Matrix Theory and Statistics
Théorie matricielle et les statistiques
(Org: **Ejaz Ahmed** and/et **Abdul Hussein** (Windsor))

S. EJAZ AHMED, U. of Windsor, Windsor, ON, Canada
Biased Estimation in Generalized Linear Models

In this talk, we consider the estimation problem for the parameters of generalized linear models which may have a large collection of potential predictor variables and some of them may not have influence on the response of interest. In this situation, selecting the statistical model is always a challenging problem. In the context of two competing models, we demonstrate the relative performances of shrinkage and classical estimators based on the asymptotic analysis of quadratic risk functions. We demonstrate that the shrinkage estimator outperforms the maximum likelihood estimator uniformly. For comparison purpose, we also consider an absolute penalty estimation (APE) approach. This comparison shows that shrinkage method performs better than the APE type method when the dimension of the restricted parameter space is large. This talk ends with real-life example showing the value of the suggested method in practice. We consider South African heart disease data, which was collected on males in a heart disease high-risk region of Western Cape, South Africa.

MOHAMED AMEZZIANE, DePaul University, Chicago, USA
Kernel Estimation and Bandwidth Matrix Selection for Multivariate Distribution Functions

Results from matrix algebra and analysis are used to minimize local and global L_2 criteria of kernel estimates of multivariate distribution functions. Under different matrix structures, optimal bandwidth matrices are derived and their selectors are shown to possess fast rates of convergence.

SHABNAM CHITSAZ, University of Windsor, 401 Sunset Avenue, Windsor, Ontario N9B 3P4
Improving the Performance of the Mean Response in Multivariate Multiple Regression Model With application to a Financial Model

Let $\mathbf{y_t}$ be the $k \times 1$ vector of excess return on k assets and let x_t be the excess return on the market portfolio at time t. The capital asset pricing model (CPAM) can be associated with the null hypothesis $H_0\colon \alpha = \mathbf{0}$ in the regression model: $\mathbf{y_t} = \alpha + x_t\beta + \epsilon_t, 1 \leq t \leq n$.

In this talk, we are interested in improving the estimation of mean response parameter by incorporating the available information about the intercept parameter vector. In this context, we suggest preliminary test and shrinkage estimation for the parameter matrix. We investigate the asymptotic relative performance of suggested estimators with the classical estimator. The detailed simulation study illustrates the feasibility of the estimation approach and evaluates their characteristics. The proposed estimation strategies are also applied to stock data for illustrative purpose.

KJELL DOKSUM, Univ. of Wisconsin, Dept. of Statistics, 1300 Univ. Ave, Madison, WI
Genomics, Matrices, and Statistical Inference

Genomics research often involves a 0-1 response such as "control" and "case" (nondisease and disease) and a number of covariates such as genetic markers (up to 1.2 million at this date), population membership, conventional covariates such as age, sex, body mass index (BMI), and treatment levels. Population geneticists have developed clever procedures for analysing data from experiments involving such genomics data. These procedures involve using what could be called inverse principal component analysis (PCA) and rely on matrix equations that connect regular and inverse PCA, where inverse PCA refers to PCA with variables exchanged with sample values. That is, inverse PCA is PCA with the design matrix transposed. We examine the properties of the procedures proposed based on this inverse PCA. The population geneticists propose intuitive models for

their data. However, in their analysis they ignore their own models and use simple Stat 101 nonparametric chi-square tests. We compare these tests with likelihood ratio tests and Bayesian procedures. More generally, we examine statistical inference appropriate for the models proposed by population geneticists.

This is joint work with Kam Tsui and Fan Yang.

NOUREDDINE EL KAROUI, Department of Statistics, UC Berkeley, 367 Evans Hall, Berkeley, CA 94720-3860, USA
On the spectral properties of large-dimensional kernel random matrices

Prompted by the recent explosion of the size of datasets statisticians are working with, there is currently renewed interest in the statistics literature for questions concerning the spectral properties of large-dimensional random matrices.

Most of the efforts so far has focused on understanding sample covariance matrices in the "large n, large p" setting, where the number of samples, n, and the number of variables in the problem, p, are of the same magnitude—say a few 10s or 100s. A basic message is that in this high-dimensional setting, the sample covariance matrix is a very poor estimator of the population covariance, especially from a spectral point of view. This naturally raises questions about the behavior of spectral methods of multivariate analysis such as the widely used principal component analysis (PCA).

The statistical learning literature has a number of "kernel analogs" to classical multivariate techniques, in which the sample covariance matrix is replaced by a kernel matrix, often times of the form $f(X_i' X_j)$ or $f(\|X_i - X_j\|)$. It is therefore natural to ask what can be said about the spectral properties of these kernel random matrices, and in particular, what is the impact of the choice of the function f and related questions.

In this talk I will discuss these questions when:

(a) the data is assumed to be genuinely high-dimensional, and

(b) when it is a noisy version of data sampled from a low-dimensional structure and the noise is high-dimensional.

In both cases, we will see that standard heuristics can lead us astray and that careful analysis yields perhaps surprising results.

SABER FALLAHPOUR, University of Windsor
Improved estimation in partially linear models with nonlinear time series errors

This paper is concerned with a partially linear regression model with serially correlated random errors which are unobservable and modelled by a random coefficient autoregressive process. We propose improved weighted semi-parametric least-square estimators for the parametric component using preliminary test and James–Stein estimation methodologies. The asymptotic properties (including asymptotic distribution bias and asymptotic distribution risk) of the proposed estimators is investigated. We show that these improved estimators perform well relative to the benchmark semi-parametric weighted least square estimators. Our simulation results strongly corroborate with our analytical results.

SHAUN FALLAT, University of Regina, Mathematics and Statistics, Regina, SK
Oppenheim's Determinantal Inequality for Certain P-matrices

In 1930, A. Oppenheim generalized Hadamard's determinantal inequality

$$\det A \leq a_{11} a_{22} \cdots a_{nn},$$

for positive semidefinite matrices A, by incorporating Hadamard (or entry-wise) products. This now famous inequality can be written as

$$\det(A \circ B) \geq (b_{11} b_{22} \cdots b_{nn}) \det A,$$

and can be used to verify that, for positive semidefinite matrices, the Hadamard product dominants the conventional product in determinant.

After reviewing a modern version of a proof of Oppenheim's inequality, we offer an extension of the class of matrices (beyond positive semidefinite) for which Oppenheim's inequality remains valid. Along the way, we define a new notion of closure under Hadamard multiplication (called duality) and concentrate on a specific kind of perturbation (called retraction).

DON FRASER, University of Toronto, Toronto, Canada
Curvature for the Bayes-frequentist disconnect

Bayesian and frequentist methods can give quite different results when applied to the same model-data information. If the parameter of interest is linear in the invariant parameterization recommended by Bayes then the Bayesian and the frequentist procedures will fully agree. But if the parameter of interest is curved then the Bayesian divergence appears and the shift is in opposite directions from the frequentist depending on the sign positive or negative of the curvature. This requires the calculation of the obvious curvature and then the recalibration to the Bayesian type curvature which is different from the Efron curvature. We give an overview and examples.

MARVIN H. GRUBER, Rochester Institute of Technology, School of Mathematical Sciences, 85 Lomb Memorial Drive, Rochester, NY 14623
A Matrix Identity and Ridge Type Estimators from Different Points of View

The matrix identity
$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$
enables the study of ridge type estimators for parameters in linear regression models from both the Bayesian and the frequentist point of view.

Ridge regression estimators can be shown to be special cases of mixed and minimax estimators by solving an optimization problem with constraints. They may also be formulated by minimizing the sum of the squares of the regression coefficients on an ellipsoid. These are frequentist arguments.

On the other hand for normal prior distributions and normal errors in the regression model they are special cases of the Bayes estimator.

By an appropriate choice of A, B, C and D in the above identity the estimators obtained by the frequentist and Bayesian argument can be shown to be algebraically equivalent. In this talk I will explain how to do this and give conditions for when this algebraic equivalence is possible.

FRANK HALL, Georgia State University, Atlanta, GA 30303, USA
Some Eigenvalue Interlacing Results on Matrices Associated with Graphs

Given a graph G, the adjacency matrix, the standard Laplacian, and the normalized Laplacian have studied extensively. In this talk, eigenvalue interlacing inequalities are given for each of these three matrices under the two operations of removing an edge or a vertex from G. Examples are provided to show that the inequalities are the best possible of their type.

CHARLES R. JOHNSON, College of William and Mary, Williamsburg, VA, USA
Matrix Equilibrants and Diagonal Scaling

For semipositive A, the $\inf$ and $\sup$ of the entry product of Ax over normalized positive vectors x is discussed and a theory developed that to facilitate computation. This the basis for a theory of diagonal scaling based upon a curious matrix/vector equation. Semipositive spectra are also discussed.

DELI LI, Lakehead University, 955 Oliver Road, Thunder Bay, ON, P7B 5E1
Necessary and suffcient conditions for the asymptotic distribution of the largest entry of a sample correlation matrix

Let $\{X_{k,i} \; ; \; i \geq 1, k \geq 1\}$ be a double array of nondegenerate i.i.d. random variables and let $\{p_n \; ; \; n \geq 1\}$ be a sequence of positive integers such that n/p_n is bounded away from 0 and ∞. In this paper we give the necessary and sufficient conditions for the asymptotic distribution of the largest entry $L_n = \max_{1 \leq i < j \leq p_n} |\hat{\rho}_{i,j}^{(n)}|$ of the sample correlation matrix $\boldsymbol{\Gamma}_n = (\hat{\rho}_{i,j}^{(n)})_{1 \leq i,j \leq p_n}$ where $\hat{\rho}_{i,j}^{(n)}$ denotes the Pearson correlation coefficient between $(X_{1,i}, \ldots, X_{n,i})'$ and $(X_{1,j}, \ldots, X_{n,j})'$. Write $F(x) = \mathbb{P}(|X_{1,1}| \leq x)$, $x \geq 0$, $W_{c,n} = \max_{1 \leq i < j \leq p_n} |\sum_{k=1}^{n} (X_{k,i} - c)(X_{k,j} - c)|$, and $W_n = W_{0,n}$, $n \geq 1$, $c \in (-\infty, \infty)$. Under the assumption that $\mathbb{E}|X_{1,1}|^{2+\delta} < \infty$ for some $\delta > 0$, we show that the following six statements are equivalent:

(i) $\lim_{n \to \infty} n^2 \int_{(n \log n)^{1/4}}^{\infty} \left(F^{n-1}(x) - F^{n-1}(\frac{\sqrt{n \log n}}{x}) \right) dF(x) = 0$,

(ii) $n\mathbb{P}(\max_{1 \leq i < j \leq n} |X_i X_j| \geq \sqrt{n \log n}) \to 0$ as $n \to \infty$,

(iii) $\frac{W_{\mu,n}}{\sqrt{n \log n}} \xrightarrow{\mathbb{P}} 2\sigma^2$,

(iv) $\left(\frac{n}{\log n}\right)^{1/2} L_n \xrightarrow{\mathbb{P}} 2$,

(v) $\lim_{n \to \infty} \mathbb{P}(\frac{W_{\mu,n}^2}{n\sigma^4} - a_n \leq t) = \exp\{-\frac{1}{\sqrt{8\pi}} e^{-t/2}\}$, $-\infty < t < \infty$,

(vi) $\lim_{n \to \infty} \mathbb{P}(nL_n^2 - a_n \leq t) = \exp\{-\frac{1}{\sqrt{8\pi}} e^{-t/2}\}$, $-\infty < t < \infty$,

where $\mu = \mathbb{E}X_{1,1}$, $\sigma^2 = \mathbb{E}(X_{1,1} - \mu)^2$, and $a_n = 4 \log p_n - \log \log p_n$. The equivalences between (i), (ii), (iii), and (v) assume that only $\mathbb{E}X_{1,1}^2 < \infty$. Weak laws of large numbers for W_n and L_n, $n \geq 1$, are also established and these are of the form $W_n/n^\alpha \xrightarrow{\mathbb{P}} 0$ $(\alpha > 1/2)$ and $n^{1-\alpha} L_n \xrightarrow{\mathbb{P}} 0$ $(1/2 < \alpha \leq 1)$, respectively. The current work thus provides weak limit analogues of the strong limit theorems of Li and Rosalsky as well as a necessary and sufficient condition for the asymptotic distribution of L_n obtained by Jiang. Some open problems are also posed.

This talk is based on my recent work joint with Professors Wei-Dong Liu and Andrew Rosalsky.

YI LI, Harvard University
The Dantzig Selector for Censored Linear Regression Models: Identifying Predictive Genes for Myeloma Disease Progression

The Dantzig variable selector has recently emerged as a powerful tool for fitting regularized regression models. A key advantage is that it does not pertain to a particular likelihood or objective function, as opposed to the existing penalized likelihood methods, and hence has the potential for wide applications. To our knowledge, almost all the Dantzig selector work has been performed with fully observed response variables. This talk introduces a new class of adaptive Dantzig variable selectors for linear regression models when the response variable is subject to right censoring. This is motivated by a clinical study of detecting predictive genes for myeloma patients' event-free survival, which is subject to right censoring. We establish the theoretical properties of our procedures, including consistency in model selection (i.e., the right subset model will be identified with a probability tending to 1) and the oracle property of the estimation (i.e., the asymptotic distribution of the estimates is the same as that when the true subset model is known a priori). The practical utility of the proposed adaptive Dantzig selectors is verified via extensive simulations. We apply the new method to the aforementioned myeloma clinical trial and identify important predictive genes for patients' event-free survival.

PETER LOLY, University of Manitoba, Winnipeg, Manitoba R3T 2N2
Eigenvalues of an Algebraic Family of Compound Magic Squares of Order $n = 3^l$, $l = 1, 2, \ldots$, and Construction and Enumeration of their Fundamental Numerical Forms.

Compound magic squares (CMSs) of order mn, whose tiled subsquares of orders m and n are also magic squares (MSs having constant row, column and diagonal linesums within each subsquare), are found back to the 10th century for the case $m = n = 3$. Interesting results follow if they are considered as matrices.

Frierson gave a simple algebraic form for compounding from the unique pattern of third order to a general $n = 9$ CMS in *The Monist* in 1907, from which he showed 6 fundamental numerical forms using the complete set of integers $1 \cdots 81$. We extend Frierson's work, finding an algebraic description of a family of associative (antipodal sum pairs $n^2 + 1$) compound magic squares of orders $n = 3^l$, $l = 1, 2, \ldots$. In doing so we have firmly established two results previously stated by Bellew (1997), 90 fundamental numerical forms for $n = 27$, as well as its generalization for all l.

The present algebra then leads to a general formula for the eigenvalues of this family, which consists of the linesum eigenvalue and l signed pairs, for rank $2l + 1$.

For $n = 9$ the 8 possible orientations of each of the 9 tiled third order subsquares give rise to 6×8^9 distinct CMSs, most with increased rank. We resolve disparate factors of 8 of Trigg (1980) and Bellew for $n = 27$ with a new result by taking account of all orders of tiled subsquares, before generalizing this for all l.

Joint work with Ian Cameron.

ABDEL-RAZZAQ MUGDADI, Southern Illinois University Carbondale
The mise and the Hellinger distance of the kernel distribution estimator of functions of observations

Let $X_1, \ldots, X_m$ be identically and independently distributed (i.i.d.) having unknown distribution function F. In this investigation, we propose a kernel method to estimate the distribution function of the function $g(X_1, \ldots, X_m)$. We derived the asymptotic mean square error for the estimator and the asymptotic mean Hellinger distance for the estimator. In addition, we propose a data-based method to obtain the bandwidth based on both the mean square error and the mean Hellinger distance.

PEHUA QIU, University of Minnesota, 224 Church St. SE, Minneapolis, MN 55455
Jump Regression Analysis and Image Processing

Nonparametric regression analysis provides a statistical tool for estimating regression curves or surfaces from noisy data. Conventional nonparametric regression methods, however, are only appropriate for estimating continuous regression functions. When a underlying regression function has jumps, functions estimated by the conventional methods are not statistically consistent at the jump positions. Recently, jump regression analysis (JRA) for estimating jump regression functions is under rapid development, because JRA has broad applications. One important application is image processing where a monochrome image can be regarded as a surface of the image intensity function which has jumps at the outlines of objects. In this talk, I will make a general introduction to the research area of JRA, and to some of its applications in image processing.

HARVEY QU, Oakland University
Optimal Designs in High-Throughput Screening

High-throughput screening (HTS) is a large-scale process that screens hundreds of thousands to millions of compounds in order to identify potentially leading candidates rapidly and accurately. There are many statistically challenging issues in HTS. In this talk, I will focus the spatial effect in primary HTS. I will discuss the consequences of spatial effects in selecting leading compounds and why the current experimental design fails to eliminate these spatial effects. A new class of designs will be proposed for elimination of spatial effects.

The new designs have the advantages such as all compounds are comparable within each microplate in spite of the existence of spatial effects; the maximum number of compounds in each microplate is attained, etc. Optimal designs are recommended for HTS experiments with one and multiple controls.

TANVIR QUADIR, Windsor

ENAYETUR RAHEEM, University of Windsor
Shrinkage Versus Lasso in Partially Linear Models

In the context of a partially linear regression model, we consider shrinkage semi-parametric estimators based on the Stein-rule. In our framework the coefficient vector may be partitioned into two sub-vectors, where the first sub-vector gives the coefficients of interest, i.e., main effects (for example treatment effects), and the second sub-vector is for variables that may or may not need to be controlled for. When estimating the first sub-vector, we may get the best estimate using the full model that includes both sub-vectors, or using the reduced model which leaves out the second sub-vector. We demonstrate that under certain conditions shrinkage estimators which combines two semi-parametric estimators computed for the full model and the reduced model outperforms the semi-parametric estimate for the full mode. Using the semi-parametric estimate for the reduced model is best when the second sub-vector is the null vector, but this estimator suffers seriously from bias otherwise. The relative dominance picture of suggested estimators is investigated. In each case we consider estimates of the nonparametric component based on B-splines. We primarily explore the suitability of estimating the nonparametric component based on B-spline, and compare the risk performance numerically with that of the kernel estimates. Further, the performance of the proposed estimators have been compared with an absolute penalty estimator through Monte Carlo simulation.

PETER XK SONG, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029, USA
Improving Quadratic Inference Functions via Covariance Matrix Shrinkage

In this talk, I will present a new construction of quadratic inference functions (QIF) that have received increasing attention in longitudinal data analysis. This new construction is based on a covariance matrix shrinkage, which ensures the QIF to be analytically valid and computationally stable when data are sparse or interaction terms are included in the marginal generalized linear models. Some numeric illustration will be given to demonstrate the proposed method.

GEORGE STYAN, McGill University, Montreal
Philatelic Sudoku Puzzles

We consider sheetlets of postage stamps with r rows and c columns featuring s distinct stamps (we do not require that rc/s be an integer) and where no particular stamp appears more than once in any single row or column and so the sheetlet defines a "Latin rectangle". The "philatelic Sudoku puzzle" is to find an $s \times s$ Latin square in which the Latin rectangle defining the sheetlet is a subregion and some blocking within the subregion is involved as with the popular "regular" 9×9 Sudoku puzzle. We let b denote the block size and so $b = 9$ in regular Sudoku. We identify six philatelic Sudoku puzzles with parameter sets $(r, c, s \,;\, b)$ as follows:

- Abkhazia 2006, marine life, $(8, 3, 12 \,;\, 4)$,
- Hong Kong 2006, musicians, $(6, 3, 6 \,;\, 3)$,
- Pakistan 2005, mushrooms, $(6, 5, 10 \,;\, 5)$,
- USA 1997, musicians, $(5, 4, 8 \,;\, 4)$,
- USA 2005, aircraft, $(5, 4, 10 \,;\, 10)$,
- USA 2007, flowers, $(2, 10, 10 \,;\, 2)$.

For each puzzle we present the solution and some interesting properties of the associated matrices.

This talk is based on Section 6 of the invited paper (with Ka Lok Chu & Simo Puntanen) entitled "Some comments on philatelic Latin squares from Pakistan", to be published in the Special Jubilee Issue of the *Pakistan Journal of Statistics*.

JI ZHU, University of Michigan
Partial Correlation Estimation by Joint Sparse Regression Models

In this talk, we propose a computationally efficient approach for selecting non-zero partial correlations under the high-dimension/low-sample-size setting. This method assumes the overall sparsity of the partial correlation matrix and employs

sparse regression techniques for model fitting. We illustrate the performance of our method by extensive simulation studies. It is shown that our method performs well in both non-zero partial correlation selection and the identification of hub variables, and also outperforms two existing methods. We then apply our method to a microarray breast cancer data set and identify a set of "hub genes" which may provide important insights on genetic regulatory networks. Finally, we prove that, under a set of suitable assumptions, the proposed procedure is asymptotically consistent in terms of model selection and parameter estimation.

This is joint work with Jie Peng, Pei Wang and Nengfeng Zhou.