UNIVERSITY OF WINDSOR

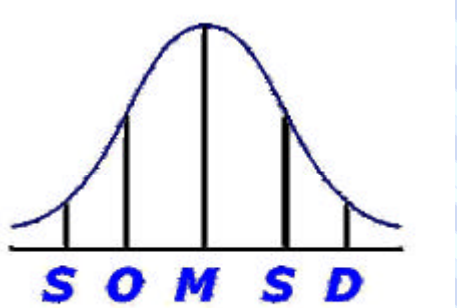
Department of Mathematics and Statistics

presents

Southern Ontario Matrices and Statistics Days

$$\begin{bmatrix} S & M \\ & O \\ S & D \end{bmatrix}$$

Program



University of Windsor
June 9-10, 2005

Organized by
S. Ejaz Ahmed (University of Windsor)
Abdul Hussein (University of Windsor)
George P. H. Styan (McGill University)

Special thanks to: WINDSOR THOMSON NELSON for their generous support.

Southern Ontario Matrices and Statistics Days

Thursday June 9

08:00-09:00 Registration

09:00-9:30 Opening Remarks

10:30-11:00 Coffee break

Session 1

9:30-10:00 George P. H. Styan, McGill University
Some comments on the life and work of Jerzy K. Baksalary (1944-2005)

10:00-10:30 Oskar Maria Baksalary, Adam Mickiewicz University
On linear combinations of various projectors

10:30-11:00 **Coffee Break**

11.00-12.00 **KEYNOTE SPEAKER**
P.K. Sen, Departments of Biostatistics, and Statistics and Operations Research, University of North Carolina at Chapel Hill
The Curse of Dimensionality in Genomics: Beyond the Euclidean Paradigm

12:00-1:30 **Lunch Break**

Southern Ontario Matrices and Statistics Days

Thursday, June 9

Session 2

1:30-2:30	**KEYNOTE SPEAKER** Simo Puntanen, University of Tampere, Tampere, Finland **Matrix tricks for linear statistical models: our personal Top Fourteen**
2:30-3:00	Shaun Fallat, University of Regina **Statistics and Total Positivity**
3:00-3:30	Yoshio Takane, McGill University **Regularization Methods in Multivariate Analysis**
3:30-4:00	Edit Gombay, University of Alberta **Monitoring Parameter Change in the Presence of Nuisance Parameters**
4:00-4:30	A.K. Gupta, Bowling Green State University
4:30-5:00	Charlie Johnson, Department of Mathematics, The College of William & Mary
5:00-5:30	Xiaoyong Wu, University of Windsor
6:00	Dinner—Location TBA

Southern Ontario Matrices and Statistics Days

Friday, June 10

Session 3

8:00-9:00	**KEYNOTE SPEAKER** Malay Ghosh, Department of Statistics, University of Florida **Ancillary Statistics: A Review**
9:00-9:30	Kjell Doksum, University of Wisconsin, Madison **Powerful Choices: Tuning parameter selection based on power**
9:30-10:00	Jörg Kaufmann, ScheringAG, Berlin
10:00-10:30	Coffee Break
10:30-11:00	Janusz Kawczak, University of North Carolina at Charlotte **On the Problem of Finding Minimax Estimators**
11:00-11:30	Steve Kirkland, University of Regina **Conditioning of the entries in Google's PageRank vector.**
11:30-12:00	Jarkko Isotalo, University of Tampere, Finland **Comparisons of OLSE and BLUE in the General Gauss-Markov Model**
12:00-12:30	Ka Lok Chu, Dawson College, Montreal **Inequalities and equalities associated with the Watson efficiency in orthogonally partitioned full rank linear models**
12:30-2:00	**Lunch Break**

Southern Ontario Matrices and Statistics Days

Friday, June 10

Session 4

2:00-3:00 **KEYNOTE SPEAKER**
Henry Wolkowicz, University of Waterloo
Symmetric Kronecker Products and Uniqueness and Existence of search directions for semidefinite programming (SDP)

3:00-3:30 S.B. Provost, University of Western Ontario
An explicit representation of the elements of certain patterned matrices

3:00-3:30 Debashis Ghosh

3:30-4:00 Abdo Alfakih, University of Windsor

4:00-4:30 Andrei Volodin, University of Regina
On the Golden Ratio, strong law, and first passage problem

4:30-5:00 Saswat Panigrahi, McGill University

5:00-5:30 Yongge Tian, University of Alberta

5:30 **Closing Remarks**

Southern Ontario Matrices and Statistics Days

Abstracts

(in alphabetical order)

Abdo Alfakih
University of Windsor

An $n \times n$ matrix $D = (d_{ij})$ is said to be a *Euclidean distance matrix (EDM)* if there exist points $p^1; p^2; \ldots ; p^n$ in some Euclidean space such that $d_{ij} = \|p^i - p^j\|^2$ for all $i; j = 1; \ldots ; n$. The Euclidean distance matrix completion problem (EDMCP) is of great interest in many different fields among them: multidimensional scaling in statistics, molecular conformations in computational chemistry, and wireless sensor networks in computer science. In this talk we discuss some problems related to the EDMCP in the context of Semidefinite programming. In particular, we present a randomized algorithm for obtaining EDM completions in low dimensional Euclidean spaces.

On linear combinations of various projectors

Oskar Maria Baksalary
Adam Mickiewicz University

During my four year long joint cooperation with Jerzy K. Baksalary, we considered three problems dealing with linear combinations of various projectors. The first of them concerns the question of when a linear combination of two projectors (idempotent matrices) is also a projector. The complete solution to this problem was obtained in 2000, at the very beginning of our cooperation. The second problem, solved in 2004, is an analogue of the first one with generalized projectors instead of projectors, where the notion of a generalized projector is understood as a matrix whose second power is equal to its conjugate transpose. Recently, we considered still another version of the above mentioned question, with projectors replaced by hypergeneralized projectors, where a hypergeneralized projector is defined as a matrix whose second power is equal to its Moore-Penrose inverse. However, investigations concerning the question of when a linear combination of two hypergeneralized projectors is also a hypergeneralized projector are more complicated than those referring to projectors and generalized projectors and the complete answer to this question remains unknown.

(Joint cooperation with Jerzy K. Baksalary)

Statistics and Total Positivity

Shaun Fallat
University of Regina

A matrix is called totally positive if all of its minors are positive. This class of matrices grew out of the work on oscillations in vibrating systems of Gantmacher/Krein, and independently by Schoenberg who was studying certain kinds of linear transformations. However, since the pioneering work of Karlin and others it has become clear that this class also arises in many facets of Statistics. I intend to highlight some of the applications of total positivity in Statistics, and make use of some recent advances on this class to reprove some implications for moment matrices.

Ancillary Statistics: A Review

Malay Ghosh
Department of Statistics
University of Florida

Ancillary statistics are one of R. A. Fisher's most fundamental contributions to statistical inference. These are the statistics whose distributions do not depend on the model parameters. However, in conjunction with some other statistics, typically the MLE, they provide valuable information about the parameters of interest.

The present article is basically a review of some of the uses and limitations of ancillary statistics. Due to the vastness of the subject, my account is, by no means, comprehensive. The topics selected re°ect my own interest, and I am very well aware that many important contributions to the subject are left out.

I have touched upon both exact and asymptotic inference based on ancillary statistics. My discussion includes Barndoff-Nielsen's p^* formula, the role of ancillary statistics in the elimination of nuisance parameters, or in finding optimal estimating functions. I have discussed also some approximate ancillary statistics. Finally, Bayesian ancillarity is also introduced.

Monitoring Parameter Change in the Presence of Nuisance Parameters

Edit Gombay, University of Alberta, Edmonton

Sequential change detection algorithms based on strong approximations will be discussed in general, then the focus will be on these methods adapted to the case of AR(p) time series model. The tests are using the efficient score vector, and the best possible rates for the approximations have been proven to be valid in this model.

A quick review of available methods is presented, and some empirical results on the comparison will be given. These show that our truncated sequential change detection methods are in some respect superior to the ones available.

(The talk is partially based on joint work with Daniel Serban.)

Comparisons of OLSE and BLUE in the General Gauss-Markov Model.

Jarkko Isotalo
University of Tampere, Finland

We consider the equality of the ordinary least squares estimator, OLSE, and the best linear unbiased estimator, BLUE, of the estimable parametric function in the general Gauss-Markov model. Some new conditions for equality of OLSE and BLUE are given based on reparametrization of the original model and the Frisch-Waugh-Lovell Theorem.

Furthermore, we consider the Watson efficiency of OLSE of the estimable parametric function. The generalization of the Watson efficiency to include singular model is given and interesting decomposition of the total efficiency is represented.

[Joint research with Ka Lok Chu (Dawson College, Montreal), Simo Puntanen (University of Tampere), George P. H. Styan (McGill University, Montreal)]

On the Problem of Finding Minimax Estimators

Janusz Kawczak
University of North Carolina at Charlotte

We are going to revisit a classical problem of finding minimax estimators for unknown parameters in commonly encountered statistical models by employing perturbation theory of linear operators. The formulation of optimization problem for a given model leads us to a specific type of spectral equation that must be solved for unspecified parameters. However, in most of the practical cases the solution set contains uncountably many choices for the minimax estimator. Whenever possible, we propose some necessary/ sufficient conditions to make the estimator unique.

Examples involving classical linear models, estimation of the states of dynamical systems and models with states linked by recursive relationship are used to demonstrate the developed methodology.

Conditioning of the entries in Google's PageRank vector.

Steve Kirkland

University of Regina

The internet search engine Google approaches the problem of ranking web pages by computing an estimate of the stationary distribution of a certain Markov chain associated with the world wide web. That stationary distribution, known as the PageRank vector, is then used to rank the importance of the corresponding web pages. In this talk, we discuss the conditioning of the entries in the PageRank vector, and our results in turn lead to error bounds for Google's computed estimate of that vector. Our techniques involve nonnegative matrices, generalized inverses, and combinatorial considerations.

Matrix tricks for linear statistical models: our personal Top Fourteen

Simo Puntanen
University of Tampere, Tampere, Finland

In teaching a course in linear statistical models to first year graduate students or to final year undergraduate students, say, there is no way to proceed smoothly without matrices and related concepts of linear algebra; their use is really essential. Our experience is that making some particular matrix tricks familiar to students can increase their insight into linear statistical models (and also multivariate statistical analysis). In matrix algebra, there are handy, sometimes even very simple “tricks” to simplify and clarify the problem treatment---both for the student and for the researcher. Of course, the concept of trick is not uniquely defined: by trick we simply mean here a central important handy result. In this report we collect together our Top Fourteen favourite matrix tricks for linear statistical models.

(Joint work with George P. H. Styan, McGill University, Montreal, Canada, and Jarkko Isotalo, University of Tampere, Finland.)

**The Curse of Dimensionality in Genomics:
Beyond the Euclidean Paradigm**

Pranab K. Sen
Departments of Biostatistics, and Statistics and Operations Research,
University of North Carolina at Chapel Hill

Genomic sequences involve an enormously large number of positions or sites (loci), reach one aving a purely qualitative categorical response (4 categories A, C, G, T for DNA nucleotides and some 20 amino acids in RNA codons). Thus, there is a very high dimensional categorical data model with the number of positions (K) often much larger than the number of sequences

(n), so that K > > n. Linear models and Euclidean space methodology may not appear to be of any significant impact in this setup. The prospect of discrete multivariate analysis incorporating the asymptotics for the likelihood function and all its ramifications (including pseudo-, quasi-, partial-, conditional-, profile-, and empirical likelihood) depends not only managing the complex categorical data models arising in this context but also on low sample size high- dimensional perspectives, and as of now, there has not been any great methodologic advances in this respect. Statistical learning or knowledge discovery and data mining tools are useful from computational point of view, but they may lack to a certain extent statistical methodologic support.

Although, it is often assumed that the positions have independent responses, in reality, inter-site stochastic dependence is very much perceptable. Hamming distance based methodology has therefore been advocated recently. They allow a dimension reduction taking into account the inter-site stochastic dependence to a certain extent. Pinheiro et al. (2000, 2005) have considered some MANOVA procedures based on variants of the Hamming distance. There has been some further research in this direction which incorporate second-order decomposability of Hoeffding's (1948) U-statistics, and yet achieving asymptotic normality of the pseudo-U-statistics arising in this context. Applications to real data have also been considered.

(Joint work with Hildete P. Pinheiro and Aluisio S. Pinheiro, University of Campinus, Sao Paulo, Brazil)

Some comments on the life and work of Jerzy K. Baksalary (1944--2005)

George P. H. Styan,
McGill University

Jerzy Baksalary was born in Pozna\'n, Poland, on 25 June 1944 and passed away in Pozna\'n on 8 March 2005. He was 60 years old. Although suffering, he remained active in his research work to the very end.

Jerzy Baksalary published extensively on matrix methods for statistics. He is the author or co-author of 170 published research publications in linear algebra and statistics, which may be classified as follows: 127 papers in peer-refereed research journals, with 45 papers in *Linear Algebra and its Applications* (LAA) and 24 in the *Journal of Statistical Planning and Inference* (JSPI), 27 solutions to problems (24 in *Image: The Bulletin of the International Linear Algebra Society*), and one each in
Econometric Theory, *The IMS Bulletin*, and *Statistica Neerlandica*), 12 chapters in research collections (conference proceedings, Festschriften, and other edited books), two journal special issues (one of LAA and one of JSPI) and two problems (both in *Image*).

In this talk, we build upon comments on the life and work of Jerzy K. Baksalary presented in the Special Memorial Session for Jerzy Baksalary organized by Oskar Maria Baksalary, Simo Puntanen, George P.H. Styan, and Gotz Trenkler at the 14th International Workshop on Matrices and Statistics (Auckland, New Zealand, 29 March--1 April 2005).

(Joint research with Oskar Maria Baksalary.)

Regularization Methods in Multivariate Analysis

Yoshio Takane
McGill University

We discuss simple regularization methods in two representative multivariate data analysis techniques, generalized (multiple-set) canonical correlation analysis (GCANO) and redundancy analysis (RA). CANO and RA subsume a number of existing multivariate data analysis techniques including multiple correspondence analysis (MCA), 2-set canonical correlation analysis (CANO), discriminant analysis (DA), and so on. We incorporate a ridge type of regularization in these techniques and demonstrate the situations in which it is most effective. Regularization is deemed important as a way of supplementing insufficient data by prior knowledge, and/or of incorporating certain desirable properties in the estimates of parameters. We also discuss some mathematical properties of a matrix operator involved in the ridge type of regularization methods and their implications.

(Joint work with Heungsun Hwang, HEC Montreal)

On the Golden Ratio, strong law, and first passage problem

Andrei Volodin
University of Regina

For a sequence of correlated square integrable random variables, conditions are provided for the strong law of large numbers to hold almost surely. The hypotheses stipulate that two series converge, the terms of which involve both the Golden Ratio and bounds on variation. An application to first passage times is provided.

Symmetric Kronecker Products and Uniqueness and Existence of search directions for semidefinite programming (SDP)

Henry Wolkowicz
University of Waterloo

Primal-dual interior-point (p-d i-p) methods for Semidefinite Programming (SDP) are generally based on solving a system of matrix equations for a Newton type search direction for a symmetrization of the optimality conditions. These search directions followed the derivation of similar p-d i-p methods for linear programming (LP). Among these, a computationally interesting search direction is the AHO direction. However, in contrast to the LP case, existence and uniqueness of the AHO search direction is not guaranteed under the standard nondegeneracy assumptions. Two different sufficient conditions are known that guarantee the existence and

uniqueness independent of the specific linear constraints. The first is given by Shida-Shindoh-Kojima and is based on
the semidefiniteness of the symmetrization of the matrix product SX at the current iterate. The second is a centrality condition given first by Monteiro-Zanjacomo and then improved by Monteiro-Todd.

In this talk, we revisit and strengthen both of the above mentioned sufficient conditions. We include characterizations for existence and uniqueness in the matrix equations arising from the linearization of the optimality conditions. As well, we present new results on the relationship between the Kronecker product and the symmetric Kronecker product that arise from these matrix equations. We conclude with a proof that the existence and uniqueness of the AHO direction is a generic property for every SDP problem and extend the results to the general Monteiro-Zhang family of search directions.

(Joint work with Levent Tuncel)

Unbiased Invariant Minimum Norm Estimation in Generalized Growth Curve Model

Xiaoyong Wu
University of Windsor

Abstract. This paper considers the generalized growth curve model $Y = \sum_{i=1}^{m} X_i B_i Z_i' + U\varepsilon$ subject to $R(X_m) \subseteq R(X_{m-1}) \subseteq \cdots \subseteq R(X_1)$, where ε splits into a number of independent and identically distributed subvectors with mean zero and unknown covariance matrix Σ. The matrices of unknown regression coefficients are B_i and the matrices X_i, Z_i and U are known, $i = 1, 2, ..., m$. An unbiased invariant minimum norm quadratic estimator of $tr(C\Sigma)$ is derived and the conditions for its optimality under the minimum variance criterion are investigated. The necessary and sufficient conditions for unbiased invariant minimum norm quadratic estimator of $tr(C\Sigma)$ to be its uniformly minimum variance invariant quadratic unbiased estimator are obtained. An unbiased invariant minimum norm quadratic plus linear estimator of $tr(C\Sigma) + \sum_{i=1}^{m} tr(D_i' B_i)$ is also given.

(Joint work with Guohua Zou, Chinese Academy of Sciences, Wenli Yang, Beijing Normal University, S. Ejaz Ahmed, University of Windsor)