15^{th} International Workshop on Matrices and Statistics

IWMS-2006
June 13 - 17, 2006, Uppsala, Sweden

ABSTRACTS

Preface

On behalf of the organizing committees we would like to give all of you a warm welcome. We hope you will enjoy the conference and stay in Uppsala.

We also express our gratitude to

Center of Biostochastics,
John Wiley & Sons, Ltd.,
Scandinavian Airlines (SAS),
Swedish Research Council,
SPSS Sweden AB,
Swedish Statistical Association,
Tampereen Yliopiston Tukisäätiö,
The Linnaeus Centre for Bioinformatics (LCB),
The MathWorks
The Swedish Research Council for Environment,
Agricultural Sciences and Spatial Planning

for their financial support. Moreover, all speakers and contributors are gratefully acknowledged.

Maya Neytcheva
Dietrich von Rosen

Content

Prediction ability for PLS

Sven Ahlinder

Volvo Technology Corp. Dept., Göteborg

Abstract

PLS, Partial Least Squares, is a regression method related to Singular Value Decomposition, SVD. Simulations has been done to investigate the prediction ablility of PLS. The method used is to take different dataset from world wide web, short, long, fat and lean, and exclude parts of it in order to test the prediction of the excluded part. Only datasets with good fit (R^2) has been used. The software used is Simca from Umetrics.com.

Different exclusions were done. Exclusion of 10%, 20% up to 90% of the data was made according to t-values with high absolute value or nominal value.

The result was that the coefficients are sensitive for exclusion of some data but not to the majority. The software Simca from Umetrics.com overestimates heavily the confidence intervals for predictions, since the predictions are very much more stable than the coefficients

Asymptotic distributions of estimators in simultaneous equation models with many instruments

Anderson, Thedore W., Kunitomo, Naoto and Matsushita, Yukitoshi

Stanford University

Abstract

The asymptotic distributions of estimators of coefficients of one equation in a simultaneous equation model with many instruments are found under alternative assumptions about the number of instruments. The asymptotic optimality of the Limited Information Maximum Likelihood estimator is shown. Graphs of some distributions will illustrate comparisons of estimators.

Characterizations of EP, normal and Hermitian matrices

Oskar Maria Baksalary[1] and Götz Trenkler[2]

[1] *Adam Mickiewicz University, Poznán, Poland*
[2] *University of Dortmund, Germany*

Abstract

Various characterizations of EP, normal and Hermitian matrices are obtained by exploiting an elegant representation of matrices derived by Hartwig and Spindelböck. One aim of the present paper is to demonstrate ity usefulness when investigating matrix identities. The second aim is to extend and generalize lists of characterizations of EP, normal and Hermitian matrices known in the literature, by providing numerous sets of equivalent conditions referring to the notions of a conjugate transpose, the Moore-Penrose inverse, and the group inverse.

Nonparametric methods in multivariate factorial designs

Arne Bathke[1] **and Solomon Harrar**[2]

[1] *University of Kentucky, USA*
[2] *South Dakota State University, USA*

Abstract

We present three different multivariate nonparametric tests: ANOVA-Dempster-Type, Lawley-Hotelling, and Bartlett-Nanda-Pillai. Asymptotic theory has been derived for two different asymptotic frameworks: the situation where the number of replications is limited, whereas the number of treatments goes to infinity; and the regular setup with large sample sizes. The nonparametric tests are based on separate rankings for the different variables. For these tests, we also discuss different approximation procedures. The finite performance of the tests is investigated through simulations. It turns out that the proposed nonparametric tests perform very well as compared to their parametric competitors, especially in the presence of outliers. Application using SAS standard procedures is demonstrated with examples corresponding to both asymptotic frameworks.

Keywords

Multivariate Analysis of Variance, Nonparametric Model, Ordinal Data.

Bidiagonal decomposition and statistical computing

Åke Björck

Linköping University, Sweden

Abstract

In numerical linear algebra matrix decompositions and elementary orthogonal transformation have played a key role since the 1960's. Decompositions like the QR and SVD are closely related to several statistical problems. In this talk we focus on the bidiagonal decomposition (BD) of a matrix, $X = UBV^T$, with U and V orthogonal. This is most commonly used as a preliminary step for the SVD. We consider here its relation to Partial Least Squares (PLS) regression, a much used tool in chemometrics.

In a seminal paper Golub and Kahan (1965) gave two mathematically equivalent algorithms for computing the BD. The first algorithm, uses Householder transformations applied alternately from left and right to X. It is a very stable algorithm and suitable for most dense problems. The second algorithm uses a coupled two-term Lanczos recurrence. This has the advantage that the matrix X is not explicitly transformed and is therefore suitable for large scale problems with X sparse. But it is very sensitive to details in the implementation.

The standard implementation of PLS is different from both algorithms above. It combines the drawback of explicitly transforming the matrix X with some of the instabilities of the Lanczos approach. We argue here that a Householder implementation of PLS should be used instead.

PLS is used also to model problems with several dependent variables together. We present a band decomposition that generalizes the bidiagonal decomposition and discuss the implementation of this.

A fully algebraic AMLI method and solution of material microstructure problems

Radim Blaheta[1], Petr Byczanski[1] and Marie Blahetova[2]

[1] *Institute of Geonics AS CR, Ostrava, Czech Republic*
[2] *Technical University of Ostrava, Czech Republic*

Abstract

A classical methodology of materials science is to investigate relations between microstructure and macroscopic properties of materials. Nowadays, the advance in both numerical methods and computers allows that the micro-macro structure relations can be simulated numerically and used for understanding properties of a considered material and for development of new materials with optimized or tailored behaviour. Nevertheless, the computational effort of such multiscale simulations is still considerable, which requires to seek efficient numerical solution methods of finite element analysis and solution of the arising algebraic systems. AMLI methods belong among potentially optimal methods for the solution of finite element systems. Standard AMLI methods exploit a geometric hierarchy of finite element meshes whereas fully algebraic AMLI methods create a hierarchy of spaces and problems algebraically, in our case by aggregation. An advantage of the algebraic approach is the possibility for adapting the subproblems to the heterogeneous microstructure, which is moreover usually described statistically in representative volumes. Further, algebraic approach allows to develop method which is both efficient and robust with respect to jumps in the material properties of the constituents. The described algebraic AMLI method uses aggregation based on element stiffness and robustness of this method is explained via local analysis of the strengthened C.B.S. inequality.

Keywords

microstructure, statistically representative volumes, homogenization, fully algebraic AMLI, multilevel preconditioning

Eigenvalues and eigenvectors of a special nonsymmetric matrix

Lennart Bondesson[1] **and Imbi Traat**[2]

[1] *University of Umeå, Sweden*
[2] *University of Tartu, Estonia*

Abstract

As a continuation to the talk in IWMS05 we present further results on a special nonsymmetric matrix with consecutive integer eigenvalues. We have generalized the matrix, initially emerged from the sampling theory. So, we consider a matrix $\mathbf{A} = (a_{ij})$: $N \times N$ defined as

$$a_{ij} = \frac{r_i}{r_i - r_j}, \ i \neq j, \ \text{and} \ a_{jj} = \sum_{i;\ i \neq j} a_{ij}, \tag{1}$$

where the r_is are distinct *real* or *complex* numbers. The eigenvalues of $\mathbf{A}$ are integers $0, 1, \ldots, N-1$. We give the expressions of the right and left eigenvectors, and also simple formulae for calculating right eigenvectors recursively.

Keywords

Integer eigenvalues, Left eigenvectors, Right eigenvectors, Recursive formulae.

References

Bondesson L., Traat I. (2005). On a matrix with integer eigenvalues and its relation to Conditional Poisson sampling. *Res. Lett. Inf. Math. Sci. 8*, 155–163, (http://iims.massey.ac.nz/research/letters).

Bondesson L., Traat I., Lundqvist A. (2006). Pareto Sampling versus Sampford and Conditional Poisson Sampling, *Scand. J. Stat. 33*, (to appear).

Small area estimation with skewed data

Hukum Chandra and Ray Chambers

University of Southampton, United Kingdom

Abstract

Small area estimation (SAE) is now common in survey sampling, with several methods proposed in the literature (Rao, 2003). However, research continues on several important practical problems. For example, in business surveys data are typically skewed and the standard approach for SAE based on linear mixed models leads to inefficient estimates. In this paper we discuss SAE techniques for skewed data that are linear following a suitable transformation. In this context, implementation of the empirical best linear unbiased prediction (EBLUP) approach (Prasad and Rao, 1990) under transformation to a linear mixed model is quite complicated. However, this is not the case with the model-based direct (MBD) approach (Chambers and Chandra, 2005), which is based on weighted linear estimators. In this paper we extend the MBD approach to skewed data using sample weights derived via model calibration (Wu and Sitter, 2001) based on a lognormal model with random area effects. Our results show that this estimator is both efficient and robust with respect to the distribution of these random effects. We also examine the robustness of the method under model misspecification. An application to real data demonstrates the satisfactory performance of the method.

Keywords

Small areas, Skewed data, Model Calibration, Expected value model, Calibrated sample weights, MBD approach, EBLUP.

Testing equality of multivariate normal populations with recursive graphical Markov structure

Sanjay Chaudhuri[1] **and Michael D. Perlman**[2]

[1] *National University of Singapore, Singapore*
[2] *University of Washington, Seattle, USA*

Abstract

The classical likelihood ratio (LR) statistic ($\equiv$ Bartlett's statistic) for testing equality of covariance matrices from $k \geq 2$ multivariate normal populations can be decomposed in two ways, corresponding to two distinct step-down decompositions of the null hypothesis. The factors of the LR statistic that appear in these two decompositions can be interpreted as conditional and unconditional LR statistics for the step-wise null hypotheses, and their mutual independence under the null hypothesis facilitates the determination of the overall significance levels. Next, these decompositions are extended to obtain the LR statistic and associated step-wise procedure for testing equality of $k \geq 2$ covariance matrices that satisfy the Markov constraints determined by a common specified acyclic directed graph (ADG $\equiv$ DAG). Finally, these results are extended to the case where the mean vectors also satisfy a common recursive linear regression constraint.

Keywords

Testing equality of covariance matrices, multivariate normal distributions, likelihood ratio test, Bartlett's test, step-down procedure, recursive graphical Markov model, acyclic directed digraph, directed acylic graph, Bayesian network.

References

Perlman, M. D. (1980). Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations. *Ann. Statist. 8* 247–263.

Chaudhuri, S and Perlman M. D. (2005). Two Step-Down tests for equality of covariance matrices. *Tech. Report No 478, Dept. of Statistics, University of Washington, Seattle.*

Matrix potentiation: some of its properties and applications

Carlos A. Coelho

New Univerity of Lisbon, Portugal

Abstract

We define the matrix $C = A^B$, where A and B are matrices. Properties of this matrix operation are studied. The definition of this matrix operation allows us to define the matrices $A^n = [a_{ij}^n]$ and $a^B = [a^{b_{ij}}]$ as results from matrix operations and yet to define the matrix logarithmic operation $B = \log_A C$, where all A, B and C are matrices.

Also based on this matrix operation we are still able to define univariate and multivariate models $\underline{Y} = X^{\underline{\beta}} \cdot \underline{\epsilon}$ and $Y = X^{\beta} \cdot E$ and the corresponding estimation procedures.

Keywords

Matrix operation, matrix potentiation, matrix logarithms, properties, applications.

Inference for random effects models associated to commutative Jordan algebras

Ricardo Covas[1] **and Joao Tiago Mexia**[2]

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

Abstract

A random effects model $\mathbf{Y} = \sum_{i=1}^{w} \mathbf{X}_i \beta_i + \mathbf{e}$ is associated to the commutative Jordan algebra $\mathcal{A}$ if the matrices $\mathbf{M}_i = \mathbf{X}_i \mathbf{X'}_i, i = 1, ..., w$ and $\mathbf{I}_n$ constitute a basis for $\mathcal{A}$. It is shown how to obtain complete sufficient statistics for such models from which UMVUE are obtained for the variance components. Moreover, pivot variables are derived from the statistics that induce probability measures in the parameter spaces, which may be used to obtain confidence intervals for the variance components. Hypothesis testing for these components is then carried out trough duality

Keywords

Jordan Algebra, Random Effects Model, Pivot Variable, UMVUE

References

Covas, R. (2003). "Semi-Bayesian Inference and Variance Components Models". Master Thesis in Statistics and Optimization, New University of Lisbon - Faculty of Science and Technology (Document in Portuguese Language).

Fonseca, M. et al. (2003). "Estimators and Tests for Variance Components in Cross Nested Orthogonal Models". Discussiones Mathematicae - Probability and Statistics 23, 173-201

Weerahandi S. (1996). "Exact Statistical Methods for Data Analysis, 2nd Print". Springer Verlag,

Optimal designs under the polynomial growth curve models

Katarzyna Filipiak and Anna Szczepańska

Agricultural University of Poznań, Poland

Abstract

Optimality of designs under a multivariate polynomial growth curve models with correlated errors is studied.
If the time points in which the characteristic is measured are fixed, then by a design we mean an arrangement of treatments on units. In such a situation we determine universally optimal designs in the Kiefer sense.
If we assume that the assignment of treatments on units is fixed and that the same characteristic is measured at q time points, then by a design we mean an allocation of time points in the experiment. Since in such a situation universally optimal design does not exist, we determine A-, D- and E-optimal designs under a univariate model with correlated errors. We study optimality with respect to the scaling of time interval and with respect to the particular types of dependence.

Keywords

Growth curve models, Universal optimality, A-,D- and E-optimality, Correlation structures.

References

Filipiak, K. and A. Szczepańska (2005). A-, D- and E-optimal designs for quadratic and cubic growth curve models with correlated errors. *Listy Biometryczne - Biometrical Letters 42/1*, 43–56.

Markiewicz, A. and A. Szczepańska (2006). Optimal designs in multivariate linear models. Submitted.

Pukelsheim (1993). *Optimal design of Experiments.* New York: Wiley.

Least squares and generalized least squares in models with orthogonal block structure

Miguel Fonseca[1], **João Tiago Mexia**[1] and **Roman Zmyślony**[2]

[1] *New University of Lisbon, Portugal,*
[2] *University of Zielona Góra, Poland*

Abstract

Besides the basic model, Kronecker products of rotated models are used to isolate the variance components as parameters of a linear model. A characterization of BLUE given by Zmyślony is applied to the different models. Generalized least squares are used to complete the estimation. Models with orthogonal block structure in which the orthogonal projection matrix on the space spanned by the mean vectors commutes with the covariance matrix will be considered. It will be shown that there are BLUE for all estimable functions and, if normality is assumed, UMVUE both for estimable functions and variance components. Binary operations will be considered for these models.

An invariance property of the Fisher information matrix for time series models

Edward J. Godolphin

Royal Holloway University of London, United Kingdom

Abstract

The Fisher information matrix F is an important concept in inferential statistics which is required for the parameter estimation method of scoring whilst F¡1 provides the Cramt́er-Rao lower bound in the sense of Ĺowner ordering. The purpose of this presentation is to derive and consider an invariance property for the elements of F when the underlying model has a parametric dependent time series structure, as described by Godolphin and Bane (2006). Several illustrations are described and the implications of the invariance property are investigated, in particular for the Stein equation approach of Klein and Spreij (2003). The generalization of this property for the Fisher information matrix of time series models with seasonal dependent structure is discussed. Godolphin, E.J. and Bane, S.R. (2006) On the evaluation of the information matrix for multiplicative seasonal time series. J. Time Ser. Anal. 27, 167-190. Klein, A. and Spreij, P. (2003) Some results on Vandermode matrices with an application to time series analysis. SIAM J. Matrix Anal. Appl. 25, 213-223.

References

Godolphin, E.J. and Bane, S.R. (2006) On the evaluation of the information matrix for multiplicative seasonal time series. J. Time Ser. Anal. 27, 167-190.

Klein, A. and Spreij, P. (2003) Some results on Vandermode matrices with an application to time series analysis. SIAM J. Matrix Anal. Appl. 25, 213-223.

Matrices and moments: perturbation for least squares

Gene H. Golub

Stanford University, USA

Abstract

Given a matrix A, $(m \times n)$ a vector b, and an approximate solution vector, we are interested in determining approximate error bounds induced by the approximate solution. We are able to obtain bounds for the perturbation using the Theory of Moments.

For an $n \times n$ symmetric, positive definite matrix A and a real vector u, we study a method to estimate and bound the quadratic form $u'F(A)u/u'u$ where F is a differentiable function. This problem arises in many applications in least squares theory e.g. computing a parameter in a least squares problem with a quadratic constraint, regularization and estimating backward perturbations of linear least squares problems. We describe a method based on the theory of moments and numerical quadrature for estimating the quadratic form. A basic tool is the Lanczos algorithm which can be used for computing the recursive relationship for orthogonal polynomials. We will present some numerical results showing the efficacy of our methods and will discuss various extensions of the method.

On the joint distribution of a linear and a quadratic form in skew normal variables

A. K. Gupta[1], T. Kollo[2] and A. Selart[2]

[1]*Bowling Green State University, USA*
[2]*University of Tartu, Estonia*

Abstract

We are going to study the joint distribution of a linear and a quadratic form of a skew normal vector $\mathbf{Z}$. A p-vector $\mathbf{Z}$ is skew normally distributed with parameters $\Omega > 0$ and $\boldsymbol{\alpha} \in \mathbf{R}^p$ (Azzalini and Capitanio, 1999) if it has the density function

$$f_{\mathbf{z}}(\mathbf{z}) = 2\phi_p(\mathbf{z};\Omega)\Phi(\boldsymbol{\alpha}'\mathbf{z}), \quad \mathbf{z} \in \mathbf{R}^p,$$

where $\phi_p(\mathbf{z};\Omega)$ is the p-dimensional normal density with zero mean vector and correlation matrix Ω and $\Phi(\cdot)$ is the distribution function of the standard normal distribution $N(0,1)$. We use notation $\mathbf{z} \sim SN_p(\Omega, \boldsymbol{\alpha})$, if the random vector $\mathbf{z}$ has p-variate skew normal density.

Consider the linear form $B\mathbf{z}$ where B is $q \times p$ and the quadratic form $\mathbf{z}'A\mathbf{z}$ where $A' = A$. The joint moment generating function of the joint distribution of the linear and the quadratic form is found and some special cases examined. Finally an approximation of the density function of the joint distribution is found in the form of an Edgeworth type expansion through the skew normal and normal distribution following the ideas from Gupta and Kollo (2003).

Keywords

Skew normal distribution, Linear and quadratic form, Density expansion.

References

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Royal Statistical Society B 61*, 579–602.

Gupta, A.K. and Kollo, T. (2003). Density expansions based on the multivariate skew normal distribution. *Sankhya 65*, 821–835.

Duality between matrix variate t and matrix variate V.G. distributions

Solomon W. Harrar[1], **Eugene Seneta**[2] **and Arjun K. Gupta**[3]

[1] *South Dakota State University, USA*
[2] *University of Sydney, Australia*
[3] *Bowling Green State University, USA*

Abstract

The (univariate) t-distribution and symmetric V.G. distribution are competing models (Madan and Seneta, 1990; Epps, 2000) for the distribution of log-increments of the price of a financial asset. Both result from scale-mixing of the normal distribution. The analogous matrix variate distributions and their characteristic functions are derived in the sequel and are dual to each other in the sense of a simple Duality Theorem. This theorem can thus be used to yield the derivation of the characteristic function of the t-distribution and is the essence of the idea used by Dreier and Kotz (2002). The present paper generalizes the univariate ideas in Section 6 of Seneta (2004) to the general MGIG (Matrix Generalized Inverse Gaussian) distribution.

Generalized inverses of partitioned matrices and matrix sums: formulas, proofs, applications, and relationships

David A. Harville

IBM Thomas J. Watson Research Center, New York, USA.

Abstract

Results on the ordinary inverses of nonsingular partitioned matrices and of nonsingular matrices of the form $\mathbf{R} + \mathbf{STU}$ are briefly reviewed, and their proofs and their applications to linear statistical models are briefly discussed. In those results, the ordinary inverse of the original matrix is reexpressed in terms of the ordinary inverses of nonsingular matrices that may be smaller in dimension and/or easier to invert. For those partitioned matrices and those matrices of the form $\mathbf{R} + \mathbf{STU}$ that are not necessarily nonsingular but that satisfy certain conditions, extensions to generalized inverses are straightforward: it is simply a matter of replacing ordinary inverses with generalized inverses. The requisite conditions are described and discussed. Meyer's (1973) formula for a generalized inverse of an arbitrary partitioned matrix is reviewed, and it is shown how that formula can be used to obtain a formula for a generalized inverse of an arbitrary matrix of the form $\mathbf{R} + \mathbf{STU}$.

Keywords

Partitioned matrix, Matrix sum, Generalized inverse, Banachiewicz inversion formula, Duncan and Woodbury inversion formulas.

References

Meyer, C. D. (1973). Generalized inverses and ranks of block matrices. *SIAM J. Appl. Math. 25*, 597–602.

Equivalence of BLUEs and of BLUPs and the role of stochastic constraints

Stephen Haslett[1] and John Haslett[2]

[1] *Massey University, New Zealand,*
[2] *Trinity College, Dublin, Ireland*

Abstract

The linear mixed model (LMM) is used in a wide variety of situations including small area estimation. It also underpins methods used to study the wider class of the Generalized Linear Mixed Model (GLMM). This paper considers best linear unbiased prediction (BLUP) and best linear unbiased estimation (BLUE) for the LMM. Widely studied special cases include the linear model (LM) with correlated or uncorrelated errors.

The paper begins by revisiting the question of when two LM share a common solution. Precisely what this question means is discussed. Our approach uses an equality constrained approach involving Lagrange multipliers. For LMM we extend LM results by the use of a further equality constraint, described by Rao and Toutenburg (1995) as a ?stochastic constraint?. This method saves considerable algebra, by providing a simple algebraic connection between results for the LM and the LMM.

Generalized inverses in stochastic modelling

Jeffrey J. Hunter

Massey University, Auckland, New Zealand

Abstract

In many stochastic models, in particular Markov chains in discrete or continuous time and Markov renewal processes, a Markov chain is present either directly or indirectly through some form of embedding. The analysis of many problems of interest associated with these models, eg. stationary distributions, moments of first passage time distributions and moments of occupation time random variables, often involves the solution of a system of linear equations involving I - P, where P is the transition matrix of a finite, irreducible, discrete time Markov chain. Generalized inverses play an important role in the solution of such singular sets of equations. In this presentation the application of generalized inverses to the aforementioned problems is considered, and will include results concerning the characterization of types of generalized inverses associated with Markovian kernels, the analysis of perturbed systems, and applications to mixing times.

Non-normality of significance test statistic in adaptive regression model

Eric Iksoon Im

College of Business & Economics, Hawaii, USA

Abstract

In the adaptive regression model where the intercept is subject to two stochastic variations: permanent and transitory ones, whether the model is stationary or non-stationary crucially depends on the statistical significance of estimate of the ML estimator (g) of the ratio of permanent to total variance (γ_0). However, in a number of empirical applications of the model, the significance test has been carried out on an unsubstantiated assumption that has a truncated normal distribution. In this paper, we derive the analytical expression of the asymptotic variance of the ML estimator which explicitly shows that the regularity conditions required for normal distribution are violated. Subsequently, a Monte Carlo study is conducted to verify the theoretical implication of the analytical asymptotic variance of g is valid. We also generate statistical tables that may be useful to practitioners of the model in testing $H_0 : \gamma_0 = 0$.

Further characterizations of linear sufficiency for a given parametric function in the general Gauss–Markov model

Jarkko Isotalo and Simo Puntanen

University of Tampere, Finland

Abstract

We consider the properties of those linear statistics $\mathbf{Ty}$, which are *linearly sufficient* for the given estimable parametric function $\mathbf{K}'\beta$ under the general Gauss–Markov model $\{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{V}\}$. The concept of linear sufficiency was first introduced by Barnard (1963), Baksalary and Kala (1981), and Drygas (1983), and then extended to concern the estimation of the estimable parametric function $\mathbf{K}'\beta$ by Baksalary and Kala (1986). Formally, Baksalary and Kala (1986) defined a linear statistic $\mathbf{Ty}$ to be linearly sufficient for $\mathbf{K}'\beta$ if there exists a matrix $\mathbf{A}$ such that $\mathbf{ATy}$ is the best linear unbiased estimator, BLUE, of $\mathbf{K}'\beta$. Baksalary and Kala (1986) proved that $\mathbf{Ty}$ is linearly sufficient for $\mathbf{K}'\beta$ if and only if the null space inclusion

$$\mathcal{N}(\mathbf{TX} : \mathbf{TVX}^{\perp}) \subset \mathcal{N}(\mathbf{K}' : \mathbf{0})$$

holds.

We present further characterizations of linear sufficiency for the given estimable parametric function $\mathbf{K}'\beta$ in the general Gauss–Markov model. Our results are based on reparametrization of the original model and the Frisch–Waugh–Lovell Theorem.

Keywords

Best linear unbiased estimation, Linear sufficiency, Reparametrized model, Frisch–Waugh–Lovell Theorem.

References

Baksalary, J.K. and R. Kala (1981). Linear transformations preserving best linear unbiased estimators in a general Gauss–Markoff model. *Ann. Statist.*, *4*, 913–916.

Baksalary, J.K. and R. Kala (1986). Linear sufficiency with respect to a given vector of parametric functions. *J. Statist. Plann. Inference*, *14*, 331–338.

Barnard, G.A. (1963). The logic of least squares. *J. Roy. Statist. Soc. Ser. B*, *25*, 124–127.

Drygas, H. (1983). Sufficiency and completeness in the general Gauss–Markov model. *Sankhyā, Ser. A*, *45*, 88–98.

A practitioners approach to statistical analysis of large data sets using MATLAB and distributed computing

Klaus Juenemann

The Mathworks

Abstract

This presentation highlights new capabilities in The MathWorks product family. A particular emphasis is put on how to use MATLAB to generate statistics for large data sets and on how to solve computationally intensive problems through distributed computing.

As a starting point A MathWorks engineer will demonstrate how to access data from a variety of sources, e.g. Excel files and databases, and quickly perform qualitative and quantitative statistical analysis on this data. In particular, new capabilities of MATLAB 7.2 and Statistics Toolbox 5.2 will be illustrated.

The next step gives a general introduction on how to develop distributed MATLAB and Simulink applications and execute them on a cluster of computers inside the MATLAB development environment. This step also includes a practical demonstration on how to transform standard MATLAB code into distributed code. The demonstration is carried out using a four computer cluster and the MATLAB Distributed Computing Engine.

Finally, a speed comparison between MATLAB and C will be discussed and demonstrated. The comparison will include vector and matrix operation, as well as scalar operations and "for loops". The whetstone benchmark for simple arithmetic, array arithmetic, trigonometric functions and math functions will be executed both using C and MATLAB.

Elliptical K-N distributions

Tõnu Kollo[1], Anu Roos[1] and Dietrich von Rosen[2]

[1] *University of Tartu, Estonia*
[2] *SLU, Uppsala, Sweden*

Abstract

We examine a class of matrix elliptical distributions - K-N distributions. Matrix $\mathbf{X} : p \times q$ is from the class of K-N distributions if its density is of the form

$$f_{\mathbf{X}}(\mathbf{X}) = \frac{1}{(2\pi)^{\frac{pq}{2}}} |\Sigma|^{-1} \exp(-\frac{1}{2}a) \left(\frac{\lambda a}{pq} + 1 - \lambda \right), \quad (1)$$

where

$$a = \text{pvec } (\mathbf{X} - \mathbf{M})'(\mathbf{V} \otimes \mathbf{W})\text{pvec}(\mathbf{X} - \mathbf{M}), \quad \Sigma = \mathbf{V} \otimes \mathbf{W},$$

and $\mathbf{M} : p \times q$, $\mathbf{V} : p \times p$, $\mathbf{W} : p \times q$ and λ are the parameters of the distribution. Given density function (1) is a density of the mixture of a Kotz and a normal distribution, both with parameters $\mathbf{M}$, $\mathbf{V}$ and $\mathbf{W}$ and weights λ and $1 - \lambda$ respectively. For Kotz distribution see Fang, Kotz & Ng (1990) or Kollo & Roos (2005).

This class of distributions is interesting and useful because an Edgeworth-type approximation to the density of the maximum likelihood estimator of the location parameter in the Growth Curve model belongs to this class (Kollo & von Rosen (2005)). An interesting fact is that the distribution may not be unimodal. We shall present the condition for unimodality as well as the first moments and kurtosis characteristics of the distribution. The marginal distributions and conditional distributions are studied.

Keywords

elliptical distribution, Growth Curve model, Kotz-type distribution, location parameter, mixture distribution, normal distribution.

References

Fang, K.-T., Kotz, S. & Ng, W. (1990). *Symmetric Multivariate and Related Distributions.* Chapman and Hall, London, New York.

Kollo, T. & Roos, A. (2005) On Kotz-type elliptical distributions. In: *Contemporary Multivariate Analysis and Experimental Designs - In Celebration of Professor Kai -Tai Fang's 65th Birthday.* Eds. J. Fan and G. Li. World Scientific, New Jersey, 271–282.

Kollo, T. & von Rosen, D. (2005). *Advanced Multivariate Statistics with Matrices.* Springer, Dordrecht.

A letter of recommendation for Tarmo Pukkila

Sergio Koreisha

University of Oregon

Abstract

If you have any additional questions please feel free to contact me.

A note on the inertia of sum–of–squares matrices in linear models

Nicklas Korsell

University of Uppsala, Sweden

Abstract

We study the *inertia* of a matrix on the form $\mathbf{N} = \mathbf{P}_1\mathbf{V}\mathbf{P}_1 - \mathbf{P}_2\mathbf{W}\mathbf{P}_2$ that often appears in the analysis of linear models. Here $\mathbf{P}_1$ and $\mathbf{P}_2$ are symmetric and idempotent such that $\mathbf{P}_2\mathbf{P}_1 = \mathbf{P}_1\mathbf{P}_2 = \mathbf{P}_1$ and $\mathbf{V}$ and $\mathbf{W}$ are symmetric positive semi-definite. This type of matrix appears for example as the matrix in the quadratic form for the difference in residual–sum–of–squares of two nested linear models. An application is given in determining of the degrees of freedom of the difference of out-of-sample forecast errors between two linear models.

Keywords

Inertia, Linear models, Quadratic forms.

The diagonal elements of a projection matrix

Kristi Kuljus

Uppsala University, Sweden

Abstract

We consider the projection matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, $\mathbf{X} : n \times p$, $n > p$, which plays a central role in regression analysis. We show that the ith diagonal element of this matrix can be expressed via $p \times p$ squared minors of $\mathbf{X}$. The Cauchy-Binet formula is used to prove the result.

Linear algebra simplifies derivation of K. Pearson's chi-squared statistic for frequency distributions

Lynn R. LaMotte

Louisiana State University Health Sciences Center, New Orleans, USA

Abstract

In his 1900 paper, K. Pearson derived the distribution of $Y'\Sigma^{-1}Y$ when $Y \sim N(0, \Sigma)$ and christened it χ^2. Then he applied it to get an approximate p-value to assess the goodness of fit of an observed frequency distribution f to a fully-specified probability distribution, resulting in the now-ubiquitous form $\sum[f_i - E_0(f_i)]^2/E_0(f_i)$. Pearson transformed f to avoid the singularity of $\mathrm{Var}(f)$, but he did not address the consequent question of invariance to the choice of the transformation. Two results in linear algebra are shown here to resolve that question, as well as making the derivations of chi-squared statistics for frequency distributions simple and direct.

Principal components selection based on random matrix theory

Nils Lehmann

University of Duisburg-Essen, Germany

Abstract

To determine the reduced dimension, i.e. the number of principal components to retain, we are used to employ heuristic approaches as given in Jackson (1991). Recent progress in eigenvalue statistics of the Wishart (or Laguerre) ensemble of random matrices (Johnstone (2001), Soshnikov (2002)) points the way to inferential treatment of this task. The theory of random matrices, cf. Mehta (1991), is an active field of its own right at the interface between matrix theory and statistics. Its elegant methods and far-reaching results fascinate statisticians and physicists as well as engineers, see Muirhead (1982), Forrester *et al.* (2003) and Tulino & Verdu (2004). A short overview of findings related to principal components analysis will be given, together with a discussion of the special case of extensively many variables (Lehmann (2005)). The latter case is encountered in various settings, prominently in gene expression analysis and proteomics.

Keywords

Principal components analysis, random matrices.

References

Forrester, P.J., Snaith, N.C., Verbaarschot, J.J.M. (2003). Developments in random matrix theory, *J. Phys. A: Math. Gen.* **36** R1-R10.

Jackson, E.J. (1991). *A users guide to principal components*, Wiley, New York.

Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis, *Ann. Stat.* **29** 295-327.

Lehmann, N. (2005). Principal components selection given extensively many variables, *Stat. Prob. Lett.* **74** 51-58.

Mehta, M.L. (1991). *Random Matrices*, 2nd ed., Academic Press, New York.

Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

Soshnikov, A. (2002). A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices, *J. Stat. Phys.* **108** 1033-1056.

Tulino, A.M., Verdu S. (2004). *Random Matrix Theory and Wireless Communications*, now Publishers, Delft.

A practitioner's note on a unit-free matrix perturbation analysis

Gregory L. Light

Providence College, Providence, USA

Abstract

This paper introduces the concept of "relative derivative", $\frac{a}{b}\frac{dy}{dx}$, as a measure of proportional sensitivity. We first define the term for $x \in \mapsto y \in$ and re-express the Taylor series; then we re-express the Taylor's theorem in ^{n}. The advantages of relative derivatives are: (1) they are unit-free and thus are more amenable to computer simulations as inputs; (2) by judicious selection of the "scaling units" (a, b) one may simplify a differential relationship or reveal new proportional relationships.

We apply relative derivatives to matrix functions [*Magnus*, *Neudec*ker, 1999], to express how the determinant, the (Moore-Penrose) inverse, the eigenvalues and eigenvectors [*Pillis*, *Neumann*, 1985; *Andrew*, *Chu*, *Lancaster*, 1993], and matrix powers [*Sebastiani*, 1996] change proportionately to some scaling units if the underlying matrix undergoes a perturbation. We also apply relative derivatives to a sensitivity analysis of the solutions to systems of linear equations [*Chandrasekaran*, *Ipsen*, 1995] and of the least-squares estimates in regression [*Malyshev*, 2003]; e.g., for the latter we have

$$\frac{db_i}{s_{b_i}} = r_{b_i b_j} \left(\frac{s_{b_j} s_{x_j}}{s_e} \right) \sum_{m=1}^{k} \left[\left(\frac{dx_{j,m}}{s_{x_j}} \right) \left(\frac{e_m}{s_e} \right) + \left(\frac{de_m}{s_e} \right) \left(\frac{x_{j,m}}{s_{x_j}} \right) \right],$$

where b_i is the estimated coefficient of the independent variable X_i, s_{b_i} is the estimated standard error of b_i, $r_{b_i b_j}$ is the sample coefficient of correlation between b_i and b_j, e_m is the residual for observation m, and s_{x_j} and s_e are respectively the sample standard deviation of the independent variable X_j and the residual standard deviation from a random sample of size k.

In summary, our approach facilitates fractional analyses; we relax the logarithmic derivative $\frac{x}{y}\frac{dy}{dx}$ [*Lundy*, *Sen*, 1995; *Kirkland*, *Neumann*, *Xu*, 2004] to a general $\frac{a}{b}\frac{dy}{dx}$; our idea is not new [*Aversa*, Pr*eiss*, 1999; *Saltelli*, *Chan*, *S* cot *t*, 2000], but it deserves better appreciation, and this paper applies it to matrix functions, showing its potential utility in multivariate analyses, likelihood theory, experimental designs [*Olkin*, 1984], etc. in statistics.

Keywords

Proportional sensitivity, relative derivative, scaling, conditioning.

A generalized propensity score approach to comparing the costs of health care episodes

Antti Liski and Reijo Sund

National Research and Development Centre for Welfare and Health, Helsinki, Finland

In order to enhance the performance of the health system there is a need for comparative information suitable for benchmarking purposes. One widely used approach in the performance assessment is the disease-based comparison of episodes of care between providers (hospitals or care districts). The basic idea is to produce performance indicators which are adjusted for confounding factors (differences in population characteristics). Costs are obviously an important performance measure, but the adjustment has turned out to be methodologically challenging. The aim of this study is to develop a new method for comparing the costs of health care episodes. The method is presented in the context of hip fracture care episodes and comparisons are made between the care districts in Finland. The data consists of the population of 16881 hip fracture patients aged 65 or older in 1998-2001 and includes information for patient characteristics to be adjusted as well as the cumulative costs of treatment at one year after the fracture.

The differences of patient characteristics between care districts were analysed by utilising a weighted multinomial logit model and the predicted values were used as a generalized propensity score. The data were sorted by the propensity score in each care district. Then the costs within each district were smoothed over the propensity score by using local first degree polynomial fitting with normal weights. This technique results in smoothed cost-curves over the scale of the propensity score. The cost-curves were integrated over the propensity score to obtain estimates for the costs of each care district.

It turns out that there are clear differences between the cost-levels of various care districts. Thus the cost variation of hip fracture treatments between care districts can not be explained by patients characteristics only, but is also attributable to different treatment practices. This is an important finding because the identification of good practices can lead to learning from them, and consequently also to savings in treatment costs.

Keywords

Hip fracture, Health care costs, Kernel smoothing, Propensity score.

References

Imbens, G. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika. 87*, 706–710.

Rosenbaum, P.R. and D.B. Rubin (1983). The central role of the propensity score in observational studies for causal effects *Biometrika 70*, 41–55.

On pseudo maximum likelihood estimation for multivariate time series models with conditional heteroskedasticity

Shuangzhe Liu[1] and Heinz Neudecker[2]

[1] *University of Canberra, Australia,*
[2] *University of Amsterdam, The Netherlands*

Abstract

We consider a general multivariate conditionally heteroskedastic model under a conditional distribution that is not necessarily normal. This model contains autoregressive conditionally heteroskedastic (ARCH) models as a special class. We use the pseudo maximum likelihood estimation method and derive a new estimator of the asymptotic variance matrix for the pseudo maximum likelihood estimator. We also study three special cases in this class, which are conditionally heteroskedastic autoregressive moving-average models, regression models with ARCH errors and models with constant conditional correlations.

Keywords

CHARMA, R-ARCH, CCC, pseudo maximum likelihood, asymptotic variance matrix, matrix differential calculus

Efficient evaluation of the residual sum of squares for quantitative trait locus mapping in the case of complete marker genotype information

Kajsa Ljungberg and Sverker Holmgren

Uppsala University, Sweden

Abstract

A core computation of many popular quantitative trait locus, QTL, mapping methods is determining the residual sum of squares, RSS, for a regression of trait values on (pseudo-)marker genotypes. A single evaluation is easily performed using the standard method QR factorization, but together the RSS computations take considerable time and often constitute the major part of the computational effort. We present an algorithm for RSS evaluation that is mathematically equivalent to evaluation via QR factorization but 10-100 times faster depending on the model and data dimensions. It can be used for all standard QTL models. Our method opens the possibility for more detailed data analysis and more extensive model comparisons.

Kiefer optimal designs in multivariate linear models

Augustyn Markiewicz

Agricultural University of Poznań, Poland

Abstract

In the literature only few results are established for design optimality in multivariate linear models, mainly on optimal choice of time points in growth curve models. We study optimality of experimental designs in multivariate linear models with a known or unknown dispersion matrix. In the case of known dispersion matrix we use the concept of Kiefer optimality while in the case of unknown dispersion matrix optimality is considered with respect to the precision in chosen estimation method of parameters of interest; for example in maximum likelihood estimation. We show relations between optimality of designs in univariate models and in their multivariate extensions.

Keywords

Multivariate linear model, Growth curve model, Estimable function, Optimal design, Kiefer optimality.

Testing the equivalence of two covariance matrices

Thomas Mathew

University of Maryland, Baltimore, MD, USA

Abstract

The problem of testing the equivalence of two means, or the equivalence of two variances of univariate populations come up in some applications, most notably in the assessment of bioequivalence between a brand name drug and a generic copy. The corresponding multivariate problems are relevant in the context of multivariate bioequivalence testing. My talk deals with the development of possible criteria to be used in order to assess the equivalence between two covariance matrices, and the derivation of tests based on these criteria. The problem will be described and investigated in the standard set up of two independent Wishart matrices. While testing the equality of two covariance matrices is quite straightforward in this context, the hypothesis of equivalence presents some difficulties. In the talk, I will describe approximate test procedures, along with numerical results on the performance of the tests. An example from multivariate bioequivalence testing will be used to illustrate the results.

Petr Mayer

Czech University of Technology

Bounds for the Perron root using the sum of entries of matrix powers

Jorma K. Merikoski and Ari Virtanen

University of Tampere, Finland

Abstract

Let $\mathbf{A} = (a_{ik})$ be a nonnegative square matrix with Perron root $r(\mathbf{A})$. We survey and pursue further the question of finding bounds for $r(\mathbf{A})$ using the sum of entries of powers of $\mathbf{A}$ and $G(\mathbf{A}) = (\sqrt{a_{ik}a_{ki}})$ respectively. We encounter three open problems.

Numerical methods for the REML method in genetic analysis of complex traits

Kateryna Mishchenko[1], Lars Rönnegård[2], Örjan Carlborg[2], Sverker Holmgren[2]

[1] *Mälardalen University, Sweden*
[2] *Uppsala University, Sweden*

Abstract

We study numerical methods for the estimation of variance components by the Restricted Maximum Likelihood (REML) method. The aim is to obtain an efficient and robust computational scheme for mapping of Quantitative Trait Loci (QTL) in genetic analysis of complex traits. Two computational subproblems are considered; The optimization procedure for maximizing the likelihood, and the evaluation of this quantity and its derivatives. Several Newton-type optimization schemes where the constraints are taken into account in different ways are studied. Also, different schemes for approximating the Hessian of the objective function are compared. We improve the efficiency for the evaluation of the gradient of the likelihood using numerical linear algebra techniques. The efficiency and robustness of the new numerical methods is compared to previously used schemes for data sets from animal experiments.

BLUE or BLUP - question about model or about estimator properties?

Märt Möls[1], Simo Puntanen[2] and Jarkko Isotalo[2]

[1] *University of Tartu, Estonia*
[2] *University of Tampere, Finland*

Abstract

In certain situations it can be difficult to decide wether a particular factor should be considered as random or fixed. Questions about underling model are not always easy to answer. However, as we will show, there exists another approach to the problem. Namely one can fix the underlying model (be it random, fixed or mixed) and then, for a given model and a given factor, discuss the properties of BLUP (Best Linear Unbiased Predictor) and BLUE (Best Linear Unbiased Estimator). We will show that even for the same underlying model both BLUE and BLUP might have desirable properties. For example sometimes one might prefer to predict random factor effects using the formula for BLUP (treat it as a random factor) and sometimes one might prefer the formula for BLUE (treat it as fixed factor) - depending on the estimator properties one desires to have. Same can be shown to hold for a fixed factor. Selecting an estimator based on its properties can be easier than deciding which underling model is the correct one. In the talk we compare the properties of BLUP and BLUE for several underlying models.

Keywords

Linear Mixed Model, BLUE, BLUP.

Shift invariant permutations in linear random factor models

Tatjana Nahtman[1] and Dietrich von Rosen[2]

[1] *University of Tartu, Estonia; Karolinska Institutet, Sweden*
[2] *SLU, Uppsala, Sweden*

Abstract

The objective is to present a comprehensive study of the covariance structures in linear models which are represented with the help of Toeplitz matrices. We outline the conditions under which the covariance matrix of a given random factor will provide certain types of reparameterizations.

The best way to describe invariance properties of random factors, including interactions, is via their covariance matrices. Because of invariance it appears that the natural quantities to study are the eigenvalues and eigenvectors. It is easy to imagine that restrictions on the factor levels will lead to singular covariance matrices with eigenvalues equal to 0. The corresponding eigenvectors then tell us what kind of restrictions can be put on the factors.

Keywords:

Covariance Structures, Invariance, Marginal Permutations, Reparameterization, Spectrum, Toeplitz matrix.

References:

Basilevsky, A. (1983). *Applied matrix algebra in the statistical sciences*. New York, North-Holland.

Nahtman, T. (2006). Marginal permutation invariant covariance matrices with applications to linear models. *Linear algebra and Its Applications*, (accepted)

On the asymptotic distribution of the 'natural' estimator of Cronbach's alpha with standardized variates under nonnormality, ellipticity and normality

Heinz Neudecker

University of Amsterdam, The Netherlands

Abstract

Following van Zyl, Neudecker and Nel (2000) we consider asymptotic properties of the 'natural' estimator of Cronbach's alpha when the variates are standardized. This means that the population correlation matrix P is the population variance matrix Σ, because now all diagonal elements of Σ are equal to unity. The 'natural' estimator $\hat{\alpha}_s = (p-1)^{-1}p[1-p(1'R1)^{-1}]$, where R is the sample correlation matrix and p is the number of items (responses). We find the asymptotic distribution of $\hat{\alpha}_s$ under nonnormality, ellipticity and normality. Use is made of a (0,1) 'duplication' matrix D. This enables us to easily switch between vec$A$ and w(A), where A is a symmetric zero-axial matrix ($A_d = 0$) and w(A) contains the 'free' elements of A.

Keywords

Asymptotic distribution, Cronbach's alpha, nonnormality, ellipticity and normality.

References

van Zyl, J.M., H. Neudecker, and D.G. Nel, (2000) On the distribution of the maximum likelihood estimator of Cronbach's alpha, *Psychometrika 65(3)*, 271–280.

Multivariate sign test and spatial median for clustered data

Jaakko Nevalainen[1], Denis Larocque[2] and Hannu Oja[1]

[1] *University of Tampere, Finland,*
[2] *HEC Montréal, Québec, Canada*

Abstract

Clustered data consists of n independent clusters $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ but the observations in the same cluster $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{im_i})$ may be dependent. Examples of such data arise when a study involves students of several school classes, or siblings of several families. Unless this dependency is carefully taken into account, the tests and estimates tend to result in falsely small p-values and estimated standard errors.

Our interest is to consider the multivariate location problem for clustered data. Multivariate sign test and spatial median are developed for this setting. In pursuit of improvement in efficiency, weighted versions of the test and the estimator are also considered. Limiting as well as finite sample efficiencies, relative to the sample mean vector and to the unweighted spatial median, are explored for different designs. We learn that the efficiency of the spatial median suffers less from intracluster dependencies than the mean vector, and also that weighting can substantially improve precision. Breakdown points of the weighted estimators are discussed. Finally, affine invariant/equivariant versions of the test/estimator are described.

References

Larocque, D., Nevalainen, J. and Oja H. (2006). A weighted multivariate sign test for cluster correlated data. Conditionally accepted to *Biometrika*.

Larocque, D. (2003). An affine-invariant sign test for cluster correlated data. *The Canadian Journal of Statistics*, 31: 437–455.

Nevalainen, J., Larocque D. and Oja H. (2005). On multivariate spatial median for clustered data. Submitted.

Inequalities: Some probabilistic, some matrix, and some both

Ingram Olkin

Stanford University, USA

Abstract

Although there have been theories of linear equations, functional equations, differential equations. and so on, there are few theories of inequations (inequalities). There are some general themes such as convexity or majorization, but mostly, inequalities are obtained by a variety of methods. It may be surprising that there is a connection between some probabilistic inequalities and matrix inequalities; these come about from the fact that quadratic forms have a probabilistic interpretation. In this survey we exhibit the connection between probabilistic inequalities and matrix inequalities. We also present some results on the mean of certain random determinants.

Matrices and Politics

Friedrich Pukelsheim

University of Augsburg, Germany

Abstract

In proportional representation systems, biproportional apportionment methods provide a novel procedure to translate electoral votes into parliamentary seats. They are applicable in systems where the whole electoral region is subdivided into smaller electoral districts, and where the districts are to receive a pre-specified number of seats. Thus the input is a vote matrix consisting of the number of votes per district (rows), and per party (columns). The output is an apportionment matrix of corresponding seat numbers. Within a district, the sum of the seat numbers must be equal to the pre-specified district seats; within a party, the sum of the seat numbers must be such that it reflects the success of the party across the whole electoral region. Biproportional apportionment methods are such, that they achieve a double proportionality, across districts relative to the censors population counts, and across parties relative to their vote totals in the electoral region. The method has been used for the first time in the February 2006 election of the Zurich City Parliament.

What kind of research would I carry out at a university?

Tarmo Pukkila

Ministry of Social Affairs and Health, Helsinki, Finland

Abstract

Ageing is one of the biggest challenges all over the world in the future. In this talk I will consider the possible implications of ageing. For example in Finland the number of people in the age group 65+ is about 800 000 today. In 25 years this age group will grow to about 1 400 000, i.e. by about 600 000 seniors. At the same time the number of working age people will decrease by about 300 000 people. Already these numbers indicate that that we might have very difficult equalities and inequalities to solve. Finland is not alone with the ageing problem. We have to take into account also the ongoing globalization process. A big and important problem is the following: How do we finance the social security in the future?

Convergence issues of some classes of IAD methods in computing Markov chains

Ivana Pultarová

Czech Technical University in Prague, Czech Republic.

Abstract

The iterative aggregation - disaggregation (IAD) methods present a multilevel approach in solution a stationary probability distribution vector of a finite discrete time Markov chain represented by an irreducible stochastic matrix B. They consist in combining some basic iteration with solution a certain coarse scale problem.

The present work introduces some new results on the convergence properties of some classes of the IAD methods. We study the methods where the basic iteration corresponds to a polynomial matrix iteration, where the polynomial p fulfils $p(1) = 1$. It is shown that when $p(t) = \alpha t + (1-\alpha)$, $\alpha \in (0,1)$, then the algorithms converges in local sense for any irreducible matrix B. The existence and uniqueness of the fixed point is proved for $\alpha \in (0,1\rangle$. In the case $p(t) = t$, the local convergence is obtained for any irreducible matrix B with a positive diagonal. In addition to it, when B contains at least one positive row, then the asymptotic convergence factor can be estimated.

New examples of divergence of the IAD methods are introduced for $p(t) = t$ or t^2. Also some observations concerning rapid convergence, i.e. obtaining the exact solution within finite number of steps, are presented.

The introduced theory is based on sparsity structures of stochastic matrices contrary to the quantitative analysis of convergence properties of the IAD methods for nearly completely reducible Markov chains. In order to find the boundary line between the local convergence and divergence in local sense, some new questions and hypothesis are formulated.

Keywords

Iterative aggregation - disaggregation methods, Markov chains, Stationary probability distribution vector.

References

Marek, I. and Pultarová, I. A note on local and global convergence analysis of iterative aggregation-disaggregation methods. To appear in Linear Algebra Appl.

Stewart, W. J. (1994). *Introduction to the numerical solutions of Markov chains*. Princeton University Press.

Tarmo Pukkila through my camera

Simo Puntanen

University of Tampere, Finland

Abstract

The photo is taken in Haikala's farm, in 1983. Persons from left: Gustav Elfving, Mrs Rao, Kaisa Huuhtanen, Helena Pukkila, C.R. Rao and Tarmo Pukkila.

On the role of the constant term in linear regression

Simo Puntanen[1], Jarkko Isotalo[1] and George P. H. Styan[2]

[1] *University of Tampere, Finland,*
[2] *McGill University, Montréal, Québec, Canada*

Abstract

In this talk we review some properties of the constant term in linear regression, that in particular come across while teaching statistics. The constant term corresponds to a variable whose observed values are all identical and hence its variance is zero. Hence students might wonder if such a variable indeed a decent variable. We go through some important geometric considerations and comment on various models and the role of the constant term therein.

Keywords

BLUE, centering, coefficient of determination, collinearity, OLSE, orthogonal projector.

A philatelic introduction to matrices and statistics

Simo Puntanen[1] and George P. H. Styan[2]

[1] *University of Tampere, Finland,*
[2] *McGill University, Montréal, Québec, Canada*

Abstract

Stamp collectors may consider the rich and fascinating world of postage stamps to be "a mirror of civilization". This was indeed a conclusion made by William L. Schaaf in the preface to his 1978 book entitled Mathematics and Science: An Adventure in Postage Stamps, exemplifying that "multitudes of stamps reflect the impact of mathematics and science on society". We agree with Schaaf's comments, and in this talk we look at postage stamps that have some connection with matrices and statistics.

The new challenges of numerical algebra

Yousef Saad

University of Minnesota, USA

Abstract

Research in numerical linear algebra is changing rapidly because of the new trends in the sciences and the world economy. For example, the current information revolution has given rise to a large number of new challenges to computer scientists and mathematicians. Many of the numerical methods in scientific computing of the past few decades were motivated by problems arising from the solution of PDEs – as a response to demands of now established industries, such as the aerospace industry (Navier-Stokes). A few of the new trends that are likely to shape the decades ahead are: (1) Nanotechnology (e.g. materials science), (2) Biology and genetics (3) information technology. In each case there are enormously challenging problems to solve which require matrix algorithms that are far more powerful than those available today. I will discuss a few of these problems. In materials science, the fundamental equation is Schrodinger and the basic problem is an eigenvalue problem with many eigenvalues. In genetics, one of the major issues is clustering, i.e., recognizing patterns in genes. In information sciences, one can mention again clustering (e.g., pattern recognition) but also the broad problem of dimensionality reduction (e.g., Principal Component Analysis) which is essential for handling huge data sets.

Multivariate analysis with fewer observations than the dimension: A Review

Muni Srivastava

University of Toronto, ON, Canada

Abstract

In DNA microarray data, gene expressions are available on thousands of genes of an individual but there are only few individuals in the dataset. Although, these genes are correlated, most of the statistical analyses carried out in the literature ignore this correlation without any justification. For example, if in one-sample problem it is found that the data supports the sphericity hypothesis about the covariance matrix Σ, that is $\Sigma = \sigma^2 I_p$ for some unknown σ^2 and $p \times p$ identity matrix I_p, then any inference on the mean vector may ignore the correlation between the genes, and use the univariate methods. However, it has not been done in the literature. Similarly, if the covariance matrix is a diagonal matrix, then univariate methods with unequal variances for the components may be used for any inference. For example, Dudoit et al. (2002) assumed that the covariance matrix is a diagonal matrix in their classification procedure without verifying that the data support this assumption. In fact, it is shown in Srivastava (2005a) that the data do not support this assumption.

Throughout this article, we shall assume that the sample size N is less than the dimension p. This in turn implies that no tests invariant under the nonsingular linear transformations exist for testing the hypothesis on the mean vector in one-sample, and mean vectors in two-sample and many samples, the so called MANOVA problem, see Lehmann (1959, p.318) or Eaton (1983). Similarly, likelihood ratio tests for the hypothesis on the covariance matrices in two or more than two populations do not exist. Thus various test criteria have been recently proposed to verify the assumptions made on the covariance matrix or matrices in two or more than two populations. Similarly, several test criteria have been proposed for the inference on the mean vectors. In this article, we review these procedures.

Criteria for block P-property

Tomasz Szulc[1] and Ludwig Elsner[2]

[1]*Adam Mickiewicz University, Poznań, Poland*
[2]*University of Bielefeld, Germany*

Abstract

The notion of block P-property has been introduced by L. Elsner and T. Szulc in [*Linear and Multilinear Algebra* 44 (1998), 1-12]. This concept has many analogies to the usual P-property and it is very useful to characterize basic matrix properties (nonsingularity, stability, Schur stability etc.) of convex combinations of finite many real matrices (see for example the papers by L. Elsner and T. Szulc [*Linear and Multilinear Algebra* 44 (1998), 301-312], [*Linear and Multilinear Algebra* 48 (2000), 1-19] and by L. Elsner, V. Monov and T. Szulc [*Linear and Multilinear Algebra* 50 (2002), 199-218]). The paper presents some criteria for the block P-property for an n-by-n real matrix. In particular, some results on a 4-by-4 real matrix are derived - they are based on the concept of strict copositivity.

Keywords

Block P-property, P-property, convex combinations of matrices, copositive matrices.

Extending the Frisch scheme for dynamic errors-in-variables problems to correlated output noise

Torsten Söderström

Uppsala University, Sweden

Abstract

The Frisch scheme is a well established method for identifying dynamic systems where both input and output measurements are corrupted by noise. The method has its root in static problems with applications to econometrics. When applied to identification of dynamic systems, it has so far been limited to the case when the output noise is white. The talk will present some possibilities for generalizing the method to handle correlated output noise.

Large deviations and exact Bahadur efficiency of the Lilliefors test of exponentiality.

Anna Tchirina

St.Petersburg Electrotechnical University, Russia

Abstract

We study the asymptotic properties of a well-known Kolmogorov – Smirnov exponentiality test with estimated scale parameter. The test statistic is defined as follows:

$$Li_n = \sup_{x \geq 0} |1 - e^{-(x/\overline{X})} - F_n(x)|$$

with F_n for the empirical distribution function and $\overline{X}$ for the sample mean.

First simulation results for this test appeared in [1]; for the later modifications and improvements see e.g. Durbin (1975), D'Agostino & Stephens (1986). The weak convergence of the corresponding empirical processes was studied in [2]. The new result presented here solves an old problem of finding the rough large deviation asymptotics of Li_n under the hypothesis of exponentiality, namely

$$\lim_{n \to \infty} n^{-1} \ln \mathbf{P}(Li_n \geq a) = k(a) = \max(k_1(a),\ k_2(a)), \quad \text{where}$$

$$k_1(a) = \sup_{x \geq 0} \inf_{\mu \geq 0} \inf_{\lambda \geq 1} \{-\mu(a + e^{-x}) + \lambda - 1 - \ln \lambda - \lambda x + \ln(e^{\mu} + e^{\lambda x} - 1)\} \quad \text{and}$$

$$k_2(a) = \sup_{x \geq 0} \inf_{\mu \geq 0} \inf_{0 \leq \lambda \leq 1} \{-\mu(a - e^{-x}) + \lambda - 1 - \ln \lambda - \lambda x + \ln(e^{-\mu} + e^{\lambda x} - 1)\}.$$

Moreover, we prove the continuity of $k(a)$ and obtain its asymptotics in the neighbourhood of the zero:

The proof is based on the new theorem from [3] and the standard techniques for supremum-type statistics.

Using the large deviation asymptotics we find the exact Bahadur efficiency of the test for various families of alternatives.

Keywords

Lilliefors statistics, Exponentiality, Large deviations, Bahadur efficiency.

References

Lilliefors, H. (1969) . On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. *JASA 64*, 387–389.

Rao, J.S. and Sethuraman, J. (1975) . Weak convergence of empirical distribution functions of random variables subject to perturbations and scale factors. *Ann.Statist. 3*, 299–313.

Tchirina, A. (2005). Large Deviations for a Class of Scale-Free Statistics under the Gamma Distribution. *J. Math. Sci. 128*, 2640–2655.

Regularization techniques in model fitting and parameter estimation

Sabine Van Huffel and Diana Sima

Katholieke Universiteit Leuven, Belgium

Abstract

Regularization is any technique that alters the formulation of an estimation problem in order to stabilize the numerical computations and to robustify the solution. We discuss several special topics involving regularization. These include new insights related to introducing regularization for linear discrete ill-posed problems in the context of the Total Least Squares problem (truncated TLS and regularized TLS). Regularization is also used for smoothing scatter data with nonparametric or semiparametric regression. New results that study the effect of a regularization constraint in the context of nonlinear semiparametric regression are presented. The considered semiparametric model formulation is motivated by a biomedical application

Variable selection issues in generalized linear models

Júlia Volaufová and Lynn R. LaMotte

School of Public Health, Louisiana State University Health Sciences Center, New Orleans, USA

Abstract

In generalized linear models as well as in any multiple regression setting, often one goal is to find a few good predictive models. It is a known fact that the all-possible-submodels approach would be the best approach if it is computationally feasible. For example, in procedures dealing with logistic regression, the algorithm developed by Furnival, G.M. and Wilson, R.W. (1974) is implemented in statistical packages such as SAS. While in a multiple regression setting the criterion is the R^2 value, by analogy, in GLM it might be the likelihood ratio χ^2 statistic. However, if one chooses a different criterion, such as the score statistic, it simplifies the calculations since it avoids refitting the model and parameter estimation for each submodel under investigation. Here several issues related to the equivalence with multiple regression and differences regarding the LR statistic are presented.

More on Projectors

Hans Joachim Werner

University of Bonn, Germany

Abstract

Projectors (i.e., idempotent matrices) are closely related to generalized inverses of matrices and play an important role in many fields such as linear algebra, probability theory and statistics. This talk gives further insights into the theory of projectors. We present a powerful necessary and sufficient condition for the Moore-Penrose inverse of a matrix to be idempotent. By means of this, we then prove a series of results for certain kinds of matrices.

Why do the simulation extrapolation procedures work in EIV?

Silvelyn Zwanzig

Uppsala University, Sweden

Abstract

Cook and Stefanski (1994) have introduced a simulation extrapolation procedure (SIMEX). The idea of the SIMEX procedures is to compensate for the effect of the measurement errors while still using the naive regression estimators. In the report Polzehl and Zwanzig (2005)we have proposed a symmetrized and an adaptive version of SIMEX. Both new procedures deliver approximations of the total least squares estimator. The quality of the approximations depends on the simulation size only. In the talk I will discuss the underlying matrix relations.

References

Cook, J.R. and Stefanski, L.A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, *89*, 1314-1328.

Polzehl, J. and Zwanzig, S. (2005). SIMEX and TLS: An equivalence result, preprint, WIAS Berlin, No 999, ISSN 0946-8633.