (see, e.g., [2]). The accuracy of p-values and adjusted power of tests were studied using a simple simulation study.

Keywords

Random coefficient model, Covariates missing at random, Approximate variance, Testing hypotheses.

References

[1] Chen, Q., Ibrahim, J.G., Chen M-H., and Senchaudhuri, P. (2008). Theory and inference for regression models with missing responses and covariates. *Journal of Multivariate Analysis*, 99, 1302–1331.

[2] Dempster, A.P., Laird, N., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1), 1–38.

Studying the performance of the jellyfish search optimiser for the application of projection pursuit

H. Sherry Zhang

The University of Texas at Austin, USA

Abstract

The projection pursuit (PP) guided tour interactively optimises a criteria function known as the PP index, to explore high-dimensional data by revealing interesting projections. The optimisation in PP can be non- trivial, involving non-smooth functions and optima with a small "squint angle", detectable only from close proximity. To address these challenges, this study investigates the performance of a recently introduced swarm-based algorithm, Jellyfish Search Optimiser (JSO), for optimising PP indexes. The performance of JSO for visualising data is evaluated across various hyper-parameter settings and compared with existing optimisers. Additionally, this work proposes novel methods to quantify two properties of the PP index – smoothness and squintability – that capture the complexities inherent in PP optimisation problems. These two metrics are evaluated along with JSO hyper-parameters to determine their effects on JSO success rate. Our numerical results confirm the positive impact of these metrics on the JSO success rate, with squintability being the most significant. The JSO algorithm has been implemented in the tourr package and functions to calculate smoothness and squintability are available in the ferrn package.

Part VI

List of participants

Participants

1. **Aurore Archimbaud**
   TBS Business School, Toulouse, France
   a.archimbaud@tbs-education.fr
2. **Emanuel Ben-David**
   US Census Bureau, USA
   emanuel.ben.david@census.gov
3. **Yoav Benjamini**
   Tel Aviv University, Israel
   ybenja@tauex.tau.ac.il
4. **Shubhabrata Das**
   Indian Institute of Management, Bangalore, India
   shubho@iimb.ac.in
5. **Lutz Duembgen**
   University of Bern, Switzerland
   lutz.duembgen@unibe.ch
6. **Martin Eppert**
   Technical University of Munich, Germany
   martin.eppert@tum.de
7. **Katarzyna Filipiak**
   Institute of Mathematics, Poznań University of Technology, Poznań, Poland
   katarzyna.filipiak@put.poznan.pl
8. **Cinzia Franceschini**
   Sassari University, Sassari, Italy
   cinziafranceschini@msn.com
9. **Jozef Hanč**
   P. J. Šafárik University, Košice, Slovakia
   jozef.hanc@upjs.sk
10. **Martina Hančová**
    P. J. Šafárik University, Košice, Slovakia
    martina.hancova@upjs.sk
11. **Stephen Haslett**
    Massey University, New Zealand; Australian National University, Australia; University of Wollongong, Australia
    s.j.haslett@massey.ac.nz
12. **Mateusz John**
    Poznań University of Technology, Poland
    mateusz.john@put.poznan.pl
13. **Xiaomeng Ju**
    New York University School of Medicine, United States
    jux01@nyu.edu

14. **Daniel Klein**
Institute of Mathematics, P. J. Šafárik University in Košice, Košice, Slovakia
daniel.klein@upjs.sk

15. **Andrzej Kosinski**
Duke University, USA
andrzej.kosinski@duke.edu

16. **Lynn Roy LaMotte**
LSU Health New Orleans, USA
lrlamo@icloud.com

17. **Shuangzhe Liu**
University of Canberra, Canberra, Australia
shuangzhe.liu@canberra.edu.au

18. **Nicola Loperfido**
Università degli Studi di Urbino Carlo Bo, Italy
nicola.loperfido@uniurb.it

19. **Augustyn Markiewicz**
Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Poznań, Poland
amark@up.poznan.pl

20. **Adam Mieldzioc**
Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Poznań, Poland
adam.mieldzioc@up.poznan.pl

21. **Monika Mokrzycka**
Institute of Plant Genetics, Polish Academy of Sciences, Poznań, Poland
mmok@igr.poznan.pl

22. **Malwina Mrowińska**
Poznań University of Technology, Poland
malwina.mrowinska@put.poznan.pl

23. **Simo Puntanen**
Tampere University, Tampere, Finland
simo.puntanen@tuni.fi

24. **Júlia Volaufová**
LSUHSC, New Orleans, USA
jvolaufova@icloud.com

25. **Dietrich von Rosen**
Swedish University of Agricultural Sciences, Uppsala, Sweden, Linköping University, Sweden
dietrich.von.rosen@slu.se

26. **Emelyne Umunoza Gasana**
University of Rwanda, Kigali, Rwanda
emygasana@gmail.com

27. **Douglas Wiens**
University of Alberta, Edmonton, Canada
doug.wiens@ualberta.ca

28. **Viktor Witkovský**
Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia
witkovsky@savba.sk

29. **Ivan Žežula**
Institute of Mathematics, P. J. Šafárik University in Košice, Košice, Slovakia
ivan.zezula@upjs.sk

30. **H. Sherry Zhang**
The University of Texas at Austin, USA
huize.zhang@austin.utexas.edu

Index