Program and Abstracts

14th International Workshop on Matrices and Statistics

29th March to 1st April 2005

Massey University, Albany Campus
Auckland, New Zealand

14th International Workshop on Matrices and Statistics

Massey University,
Auckland, New Zealand

29th March to 1st April 2005

Program and Abstracts

Edited by Paul Cowpertwait and Jeff Hunter

This workshop has been endorsed by the following Societies:

The International Statistical Institute

IWMS 2005 is a Satellite Conference to the ISI 55th Biennial Session, Sydney April 5 – 12, 2005.

The International Linear Algebra Society

Sponsors

The International Organizing Committee and the Local Organizing Committee wish to acknowledge the support provided by the following sponsors:

Royal Society of New Zealand
International Conference Fund provided by the NZ Government through the Ministry of Research, Science and Technology ISAT Linkages Fund)

New Zealand Institute of Mathematics and its Applications

Statistics New Zealand

SAS
Supporting the SAS Lecturer, Prof S Searle, and the SAS Reception

Nokia
Supporting the Workshop and the Nokia Lecturer, C R Rao

Massey University
Supporting C R Rao as a Distinguished Visitor to both the Palmerston North and Albany Campuses

The New Zealand Statistical Association
Supporting the NZSA Visiting Lecturer for 2005, C R Rao with travel in New Zealand to Otago University, University of Canterbury, Victoria University of Wellington, Massey University at Palmerston North, University of Waikato , University of Auckland, and Massey University at Auckland

The New Zealand Mathematical Society
Supporting the NZMS Lecturer, E Seneta

International Organising Committee

Prof George P.H. Styan (Chair), McGill University, Montréal, Canada
Dr Simo Puntanen, University of Tampere, Finland
Prof Dr Hans Joachim Werner

Local Committee

Institute of Information and Mathematical Sciences, Massey University
Prof Jeffrey J Hunter – Chair
Freda Anderson – Secretary
Paul Cowpertwait - Abstracts and Editor
Claire Jordan - Accommodation, Conference Dinner, Catering
Marie Fitch – Catering
Nikki Luke - Local arrangements
Barry McDonald – Transport
Beatrix Jones and Danny Walsh – Excursion
Stephen Ford – Webmaster

Supporting ex-officio members

Professor Stephen Haslett, Massey University - Representing the Statistics Research Consulting Centre, Massey University, Palmerston North
Dr Harold Henderson, AgResearch, Ruakura - Representing the New Zealand Statistical Association
Professor Robert McKibbin - Representing the Institute of Information and Mathematical Sciences, Massey University, Albany Campus, Auckland
Mr Eddie Davis - Representing the Royal Society of New Zealand International Conference Fund with funding provided by the Government through the Ministry of Research, Science and Technology ISAT Linkages Fund

Workshop Booklet

The purpose of the conference is to stimulate research and, in an informal setting, to foster the interaction of researchers in the interface between statistics (including applied probability) and matrix theory. The Workshop will provide a forum through which statisticians may be better informed of the latest developments and newest techniques in linear algebra and matrix theory and may exchange ideas with researchers from a wide variety of countries.

International Workshops in Matrices and Statistics are held annually in different locations around the world. This Workshop is the fourteenth in the series with the Workshops having been held at the following locations:

1. Tampere, Finland, August 1990
2. Auckland, New Zealand, December 1992
3. Tartu, Estonia, May 1994
4. Montréal, Québec, Canada, July 1995
5. Shrewsbury, England, July 1996
6. Istanbul, Turkey, August 1997
7. Fort Lauderdale, Florida, USA, December 1998
8. Tampere, Finland, August 1999
9. Hyderabad, India, December 2000
10 Voorburg, The Netherlands, August 2001
11. Lyngby, Denmark, August 2002
12. Dortmund, Germany, August 2003
13. Bedlewo, Poland, August 2004

March 2005 marks the 200th anniversary of Legendre's appendix on the method of least squares

Contents

Conference Program

Tuesday March 29

8.00 – 9.20	**Registration** – Foyer of Quad B Building, Massey University, Albany Campus Auckland
9.20 - 9.30	**Maori Powhiri**
	The powhiri is the Maori welcome ceremony. It removes the tapu of the Manuhiri (visitors) to make them one with the Tangata Whenua (Home people) and is a gradual process of the Manuhiri and the Tangata Whenua coming together.
	International and New Zealand visitors to assemble in the Quadrangle between Quad A and Quad B Buildings. Massey University participants to assemble in Quad B 5.
9.30	**Response** on behalf of the Tangata Whenua - Jeff Hunter
	Waiata
	Response on behalf of the Manuhiri - George Styan
9.35	**Opening Remarks** by Jeff Hunter, Chair of Local Organising Committee
9.40	**Welcome to Massey University** by Prof John Raine, Deputy Vice Chancellor, Massey University, Auckland

Session 1:	Keynote Lecture 1
09.45	**Welcome to the Massey University Distinguished Lecturer** by Robert Anderson, Pro-Vice Chancellor, College of Sciences, Massey University
	Introduction of the Nokia Lecturer - George Styan, Chair of the International Organizing Committee
09.50 – 10.50	**Nokia Lecture – Eberly Professor Emeritus C R Rao**, Pennsylvania State University, USA *Statistical Proofs of Matrix Theorems*
11.00 – 11.30	**Morning Tea** - Foyer to Quad B

Session 2: Invited Lectures
Chair: Alan Lee

11.30 – 12.00	**Tõnu Kollo**, University of Tartu, Estonia and D. von Rosen, Swedish University of Agricultural Sciences, Sweden *Approximation of the Parameter Distributions of Growth Curve Model*
12.00 – 12.30	**Nye John**, Waikato University, Hamilton, New Zealand *Inverse of the Information Matrix*
12.30 – 13.00	**Alexander Kukush**, Kiev National Taras Shevchenko University, Ukraine *Invariant estimator in a quadratic measurement error model*
13.05 - 14.00	**Lunch** - Staff Lounge, Study Centre

Session 3: Invited Lectures
Chair: Nye John

14.00 – 14.30	**Götz Trenkler**, Dortmund University, Germany *On the commutativity of orthogonal projectors*
14.30 – 15.00	**Stephen Haslett**, Massey University, New Zealand and John Haslett, Trinity College Dublin *What are the residuals for the linear model?*
15.00 – 15.30	**Joachim Werner**, Univ of Bonn, Germany and Ingram Olkin, Stanford University *On permutations of matrix products*
15.30 – 16.00	**Afternoon Tea**

Session 4: Contributed Lectures
Chair: Alexander Kukush

16.00 – 16.20	**B. Jones**, Massey University, Auckland and M. West, Duke University, N.C., U.S.A. *Covariance decomposition for Gaussian graphical models*
16.20 – 16.40	**Burkhard Schaffrin**, Ohio State University, Columbus, Ohio, USA *On the optimal choice of the regularization parameter through variance ratio estimation*
16.40 – 17.00	**Song-Gui Wang** and **Zhong-Zhen Jia**, Beijing University of Technology, Beijing, China *Estimating the covariance matrix by spectral decomposition approach in linear mixed model*
17.00 – 17.20	**Mike Doherty**, Statistics New Zealand, Wellington. *Partially diffuse starting values in state-space models*
17.30 – 20.30	**"SAS Reception"** - Opening Reception held in the Staff Lounge, Study Centre

Wednesday March 30

	Session 5: Keynote Lecture 2
09.00	**Introduction to the SAS Lecturer – Joachim Werner**, Vice Chair of the International Organizing Committee
09.05 – 10.05	**SAS Lecture – Professor Emeritus Shayle Searle**, Cornell University, Ithaca, New York, U.S. *Reflections on a fifty year random walk midst matrices and statistics*
10.05	**Workshop Photograph**
10.20 – 10.40	**Morning Tea** - Foyer to Quad B
	Session 6A: Invited Lectures Chair: Stephen Haslett
10.40 – 11.10	**C R Rao**, Pennsylvania State University, U.S *Anti eigen and singular values of a matrix and applications to problems in Statistics*
11.10 – 11.40	**Karl Gustafson**, University of Colorado, Boulder, Colorado, U.S. *The geometry of statistical efficiency*
11.40 – 12.10	**Garry Tee**, University of Auckland, New Zealand *Eigenvectors of block circulant matrices*
Session 6B:	Contributed Lectures
12.10 – 12.30	**Imbi Traat** and **Tatjana Nahtman** University of Tartu, Estonia *A matrix with consecutive integer eigenvalues*
12.30 – 12.50	**Kimmo Vehkalahti**, University of Helsinki, Finland *Leaving useful traces when working with matrices*
	Excursion
13.00 –	Boxed Lunch Pickup. Participants to assemble outside the Atrium Building for the Bus to the Ferry.
13.15 –	Bus departs for Excursion to Waiheke
14.00 –	Ferry leaves for Waiheke.
	The excursion will explore Waiheke Island. Participants will get to choose among several attractions, including wineries, beaches, a boutique olive oil production facility, and the village shops and galleries.

Thursday March 31

	Session 7: Keynote Lecture 3
09.00	**Introduction to the New Zealand Mathematical Society Lecturer - Associate Professor Mick Roberts**, President of the NZMS
09.05 – 10.05	**NZMS Lecturer – Professor Eugene Seneta**, University of Sydney, Australia. *Coefficients of Ergodicity in a Matrix Setting*
10.05 – 10.30	**Morning Tea** - Foyer to Quad B
	Session 8A: Invited Lectures Chair: Tonu Kollu
10.30 – 11.00	**Estate Khmaldze**, Victoria University of Wellington, New Zealand *Inverse matrices, Volterra operators and innovation processes: application to statistics*
11.00 – 11.30	**Anyue Chen**, University of Greenwich, U.K. *Asymptotic Birth-Death Processes: A Matrix Analysis Approach*
11.30 – 12.00	**Moshe Haviv**, The Hebrew University of Jerusalem, Israel *On singularly perturbed Markov chains*
Session 8B:	Contributed Lectures
12.00 – 12.20	**D. Alexander** and **G. Jones**, Massey University, Palmerston North, N.Z. *Convergence Properties of Alternating Markov Chains*
12.20 – 12.40	**Jeff Hunter**, Massey University, Auckland, N.Z. *Simple procedures for finding mean first passage times in Markov chains*
12.40 – 13.00	**G. Jones**, Massey University, Palmerston North, N.Z. *Properties of transition matrices for chain binomial models*
13.05 – 14.00	**Lunch** - Staff Lounge, Study Centre
	Session 9: Jerzy K. Baksalary Memorial Session Organised by: Oskar Maria Baksalary, George Styan, Simo Puntanen, and Götz Trenkler Chair: George Styan
14.00 – 15.30	**George Styan,** Mc Gill University, Montreal Canada; **Simo Puntanen**, Univ Tampere, Finland and **Götz Trenkler**, Dortmund University, Germany *Comments on the life and work of Jerzy K. Baksalary* (1944 - 2005)
15.30 – 16.00	**Afternoon Tea**

Session 10: Contributed Lectures*
Chair: Moshe Haviv

16.00 – 16.20	**W. Sakamoto,** Osaka University, Japan *Diagnosing non-linear regression structure with power additive smoothing splines*
16.20 – 16.40	**Eric Iksoon Im,** University of Hawaii at Hilo, USA *Hessian Equivalence to Bordered Hessian*
16.40 – 17.00	**C. M. Cuadras,** University of Barcelona, Spain. *Continuous canonical correlation analysis*
17.00 – 17.20	**Lakshmi Narasimhaiah,** Adhiyamaan College of Engineering, Tamil Nadu, India; Kishore Hoysal, Islamiah Institute of Engineering, Bangalore, India. *Model for students expected performance level through varying control limits in relation to Power of Valuation*

19.00 – 24.00	**Workshop Dinner**, Spencer on Byron
19.00	Pre dinner drinks
19.30	Conch call to guests Hato Petera Maori Cultural Group
20.00	Dinner - New Zealand Theme

Friday April 1

Session 11: Keynote Lecture 4

09.00	**Introduction to the Keynote Lecturer - Simo Puntanen,** Member of International Organizing Committee
09.05 – 10.05	**Keynote Lecture – Emeritus Professor George Seber**, University of Auckland, New Zealand. *Things my mother never told me about Matrices*
10.05 – 10.30	**Morning Tea** -foyer to Quad B

Session 12: Invited Lectures
Chair: Simo Puntanen

10.30 – 11.00	**S Ejaz Ahmed**, University of Windsor, Canada. *Approximation Assisted Estimation of Eigen Vectors Under Quadratic Loss*
11.00 – 11.30	**Alan Lee** and **Alastair Scott**, University of Auckland, New Zealand *Semi-parametric Efficiency, Projection and the Scott-Wild Estimator*

---

* Paper presented by title: Karuthan Chinna, University of Technology MARA, Malaysia (with Parthasarathy Balachandar, Multimedia University, Cyberjaya, Malaysia). *Modelling Multivariate Meta-Analysis using Bootstrap Resampling Techniques.*

Approximation assisted estimation of eigen vectors under quadratic loss

S. Ejaz Ahmed

University of Windsor

The estimation of eigen vectors of covariance matrix is considered in the presence of prior information regarding the parameter vector of interest. This information in practice will be available in any realistic problem. Like statistical models underlying the statistical inferences to be made, the prior information will be susceptible to uncertainty and the practitioners may be reluctant to impose the additional information in the estimation process. However, a very large gain in precision may be achieved by judiciously exploiting the information. In this talk, I will discuss how to combine sample and non-sample information. The estimators based on the shrinkage and other rules are proposed. The expressions for the risks under quadratic loss of the estimators are derived and compared with the benchmark estimator. For illustration purposes, the method is applied to real datasets.

Convergence properties of alternating Markov chains

D. Alexander and G. Jones

Institute of Information Sciences and Technology,
Massey University, New Zealand

Suppose we have two Markov chains defined on the same state space. What happens if we alternate them? If they both converge to the same stationary distribution, will the chain obtained by alternating them also converge? Consideration of these questions is motivated by the possible use of two different updating schemes for MCMC estimation, when much faster convergence can be achieved by alternating both schemes than by using either singly (Jones, 2004).

References

Jones G. (2004). Markov chain Monte Carlo Estimation for the Two-Component Model, *Technometrics*, **46**, pp99-107.

Asymptotic birth-death processes: A matrix analysis approach

Anyue Chen

University of Greenwich

A denumerable Markov Process with the state space Z_+ is called an asymptotic birth-death process if there is a finite subset G of Z_+ for which the restriction of the infinitesimal generator of the process, the so called q-matrix Q, to $Z_+ \backslash G$ is a birth-death q-matrix. Matrix analysis approach is applied to such structure. Using this approach, the close relationship between asymptotic birth-death processes and the well-developed birth-death processes is revealed. Regularity criteria for such process are established. Properties of such structure are investigated. In particular, conditions for recurrence and positive recurrence are obtained. Equilibrium distributions are given for the ergodic case. Some useful information regarding the transient asymptotic birth-death processes is also provided. The probabilistic interpretations of the results are explained.

Modeling multivariate meta-analysis using bootstrap resampling techniques

Karuthan Chinna[1], Parthasarathy[2], Balachandar[3]

[1]*Faculty of Information Technology and Quantitative Sciences,*
University Technology MARA, Malaysia (karuthanchinna@hotmail.com)
[2,3] *Multimedia University, Cyber Jaya, Malaysia.*

Meta-analysis is a powerful analytical tool in research synthesis for effective evidence based decision making. The end points of several studies are pooled together in a systematic way to derive an overall statistic. The extended multivariate meta-analysis involves the synthesis of studies that may report several related variables. The outcomes, the predictors or both the outcomes and the predictors may be multivariate in nature. In such situations, the variance-covariance matrix structure needs to be taken into account for effective pooling of the data. The bootstrap techniques are distribution free, computer intensive and they work very well with minimal assumptions. In bootstrap approach, the studies are first replicated using appropriate weights. This collection of replicates is considered as the bootstrap population. From this population, repeated samples are taken and for every sample the statistics of interest are computed. This innovative meta-analytical method not only gives better insight to the area of research but also simplifies the mathematical complexity. In this paper we discuss the advantages of bootstrap resampling techniques in multivariate meta-analysis.

Decomposing the Watson efficiency in partitioned linear models

Ka Lok Chu[1], Jarkko Isotalo[2], Simo Puntanen[2]
and George P.H. Styan[3]

[1]*Dawson College, Montréal, Canada;* [2]*University of Tampere, Tampere, Finland*
[3]*McGill University, Montréal, Canada*

While considering the estimation of regression coefficients in a partitioned weakly singular linear model, Chu, Isotalo, Puntanen & Styan (2004; in press) introduced a particular decomposition for the Watson efficiency of the ordinary least squares estimator. This decomposition presents the "total" Watson efficiency as a product of three factors. In this paper we give new insight into the decomposition showing that all three factors are related to the efficiencies of particular sub-models or their transformed versions. Moreover, we prove an interesting connection between a particular split of the Watson efficiency and the concept of linear sufficiency. We shortly review the relation between the efficiency and specific canonical correlations. We also introduce the corresponding decomposition for the Bloom-field-Watson commutator criterion, and give a necessary and sufficient condition for its specific split.

Continuous canonical correlation analysis

C. M. Cuadras

Department of Statistics, University of Barcelona, Spain.

We obtain the eigenvalues and eigenfunctions of a covariance kernel with respect to another kernel, related to the Cuadras-Auge copula, the survival copula for the Marshall-Olkin distribution. Then we obtain the set of canonical correlations and functions for this copula and prove that they have continuous dimensionality. The maximum correlation is the dependence parameter and the canonical function is the Heaviside distribution.

Partially diffuse starting values in state space models

M. Doherty

Statistics New Zealand, Wellington

I will discuss when two specifications of partially diffuse starting values give the same results in all state space models. I will also discuss some of the difficulties in giving a clean and simple justification of the results.

The geometry of statistical efficiency

K. Gustafson

Department of Mathematics, University of Colorado, USA

We will place certain parts of the theory of statistical efficiency into the author's operator trigonometry, thereby providing new geometrical understanding of statistical efficiency. Important earlier results of Bloomfield and Watson, Durbin and Kendall, Rao and Rao, will be so interpreted.

What are *the* residuals for the linear model?

John Haslett[1] and Stephen Haslett[2]

[1]*Trinity College Dublin*

[2] *Institute of Information Sciences and Technology,*

Massey University, New Zealand

We consider residuals for the linear model with general covariance structure. In contrast to the situation where observations are independent, there is no unique answer to the question posed in the title. We propose a taxonomy and draw attention to three quite distinct types of residual: the marginal residual, the conditional residual and the innovation residual. We adopt a very broad perspective including miixed models, time series, and smoothers as well as models for spatial and multivariate data. We concentrate on defining the three different types of residual and discusing their interelationships.

On singularly perturbed Markov chains

Moshe Haviv

Department of Statistics
The Hebrew university of Jerusalem, 91905 Jerusalem Israel

Nothing structurally happens in case that transition probabilities in an ergodic Markov chain are slightly perturbed as all measures are continuous with the perturbation. This is not the case when the state space can be decomposed into a number of ergodic classes plus a number of transient states, and when the perturbation coupled them all. For example, the stationary distribution is now uniquely defined and mean passage times between states belonging to different classes are now well defined. The talk will survey some results which are old and some which are under construction. Special emphasis will be given to the issue of time scales.

Simple procedures for finding mean first passage times in Markov chains

Jeffrey J. Hunter

Institute of Information and Mathematical Sciences

Massey University, Auckland, New Zealand

The derivation of mean first passage times in Markov chains involves the solution of a family of linear equations. By exploring the solution of a related set of equations, using suitable generalized inverses of the Markovian kernel I - P, where P is the transition matrix of a finite irreducible Markov chain, we are able to derive elegant new results for finding the mean first passage times. As a by-product we derive the stationary distribution of the Markov chain without the necessity of any further computational procedures. Standard techniques in the literature, using for example Kemeny and Snell's fundamental matrix Z, require the initial derivation of the stationary distribution followed by the computation of Z, the inverse of $I - P + \mathbf{e}\pi^T$ where $\mathbf{e}^T = (1, 1, \ldots, 1)$ and π^T is the stationary probability vector. The procedures of this paper involve only the derivation of the inverse of a matrix of simple structure, based upon known characteristics of the Markov chain together with simple elementary vectors. No prior computations are required. Various possible families of matrices are explored leading to different related procedures. Applications to perturbed Markov chains are discussed.

Hessian equivalence to bordered Hessian

Eric Iksoon Im

College of Business & Economics, University of Hawaii at Hilo, USA

The second-order sufficient conditions are traditionally given in terms of bordered Hessian matrix for constrained optimization problems, whereas they are given in terms of Hessian matrix for unconstrained problems. This dichotomy appears to bother some practitioners of optimization. In this paper, we show that the second order conditions can be stated solely in terms of Hessian matrix for both constrained and unconstrained cases, thus providing a unified criterion for testing the second–order conditions.

Comparison of the ordinary least squares predictor and the best linear unbiased predictor in the general Gauss-Markov model

J. Isotalo and S. Puntanen

Department of Mathematics, Statistics and Philosophy, University of Tampere, Finland

In this paper we consider the prediction of new observations in a general Gauss-Markov model. We compare properties two alternative predictors: the best linear unbiased predictor, BLUP, and the ordinary least squares predictor, OLSP. In particular, we focus on questions related to equality of two predictors BLUP and OLSP, and to the efficiency of the ordinary least squares predictor with respect to the best linear unbiased predictor.

Inverse of the Information Matrix

J. A. John

Department of Statistics, University of Waikato

When perturbing a regression model, to identify outliers and influential observations, most computer packages use updating procedures to find the inverse of the information matrix. Consider a linear model partitioned into two components, one of which is perturbed and the other left unchanged. In the talk we shall examine how the updating procedure can be extended to the inverse of the information matrix of the perturbed component. We shall also briefly discuss a three component model where the third component changes as the first component is perturbed. The results have important applications in the construction of efficient experimental designs.

Covariance decomposition for Gaussian graphical models

B. Jones[1] and M. West[2]

[1] *Institute of Information and Mathematical Sciences, Massey University Albany, New Zealand*

[2] *Institute of Statistics and Decision Sciences, Duke University, USA*

This talk will examine a new strategy for interpreting Gaussian graphical models, one which potentially will provide new knowledge about how variables interact. This strategy is based on decomposition of the covariance between two variables into a sum of path weights for all paths connecting the two variables in an undirected graph. Extensions to graphs initially inferred as acyclic directed graphs are also considered. This decomposition is derived using basic identities from linear algebra, and is feasible for very large numbers of variables if the corresponding precision matrix is parse. The resulting path weights can be used to highlight the most important intermediaries between two correlated variables. The talk will be illustrated with an example using gene expression data.

Some properties of transition matrices for chain binomial models

G. Jones

Institute of Information Sciences and Technology, Massey University, New Zealand

A chain binomial model is a Markov chain with a transition matrix whose rows are binomial probabilities. Two such chains will be presented and illustrated with possible applications. The talk will focus in particular on some interesting properties of the transition matrices.

Innovations, incomplete Fisher information matrices and empirical processes

Estate V. Khmaladze

Victoria University of Wellington, New Zealand

Suppose $\Xi = (X_1, \ldots, X_n)$ is an n-dimensional vector of independent random variables, each with normal distribution $N(0,1)$. Its projection $\hat{X} = (X_1 - \bar{X}, \ldots, X_n - \bar{X})$, where $\bar{X} = \sum_{i=1}^{n} X_i / n$, into the subspace orthogonal to I = (1,1...,1) has coordinates which are dependent random variables. However, let us couple $\hat{X}$ with the filtration $\{\Phi_k\}_{k=1}^{n}$ where each σ-algebra is generated by its first k coordinates:

$$\mathrm{F}_k = \sigma\{X_1 - \bar{X}, \ldots, X_k - \bar{X}\}, \qquad k = 1, \ldots, n,$$

and let us consider the so called *innovations* for $(\hat{X}, \{F_k\}_{k=1}^{n})$:

$$Y_k = X_k - \bar{X} - E[X_k - \bar{X} \mid \mathbb{F}_{k-1}] = X_k - \frac{1}{n-k+1} \sum_{i=k}^{n} X_i$$

This conditioning on the "past" gives us the vector $\Psi = (Y_1, Y_2 \ldots, Y_n)$ again with independent coordinates. At the same time Ψ is in one-to-one correspondence with $\hat{X}$. So, it carries the same statistical "information", but has a simpler distribution than $\hat{X}$.

This construction in continuous time framework of empirical processes leads to the following substantial gain. Let $P_n(\cdot)$ be empirical distribution function of some sample and $P_n(\cdot, \theta)$ be some hypothetical distribution function, depending on the parameter θ, which has to be estimated from the sample. The *estimated* empirical process

$$\hat{\upsilon}_n(x) = \sqrt{n}(P_n(x) - P(x, \hat{\theta}))$$

converges in distribution to projection $\hat{\upsilon}$ of Brownian motion (see, e.g., (Khmaladze, 1979) (Khmaladze, Koul, 2004)). It is not convenient to use functionals $\psi(\hat{\upsilon}_n)$ as test statistics because the distribution of $\psi(\hat{\upsilon})$ is not easy to find. However, by considering

$$\mathrm{F}_x = \sigma\{\hat{\upsilon}(y), y \le x\}$$

we can instead construct the innovation process $\omega(\cdot)$ for the process $\{\hat{\upsilon}(x), F_x, -\infty < x < \infty\}$ which will be simply a Brownian motion and, at the same time, will be in one-to-one correspondence with $\hat{\upsilon}(\cdot)$. Thus, empirical version $\omega_n(\cdot)$ of $\omega(\cdot)$ will preserve all statistical "information" which lies in $\hat{\upsilon}_n(\cdot)$ and the distribution of the functionals $\psi(\omega(\cdot))$ will be relatively easy to find.

In the construction of the innovation process $\omega_n(\cdot)$ the inverse of *incomplete extended Fisher information matrix* $C(\cdot,\theta)$ plays a central role. Namely, if $p(x,\theta)$ is a pdf of $P(x,\theta)$ consider the extended score function

$$q(x,\theta)^T = [1, \frac{\partial \ln p(x,\theta)}{\partial \theta}].$$

Using it construct the matrix

$$C(z,\theta) = \int_z^\infty q(x,\theta) q(x,\theta)^T p(x,\theta) dx.$$

Then the innovation process $\omega_n(\cdot)$ is

$$\omega_n(x) = \sqrt{n}\left[P_n(x) - \int_{-\infty}^x q(z,\theta)^T C^{-1}(z,\theta) \int_z^\infty q(y,\theta) P_n(dy) P(dz,\theta) \right].$$

For some families the matrix $C(z,\theta)$ is degenerate for some z but one can prove that ω_n does not depend on the choice of the generalized inverse (see, e.g., Zigroshvili(1998)).

To the best of the author's knowledge, the question of serious applications of this approach to linear models remains open.

References

Khmaladze, E.V. (1979). The Use of Omega-square Tests for Testing Parametric Hypotheses. *Theory Probab. Appl.*, **24**, N2, 283-302.

Khmaladze, E.V., Koul, H.L. (2004). Martingale Transforms Goodness-of-Fit Tests in Regression Models. *Ann.Statist.*, **32**, N3, 995-1034.

Zigroshvili, Z.P. (1998). Generalised inverse and martingale transform for empirical processes. *Georgian Mathemat. J.*, **6**, No.4.

Approximation of the parameter distributions of growth curve model

T. Kollo [1] and D. von Rosen [2]

[1] *Institute of Mathematical Statistics, University of Tartu, Estonia*

[2] *Department of Engineering and Biostatistics,*

Swedish University of Agricultural Sciences, Sweden

We are going to approximate density functions of the maximum likelihood estimators of the parameter matrices B and Σ in the Growth Curve Model:

$$X = ABC + \Sigma^{1/2} E$$

with known matrices A and C. In both cases we rely on a general multivariate density expansion introduced in Kollo, von Rosen (1998). The main interest is focused to the parameter B. When approximating density of the parameter matrix B a mixture of two matrix distributions is obtained. One of the distributions is normal and the other one an elliptical distribution from the class of Kotz-type distributions (Fang, Kotz, Ng, 1990, p. 69). The first term in the remainder of the density expansion is of order $O(n^{-2})$ what refers to a good approximation.

We are concentrating to the mixture of the two distributions which appear in the expansion. The basic characteristics are examined for both, the Kotz type distribution and the mixture of interest. The program of the workshop will be of interest to anyone concerned with matrices, statistics, data analysis, and computational issues. For references, use the style of the examples below. The references should be in alphabetical order.

References

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

Kollo, T., von Rosen, D. (1998). A unified approach to the approximation of multivariate densities. *Scand. J. Statist.*, pp93-109.

Invariant estimator in a quadratic measurement error model

A. Kukush

Kiev National Taras Shevchenko University, Ukraine.

An adjusted least squares (ALS) estimator is derived that yields a similarity invariant and consistent estimate of the parameters of multivariate implicit quadratic measurement error model (IQMEM). Consistency means that the estimate converges to the true value of the parameter, as the sample size tends to infinity. In addition, a consistent estimator for the measurement error noise variance is proposed. Important assumptions are: (1) all errors are i.i.d. and (2) the error distribution is rotation-invariant. The estimators for the quadratic measurement error model are used to estimate consistently conic sections and ellipsoids.

In the IQMEM, the ordinary least squares (OLS) estimator is inconsistent, and due to the non-linearity of the model, the orthogonal regression (OR) estimator is inconsistent as well. Simulation examples, comparing the ALS estimator with the OLS method and the OR method, are discussed for the ellipsoid fitting problem.

The results are joint with Prof. S.Van Huffel and Dr. I.Markovsky (Belgium), and Dr. S.Shklyar (Ukraine). The consistency is shown in Kukush et al. (2004), and the numerical algorithm is proposed in Markovsky et al. (2004).

References

Kukush A., Markovsky I., and Van Huffel S. (2004). Consistent estimation in an implicit quadratic measurement error model. *Computational Statistics and Data Analysis*, pp.123-147.

Markovsky I., Kukush A., and Van Huffel S. (2004). Consistent least squares fitting of ellipsoids. *Numerische Mathematik*, pp.177-194.

Semi-parametric Efficiency, Projection and the Scott-Wild Estimator model

Alan Lee and Alastair Scott

University of Auckland, New Zealand

In situations where data are expensive to collect, it is important to get the "biggest bang for the buck" and use statistical methods that make maximum use of the information in the data. In this talk, we discuss how this idea is made precise in statistical theory by the concept of efficiency. We deal with efficient estimation in parametric and semi-parametric models for both single and multiple samples, and describe the notion of the information bound, using the geometric notion of projection. Our results are applied to regression models for data collected under case-control and two-stage sampling. In particular, we show that the methods proposed by Alastair Scott and Chris Wild for the analysis of generalized case-control data have full semi-parametric efficiency, and hence are the best possible.

Model for students expected performance level through varying control limits in relation to power of valuation

Lakshmi Narasimhaiah[1] and Kishore Hoysal[2]

[1]*Dept of Civil Engineering, Adhiyamaan College of Engineering, Dr. MGR Nagar, Hosur, Pin-635109, Tamil Nadu, India.*

[2]*Dept of Mechanical Engineering, Islamiah Institute of Engineering, Bangalore, Karnataka, India.*

To date no attempt has been made to understand and measure the performance of student in the view of student's logics. This paper first of its kind in technical education based on basic statistical tools which defines a continuously varying upper and lower limits of performance assuming mid point of the limits as ideal score. Variability is more means more chances of errors. Lower the performance level of students leads to bigger range for scores obtained. Highest level of performance will lead the variability to zero. Valuation of any answer scripts involves human factor. Here we have defined a new parameter termed as "Power of Valuation" which gives an insight in to the whole process of valuation which is tangible and every valuer is identified with his power of valuation in terms of numerical value.

Statistical proofs of matrix theorems

C.R. Rao

Pennsylvania State University, U.S.A.

Matrix algebra is extensively used in the study of linear models, multivariate analysis and optimization problems. It is interesting to note that the matrix results needed to prove statistical propositions can themselves be deduced using some statistical results which can be derived without using matrix algebra. The results are based on Fisher information and its properties which can be established without using matrix results.

Anti-eigen and anti-singular values of a matrix and applications to problems in statistics

C.R. Rao

Pennsylvania State University, U.S.A.

Let A be $p \times p$ positive definite matrix. A p-vector x such that $Ax = \lambda x$ is called an eigenvector with the associated with eigenvalue λ. Equivalent characterizations are:

(i) $\cos\theta = 1$, where θ is the angle between x and Ax.

(ii) $(x' Ax)^{-1} = x A^{-1} x$.

(iii) $\cos\Phi = 1$, where Φ is the angle between $A^{1/2}x$ and $A^{-1/2}x$.

We ask the question what is x such that $\cos\theta$ as defined in (i) is a minimum or the angle of separation between x and Ax is a maximum. Such a vector is called an anti-eigenvector and $\cos\theta$ an anti-eigenvalue of A. This is the basis of operator trigonometry developed by K. Gustafson and P.D.K.M. Rao (1997), *Numerical Range: The Field of Values of Linear Operators and Matrices*, Springer. We may define a measure of departure from condition (ii) as $\min[(x' Ax)(x' A^{-1}x)]^{-1}$ which gives the same anti-eigenvalue. The same result holds if the maximum of the angle Φ between $A^{1/2}x$ and $A^{-1/2}x$ as in condition (iii) is sought. We define a hierarchical series of anti-eigenvalues, and also consider optimization problems associated with measures of separation between an r ($< p$) dimensional subspace S and its transform AS.

Similar problems are considered for a general matrix A and its singular values leading to anti-singular values.

Other possible definitions of anti-eigen and anti-singular values, and applications to problems in statistics will be presented.

Diagnosing non-linear regression structure with power additive smoothing splines

W. Sakamoto

Graduate School of Engineering Science, Osaka University, Japan

The additive model and its extended versions have been popular and are included in some statistical softwares (Hastie and Tibshirani, 1990). However, its tight assumptions such as additivity may often give unsatisfactory fit. A power-additive smoothing spline (PASS) model, in which the power transformation (Box and Cox, 1964) is applied to response variables in the additive model, is proposed to diagnose additivity on explanatory variables. It is a nonparametric extension of the linear models with the power-additive transformation considered by Draper and Hunter (1969) and Goto (1995). The power and smoothing parameters are selected by maximizing the marginal log-likelihood in the context of an empirical Bayes approach. The marginal log-likelihood, which involves multi-dimensional integrals, is approximated by using the second-order Taylor expansion of the penalized log-likelihood around its maximum point. The approximated marginal log-likelihood is represented using the determinant of the coefficient matrix in the BLUP equations for a mixed model. A simulation experiment shows that the PASS model suggests an appropriate transformation which aims the additivity.

References:

Box G. E. P. and Cox D. R. (1964). An analysis of transformations (with discussion). *J. Roy. Statist. Soc. Ser.* B, 26, pp211--243.

Draper N. R. and Hunter W. G. (1969). Transformations: some examples revisited. *Technometrics*, 11, pp23--40.

Goto M. (1995). Double power transformations and their performances. *Proceedings of the International Conference on Statistical Methods and Statistical Computing for Quality and Productivity Improvement* (Seoul, Korea), Vol. I}, pp386--397.

Hastie T. and Tibshirani R. (1990). *Generalized Additive Models*. Chapman and Hall, London.

Sakamoto W. (2003). Exploring nonlinear structure with nonparametric regression. *Proceedings of the 54th Conference of the ISI* (CD-ROM).

On the optimal choice of the regularization parameter through variance ratio estimation

Burkhard Schaffrin

Dept. of Civil and Environmental Engineering and Geodetic Science,

The Ohio State University, Columbus, Ohio, USA

A persistent problem in Tykhonov/Phillips regularization for the treatment of ill-conditioned linear problems is the optimal choice of the regularization parameter. Here we propose an efficient algorithm which adopts formulas from variance component estimation to compute a certain variance ratio iteratively that, after convergence, may serve as "optimal" regularization parameter. During each iteration step, the trace of a matrix with the size of the unknowns has to be evaluated.

Recollections from a fifty-year random walk midst matrices, statistics and computing

Shayle R. Searle

Department of Statistical Science

Cornell University, Ithaca, N.Y. 14850, U.S.A.

A brief and personal overview is given of some developments in matrix algebra, statistics and computing during the years of my participation in these activities from 1945 – 2005.

Things my mother never told me about matrices

G. A. F Seber

University of Auckland, New Zealand

Linear algebra is one of our most useful tools in statistics, and in science generally. The question then is: What should we teach our undergraduate and our graduate students about linear algebra, and how are we going to teach it? Furthermore, what do we need to know if we are going to do research in statistics and what resources do we need? If we are in a statistics or biostatistics department, which has separated or has always been separated from the mathematics department, should we teach our own linear algebra or should we delegate it to the mathematicians? I hope to consider some of these questions in my talk.

As far as research tools are concerned, I am currently writing a book tentatively entitled, "*A matrix handbook for statisticians*". This book is for research people who may want to use the book like a compendium or encyclopaedia to look up a particular result and also obtain some references to applications. In my talk I hope to give an overview of the subject as I currently see it. In my book, proofs will be referenced but omitted to keep down the book size. This work is still in progress so I am very interested in receiving reprints and references to useful material and talking to people about the subject.

Coefficients of ergodicity in a matrix setting

E. Seneta

School of Mathematics and Statistics, University of Sydney, NSW 2006. Australia

If A is an $n \times n$ real matrix, with equal row sums a, we shall be concerned with the quantity:

$$\tau_p(A) = \sup \{ \| \delta^T A \|_p : \| \delta^T \|_p = 1,\ \delta^T \mathbf{1} = 1 \},$$

where optimization is over real-valued vectors $\delta = \{ \delta_i \}$, and $\| \delta^T \|_p = \{ \sum_{i=1}^{n} |\delta_i|^p \}^{1/p}$. In the case $p = 1$ it has explicit form $\tau_1(A) = \frac{1}{2} \max_{(i,j)} \sum_{s=1}^{n} |a_{is} - a_{js}|$ and it bounds eigenvalues of A other than a. When A is a stochastic matrix P, so that $a=1$, coefficients such as $\tau_p(P)$, have been used for investigating weak convergence in finite inhomogeneous Markov chains (products of stochastic matrices), because of the submultiplicative property of such coefficients: hence the name "coefficients of ergodicity". $\tau_1(P)$ is the Markov-Dobrushin coefficient of ergodicity. It is the spectral bounding property of this coefficient, however, which is the focus of this presentation. Statistical uses of it to be discussed include (1) its bounding the maximal correlation of a finite bivariate distribution; and (2) its measuring relative sensitivity, under perturbation of the transition matrix P, of the stationary distribution of a finite homogeneous Markov chain.

The historical origin of the Markov-Dobrushin coefficient, and some remarks (needing further development) on its relation to coefficients such as Poincaré's and Cheeger's, were given in Seneta (1996). Other aspects of ergodicity coefficients could well be surveyed. These include work published since the appearance of Seneta (1981), on explicit forms of coefficients in terms of matrix entries, on more general definitions of coefficients, and on application to matrices with structure other than that specified above for A. The author's own list of publications since 1973 on ergodicity coefficients, and publications and books relating to this topic which he has been sent by review journals and colleagues since 1981, will be made available, and will provide further guidance in this respect from their titles. More extensive literature search via MathSciNet will turn up further literature. Additions to the cited bibliography are welcome: please contact the author at the Workshop, or later by email (eseneta@maths.usyd.edu.au).

References

Seneta, E. (1981). *Non-negative Matrices and Markov Chains*, (2nd Edition) Springer-Verlag, New York.

Seneta, E. (1996). Markov and the Birth of Chain Dependence Theory, *International Statistical Review*, **64**, 255-263.

Inequalities and equalities associated with the Watson efficiency in orthogonally partitioned full rank linear models

George P. H. Styan
McGill University, Montréal

We consider partitioned linear models where the model matrix $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$ has full column rank, and concentrate on the special case where $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$, when we say that the model is *orthogonally partitioned.* We assume that the underlying covariance matrix is positive definite and find conditions for which the total Watson efficiency of ordinary least squares exceeds, equals or is less than the product of the two subset Watson efficiencies, i.e., the product of the Watson efficiencies associated with the two subsets of parameters in the underlying partitioned linear model.

We introduce the notions of *generalized efficiency function, efficiency factorization multiplier* and *determinantal covariance ratio*, and obtain several inequalities and equalities. We give special attention to those partitioned linear models for which the total Watson efficiency of ordinary least squares equals the product of the two subset Watson efficiencies. A key characterization involves the relationship between a certain partial correlation coefficient and the associated ordinary correlation coefficient.

We illustrate our findings with two real data sets: Delozier's lathe data [S. Weisberg, *Applied Linear Regression*, 2nd ed., Wiley, 1985] and Worsley's fMRI data [Worsley et al., *NeuroImage*, 15, 1-15, 2002].

[Joint research with Ka Lok Chu (Dawson College, Montréal), Jarkko Isotalo and Simo Puntanen (University of Tampere), following results by Chu (Ph.D. thesis, McGill University, 2004) and by Chu, Isotalo, Puntanen and Styan (*Inequalities and Applications*, Th.M. Rassias, ed., Springer, to appear).]

Eigenvectors of block circulant matrices

Garry J. Tee

Department of Mathematics, University of Auckland, New Zealand

The eigenvectors and eigenvalues of block circulant matrices had been found for real symmetric matrices with symmetric submatrices, and for block circulant matrices with circulant submatrices. The eigenvectors are now found for general block circulant matrices, including the Jordan Canonical Form for defective eigenvectors.

A Matrix with Consecutive Integer Eigenvalues

I. Traat and T. Nahtman

Institute of Mathematical Statistics, University of Tartu, Estonia

A nonsymmetric matrix $\mathbf{A} = diag(\mathbf{1}^T \mathbf{C}) + \mathbf{C}$: N×N, where $c_{ij}=p_i(1-p_j) / (p_i-p_j)$, $c_{ii} = 0$, and $0 \leq p \leq 1$ are distinct, is remarkable due to its eigenvalues. They are consecutive integers 0 to $N-1$. The matrix has applications in sampling theory where one is looking for the inclusion probabilities of Conditional Poisson sampling design with fixed sample size n, given the inclusion probabilities p_i of Poisson sampling design (Bondesson et al., 2004). The eigenvector of **A** corresponding to the eigenvalue $n-1$ gives solution to the problem. The findings of authors are presented in the lecture. The authors feel that similar matrices must have been considered earlier in some other context and hope to get references.

References

Bondesson L., Traat I., Lundqvist A. (2004), Pareto Sampling versus Sampford and Conditional Poisson Sampling, *Research Report No. 6, ISSN 1401-730X, Department of Mathematical Statistics, Umeå University*, pp 32.

On the commutativity of orthogonal projectors

G. Trenkler

Department of Statistics, University of Dortmund, Germany

It is well-known that the Moore-Penrose inverse of the product of two orthogonal projectors is idempotent. Moreover, this product is an orthogonal projector if both factors commute. Basing on these facts and a useful representation of idempotent matrices, a number of results concerning the commutativity of orthogonal projectors are achieved. Special attention is paid to the group inverse and the parallel sum.

Leaving useful traces when working with matrices

K. Vehkalahti

Department of Mathematics and Statistics, University of Helsinki, Finland

Matrix computations are studied in the context of multivariate statistical analysis. Examples related to measurement and factor analysis are presented, focusing on the way of working. How properly does the working process get documented, and what sort of traces are left behind, are relevant questions in any area of research. Leaving useful traces may save a considerable amount of time by helping the researcher in backtracking and managing multiple projects. It also provides better possibilities for other researchers to comprehend the points of a study. These principles are demonstrated in practice with Survo software.

Estimating the covariance matrix by spectral decomposition approach in linear mixed model

Song-Gui Wang and Zhong-Zhen Jia

College of Applied Sciences, Beijing University of Technology, Beijing, 100022, China

For balanced linear mixed effects models in which random effects are multi-way classification, the covariance matrix Σ of observation vector can be written as the spectral decomposition form: $\Sigma = \sum_{i=1}^{k} \lambda_i M_i$, where the λ_i's \$ are the distinct eigenvalues of Σ and the M_i's \$ are a known completed orthogonal set of projection matrices determined by the design matrices corresponding to random effects. The λ_i's are all linear combination of unknown variance components. The estimable condition of the λ_i's is given. By using above decomposition and unified theory of the least squares which is so-called spectral decomposition estimate (SDE) approach, we can obtain estimate λ_i^* which have independent χ^2 distribution with the degree freedom $m_i = tr(M_i)$. Thus $\Sigma^* = \sum_{i=1}^{k} \lambda_i^* M_i$ is an estimate of Σ. Under the loss function $L(\Sigma^*\Sigma) = tr(\Sigma^* - \Sigma)^2$, the risk of $R(\Sigma^*) = E\{L(\Sigma^*,\Sigma)\}$ is obtained. For the random effect model, it is proved that three estimates (SDE , ANOVAE and MINQUE) of Σ have the same risk function.

On permutations of matrix products

Hans Joachim Werner[1] and Ingram Olkin[2]

[1] *Department of Statistics, Faculty of Economics, University of Bonn, Germany.*

[2] *Department of Statistics, Stanford University, USA.*

It is well-known that *trace(AB)* ≥ 0 for nonnegative definite A and B. However, *trace(ABC)* can be positive, zero or negative, even when *C* is nonnegative definite. The genesis of the present investigation is consideration of a product of square matrices $A=A_1A_2, \ldots, A_n$. Permuting the factors of *A* leads to a different matrix product. We are interested in conditions under which the spectrum remains invariant.

Publication of Papers

The Editors of "Research Letters in Information and Mathematical Sciences" (RLIMS) will publish a special issue of RLIMS of papers delivered at the Workshop which will be made available on the web - see http://iims.massey.ac.nz/research/letters/. A printed version of the special RLIMS issue will be made available to all participants.

Note that RLIMS is a preprint series. After editing, articles are published online and can be referenced. Copyright remains with the authors. All participants will be able to submit their manuscript to a refereed journal for publication.

Manuscripts for this special issue of RLIMS should be submitted, electronically to Dr Paul Cowpertwait <P.S.Cowpertwait@massey.ac.nz> by June 30, 2005 using one of the templates (Word or Latex) available at http://iims.massey.ac.nz/research/letters/instructions.html

"International Statistical Review" (ISR)
The IOC have agreed to an arrangement whereby participants delivering "synthesis and review" type statistical papers will, subject to the usual referring process, be given strong consideration for having their manuscript published in the ISR. This is a review journal, and does not cater for narrowly specialized papers. Publication will be fast once the referees' reports are positive. Participants are recommended to consult the "Statement of Editorial Policy" from the Editors of ISR (below). *Participants presenting narrowly specialized technical papers should consider making a submission to "Linear Algebra and its Applications" or the "Electronic Journal of Linear Algebra".*

INTERNATIONAL STATISTICAL REVIEW

Statement of Editorial Policy

The *International Statistical Review (ISR)* is the flagship journal of the International Statistical Institute and of its constituent sections (the Bernoulli Society for Mathematical Statistics and Probability, the International Association for Official Statistics, the International Association for Statistical Computing, the International Association for Statistical Education and the International Association of Survey Statisticians). The *ISR* is widely circulated and subscribed to by individuals and institutions in all parts of the world.

The main aim of the *ISR* is to publish papers of an expository, review, or tutorial nature that will be of wide interest to readers. Such papers may or may not contain strictly original material. All papers are refereed.

The *ISR* has two Editors, in order to demonstrate its commitment to the whole field of statistics, widely interpreted. Eugene Seneta of the University of Sydney, Australia, is primarily responsible for papers in the broad area of mathematical and theoretical statistics and probability, as well as computational statistics, statistics as applied in the physical, biological, medical and environmental sciences, industry and commerce, history of statistics and the teaching of statistics. Asta Manninen of the City of Helsinki, Urban Facts, Finland, is primarily responsible for the areas of

official and government statistics and public policy, demography and population studies, banking and finance, the social sciences, survey statistics, as well as for papers of broad public interest.

Review papers are the main raison d'être of the *ISR*, but the kind of review that the Editors would wish to encourage is *not* purely bibliographic. Readers of the *ISR* will find 'critical reviews' far more useful and these are very strongly encouraged. A critical review is one that provides an introduction to a field, pointers to key original references, and clear and interesting insights and comments both about past work and about future directions for research and applications. A good critical review will be accessible to non-specialists while being stimulating and interesting to experts. It will of necessity be something of a 'personal view' of a subject while of course retaining scientific integrity and giving full credit to original sources of cited work. There is no need for a critical review to attempt exhaustive coverage of the field (provided it does not claim to do so!) and careful direction towards key references is more important than any attempt at an exhaustive bibliography.

Broadly based papers of wide interest that contain original material are very much welcomed. However the Editors do not wish to publish (even really excellent) technically original papers that are accessible or interesting only to a small group of specialists.

Papers on the history of statistics and probability are welcomed provided they are of wide interest and preferably if they convey insights of current relevance.

Authors and referees of all papers should bear in mind that many of the readers of the *ISR* (particularly in the developing world) do not have easy access to libraries or to other journals, and therefore it is hoped that papers will be as self-contained as possible, while of course giving proper bibliographic credit.

Eugene Seneta and Asta Manninen, Joint Editors
September, 2003.

List of Participants

Ahmed, S. Ejaz	University of Windsor, Canada	seahmed@uwindsor.ca
Anderson, Robert	Massey University, Palmerston North, New Zealand	R.D.Anderson@massey.ac.nz
Bekker, Andriette	University of South Africa, South Africa	bekkea@unisa.ac.za
Chen, Anyue	University of Greenwich, London, UK	a.chen@gre.ac.uk
Chinna, Karuthan	University Technology MARA, Malaysia	karuth@tm.net.my
Cowpertwait, Paul	Massey University, Auckland, New Zealand	P.S.Cowpertwait@massey.ac.nz
Cuadras, Carles M.	University of Barcelona, Spain	ccuadras@ub.edu
Doherty, Mike	Statistics New Zealand, Wellington, New Zealand	mike.doherty@stats.govt.nz
Gustafson, Karl	University of Colorado, Boulder, Colorado, USA	gustafs@euclid.colorado.edu
Haslett, Stephen	Massey University, Palmerston North, New Zealand	s.j.haslett@massey.ac.nz
Haviv, Moshe	The Hebrew University of Jerusalem, Israel	haviv@pluto.mscc.huji.ac.il
Henderson, Harold	AgResearch Ltd, Hamilton, New Zealand	harold.henderson@agresearch.co.nz
Hitose, Yuichi	University of Auckland, New Zealand	yuichi@stat.auckland.ac.nz
Hoysal, Kishore	Islamiah Institute of Engineering, India	shore@rediffmail.com
Hunter, Jeff	Massey University, Auckland, New Zealand	J.Hunter@massey.ac.nz
Im, Eric	University of Hawaii at Hilo, USA	eim@hawaii.edu
Islam, Rabiul	Charles Sturt University, Australia	rislam@csu.edu.au
Isotalo, Jarkko	University of Tampere, Finland	jarkko.isotalo@uta.fi
Jia, Zhong-Zhen	Beijing University of Technology, China	jiazz88@yahoo.com.cn
John, Nye	University of Waikato, Hamilton, New Zealand	nye@stats.waikato.ac.nz
Jones, Beatrix	Massey University, Auckland, New Zealand	m.b.jones@massey.ac.nz
Jones, Geoff	Massey University, Palmerston North, New Zealand	g.jones@massey.ac.nz
Jordan, Claire	Massey University, Auckland, New Zealand	C.A.Jordan@massey.ac.nz
Khmaladze, Estate	Victoria University of Wellington, New Zealand	Estate.Khmaladze@mcs.vuw.ac.nz
Kollo, Tonu	University of Tartu, Estonia	tonu.kollo@ut.ee
Kukush, Alexander	Kiev National Taras Shevchenko University, Ukraine	alexander_kukush@univ.kiev.ua
Lee, Alan	University of Auckland, New Zealand	lee@stat.auckland.ac.nz
Littlejohn, Roger	AgResearch Ltd, Invermay, New Zealand	roger.littlejohn@agresearch.co.nz
Matsubara, Kazuo	Osaka University of Economics, Japan	matubara@osaka-ue.ac.jp
McDonald, Barry	Massey University, Auckland, New Zealand	B.McDonald@massey.ac.nz
Narasimhaiah, Lakshmi	Adhiyamaan College of Engineering, Bangalore , India	
Puntanen, Simo	University of Tampere, Finland	simo.puntanen@uta.fi
Rao, C.R.	Pennsylvania State University, USA	crr1@psu.edu
Ruggiero, Kathy	CSIRO, Canberra, Australia	kathy.ruggiero@csiro.au
Sakamoto, Wataru	Osaka University, Japan	sakamoto@sigmath.es.osaka-u.ac.jp
Schaffrin, Burkhard	Ohio State University, Columbus, USA	schaffrin.1@osu.edu
Schliep, Klaus	Massey University, Palmerston North, New Zealand	k.p.schliep@massey.ac.nz
Scott, Alastair	University of Auckland, New Zealand	a.scott@auckland.ac.nz
Searle, Shayle	Cornell University, Ithaca, New York, USA	ssearle@webtv.net
Seber, George	University of Auckland, New Zealand	seber@stat.auckland.ac.nz
Seneta, Eugene	University of Sydney, Australia	eseneta@maths.usyd.edu.au
Styan, George	McGill University, Montreal,, Canada	styan@math.mcgill.ca
Tee, Gary	The University of Auckland, New Zealand	tee@math.auckland.ac.nz
Thrush, Tarsha	Fonterra, Whangarei, New Zealand	tarsha.thrush@fonterra.com
Traat, Imbi	University of Tartu, Estonia	imbi.traat@ut.ee
Trenkler, Gotz	University of Dortmund, Germany	trenkler@statistik.uni-dortmund.de
Vehkalahti, Kimmo	University of Helsinki, Finland	Kimmo.Vehkalahti@helsinki.fi
Wake, Graeme	Massey University, Auckland, New Zealand	G.C.Wake@massey.ac.nz
Walsh, Daniel	Massey University, Auckland, New Zealand	d.c.walsh@massey.ac.nz
Wang, Song-Gui	Beijing University of Technology, China	wangsg@bjpu.edu.cn
Werner, Joachim	University of Bonn , Germany	hjw.de@uni-bonn.de

Photograph of Participants from the last IWMS meeting

Participants of the 13th International Workshop in Matrices and Statistics held at Bedlewo, Poland, 18 – 21 August 2004.

Photograph by Hazel Hunter

Some Biographical Material on the Keynote Speakers

The Workshop organizers are grateful to Professors C R Rao, Shayle Searle, George Seber, and Eugene Seneta for their acceptance of the invitations to deliver the keynote talks.

We have included in this booklet the following biographical information

Professor C R Rao
Short biography of C R Rao, F.R.S.
"A Conversation with C R Rao, Interview by Morris DeGroot, in *Statistical Science*, Vol 2, 1987. Reprinted with permission of the Institute of Mathematical Statistics.

Professor Shayle Searle
"Shayle Searle's Contributions to the Evolution of the SAS System" by Robert N. Rodriguez, Russell D. Wolfinger, Randall D. Tobias, SAS Institute Inc., SAS Campus Drive, Cary, NC 27513 USA. Provided by the SAS Institute, Cary, NC, U.S.A.

Professor Eugene Seneta
"Pitman Medal Awarded to E. Seneta." *Australian and New Zealand Journal of Statistics*, 40(4), *pp*385-387, 1998. Reprinted with permission from Blackwell Publishing Ltd.

Professor George Seber
A short biography of George Seber

Short Biography of C R Rao, F.R.S.

Padma Vibhushan awardee

C. R. Rao is among the world leaders in statistical science over the last six decades. His research, scholarship and professional services have had a profound influence on theory and applications of statistics. Technical terms such as, *Cramer-Rao inequality, Rao-Blackwellization, Rao's Score Test, Fisher-Rao Theorem, Rao distance, and orthogonal arrays (described by Forbes Magazine as "new manthra" for industries)* appear in all standard books on statistics. Two of his papers appear in *Breakthroughs in Statistics* in the last century. C R Rao is the author of 14 books and about 350 research papers. Three of his books have been translated into several European and Chinese and Japanese languages.

Rao received MA degree in mathematics from Andhra University and MA degree in statistics from Calcutta University and started doing research at the ISI (Indian Statistical Institute). He was invited to work on a project at the Cambridge University, UK, which required the statistical methodology he developed at the ISI. Based on the work he did he earned his Ph.D. and Sc.D. degrees of the Cambridge University, and up to date he has received 29 Honorary Doctoral degrees from universities in 17 countries in 6 continents. He held several important positions, as the Director of the Indian Statistical Institute, Jawaharlal Nehru Professor and National Professor in India and University Professor, University of Pittsburgh and Eberly Professor of Statistics and Director of the Center for Multivariate Analysis, Pennsylvania State University in USA. At the age of 84 years, he is still active doing research in statistics, attending conferences and giving lectures.

As the Head and later Director of the Research and Training School at the Indian Statistical Institute for a period of over 40 years, Rao developed research and training programs and produced outstanding students which *"put India not far from the center of the statistical map of the world"*. He supervised the research work leading to Ph.D. degree of 50 students who in turn produced 250 Ph.D.s. During this period he also directed the training programs at the International Statistical Educational Center (ISEC) which led to the development of statistics in the South East Asian region. Rao was the Chairman of a UN Committee, which examined the demand for statistical personnel in Asian countries and recommended the establishment of an Institute for statistical development in South East Asia. On the basis of his recommendation ASI (The Asian Statistical Institute) now known as Statistical Institute for Asia and Pacific was established in Tokyo to provide training to statisticians working in government and industrial organizations.

Times of India dated 31 December 1988 chose C R Rao as one of the 10 top scientists of India; the list includes the outstanding scientists, S N Bose, S Ramanujan, Harishchandra, H Khurana, C V Raman, S Chandrasekhar, Salim Ali and G N Ramachandran.

For his pioneering contributions to statistical theory and applications, Rao received numerous awards. He has been elected to the National Academy of Sciences, USA, American Academy of Arts and Science, Royal Society (UK Akademy of Sciences, FRS), Indian National Science Academy, Lithuanian Academy of Sciences and Third World Academy of Sciences. He was made an Honorary Member of the International Statistical Institute, International Biometric Society, Royal Statistical Society (UK), Finnish Statistical Society, Portuguese Statistical Society, Institute of Combinatorics and Applications and Wold Innovation Foundation, and Honorary Life Fellow of King's College, Cambridge, UK. He has been the President of all prestigious statistical associations, the International Statistical Institute, Institute of Mathematical Statistics, USA and the International Biometric Society. He was inducted into

the Hall of Fame of the National Institution for Quality and Reliability, Chennai Branch, for his contributions to industrial statistics and the promotion of quality control programs in Indian industries. At the Berlin conference of the International Statistical Institute held in 2003, Rao received the prestigious *International Mahalanobis Prize for lifetime achievement in statistics and the promotion of best statistical practice.*

He received numerous medals: Gold Medal of Calcutta University, Wilks Medal of the American Statistical Association, Wilks Army Medal, Guy Medal in Silver of the Royal Statisttical Society, Megnadh Saha Medal and Srinivasa Ramanujan Medal of the Indian National Science Academy, J.C.Bose Gold Medal of Bose Institute and Mahalanobis Centenary Gold Medal of the Indian Science Congress. He received Batnagar award of CSIR (Council of Scientific and Industrial Research, India).

Rao was honored by the President of the USA with the prestigious National Medal of Science "*as a prophet of new age*" with the citation, "*for his pioneering contributions to the foundations of statistical theory and multivariate statistical methodology and their applications, enriching the physical, biological, mathematical, economic and engineering sciences*".

The Government of India honored him with the second highest civilian award, *Padma Vibhushan* for "outstanding contributions to Science and Engineering / Statistics", and also instituted a cash award in honor of C R Rao, *"to be given once in two years to a young statistician for work done during the preceding 3 years in any field of statistics"*.

International conferences were held in USA, India, Canada and Switzerland and special issues of the prestigious journals like *Statistical Planning and Inference, Linear Algebra and its Applications, Sankhya* and several festschrift volumes were published in his honor.

For his outstanding achievements Rao has been honored with the establishment of an Advanced Institute of Mathematics, Statistics and Computer Science (AIMSCS) named after him, in the Osmania University campus, Hyderabad, India.

Professor Rao receiving the prestigious National Medal of Science from the U.S. President

Statistical Science
1987, Vol. 2, No. 1, 53–67

A Conversation with C. R. Rao

Morris H. DeGroot

Calyampudi Radhakrishna Rao was born on September 10, 1920, in Hadagali, Karnataka State, India. He received an M.A. in mathematics from Andhra University in 1940; an M.A. in statistics from Calcutta University in 1943; a Ph.D. from Cambridge University in 1948, with a thesis entitled "Statistical problems of biological classification"; and an Sc.D. from Cambridge in 1965 on the basis of his published work in statistics as a whole. He joined the Indian Statistical Institute as a statistician in 1944 and became a professor in 1949, the director of the Research and Training School in 1964, the secretary and director of the Institute in 1972, and the Jawaharlal Nehru Professor in 1976. In 1979 he was appointed a University Professor in the Department of Mathematics and Statistics at the University of Pittsburgh. He has been the President of the International Statistical Institute, the Institute of Mathematical Statistics, the International Biometric Society, and the Indian Econometric Society; and since 1964, an editor of *Sankhyā*. He is a Fellow of the Royal Society (FRS) of the United Kingdom; a Foreign Honorary Member of the American Academy of Arts and Science; an Honorary Member of the International Statistical Institute; an Honorary Fellow of the Royal Statistical Society; an Honorary Life Member of the Biometric Society; a Fellow of the Indian National Science Academy and Indian Academy of Sciences; an Honorary Fellow of King's College, Cambridge; and a Founder Fellow of the Third World Academy of Sciences, Trieste. He has received honorary degrees from Andhra University, India, 1967; Leningrad University, U.S.S.R., 1970; Delhi University, India, 1973; Athens University, Greece, 1976; Osmania University, India, 1977; Ohio State University, U.S.A., 1979; Universidad Nacional de San Marcos, Peru, 1982; University of the Philippines, 1983; and University of Tampere, Finland, 1985.

The following conversation took place in his office at the University of Pittsburgh one morning in November 1985.

"THERE WERE NOT MANY OPPORTUNITIES FOR THOSE WITH A DEGREE IN MATHEMATICS TO GET A JOB"

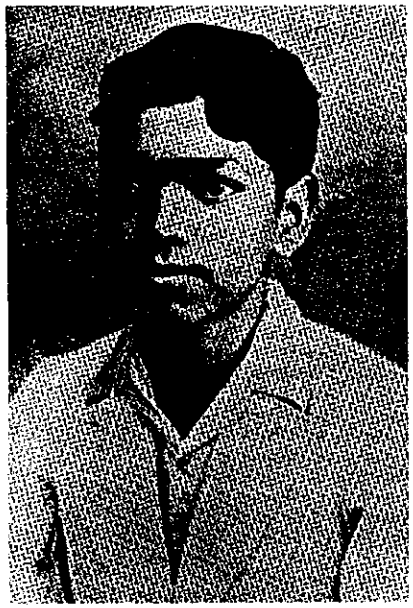
DeGroot: To start at the beginning, tell me a little about your childhood in India, what your home life was like, and how you came to get interested in statistics.

Rao: Well, I come from a family of a kind of landed aristocracy. They had a lot of property and never cared to go and study and try to make a living on the knowledge that they acquired in their studies. However, at one stage they were interested in the legal profession and they all became lawyers and judges and so on.

DeGroot: Who do you mean by "all"?

Rao: My ancestors, my relations. I was, so to say, the first one to get interested in science. I started off with mathematics, but then the war broke out when I was just finishing my master's degree in mathematics and there were not many opportunities for those with a degree in mathematics to get a job. So I tried to get a job in the military service. I applied to the Department of Survey, which is sort of a party to any military expedition. I was disqualified because I was too young and they didn't want me, but that took me to a place called Calcutta. I belonged to the South but I had gone to Calcutta for the interview to get into the military service. At that time I came to know of the Indian Statistical Institute which had been established by Mahalanobis, so I just casually visited the place and talked to people there. They said that statistics was a new subject and if there are no opportunities for a job with a mathematics background, why not come and study statistics.

So I applied for a course of studies at the Indian Statistical Institute and I was admitted. At that time there were no courses in statistics in the universities.



C. R. Rao at age 15

Some universities used to teach statistics as a part of the undergraduate program but not at the postgraduate level. The Indian Statistical Institute was the first place to have what they called a one-year training course in statistics. I took that course, but by the time the course ended Calcutta University had started a master's program in statistics. Mahalanobis said to me, "You join the university and get a formal degree." Since a formal degree is always useful in India as a passport to get jobs, I joined Calcutta University and got a master's degree in statistics. I am one of five first batch students to be awarded an M.S. degree in statistics by an Indian University. So that's how it all began.

"FISHER SAID I MUST WORK ON HIS MICE"

DeGroot: When did you then go to Cambridge and how did that come about?

Rao: After I finished my master's degree, Mahalanobis asked me to work on anthropometric data which had been collected by Dr. D. N. Mazumdar, a well-known anthropologist, in the state of Uttar Pradesh during the 1941 Indian population census. The data were sent to the Indian Statistical Institute for analysis; at that time Mahalanobis had already invented the Mahalanobis distance for studies in anthropology. He assigned the project to me. So my first experience was analyzing data collected on some five thousand individuals, with twenty characters measured on each individual. I produced a report using the Mahalanobis D^2 (distance) in the analysis, and Mahalanobis was very happy with that.

Right at the time when I finished that report, he got a telegram from the anthropological museum in Cambridge saying that a British expedition from Africa had brought a large number of bones and stones and so on. Would he please send one of his students to analyze the data. Mahalanobis replied, "Of course. I have one who is very good in statistics, but he does not know anthropology. So I'll send an anthropologist along with Rao to help you with this project." So we went there and we found a very interesting collection of bones. They had dug out ancient graves, about 1000 years old, from a place called Jebel Moya in North Africa, and brought the skeletal material to the museum in Cambridge. I was actually employed in the Cambridge museum for a couple of years to work on the skeletal material. Incidentally I joined King's College, where Mahalanobis used to be when he was a student, and I also registered for a Ph.D. degree under R. A. Fisher. But Fisher said I must work on his mice, whether it leads to a thesis or not. So I was doing some experiments on mice for R. A. Fisher which involved the mapping of chromosomes, in addition to my work at the Anthropological Museum.

DeGroot: He actually was treating live mice in his experiments?

Rao: Oh, yes. He had a laboratory where he was breeding mice because he was mapping their chromosomes. Lots of mice, maybe thousands. I think we were some ten people working there at that time. He asked me to study the linkage of four genes on a chromosome and find the distances between them. So I was doing that project for R. A. Fisher in the evenings, and in the daytime I was working in the anthropological museum. After two years we wrote the report on this skeletal material which is published as a book by the Cambridge University Press, *Ancient Inhabitants of Jebel Moya* [with R. K. Mukherji and J. C. Trevor, 1955]. The Cambridge anthropologist and my Indian colleague who was an anthropologist also got their Ph.D. theses out of that same project. I think my Indian colleague wrote on the anthropological technique and interpretation of results. The Cambridge anthropologist was also a sociologist, so he looked at the artifacts along with the bones. He made his own conclusions and we put everything together.

DeGroot: And your Ph.D. dissertation?

Rao: My Ph.D. dissertation was based on the theoretical work I did in connection with this project, which is partly in the book. But I had some additional results in my thesis. Fisher's discriminant function was used for classification within two categories, and I developed a method for many categories. I also developed another test. We had to do all the multivariate computations by hand at that time. If I had fifty characters it would be impossible to use all of them, so selection of characters was a major problem. I invented a test of whether it is really necessary to take so many characters. I said, "All right, given a set of measurements, is there further information in an additional set of measurements?" So this is the kind of test which I developed and which went into the thesis. R. A. Fisher was my supervisor. Wishart was an examiner.

"NOW I HAVE THE PRIVILEGE OF WALKING ON THE LAWNS"

DeGroot: Was Fisher much involved in your work?

Rao: Well, he thought my test for additional information was very good because it was an extension of a test he developed in discriminant analysis. That is, if somebody specified the discriminant function, he could examine whether that was the true discriminant function or not. This really involved the idea that the given discriminant function carries all the information; given that function, nothing else is needed. So when I extended that test to a given set of measurements, not necessarily one discriminant function, Fisher was happy. I also mentioned to Fisher that I was working on the problem of classification into more than two categories. I used the Bayesian approach to the problem. He asked me to work along a different line, but I used essentially Bayesian techniques. [Laughs] That's the appropriate thing to do in problems of that kind. Later on I described what you should do when you could not know the prior probabilities. I also created a kind of doubtful region where you can take the position that you are unable to say whether an observation belongs to this group or that group, and described how to operate with that region. That is an abstract of what I did for my thesis.

DeGroot: With both Fisher and Wishart at Cambridge at that time, there must have been a lot of statistical activity around the university.

Rao: That is true, but Fisher was mostly interested in mathematical genetics. So I took various courses in mathematical genetics. I had enough knowledge when I came back to India to direct research work by my students in mathematical genetics. I actually produced three Ph.D.'s in mathematical genetics at that time, very valuable people, but unfortunately they are all working in this country. And Wishart of course was in the school of agriculture. Bartlett was there for a while; then he went to Manchester and Daniels came. About that time the statistical laboratory in Cambridge was established for training statisticians. We had students like Durbin and Bailey and some others. Most of the statisticians at that time were being trained in the statistical laboratory at Cambridge, but it just started in '47 or '48.

DeGroot: Was it unusual for an Indian to be studying at Cambridge at that time?

Rao: No, lots of Indians used to go to Cambridge. It was the place that Indians went for their higher education. But I didn't study in Cambridge in that sense. I was working in Cambridge at the museum, but formally attached to King's College as a research scholar. Later on, King's College honored me by electing me a life fellow. When I was formally enrolled as a research student there, I was not allowed to walk on the lawns in King's College.

DeGroot: I've seen those signs: "Only fellows are allowed to walk on the grass."

Rao: Right. But now I have the privilege of walking on the lawns, and whenever I go there I can stay in the college, have a free dinner, and probably drink as much wine as I want. [Laughs] There are only eleven life fellows at any time. Only if somebody dies do they elect another one.

"ONE STUDENT ASKED, 'WHY DON'T YOU PROVE IT FOR FINITE SAMPLES?'"

DeGroot: Some of the most famous results in statistics have your name on them. One example is the Cramér-Rao inequality, as it is commonly known. I know that Cramér and you did not work on this together. Could you tell me how it came about that your names are paired on this result in this way?

Rao: After I completed my master's degree in 1943, I was asked to teach my juniors in the master's course. There was a shortage of teachers. I think R. C. Bose and S. N. Roy were the two main teachers, and there were a few others.

DeGroot: This was at the Indian Statistical Institute?

Rao: At both the Institute and Calcutta University. You see, when Calcutta University started the master's program they had no teachers, so they borrowed all the teachers from the Indian Statistical Institute. Actually, the lectures were held in the Indian Statistical Institute. Many of our teachers did not know much statistics. They were trying to learn and teach the students. So when I had my formal

C. R. Rao and Harald Cramér, 1978

education, they thought I would be a good teacher and I was asked to teach. I was giving a course on estimation covering all the large-sample properties of maximum likelihood. I proved the asymptotic inequality of R. A. Fisher that the asymptotic variance is not smaller than one divided by the information. Then one student asked, "Why don't you prove it for finite samples?" So I went back home and worked all night. The next day I came up with this inequality, using unbiasedness and the Schwarz inequality. I proved this result in the class in January 1944. Once I started all this, I looked at sufficient statistics. Of course the maximum likelihood estimate is a function of sufficient statistics. Is there an independent small sample justification for sufficient statistics? I thought I must be able to prove some exact properties, and that's how the Rao-Blackwell theorem came up. Which I did two years before Blackwell.

DeGroot: So your work on that type of improvement of an estimator really grew out of your work on inequalities.

Rao: Yes, they are all in the same paper. ["Information and accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.* **37** (1945) 81–91.] That's in 1945. Blackwell's work was in 1947. I saw Cramér's work on the information inequality when his book on *Mathematical Methods of Statistics* came out in 1946. That is sort of a basic paper mostly arising out of my teaching that course on estimation. I also developed the differential geometric method. I tried to connect Fisher information with a metric in a Riemann space. So my differential geometric approach to estimation is also in the same paper.

DeGroot: That's quite a paper. This differential geometry approach to estimation is now becoming of interest to many statisticians.

Rao: Yes. It's so difficult, nobody liked to work on problems of that kind. I am gratified to see that the geodesic distance I defined using the Fisher information metric is now referred to as Rao distance. Now to complete the story, Fisher came out with second-order properties of a maximum likelihood estimator. That's a very difficult concept and Fisher's investigation lacked some details. Since very difficult computations were involved I tried to modify his definition of second-order efficiency and came out with a quantity which characterizes maximum likelihood estimators in a sense if you go to second-order terms by expanding the asymptotic variance. In the second-order terms, the maximum likelihood estimator has the smallest coefficient. Now that coefficient has been termed the Fisher-Rao inequality by Efron.

DeGroot: That notion of second-order efficiency is very closely associated with you.

Rao: It really started with Fisher. What I did was try to correct Fisher's computations, and also modify his criterion of the minimum loss of information because it is very difficult to work with. That's an exact property which does not really hold, but asymptotically to the second order that idea works.

"FISHER WAS CRITICAL ABOUT NEYMAN"

DeGroot: Who do you feel have been the major influences both on your life and on your career?

Rao: I would say Professor Mahalanobis in India. His whole idea was that statistics must have a purpose. So he said, don't try to solve problems in mathematical statistics. First look at the data and see what the problems are, and then try to develop the appropriate methodology for that kind of data. The next is, of course, R. A. Fisher. When I went to work with him, I first told him, "I have studied mathematical statistics; I have also done some research. Would you give me a problem that I could work for my thesis?"—as students do in this country. Fisher said, "No, the problem must be yours. I can help you to solve the problem." [Laughs]

Also, I wrote a paper in mathematical genetics and said that the method I developed was superior to one that somebody else gave, but Fisher said, "I won't read your paper unless you do some computations. Gather the data and compare what your method gives with what the other method gives, and then come with a paper and I shall read it." This is the kind of advice which I think was very useful later in my consultation work and also in my own research in mathematical statistics. I always develop methodology from the data given to me for analysis rather than look at other's work and try to extend it in terms of mathematics.

I used to meet Wishart frequently when I was in Cambridge because we used to discuss lots of problems. That was also useful, but the two people who provided inspiration for me are R. A. Fisher and Mahalanobis. Of course, at that time Neyman's work was becoming popular and everybody had to read that. So I read a lot of work that Neyman did and I think he was also responsible for my getting into the current research at that time in testing of hypotheses.

DeGroot: Did Fisher ever discuss Neyman's work with you?

Rao: Yes. He was critical about Neyman. Whenever I had discussions and wanted to use Neyman's ideas in solving some of the problems, Fisher always discouraged me from doing that. Actually, I now feel that I could have done some work which is currently popular if I had used Neyman's ideas in those days. I was not reluctant to use Neyman's ideas, but I did not probably because of Fisher's influence.

DeGroot: Was Fisher generally pretty good about reading things that you would give him?

Rao: Oh, yes. He would take great interest. And then he would read and comment on it. He was also a very exciting teacher. Not many could follow what he was teaching; but if one tried to see what he was saying, he was an inspiring teacher. I also used to attend his seminars and they were very good. But most of the time when I was at Cambridge I used to help him in his genetic experiments.

"LINNIK SUGGESTED THAT WE ALL WRITE A BOOK TOGETHER ON CHARACTERIZATION PROBLEMS"

DeGroot: You've published, according to my count, nine books and about 230 papers. Your book *Linear Statistical Inference and Its Applications* is of course widely known. [Second Edition, John Wiley, New York, 1973] I've always felt that the title was something of a misnomer because that book really covers much more than linear statistical inference. In fact, it covers much of what is in a standard graduate course in statistical theory.

Rao: Actually, I had a different title but the referees suggested this title because it would compete better with other books with similar titles at that time. So I said, "OK. From the commercial point of view, if that title will sell very well then you may choose that title." [Laughs] It had a different title to begin with.

DeGroot: What was your original title?

Rao: Advanced Statistics or something like that. Actually, when I wrote my first book and called it *Advanced Statistical Methods in Biometric Research* [First Edition, John Wiley, New York, 1952; Reprint, Haffner, New York, 1974], I really summarized what I did in Cambridge on the anthropological material. Most of the examples there are all from anthropometric data. People wondered how I suddenly introduced data on head length and head breadth and so on in between theoretical discussions. But the major emphasis in that book is the applications of statistical methodology to univariate and multivariate data. It's slightly more theoretical than R. A. Fisher's book, but because the title said biometric methods the statisticians were not attracted to it.

DeGroot: Yes, that's another book whose title is much narrower than the content of the book—where you cover much more material.

Rao: Well, I thought I was discussing some biological problems and so the title was appropriate.

DeGroot: I remember using that book when I was a student just for learning about statistical theory generally.

Rao: My *Linear Statistical Inference* is really only a mathematical version of that book, I think.

DeGroot: I see. Well, have you been giving any thought to a new edition of *Linear Statistical Inference*?

Rao: When I try to review the book, I find that much has happened since the second edition came out, so I was thinking of working on a third edition not as a single volume but maybe as a couple of volumes. It depends on how much time I have because in this country there is a lot of pressure for writing papers, getting grants, writing more papers for the grants, and so on. [Laughs] So I don't have much time to think of revising the book. But I have given a variety of courses in this country and made some notes which would be useful in making a third edition of the book.

DeGroot: I'm also curious about your book with Kagan and Linnik, *Characterization Problems of Mathematical Statistics* [John Wiley, New York, 1973]. That book was published first in Russia and I first saw it in the Russian language edition when Kagan sent me a copy. Then subsequently it was translated into English. How did that collaboration with the Russians come about?

Rao: I think I was interested in characterization problems right from my student days. There was a paper by Ragnar Frisch in which the following problem was proposed: If there are two random variables X and Y where $X = \psi + \eta$ and $Y = \psi + v$, so there is a common variable ψ, under what condition is the regression of X on Y linear? X and Y have only one common variable ψ; the other variables η and v are independent. I gave a general solution to this problem in my thesis as a partial requirement for the M.S. degree. This was ultimately published in *Econometrica* in 1947 [**15** 245–249; Correction, **17** 212].

Then when I had my research students, I gave them problems on characterization for their thesis work. I was doing some work on characterization right before

C. R. Rao, A. N. Kolmogorov, and P. C. Mahalanobis, Calcutta, 1962

Linnik's first visit to India in 1955. I told him what I was doing, and being a mathematician he liked the idea. When he went back he started working on very complicated characterization problems using ideas of number theory and so on. It was very surprising. I was visiting the Soviet Union frequently at that time and we used to meet. I attended the first mathematics conference held in the Soviet Union after the war in 1956. At that time I met other acquaintances of Linnik. Linnik introduced me to his student Kagan subsequently. As I said, at that time Linnik used number theory ideas and he was trying to characterize the normal distribution by the identical distribution of two linear statistics. Then I took Linnik's ideas and tried to extend his methods to problems involving linear statistics but under different conditions and so on. During a subsequent meeting with Linnik, he suggested that we all write a book together on characterization problems in mathematical statistics. We decided how many chapters I would write and what Kagan and Linnik would write. I think within a year or so, we put our material together. I wrote in English and they translated it into Russian. They kept me informed of what they were doing.

DeGroot: So you would communicate with each other in English.

Rao: Yes. Linnik knew English and Kagan knows English. We formed a good team and were able to complete the work within a year or so.

"YOU CAN SAY ANYTHING IN AN ABSTRACT"

DeGroot: Linnik must have died just about the time that the book came out.

Rao: Yes, he died in June '72. Linnik and Kagan also translated my *Linear Statistical Inference*.

DeGroot: That's very interesting. They were also influential in translating my decision theory book into Russian.

Rao: There's one interesting thing I must tell you about the translation of my *Linear Statistical Inference*. I had made the comment that in the Fisher-Behrens problem, there is no similar-region test when σ_1 and σ_2 are different. Linnik had proved that in a certain class a similar-region test actually exists. So Linnik said my statement was wrong and they just modified it in the Russian translation. In this connection, when Wilks was alive I discovered that he had an abstract in *The Annals of Mathematical Statistics* where he claimed that no similar-region test exists. So I wrote to Wilks in the 50s or so, asking him to please send me the proof or tell me how he had done it. Wilks replied that this was done during the wartime, that his office was moved after the war and the whole folder containing the proof of this proposition was lost in the process, and that he does not recall how he did it. You can look at the abstract by Wilks [**11** 475–476, 1940.]

DeGroot: Wow!

Rao: There was another incident recently in which somebody claimed priority because he had mentioned a result slightly less general than mine in an abstract in the *Annals*. You can say anything in an abstract. If it is right, you can claim credit and priority.

DeGroot: Yes. Take a chance; maybe it will be right. There is no serious screening of abstracts. I think that's OK, as long as everyone realizes that the results are not necessarily correct or original.

Rao: Actually when that person wrote the full paper on the basis of the abstract, I was a referee and it turned out that this result was also not correct as stated.

DeGroot: Do you still visit the Soviet Union?

Rao: I was there last year for the conference at Vilnius. As Director of the Indian Statistical Institute, I was very much involved in the exchange of scientists between the Soviet Academy and the Indian Statistical Institute. In fact, we had a separate agreement between us for the exchange of up to eight scientists every year on either side. So we send our people and they still come to the Institute.

DeGroot: What is your impression of the political situation in the Soviet Union? Have you seen any changes through the years?

Rao: Well, there was some change since the fifties. I think that in matters of science they now just do the same thing as the rest of the world.

DeGroot: Do you see Kagan when you are there? I know he has been trying to emigrate for several years.

Rao: Not this time. He was not at the conference that I attended. The last I saw him was in 1976. But I am in correspondence with him. We still try to collaborate a bit. I used to like the Leningrad group because whenever I was there I would give seminars and they would tell me what they were doing. I used to give them some problems and they would solve them immediately. [Laughs] So there were some papers on problems posed by Rao. It was quite a good group. Ibragimov and others were there. I am an honorary professor of Leningrad University. That came along with the honorary doctorate I got from there. So I always visit the university when I go there.

DeGroot: Do you have royalties in the Soviet Union?

Rao: Yes, I have royalties for the translation of *Linear Statistical Inference* and for *Characterization Problems*.

DeGroot: I found it difficult to spend my royalties when I was there. How do you spend yours?

Rao: Generally when I go there I just call my friends and give them a party. [Laughs] I still have unspent money there because when I go as a visitor the Soviet Academy pays all my expenses. So this money is in the bank. By the way, they give interest on money. I put the money in the bank and the second time I went to find out how much money I had, I found some amount of money added in red ink. They said that was my interest.

"IF WE HAVE GOOD VISITORS, ONE CAN WORK FASTER"

DeGroot: What are your favorite publications among your many books and papers? Do you have any particular favorites that you enjoyed doing, and which ones do you think have been the most influential?

Rao: I enjoyed doing my applied work, especially the applications in anthropology, but that work gave rise to the development of methodology in multivariate analysis. I think the papers where the work I have done was followed up by many others are my early papers on estimation. Then there are papers on second-order efficiency, and now the papers on differential geometry that are coming up. So I am happy because somebody is following up the work I initiated. A second set of papers I like are mostly in the analysis of repeated measurements and in singular linear models, i.e., when the design and covariance matrices are deficient in rank. I developed generalized inverses of matrices for dealing with such problems. My work on orthogonal arrays is widely used in industrial experimentation, and has led to important inequalities in coding theory. I found that special courses on orthogonal arrays are given in this country. In 1947 (*Proc. Cambridge Philos. Soc.* 44 50–57) I introduced two general asymptotic test criteria called score tests for simple and composite hypotheses as alternatives to Wald's tests. I find that my score test for composite hypotheses has become entrenched in the econometrics literature under a fancier name, the Langrangian Multiplier Test. So those are a few papers which I like and which have received some attention.

DeGroot: What are your current research interests?

Rao: I am working on two types of problems: One is prediction, not in time series but in repeated measurements like growth models. I think I have done some fairly useful work for applications in practice. And now I am working on what is really discrimination but what electrical engineers call signal detection. Problems like when one can recognize that a real signal is coming and not just pure noise.

DeGroot: That's a classical problem.

Rao: Yes, that's right. But there are some new problems that arise in this connection which we are currently solving. There is also another problem. For instance, in a discrimination problem we essentially estimate a discriminant function on the basis of data, and then we use this for discriminating the future observations. So the real performance of the estimated discriminant function has not been properly studied in the past, because when they studied the properties of the discriminant function they always studied them on the average; that is, the properties if you would go on estimating the discriminant function each time with fresh data, and then apply these estimates. But in a real problem you get stuck with a particular estimate of the discriminant function for use in discriminating future observations.

DeGroot: You really want the properties conditionally on the given estimate.

Rao: Yes, and to know how it's going to perform in the future—how we can estimate its performance from the available data itself.

DeGroot: Sounds Bayesian to me.

Rao: Yes. So that's the new type of problem which we are trying to work on, and already we have written some papers on that subject.

DeGroot: Who are the "we"?

Rao: Visitors who come here who are assigned to the project on signal detection that we have. So I am doing this work in collaboration with visitors because that way we can work faster. Of course, the problems are conceived and some groundwork is done before the visitors come. To write, say, two papers in a month or two months is not unusual although it is strenuous. Probably one writes a paper once in four months or six months in the normal course. But if we have good visitors who are good in mathematics and have similar interests, then one can work faster.

"THE PRIME MINISTER USED TO LIKE US"

DeGroot: How did you get to know Nehru and Indira Gandhi?

Rao: Well, statistics was originally with the Prime Minister's office; that is, all decisions about the Indian Statistical Institute and the development of statistics were taken directly by the Prime Minister for a long time. It was not under any particular ministry. Probably no ministry wanted statistics. So whenever we wanted to discuss problems or ask for fresh grants or make proposals for the development of statistics in India, we used to go to the Prime Minister. Jawaharlal Nehru was the Prime Minister at that time. So we met him very often to discuss problems of statistics and he was very helpful. Actually, Jawaharlal Nehru was responsible for recognizing the Indian Statistical Institute as an institute of national importance. He moved a bill in the parliament declaring the Indian Statistical Institute an institute of national

C. R. Rao and Indira Gandhi, 1967

importance and giving it a charter to award degrees. Formerly we had been training statisticians, giving them courses both in the theory and the applications of statistics, but we were not allowed to give degrees. We used to give a diploma, but after this bill was passed in the parliament we started giving B.S., M.S., and Ph.D. degrees in statistics.

DeGroot: When was that?

Rao: Oh, this was in '60 or '61. Before that, when we did not have the charter to give degrees, all of our students who were doing research with us had to submit their theses to some university in India, either Calcutta University or Bombay University. We were recognized as supervisors for theses, but the Institute could not give them degrees. The Prime Minister used to like us, and he gave us a lot of support by giving us as much money as we wanted for educational, research, and consulting purposes. And then after Nehru died, Indira Gandhi took his position and we continued our association with her. She knew Mahalanobis and the Indian Statistical Institute he founded because of her father's relationship with the Institute and the help he had given to the Institute.

DeGroot: Have you had other relationships with the Indian government?

Rao: Actually, the Indian Statistical Institute was responsible for drafting one of the five-year plans.

DeGroot: The economic plans?

Rao: Yes. Much of the second plan was drafted by the Indian Statistical Institute under the guidance of Mahalanobis.

DeGroot: When was that?

Rao: 1954 or so. So we continued our association with the planning commission, and the Prime Minister is the chairman of the planning commission. There are a large number of projects which we did for the planning commission. As a matter of fact, the Indian Statistical Institute branch at Delhi was located in the planning commission for a long time. It's only recently, when we built a new campus for the Indian Statistical Institute which was opened by Indira Gandhi, that we moved from the planning commission to our own premises. So we were very much attached to the government because of our involvement in working on projects for the planning commission. And I used to work on some committees on statistics formed by the Prime Minister.

DeGroot: Do you know the present Prime Minister?

Rao: I don't know him because he was not in

politics when we were in contact with the Prime Minister's office. Occasionally we saw his brother, Sanjay, but not Rajiv.

"THE WORKERS HAD WRONG IDEAS ABOUT COMPUTERS"

DeGroot: I'm sure you've had many interesting and fascinating experiences during your time at the Indian Statistical Institute. I remember hearing about some of the problems that you had when you took over as director in Calcutta. Could you tell me a little bit about your experiences during that period?

Rao: The Indian Statistical Institute is a very large organization. It employs 3000 people. There are lots of divisions engaged in working on different kinds of problems. Our idea was that everybody must use statistical ideas in their investigations although they are working on their own scientific problems. So that was a difficult job, to keep in touch with the various scientists working on different kinds of problems, to see to what extent they were using statistical methods in their investigations and to give them proper advice. There was also the problem of bringing scientists working in different disciplines together to work on some kind of common projects. For instance, when I sent a team for a demographic survey, I included some physical anthropologists who would go and do the measurements, sociologists to do studies on families and relationships, and seriologists to take blood samples and do the blood grouping and so on. So this was the most difficult problem and I succeeded in most cases and failed in other cases.

We had some trouble when we brought in the computers. The Indian Statistical Institute was the first organization in India to acquire digital computers. We had one computer brought from the Soviet Union in '55 and another digital computer brought from Great Britain. But the workers in India did not like our introduction of computers in a big way. They had wrong ideas about computers. They thought it would affect their employment opportunities.

DeGroot: Are the workers unionized?

Rao: Oh yes, we have very strong unions in every organization in India. Let me continue. I had developed an interest in computers when I was working in Urbana-Champaign at the University of Illinois in 1953–1954. I think that university had the first digital computer, called Illiac, and I used to work on that computer. I took a course on programming. So I was probably the first programmer of India. I was one of the few who could do programming using machine language. The University of Illinois gave me two students to develop computer programs for statistical methods at that time. So I made them work, and in order to understand those programs I learned computer programming myself. Nowadays it's easy, but at that time to use sixteen instructions to write the whole formula was a difficult problem. So I had a lot of interest in computers and I really wanted to develop computers in India, which was not possible under the conditions prevailing in India. That was a big disappointment for me. We lost the opportunity to do certain kinds of research or develop methodology based on complex computations.

DeGroot: I seem to remember that at about that time you moved from Calcutta to Delhi.

Rao: No, I moved from Calcutta to Delhi much later.

DeGroot: It was unrelated?

Rao: It was unrelated. Because I was not finding much time for doing research in Calcutta, I moved over to Delhi and asked somebody to share the administrative responsibilities in Calcutta. It was purely to get time to do research that I moved to Delhi.

"THE INDIAN STATISTICAL INSTITUTE IS REALLY A FANTASTIC PLACE"

DeGroot: What are some of your more pleasant recollections of your tenure there?

Rao: The Indian Statistical Institute is really a fantastic place. It used to attract famous scientists from all over the world. For instance, Norbert Wiener worked at the Institute for six months in, I think, '52 or '53. And J. B. S. Haldane was a regular employee of the Indian Statistical Institute. He was there for five years, and I was formally head of the department where Haldane was working. Mrs. Haldane also used to work there. A number of famous economists visited the Indian Statistical Institute: Ragnar Frisch, Simon Kuznets, Richard Stone, who got the Nobel Prize in Economics last year; and some of the presidential advisors in economics from the U.S.A. That tall economist, John Kenneth Galbraith, spent three months at the Indian Statistical Institute. On occasions I used to go for a walk with him. He is so tall I had to look up in order to carry on a conversation with him. Oscar Lange from Poland also visited the Indian Statistical Institute. And the Indian Statistical Institute was declared a show piece which every visitor to Calcutta must visit. So Kissinger came to the Indian Statistical Institute. I have nice photos shaking hands with Kissinger and Mrs. Kissinger, and taking him around the Institute and explaining to him what we are doing. Premiers of all countries would come—Ho Chi Minh, Chou En-lai. It was quite an interesting place. And lots of scientists; at any point of time there would be ten or twelve foreign scientists visiting the Institute.

DeGroot: I remember your telling me that it's the only statistical institute in the world that has a dinosaur in it.

Rao: Oh yes, there is a geological department, and dinosaurs were discovered for the first time in India by the Indian Statistical Institute. We have a geological museum. By the way, Jawaharlal Nehru was a geologist. He studied geology when he was a student at Cambridge, so he had a lot of interest in geology. He was very much excited when we found the dinosaur. And so when the first bone, the dinosaur's thigh bone, was discovered somewhere in South India we took the dinosaur bone all the way to Delhi to show it to Nehru.

DeGroot: There seems to be a tradition in India of mathematically-talented young people going into statistics rather than into other branches of mathematics proper. Is that a correct impression, and if so, how do you think that came about?

Rao: I think that was true only for some time when there were no jobs for those who studied mathematics. They used to study mathematics just for the love of mathematics. Generally, Indians are more abstract minded; they are not practical-oriented people. That's the reason why Indians did not contribute much to science in the past. But physics, metaphysics, philosophy, mathematics, they just love, because they think that anything can be proved by argument. [Laughs] Of course, in mathematics there are axioms, and therefore argument is relevant. But not in areas where there are no fundamental axioms. But when they found that there were no good jobs in mathematics and that statistics was a developing field, many mathematicians thought it logical to study statistics! And that's the reason why a lot of good statisticians developed in the late '40s and '50s in India.

DeGroot: Is the situation still the same?

Rao: No, nowadays we are not getting such good students for statistics as we used to get in the past. Probably there are areas of applied physics and chemistry which are more attractive now. Some students still go for mathematics. So it was only a temporary phenomenon that all good mathematicians came to statistics. That was true in this country, too.

"I THOUGHT I SHOULD VISIT THE STATES FOR JUST ONE YEAR"

DeGroot: What prompted you to accept a faculty position in the United States here at the University of Pittsburgh in 1979? You had, of course, visited the United States often.

Rao: Yes, several times. And I also worked briefly at various universities in the United States. Actually, I had retired from the Indian Statistical Institute, but they gave me another position—a special chair called the Jawaharlal Nehru Professorship.

DeGroot: That chair was awarded upon your retirement as director?

Rao: Yes. It was approved by the Prime Minister, Indira Gandhi. But after retirement I thought I should visit the States for just one year. So I took a visiting professorship at the Ohio State University, Columbus. At that time my son, who was studying engineering in India, came to the United States to visit me. I was briefly in Pittsburgh, and I think he met the Dean and the Dean said that he could continue his engineering courses here. He already had done three years in India. So he met the engineering faculty; they gave him credit for what he had done and told him that if he studied for another two years here, he would get the same kind of degree, a Bachelor's in Engineering, that he would get in India. So he accepted admission here and did not go back. A few days later the Dean sent word through the chairman of the Math Department here asking whether I would take a job here. One of the attractions was that I wouldn't have to pay tuition for my son if I took the job. So I said maybe, temporarily for two years, I will accept, because I have my job in India; I have to get back. He said, "You take the job. I'll give it to you on a permanent basis. You can go back whenever you like, or you can work part of the time here and part of the time in India."

DeGroot: That sounds like an offer you couldn't refuse. Have you been dividing your time going back and forth?

Rao: Not much. I have been spending most of my time in Pittsburgh because this department itself needed some development. I think that when I joined they didn't have many graduate students in statistics. I even heard that they were going to close down their graduate studies in statistics. But over the course of the past five years we tried to develop advanced courses in statistics, and to increase the number of graduate students, and also to write some papers. I think the department is now very well rated.

"I USED TO WATCH WHAT THE GARDENERS WERE DOING AND DIRECT THEIR WORK"

DeGroot: During the years that you have been in Pittsburgh, I've come to learn of your many talents outside of statistics such as gardening, photography, and cooking. Let's take these one at a time. Would you tell me about your interest in gardening?

Rao: Well, I've always loved gardening. I used to work in my garden in India. The Indian Statistical Institute has a big campus and there's a lot of area for gardening. I think they have about 50 gardeners working there. So I used to go around early in the morning to watch what they are doing and to direct their work. I used to advise them on the type of flower plants they should grow, give them a design for planting trees, and indicate where they should grow vegetables. So I always took an interest in gardening. But I never

allowed anybody to cut the flowers from the plants, so everybody had a grudge against me. [Laughs] They used to come stealthily and cut some flowers in the evenings. But I used to go down to the garden several times a day—in the morning, at lunch time, and in the evening. So when I came here I found a small space, but it is too small a place for doing gardening compared to what I had at the Indian Statistical Institute.

DeGroot: A small space in your backyard here?

Rao: Yes. I tried to grow some exotic vegetables from India, oriental vegetables. It was very successful, and I think I had better crops here than I had in India. That is because I used fertilizers here. We don't have good fertilizers in India; I used farmyard manure there.

DeGroot: Do you perform statistical experiments in your garden and analyze the data?

Rao: Yes, I used to. Here also I tried one experiment, but in India I generally experimented. For instance, one of the experiments which I was doing in collaboration with my colleague T. A. Davis at the time of my retirement was the following: The cowpea is a creeper. It is a plant which twines on the right when it grows. That is, it grows up around a pole in a right-spiral. At the Indian Statistical Institute, Davis had already done some work on the right-handed and left-handed palms. Some palms grow in a right-spiral and some grow the other way, i.e., the positions of successive leaves as a palm grows form a right-handed or a left-handed spiral. It's a very nice phenomenon. You know how a flower unfolds. For some flowers the petal to the right will be below the petal to its left, but in some flowers it's the other way, the petal to the right is above the petal to the left. That shows that some flowers unfold in a clockwise direction and some flowers unfold the other way. So there are left flowers and right flowers. We were very much interested in knowing what causes this difference among flowers. So we used to count. We had a team of students who used to go and count how many flowers on a plant are of the left and right kinds.

DeGroot: On the same plant?

Rao: On the same plant; even on the same branch. So there is some random mechanism going on which causes this. Davis made these observations on thousands and thousands of flowers and also on palm trees for left-right spirality. We also studied the yields of these palm trees, and we found that the left-handed palms gave greater yield than the right-handed palms.

So we said kill all the rightists. By selective plantation, choosing only the left-handed palms from the nursery and growing them, you can possibly increase the yield by 5 to 10%. This creeper, the cowpea, gave us a lot of trouble. It always used to grow on the right side. We grew a large number of plants and found that all of them were growing in the right-spiral. And I was told there are some creepers which always grow in the left-spiral. So we used to go early in the morning and pull the growing tips from the right to the left direction and tie them with ropes to the poles to force the plants to change their habit. Sometimes we used to stretch the growing top straight and tie it up and see the next morning where it had grown. But it always used to have a right tendency. We tried to measure the yields of the cowpeas that were forced to go left but unfortunately some damage might have been caused to the plants because we were handling them. So we did not think that the data we collected gave any evidence that forcibly making a plant grow on the left increased the yield. [Laughs] I had a garden where we used to do experiments of this kind right in front of the place where I was staying. It is quite exciting to do things like this and to get students interested in this phenomenon.

"WE USED TO DO ALL THE COOKING FOR DAYS AND DAYS"

DeGroot: What about your interest in photography?

Rao: Oh, I have some interest in photography. Some of my pictures are very good; I used to win some prizes by submitting them in competitions.

DeGroot: Do you just take pictures where you find them?

Rao: Yes, wherever. I don't really have time to especially go on photographic expeditions. When I just go out and feel like taking my camera, I take pictures. But not on any planned basis. If I had done that, maybe I would have made a lot more pictures; but I never did.

DeGroot: Do you usually take your camera with you when you travel?

Rao: Not generally, but sometimes I take it. It depends on my mood. I also took a lot of movies when I was young. But it's a very expensive hobby. I could do that only when I came to this country and gave a lecture and earned $50 or $100. [Laughs] Of course in India I could never take pictures because we don't have the proper film; and developing and printing costs are very high. And I didn't have a good camera. Well, I took all my good pictures with very cheap cameras costing five to ten dollars. We really don't have opportunities for this kind of hobby back in India.

DeGroot: So this is more of an American development with you?

Rao: I don't know the current statistics; an average American takes a thousand pictures a year or something like that. [Laughs] That's how Kodak flourishes.

C. R. Rao at the University of Tampere after receiving an honorary doctorate, May 1985

But I am also interested in cultural activities like dance and music. We put our daughter in dance school very early because of our own interest in dancing.

DeGroot: I've seen her dance and she's an excellent dancer.

Rao: Yes, she's a very good dancer. She runs a dance school in Buffalo.

DeGroot: Does she do any Western dancing?

Rao: She does a little bit of Western dancing but mainly she does Indian classical dancing. I developed a new dance academy in Delhi and was its president. We covered some neglected areas of dancing.

DeGroot: So you are quite knowledgable about Indian classical dance and Indian classical music.

Rao: More dance than music, because dance also has the visual effect.

DeGroot: What about Indian classical cooking?

Rao: Well, traditionally I used to like cooking because I used to help my mother cook. My mother also had other interests and she used to visit other places. At that time there were only boys and my father in the house, and we used to do all the cooking for days and days. My wife comes from a family where there was only one girl and five brothers, so she was very much pampered and did not know cooking. So all the cooking she now knows, she learnt from me. [Laughs] My son is a very good cook. So it seems to me that there's some gene in the family for this culinary art. I think basically that cooking is an art, so I think that it's an interest in art.

DeGroot: Are any of your brothers or sisters interested in statistics or other mathematical things?

Rao: My brothers and sisters were all basically very intelligent people. I had an elder brother who was a doctor who was phenomenal. He had a wonderful memory. If you gave him 150 names, he could just read them once and then repeat all those names. One brother was an engineer, quite a good one. You see, we are a family of ten, so many of them died. Another was in commerce. One of my sisters was a poet; she wrote poetry in Telugu. And one was really a business woman; she did a lot of business and made a lot of property along with her husband. They didn't have any formal education because girls did not study years ago in India. So it was all an innate thing. My sister who was a poet just studied by herself and started writing poetry. But they were not mathematicians, except for the engineer who had to study some mathematics.

"THE TWO GREAT METHODOLOGIES IN STATISTICS ARE SAMPLE SURVEYS AND DESIGN OF EXPERIMENTS"

DeGroot: Are there other things that you like to do when you are not doing statistics?

Rao: Well, as I grow old I find that there's not much time to do many things. But in the next two or three years I may be visiting the third world countries and helping them in developing research in statistics. This is an assignment that the Third World Academy of Science, of which I am founder fellow, is trying to give me. I have not replied to their letter yet because I am not sure how my health will be in the coming years. But this is something which I would like to do because I know the third world countries are not developing basic research. They are dependent on foreign countries for all technological improvements, which is not good in the long run. That's the reason why we founded the Third World Academy. In addition, I may have some assignments in India which would keep me busy and occupied for a long time to come.

DeGroot: What do you think are the important trends in contemporary statistics? Where do you see the field going? Where do you think it should be going?

Rao: Those are questions about which I am con-

stantly thinking. I believe the two great methodologies in statistics are sample surveys, which is essentially collecting existing information, and design of experiments, where you generate observations to provide information on some given questions. Different types of data analyses are, of course, then applied depending upon what the statistician thinks is the right thing to do. They are not as fundamental as the data which are collected through principles of design and sample surveys. If data are good, results should be obvious; the analysis is only to convince somebody that there is a real difference and so on.

DeGroot: So you are saying that with a good design or a good survey, the answer should be obvious without doing any complicated analysis.

Rao: The analysis is only to make sure that we are not deceived by what we are seeing. Much of it should be obvious by looking at the data or by simple analysis. Even graphics is very important. That's the reason why R. A. Fisher emphasized graphical representation of data as part of statistical analysis. Although many people are not aware of this, the very first chapter of his book, *Statistical Methods for Research Workers*, is graphics. Nobody reads that although it is a very nice chapter. So I think we should be doing more graphical analysis, especially with the help of the computers we have. We can do things now, which we could not previously do, with transformations of variables and graphics. Further research in sample surveys and design of experiments to refine the already refined methods in these two areas would be helpful.

Then we should be developing some routine methods of statistics for applications, like the quality control methods and what Shewhart did in industry: a simple and efficient way of presenting data which enables them to see whether everything is OK or something is going wrong, or whether there is scope for improvement. There is a great need for developing simple techniques for routine applications in other areas to improve the overall efficiency of goods we produce and services we offer.

"HOW DOES THE DOCTOR DECIDE WHETHER THERE IS SOMETHING PATHOLOGICAL OR NOT?"

DeGroot: The medical area is one example.

Rao: Yes, there are many challenging problems in applications of statistics to medicine. We don't seem to have done much in that area. So this is where we should be working, along with the medical doctors, to try to see how statistics can be applied to improve diagnosis, treatment, monitoring the effect of drugs on the patients, and so on. Statistics is used in a slightly different way in problems of bioassays and the screening of drugs. But I think a lot remains to be done in medical diagnosis. I don't know whether I told you this story or not. When I was working at Johns Hopkins—that's the place where many famous doctors are—I went to a doctor because I felt some uneasiness in my stomach. He made a large number of tests and asked me what was wrong with me. I said that the Indian doctors told me that the food I take does not stay in the stomach long enough. It gets into the intestines within half an hour but normally, in other cases, it takes one-and-a-half hours or so.

So he said "The fact that the food stays only for half an hour doesn't mean there is something wrong with you. You know, if you take a set of normal individuals and find how long the food stays in the stomach for each individual, that's a variable and it is normally distributed." *He* tells *me*. [Laughs] So the fact that you are a member of that normal group having a small value doesn't mean there is something wrong with you. If a person is five feet tall, that's not pathological; some healthy individuals are five feet high and some are six feet high, and so on. So if it's not by looking at the battery of tests, how does the doctor decide whether there is something pathological or not? Suppose that with all those abnormal values, the patient looks all right. Does that mean there is something wrong with him? So it's probably a combination of the tests and the doctor's evaluation of the patient's general condition.

As soon as you go to the doctor he asks what is wrong with you, and he takes a case history. Are you all right? Are you able to think? Are you able to write papers? Are you able to walk? Are you able to drink? You see, if you do all this with your bad cholesterol, then it's OK, that's normal with you. So diagnostic tests like measurements on the blood chemistry are not by themselves enough to make a diagnosis. The patient's condition must also be an input into it. The question is how to put them together and come up with a diagnosis. It's an art for the doctor. Probably in the past nobody made any tests. If you went and told the doctor that you had a fever and other things, he felt your pulse and prescribed a medicine. So the general condition of the individual is probably far more important to the diagnosis than what the diagnostic tests reveal. So I am working on this kind of problem: how to put together the results of tests and the doctor's judgment of the patient's case history, and other prior evidence, which may be personal and subjective, and come to the right diagnosis.

DeGroot: That's great.

Rao: So there should be more routine applications of statistics wherever they are needed, and we should be getting into new areas where the classical approaches really do not work.

DeGroot: I think many statistical techniques have become standard simply because they were easy to apply, just the reason you are talking about, regardless of whether they were always relevant.

Rao: They all arose from applications in biology, but now that stage is past. I don't think we are very successful with statistical methods in psychology or even in economics. Possibly what is wrong with the economists is that they are not trying to refine their measurements or trying to measure new variables which cause economic changes. That is far more important than dabbling with whatever data are available and trying to make predictions based on them. So there's a lot of work that needs to be done in the soft sciences such as psychology and economics.

DeGroot: What does the future hold for C. R. Rao? You mentioned the possibility of traveling in third world countries.

Rao: Well, I will try to continue research to the extent possible with the facilities available wherever I am. And I like teaching and talking to students, trying to inspire them, giving them new problems—so probably I'll continue to do that. And if the Third World Academy forces me to go out and give lectures, I guess I could do that. But I do not know. I would like to continue the type of work I have been doing before, but probably at a slower pace than before.

DeGroot: I'm sure that in whatever you do, you will be as highly successful as you always have been. Thank you, C. R.

Shayle Searle's Contributions to the Evolution of the SAS System

Robert N. Rodriguez, Russell D. Wolfinger, Randall D. Tobias, Oliver Schabenberger
SAS Institute Inc., SAS Campus Drive, Cary, NC 27513 USA

Prepared for Conference in Honor of Shayle R. Searle on the Occasion of his Retirement, Cornell University, Ithaca, New York, August 9–10, 1996

Updated March 2005, for the 14th International Workshop on Matrices and Statistics, Massey University, Auckland

Abstract

For over twenty years, SAS statistical software has grown and evolved in response to the needs of its audience. One of the most influential members of this audience has been Shayle Searle. Our presentations acknowledge Shayle's many contributions to SAS users and to the developers of the linear models procedures in the SAS System.

Contributions to SAS Users (RNR)

This celebration of Shayle's career coincides with the 20th anniversary of SAS Institute, and so it is particularly appropriate to look back on his many contributions to the evolution of SAS software for linear models. Figure 1 provides a chronology of these activities.

Since the early 1970s, when Shayle's book *Linear Models* was published, there has been substantial growth in the SAS System for statistical analysis. The SAS72 manual, which was the only SAS documentation available at the time, contained 140 pages for statistical procedures, and it was first distributed to around 60 sites. The software itself consisted of around 35,000 lines of code. Today, the documentation for SAS/STAT software spans over 2800 pages, the code is measured in millions of lines, and the number of SAS users is estimated at over 3.5 million across 60 countries.

Likewise, the number of SAS employees has grown during this period. Figure 2 shows the building on Hillsborough Street in Raleigh, North Carolina, which was the Institute's home until 1980, when it moved to the present-day SAS campus in Cary.

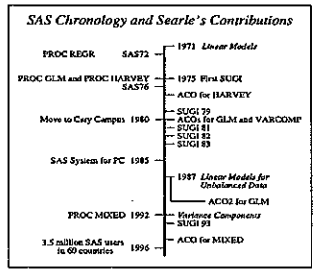


Figure 1: SAS Chronology and Searle's Contributions

Currently, 2100 employees are located at the Cary campus, which is shown in Figure 3, and the Institute has over 4000 employees worldwide.

Shayle was present at the first SAS Users Group International (SUGI) conference, which took place in Kissimmee, Florida, in 1975. Linear models and variance components analysis were topics of primary interest to the attendees. The GLM procedure was being developed by Dr. James Goodnight, the Institute president, and there

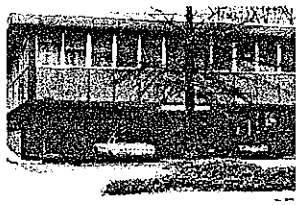


Figure 2: SAS Office Prior to 1980

was considerable discussion concerning the appropriate analysis of variance for unbalanced data. Both Shayle and Jim recall that the participants could not agree about which types of sums of squares the procedure should provide in its default output. In the end, Jim declared that he would include all of them. Although this resolved the discussion, it left open some important questions concerning the interpretation and applicability of the various sums of squares.

Fortunately, Shayle continued to think about these issues, and his expository presentations at subsequent SUGI conferences have provided valuable illumination for many SAS users. (Of course, many other individuals have also made significant suggestions and expository contributions in this area, as indicated in the Acknowledgments section of the *SAS/STAT User's Guide*.)

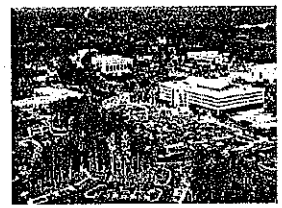


Figure 3: SAS Campus in 1996

I have not been able to locate any photographs taken during the Kissimmee meeting. However, Figure 4 shows Jim and Shayle resuming their discussion at the 1983 SUGI conference.

Figure 4: Jim Goodnight and Shayle Searle

Shayle was one of the earliest members of the statistics profession to recognize both the disadvantages and the advantages of the computer as a tool for statistical work. A typical starting point for Shayle's SUGI papers was the observation that "understanding something that has been computed is more of a problem than computing what we understand"; see Searle (1983). These papers addressed the need for better understanding of computer output, and they provided highly detailed explanations of linear models computations that made use of numerically simple examples, which are a hallmark of his expository writing.

Tutorials like these continue to be of benefit to SAS users, and so a brief review of Shayle's SUGI papers is appropriate at this point. His 1979 paper noted that users often compute F-statistics without first formulating a valid hypothesis. This backwards approach then leads to the question, "What hypothesis is it that I am testing?" This paper explains the derivation and use of the four types of estimable functions in the GLM procedure, and it clarifies the limitations and utility of GLM output for various situations involving the two-way crossed classification model with unbalanced data.

Shayle's 1981 paper dealt with quirks in linear model computations, special issues involving the HARVEY, GLM, and VARCOMP procedures, and the notion of "least squares means" (the beginning of a path pursued later in this presentation by Randy Tobias.) Shayle's 1982 paper (with G. F. S. Hudson) returned to the problem of understanding the hypotheses that are tested by the GLM

procedure, and it explored the nature and limitations of the hypotheses chosen for Type IV sums of squares.

Shayle's 1983 paper advocated broader use of analysis of covariance, and it covered situations involving several covariates, intraclass slopes of all forms, and unbalanced data. Shayle's 1993 paper discussed the history of analysis of variance computations, and it distilled his views on the analysis of unbalanced data which are presented at length in his 1987 book *Linear Models for Unbalanced Data*. This paper is particularly helpful for its recommendations concerning the use of the GLM procedure with unbalanced data, and it distinguishes between the all-cells-filled case and the some-cells-empty case.

Having attended all of Shayle's SUGI presentations, I can personally testify that they de-mystified difficult concepts and equipped attendees to make more effective use of the software. I recall returning from the 1983 conference to the General Motors Research Laboratories, where I worked at the time, and being asked on short notice to analyze a set of automobile accident data in response to a question posed by a top-level GM executive about the relative safety of American and Japanese cars. The analysis proved to be an immediate application of what I had learned from Shayle's talk!

SAS users have also benefited greatly from Shayle's annotated computer outputs (ACOs); see Figure 1. This series, which is available through the Biometrics Unit at Cornell, complements Shayle's books and our SAS manuals by providing detailed explanations of the output from various linear models procedures. The most recent addition is a 1995 ACO on the MIXED procedure.

During my years as a SAS user and developer, some of the lasting lessons that I have learned from Shayle are that

- users should not rely on algorithms that are detached from data specifics
- standard output can be hard to understand, and consequently labeling and terminology are critical. In software documentation, simple examples and detailed explanations are essential.
- statistical software needs to provide much more than statistical computing!

Shayle, thank you for all your lessons. I wish you a happy, productive retirement, and speaking on behalf of the many SAS users whom you have helped, I hope that you will continue to find opportunities to teach us.

Shayle Searle's Influence on a Young Statistician (RDW)

"The application of generalized inverse matrices to linear statistical models is of relatively recent occurrence." Thus begins Searle's classic 1971 text *Linear Models*, a text that also marked the beginning of my formal graduate training in statistics. As a quiet but confident new graduate student at NC State in 1985 with a B.S. in mathematics, I felt my previous training in linear algebra would be perfectly adequate to handle my initial coursework in linear statistical modeling. I was mistaken. Having never even heard of a generalized inverse, Shayle Searle had me scrambling with the very first sentence of his book.

Soon thereafter Searle defines a "...generalized inverse of a matrix A as any matrix G that satisfies the equation AGA = A." A little later we read that this definition "...does not define G as 'the' generalized inverse of A but as 'a' generalized inverse..." and we are then launched directly into a formal treatment of the various kinds generalized inverses and their properties. Such is the style of Searle, getting directly to the point at hand and applying it in as practical a way as possible. Extra descriptions and precise examples are included freely in attempt to make the point perfectly clear.

This attempt was successful in my case, but not without many hours of pondering and discussion with instructors and fellow students over such concepts as consistency, definiteness, quadratic forms, estimability, testability, sum-to-zero constraints, and last (but certainly not least!), the mixed model. Mine was a baptism by fire into the wonderful theory of linear models, and it has truly formed the foundation upon which most of my present statistical activity rests.

It has now been about 11 years since that intense first semester, and yet Searle's work continues to be influential. A case in point: In December of 1994, Bob, Randy, and I were asked by the journal *Statistics and Computing* to respond to a new article by John Nelder entitled "The statistics of linear models: back to basics." One quote from that article that certainly caught our attention was the following (with Greek symbols referring to a standard two-way ANOVA model):

> ...the Type III SS used in SAS (SAS Institute Inc., 1985) uses a quadratic form whose noncentrality parameter is symmetric in the α_i; this SS (i) loses power if used in any test, and (ii) is obtained by constraining the γ_{ij} margins to zero. It thus corresponds to an uninteresting hypothesis.

How to respond to such a criticism, considering that Type III tests form some of the principal default output from PROC GLM and PROC MIXED? Fortunately we had in hand *Linear Models for Unbalanced Data* (Searle 1987). It not only provides a thorough discussion of the four types of sums of squares (to which we refer in our reply), but Section 4.6 even goes so far as to refer to the Type III test for the two-way model as "..interesting, reasonable, useful..." In addition to this (and unbeknownst to us), Shayle composed his own reply to the article, including a nice discussion of the whole issue (see the June 1995 issue of *Statistics and Computing* for both replies).

Searle also makes the following remark in his reply to Nelder:

> And why...does 'model selection and prediction' have any connection whatever with trying 'to justify Type III sums of squares.'? I don't get it at all.

After a personal discussion with Nelder in June, it appears that Shayle has gotten to the heart of the issue with this query. The resolution is that Nelder is criticizing the use of Type III tests during the *model selection* phase of statistical analysis, whereas they may be appropriate during the *inference* phase once a model has been selected. As the dust is settling, we are grateful to have Searle's insight in these matters.

I had the privilege of meeting Shayle just before Nelder's article appeared as Shayle was working on a new Annotated Computer Output (ACO) about PROC MIXED. My wife Liz and I breakfasted with him one morning during the ASA meetings in San Francisco. We had barely begun discussing the ACO when Shayle surprised both us and our waiter by demonstrating what he referred to as "the proper way to serve tea." An adjoining remark was along the lines of "These #*%@ Americans never get it right." That taken care of, Shayle proceeded to provide me with a fully laundry list of suggestions regarding the output and labels from PROC MIXED. He was particularly critical of the output labeled "Mixed Model Equations", that not only had a mistake in it, but, as Shayle put it, "aren't a set of equations at all" because the final row of the output corresponded to the y-partition of the matrix that PROC MIXED sweeps. If you run PROC MIXED today with the MMEQ option you'll see that this row no longer appears.

A few years ago, we helped Shayle typeset the new ACO, which is now available from the Biometrics Unit at Cornell. During the course of this project, we ordered a complete set of the ACOs for reference. It was then I discovered, with much chagrin, the existence of the solutions manual to *Linear Models*. The chances are pretty good I would have emptied my somewhat meager bank account during that first semester of graduate school to obtain a copy of this book. Anyway, the ACOs have been very interesting, and since interacting with Shayle, I've been able to incorporate several more of his "suggestions" into PROC MIXED and plan to include more in the future. It's been quite a privilege working with him.

With the collective experience assembled in this room, I can't resist posing a question for discussion. It's one that has been recently sent to me by a colleague who has been running some interesting simulation studies. The question is this: Should variance component estimates in PROC MIXED be allowed to go negative by default? (They are currently constrained to be nonnegative by default but can be left unconstrained by using the NOBOUND option.) What piqued my interest in this question is that my colleague's simulation results show that the F-tests for fixed effects which are nested within a particular random effect are much too conservative in cases where this random effect's variance component is constrained to be nonnegative. In other words, PROC MIXED isn't printing true F-tests in this case. On the other hand, traditional precedents and model definitions argue against the unconstrained default.

Some of Shayle's thoughts on the problem are on pages 129-131 of *Variance Components* by Searle, Casella, and McCulloch (1992), included six possible courses of action to take when one encounters a negative ANOVA variance component. Shayle, what should PROC MIXED do by default?

Let me also say that *Variance Components* has been helpful to me in my recent work on the Bayesian analysis of variance component models. The exact expressions for the second derivatives of the likelihood were invaluable in confirming my calculations of Jeffrey's prior, and they even revealed an error in equation (5.2.10a) of Box and Tiao (1973).

Shayle, I hope your retirement is in name only, because your work tends to provide me with answers when I need them most; however, I seem to be lagging behind your ideas by about a decade. Thank you for all of the wonderful contributions you've made to statistics!

Recent Developments for Least Squares Means (RDT)

This section of our presentation describes some recent enhancements of the GLM and MIXED procedures for least squares means. Conceptually, the starting point for this

work is the definition for least-squares means provided by Searle, Speed, and Milliken (1980) and Searle (1981).

SAS Facilities for Comparing Independent Means

Multiple comparisons in the analysis of group means have long been included in the statistical tools of the SAS System, via the MEANS statement in the GLM procedure. The first multiple comparisons procedure (MCP) to be implemented was the Waller-Duncan Bayesian test (Waller and Duncan 1969). In the late 1970's many standard MCPs were added, including methods for all pairwise comparisons (Gabriel 1978, Hochberg 1974, Scheffe 1953, Sidak 1967,Tukey 1953), comparisons with a control (Dunnett 1955), and a variety of multiple stage tests (Duncan 1955, Einot and Gabriel 1975, Newman 1939, Ryan 1960, Welsch 1977). Each of the tests was requested by including an option in the MEANS statement, usually the name of the principal researcher responsible for the test. For example,

```
MEANS TMT / BON SIDAK SCHEFFE TUKEY
            GABRIEL GT2 DUNNETT
            DUNNETTU DUNNETTL SNK
            REGWQ REGWF WALLER
            and a few more ...;
```

In addition to the above facilities in the GLM procedure, the SAS Users Group Supplemental Library included the RSMCB procedure, contributed by Aubuchon, Gupta, and Hsu (1985), for ranking and selection and multiple comparisons with the best.

Limitations

The GLM procedure is a premier statistical tool for analyzing unbalanced designs (Milliken and Johnson 1984). However, MCPs usually do not apply in unbalanced cases because most MCPs require independent means with equal variances. The second part of this assumption implies that many of the procedures mentioned above are only defined for groups of equal size. The GLM procedure usually gets around this constraint by using the harmonic mean of the sample sizes when necessary for calculating critical values (Kramer 1956). The independence assumption is more fundamental. While the arithmetic means are always independent under the usual linear model assumptions, with unbalanced data they may not even estimate any model quantities of interest.

For example, consider the unbalanced two-way design given in Table 1. No matter how you look at it, this data exhibits a strong effect due to the columns (F-test $p < 0.0001$) and no significant interaction between rows and columns (F-test $p > 0.7$). But the lack of balance affects how the row effect is interpreted. In a main-effects-only model, there are no significant differences between the row means themselves (Type I F-test $p > 0.7$), but there are highly significant differences between the row means corrected for the column effects (Type III F-test $p < 0.0001$).

17, 28, 19, 21, 19	43, 30, 39, 44, 44	16
21, 21, 24, 25	39, 45, 42, 47	19, 22, 16
22, 30, 33, 31	46	26, 31, 26, 33, 29, 25

Table 1. Unbalanced Two-Way Design

The *least squares means* (LS-means) for rows are in effect the column-corrected row means. In the same way that the Type I F-test tests for differences between the arithmetic row means, the Type III F-test assesses differences between the LS-means; see Table 2. This table clarifies the difference between the results of the Type I and Type III tests.

The LS-means are defined in the SAS/STAT documentation as "the expected value of ... means that you would expect for a balanced design." A much clearer and more explicit definition is given by Searle, Speed, and Milliken (1980): estimates of population marginal means (PMM's) for a balanced population. This definition makes a clear distinction between the parametric functions being estimated and the statistics used to estimate them.

Row	Arithmetic Mean	Least-squares Mean
1	29.1	26.1
2	29.2	28.4
3	30.2	35.5

Table 2. Means for Unbalanced Two-Way Design

It is clear that multiple comparisons on the arithmetic row means in this case would be misleading. The row LS-means themselves are natural candidates for multiple comparisons, but for unbalanced designs such as the one above, they are correlated, and few of the standard MCPs apply to correlated means. Even in balanced designs, LS-means may be correlated if there are covariates. Furthermore, note that all of the above assumes a general *fixed-effects* linear model with independent and

identically distributed (IID) noise. In the context of *mixed models*, where the noise is no longer assumed to be IID, the problem of correlated estimates comes up even when only arithmetic means are being considered.

The GLM procedure is often used to analyze highly unbalanced data. The fact that it offered multiple comparisons only of group averages was a severe incongruity. For example, Hsu (1989) discussed "the misleading inference given by the MEANS option of SAS PROC GLM in general linear models ...". What was called for was multiple comparisons for correlated quantities that would cover LS-means in both mixed and unbalanced fixed-effect models.

A Unified Theory of Multiple Comparisons

Eventually I did find a sufficiently general methodology for multiple comparisons, due mainly to Jason Hsu at Ohio State University. The set-up is described in Hsu (1996): Given estimates $\hat{\theta}_i$, $i = 1, \ldots, k$ of parameters θ_i, $i = 1, \ldots, k$, it is desired to make inferences about a *fixed* subset $\mathcal{S}$ of contrasts between the parameters. The variance-covariance matrix $\sigma^2 C$ for the $\hat{\theta}_i$ is assumed to be known up to a constant, for which an independent estimate $\hat{\sigma}^2$ is available. Inference for an individual difference $l'\theta$ (expressed as a linear combination of the parameters) is based on the standardized estimated linear combination

$$t_l = \hat{\sigma}^{-1}\frac{l'(\hat{\theta}-\theta)}{\sqrt{l'Cl}},$$

in such a way that the over-all error rate is controlled. The error rate is

$$P(\max_{l\in\mathcal{S}} t_l > d_1^{\mathcal{S}}) \quad \text{and} \quad P(\max_{l\in\mathcal{S}} |t_l| > d_2^{\mathcal{S}}) \qquad (1)$$

for one-sided and two-sided inference, respectively. Hypothesis testing calls for calculating the error rate for the observed value of $\max_{l\in\mathcal{S}} t_l$, while confidence intervals require fixing the probability and solving for $d_1^{\mathcal{S}}$ or $d_2^{\mathcal{S}}$, as appropriate.

Let L be the matrix whose rows are the linear combinations in $\mathcal{S}$. Assuming Gaussian noise, the t_l's collectively have a multivariate (central) T distribution. The probabilities in (1) usually must be computed numerically. The full integration requires a $|\mathcal{S}|$-fold multiple integral in general, which is infeasible for $|\mathcal{S}| > 3$ or so. However, when $L'CL$ has a special form called *one-way structure* by Hsu (1996), the multiple integral collapses to a double integral. Examples of situations with one-way structure are

- when the $\hat{\theta}_i$ are arithmetic means of independent groups (that is, a one-way analysis)
- when $\mathcal{S}$ consists of differences between one "control" group and the rest (Dunnett's test)
- when $\mathcal{S}$ consists of all (ordered) pairwise differences, and all groups are the same size (Tukey's test)

If numerical computation of (1) is not feasible, various bounds (Bonferroni, Sidak 1967, Worsley 1982) or approximations (Hsu 1996) can be employed.

Thus, by distinguishing between what needs to be computed (the probabilities (1) for various sets of contrasts $\mathcal{S}$) and how it is to be computed (exactly or approximately or conservatively), Hsu provides a framework for a general computational tool. Our implementation matched this distinction; one option is used to specify the general class of MCP to perform (all pairs *versus* comparisons with a control), and another specifies the computational approach.

There were two primary candidates for these facilities: the GLM procedure and the newer MIXED procedure for general linear modeling with non-trivial error structures. The latter case is complicated by the fact that more than one component of the covariance matrix must be estimated. We made the assumption that the covariance was known up to a χ^2-distributed constant (possibly with non-integer degrees of freedom), primarily because there seemed no other way to proceed in general. Simulations indicate that MCP inference made under this assumption has characteristics similar to that for mixed model F-tests based on asymptotics, although a more complete study is in order.

Extensions

While we were working on LS-means we made a number of other enhancements. Our starting point was the Searle, Speed, and Millikin definition of LS-means as the marginal means over a balanced population. As such, LS-means may not be useful if "a balanced population" is unreasonable in a particular application. A new option allows users to specify the population over which the expected marginal means are to be computed. Similarly, covariates had been handled by setting them to their mean values in the analysis dataset, and a new option allows users to set their values directly. A third new option allows users to test for "simple effects" of one or more factors in an interaction. None of these enhancements are directly related to MCP, except that all of the new features taken together go a long way toward making it possible to approach the entire analysis of unbalanced designs from the point of view of LS-means.

Multiple-comparisons-with-the-best or MCB is a type of inference connected with bioequivalence testing, subset selection, and indifference zone selection (Hsu 1996). While not directly covered by the MCP enhancements for LS-means mentioned above, it is related to Dunnett's and Tukey's tests. Therefore, one can write a SAS program which uses the results of these tests to construct MCB confidence intervals and p-values. A macro program to do this in general has been added to the SAS sample library, making this important type of inference generally available.

Conclusion

The addition of multiple comparisons for correlated quantities is a significant enhancement to SAS statistical software. It not only provides multiple comparisons for the full range of linear models that can be fitted with the GLM procedure, but it also leads to an approach to multiple comparisons for mixed models with non-trivial error structures. Directions for further development include

- adding more fundamental types of comparisons, in addition to all pairs and comparisons with a control, and perhaps even a facility for specifying arbitrary comparisons of the model parameters
- adding facilities for multiple-stage tests of correlated means
- providing more techniques that build on the basic computations already available, as do the macros for MCB

This experience demonstrates that a fast way of getting a new statistical technique into commercial software is to make it general enough to convince the developer that he won't have to do the job all over again very soon. The work of Searle, Speed, and Millikin (1980) defining LS-means and the work of Hsu (1996) characterizing multiple comparisons in general were crucial for this development.

2005 Addendum: Further Development for Least Squares Means

The nine years since the 1996 Searle Conference have seen continued development for analysis of least-squares means in SAS. We have found LS-means to be a fundamental tool for asking and answering statistical questions not only about the linear models for which Searle, Speed, and Milliken (1980) and Searle (1981) originally proposed them, but also for mixed, generalized linear, and generalized linear mixed models. We take our cue for this update from the "directions for further development" we discussed in 1996.

Note that all of the enhancements discussed here are implemented in the new GLIMMIX procedure for generalized linear mixed models, available for download to users of SAS 9.1 at `support.sas.com`. In generalized and generalized linear mixed models, LS-means play a pivotal role in estimation and inference. Because of the nonlinearity of the link function, estimating and testing linear combinations of parameter estimates is only meaningful on the linked scale, for which the model effects are assumed to be additive. Because expectations and variances are related in such models, and because data are marginally correlated in mixed models, arithmetic means on the linear scale have no use. In this context, LS-means are crucial to formulating appropriate hypotheses about group comparisons. Simple transformations of these contrasts produce other important inferential quantities, for example, the exponentiation of LS-mean differences in models with logit link produces odds ratios.

More Types of Comparisons

Analysis of Means

In addition to all-pairwise and one- and two-tailed control comparisons, we added the DIFF=ANOM option to the LSMEANS statement of PROC GLIMMIX, for comparing each LS-mean with the average of the LS-means. In the context of one-way analysis of variance, these comparisons, suitably adjusted for multiplicity, are the basis of the *analysis of means* (Ott, 1967). An analysis of means is usually condensed into a graphical display that represents each mean by how much it differs from the overall mean, with decision limits that say whether this difference is significant. Such a plot is also available in PROC GLIMMIX; Figure 5 gives an example of an Analysis of Means plot for the comparison of 16 plant varieties (entries) in a randomized block design where the outcome variable has a binomial distribution.

```
proc glimmix;
   class block entry;
   model y/n = entry / dist=binomial;
   random block;
   lsmeans entry / plot=anomplot
                   adjust=nelson;
run;
```

The decision limits are computed using the distribution described by Nelson (1982), using a factor-analytic covariance approximation described in Hsu (1992). With

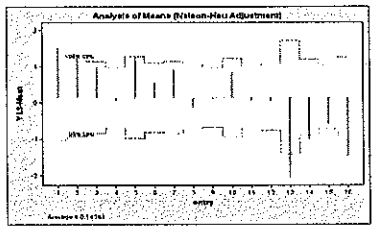


Figure 5: Analysis of Means for Binomial Data

Gaussian data the upper and lower decision limits would be horizontal lines, since the design is balanced. Because the variability of the binomial distribution is tied to its mean, for this data the decision limits vary with the magnitude of the LS-means, here shown on the logistic scale.

Arbitrary LS-mean Contrasts

Besides the "canned" comparison types available via the LSMEANS statement, we added an LSMESTIMATE statement to PROC GLIMMIX, to provide for custom hypothesis tests among the least-squares means. In contrast to the hypotheses tested with the ESTIMATE or CONTRAST statements, the LSMESTIMATE statement allows you to estimate arbitrary linear combinations of the least-squares means. Multiple-row sets of coefficients are permitted, along with multiplicity adjustments for the associated p-values and confidence intervals.

Multiple-Stage Tests

The multiplicity adjustments for LS-means comparisons which we reported in 1996 were all *single-stage* procedures. They have the advantage that they yield simultaneous confidence intervals for all comparisons. However, by sacrificing the ability to get simultaneous confidence intervals, you can use multiple-stage tests to improve power while still protecting the familywise error rate. We have added the STEPDOWN option for all multiplicity adjustments to do just that, implementing the ideas of Shaffer (1986) and Westfall (1997).

Building on Basic Computations

In 1999, we published a book on multiple comparisons and multiple tests in SAS (Westfall et al. 1999). As planned, this book was going to cover the traditional MEANS comparisons, plus some home-grown macros for special tests. However, it soon became apparent that the versatility of LS-means for handling unbalanced and/or correlated data made them the right basic tool to use. Using LS-means in conjunction with SAS/IML matrix programming and the (then) new ODS output delivery system, we published very general macros for computing single-stage and multistage adjusted LS-mean comparisons (`%SimIntervals`, `%SimTests`) and for assessing the power of multiple comparisons procedures (`%SimPower`).

References

Aubuchon, J.C., Gupta, S.S., and Hsu, J.C. (1985), "Interfacing PROC RSMCB with the SAS System," Technical Report 313, Department of Statistics, Ohio State University, Columbus, OH.

Box, G. E. P. and Tiao, G. (1973), *Bayesian Inference in Statistical Analysis*, John Wiley & Sons, New York.

Duncan, D.B. (1955), "Multiple Range and Multiple F Tests," *Biometrics*, 11, 1-42.

Dunnett, C.W. (1955), "A Multiple Comparisons Procedure for Comparing Several Treatments with a Control," *Journal of the American Statistical Association*, 50, 1096-1121.

Einot, I. and Gabriel, K.R. (1975), "A Study of the Powers of Several Methods of Multiple Comparisons," *Journal of the American Statistical Association*, 70, 351.

Gabriel, K.R. (1978), "A Simple Method of Multiple Comparisons of Means," *Journal of the American Statistical Association*, 73, 364.

Hochberg, Y. (1974), "Some Conservative Generalizations of the T-method in Simultaneous Inference," *Journal of Multivariate Analysis*, 4, 224-234.

Hsu, J.C. (1989), "Multiple Comparison Procedures," American Statistical Association Tutorial Notes.

Hsu, J. C. (1992), "The Factor Analytic Approach to Simultaneous Inference in the General Linear Model," *Journal of Computational and Graphical Statistics*, 1, 151-168.

Hsu, J.C. (1996), *Multiple Comparisons: Theory and Methods*, Chapman and Hall.

Hudson, G. F. S. and Searle, S. R. (1982), "Hypothesis Testing with Type IV Sums of Squares of the Computer Routine SAS GLM", *Proceedings of the Seventh Annual SAS Users Group International Conference*, SAS Institute Inc., Cary, NC, 521-527.

Kramer, C.Y. (1956), "Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications," *Biometrics*, 12, 307-310.

Milliken, G.A. and Johnson, D.E. (1984). *Analysis of Messy Data*, Chapman and Hall.

Nelson, P. R. (1982), "Exact Critical Points for the Analysis of Means," *Communications in Statistics*, 11, 699-709.

Newman, D. (1939), "The Distribution of the Range in Samples from a Normal Population," *Biometrika*, 31, 20-30.

Ott, E. R. (1967), "Analysis of Means—A Graphical Procedure," *Industrial Quality Control*, 101-109. Reprinted in *Journal of Quality Technology*, 15 (1983), 10-18.

Ryan, T.A. (1960), "Multiple Comparisons in Psychological Research," *Psychological Bulletin*, 57, 318-328.

Scheffe, H. (1953), "A Method for Judging All Contrasts in the Analysis of Variance," *Biometrika*, 40, 87-104.

Searle, S. R. (1971), *Linear Models*, John Wiley & Sons, New York.

Searle, S. R. (1979), "Relationships between the estimable functions of SAS GLM output for unbalanced data and the hypotheses tested by traditional-style F-statistics," *Proceedings of the Fourth Annual SAS Users Group International Conference*, SAS Institute Inc., Cary, NC. 196-207.

Searle, S. R. (1981), "Quirks in Linear Model Calculations with Unbalanced Data", *Proceedings of the Sixth Annual SAS Users Group International Conference*, SAS Institute Inc., Cary, NC. 38-45.

Searle, S. R. (1983), "Analysis of Covariance Using SAS GLM", *Proceedings of the Eighth Annual SAS Users Group International Conference*, SAS Institute Inc., Cary, NC. 747-759.

Searle, S. R. (1987), *Linear Models for Unbalanced Data*, John Wiley & Sons, New York.

Searle, S. R. (1993), "Analysis-of-Variance Computing: Then and Now, With Reference to Unbalanced Data", *Proceedings of the Eighteenth Annual SAS Users Group International Conference*, SAS Institute Inc., Cary, NC. 1077-1088.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992), *Variance Components*, John Wiley & Sons, New York.

Searle, S.R., Speed, F.M., and Millikin, G.A. (1980), "Population Marginal Means in the Linear Model: An Alternative to Least Squares Means," *The American Statistician*, 34, 216-221.

Shaffer, J. P. (1986), "Modified Sequentially Rejective Multiple Test Procedures," *Journal of the American Statistical Association*, 81:329-335.

Sidak, Z. (1967), "Rectangular Confidence Regions for the Means of Multivariate Normal Distributions," *Journal of the American Statistical Association*, 62, 626-633.

Tukey, J.W. (1953), "The Problem of Multiple Comparisons," Unpublished manuscript.

Waller, R.A. and Duncan, D.B. (1972), "Computations of Bayesian t-Values for Multiple Comparisons," *Journal of Statistical Computation and Simulation*, 75, 169-172.

Welsch, R. E. (1977), "Stepwise Multiple Comparison Procedures," *Journal of the American Statistical Association*, 72, 359.

Westfall, P. H. (1997), "Multiple Testing of General Contrasts Using Logical Constraints and Correlations," *Journal of the American Statistical Association*, 92:299-306.

Westfall, P. H., Tobias, R. D., Rom, D., Wolfinger, R. D., and Hochberg, Y. (1999), *Multiple Comparisons and Multiple Tests Using the SAS System*, SAS Institute Inc., Cary, NC.

Worsley, K. J. (1982), "An Improved Bonferroni Inequality," *Biometrika*, 69, 297-302.

Austral. & New Zealand J. Statist. **40**(4), 1998, 385–387

PITMAN MEDAL AWARDED TO E. SENETA

Professor Eugene Seneta from the University of Sydney was awarded the Pitman Medal of the Statistical Society of Australia for 1998 in recognition of his outstanding contributions to probability theory and statistics, and particularly to stochastic processes and the history of statistics. The medal was presented at the 14th Australian Statistical Conference in July 1998 by Professor Des Nicholls, President of the Society.

Eugene Seneta was born in 1941 in Stary Sambor, Ukraine. He started school in post-war Germany and in 1949 he and his family emigrated to Adelaide where he continued his schooling. Eugene completed his Bachelor of Science degree at The University of Adelaide with First Class Honours in Pure and Statistical Mathematics in 1964, followed by a Master of Science degree by research in 1965 from the same institution. Eugene's M.Sc. thesis, entitled 'Transient Behaviour in Finite Absorbing Markov Chains', was written under the supervision of John Darroch. He then joined the stream of people working on stochastic processes at The Australian National University (ANU) where he was awarded a Ph.D. for his thesis on 'Topics in the Theory and Applications of Markov Chains' in 1968.

Eugene lectured at the ANU from 1964 until 1979, being promoted to the position of Reader in 1974. In 1979 he took up his current position as Chair of Mathematical Statistics at The University of Sydney. He has held visiting positions at Cambridge University, Imperial College, Princeton University, Virginia Polytechnic Institute and State University, Colorado State University, The University of Virginia, The University of Chicago, and at L'École des Hautes Études en Sciences Sociales, Paris.

Eugene has established a major international reputation for his contributions in several fields. His early work on quasi-stationary distributions for Markov chains and his work on the asymptotics of branching processes secured his place as a leader in the area of stochastic processes. In recent years the topic of quasi-stationarity in absorbing processes has seen a vigorous revival, largely in the context of birth-and-death processes, and his papers with John Darroch (1965) and David Vere-Jones (1966) are cited for motivation. The latter paper marked the beginning of Eugene's interest in Vere-Jones's R-theory of infinite non-negative matrices, which he then applied to the study of finite approximations to infinite non-negative matrices.

The keystone to Eugene's influential results in the theory of discrete branching processes was the application of the iteration-theoretic/functional equation approach, exposited in the first paper to be published in *Advances in Applied Probability* at its inception in 1969. This emphasised his possibly best-known result, which had appeared a year earlier in the *Annals of Mathematical Statistics*, and which showed that if the offspring mean exceeds unity but is finite, there always exists a sequence of norming constants for the Bienaymé–Galton–Watson process which give a non-degenerate limit distribution. These constants are generally called the Seneta constants. The result is an ultimate form incapable of improvement in this setting. Chris Heyde showed, very soon afterwards, that the structure of the constants was such that convergence was almost sure. Other influential directions in Eugene's work concerned the intimate and natural connection between the theory of discrete branching processes and the theory of regularly varying functions (1971), and work initiated in a paper with Chris Heyde in 1972 on inference for branching processes with immigration. The authoritative monograph on branching processes by Athreya & Ney (1972) is replete with references to his papers.

Eugene is the author or coauthor of more than 150 papers in international journals and has made several contributions to the *Encyclopedia of Statistical Sciences*. His books, *Regularly Varying Functions* (Springer, 1976; 2nd edn 1985 (in Russian)) and *Non-Negative Matrices* (1973, 2nd edn 1981) are widely quoted and are standards in their fields.

In *Non-Negative Matrices*, Eugene's treatment of coefficients of ergodicity as a tool for treating inhomogeneous products of non-negative matrices led on to a highly influential synthesis on this topic, in a paper of 1979, that illuminated the use of coefficients of ergodicity for spectrum localization. His result, in 1977, that weak and strong ergodicity for backwards inhomogeneous products of stochastic matrices are equivalent, had direct significance for the model of continuing interchanges of information leading to consensus. More recently he showed how the Markov–Dobrushin coefficient of ergodicity is appropriate and easy to use as a condition number measuring the stability of a stationary distribution of a finite Markov chain under perturbation of a transition matrix.

Eugene's first paper written as a sole author, 'Quasi-stationary distributions and time-reversion in genetics', was accorded the unexpected honour of being read on his behalf before the Royal Statistical Society by Peter Whittle. In 1977, his paper in *Genetics* showed how fixed population-size Markov chain models could be generalized to randomly fluctuating population size by considering proportions of types of individuals, and resorting to martingale theory, a theme that has since been extended by others.

One of his many contributions to the history of probability and statistics was the book *I.J. Bienaymé: Statistical Theory Anticipated* (Springer, 1977), coauthored with Chris Heyde. This book was, in effect, a history of probability and statistics in the 19th and early 20th centuries. Eugene is an acknowledged expert on work of this period in France and the Russian Empire. His writings have revealed, in particular, how attribution and perception in science may be a highly politicised activity. They have focused on the probabilistic and statistical work of Pascal, Cauchy, Bienaymé, Markov, Chebyshev, Chuprov, Nekrasov and Sleshinsky.

His attempt, with his former student Kathy M. Kang in 1980, to clarify path analysis from a mathematically consistent standpoint drew wide attention and favourable response from mathematically-minded statisticians. The vg (variance gamma) model in financial mathematics, sometimes called the Madan–Seneta process, proposed by Dilip Madan and Eugene in 1990, is in use on Wall Street.

Less well known, perhaps, is Eugene's service to the wider community: he provided a theoretical basis for and oversaw the production of an algorithm for scaling Higher School Certificate marks in the early 1980s. Building on a cruder algorithm in use by the NSW Board of Senior School Studies which had withdrawn from the activity, he worked with programmers at the University of Sydney and instituted a scheme that remains the essence of today's NSW Tertiary Entrance Rank. The Victorian scaling procedure also uses some of the particular details worked out by Eugene.

Eugene's contributions to Australian research were recognised by his election in 1985 to Fellowship of the Australian Academy of Science.

In an age of increasingly superficial learning, Eugene Seneta is an extraordinary rarity: a true scholar. His deeply committed scholarship informs penetrating analyses in matters both mathematical and historical that stand with the best in the world.

A Short Biography of George Seber

George Seber took his Bachelor's degree in mathematics and physics and his Master's degree in mathematics at Auckland University and then continued his graduate studies under a Commonwealth Scholarship at Manchester University, England in 1960. After completing a PhD in Statistics there he spent two years as an Assistant Lecturer in Statistics at the London School of Economics before joining the Auckand Mathematics Department in 1965. In 1971 he was invited to a personal Chair in Biometrics at Otago University but later returned to Auckland in 1973 to take up the first Chair in Statistics and the Headship of a newly created Statistics Unit within the Mathematics Department. He was involved with establishing a new Department of Statistics in 1994. He was elected a Fellow of the Royal Society (NZ) in 1997, received a Distinguished Statistical Ecologist Award by the International Association for Ecology in 1998, and was awarded the Hector medal in Science from the NZ Royal Society in 1999. In 2004 he was made a life member of the NZ Statistical Association.

He has enjoyed teaching at all levels and writing books. Currently he is author or co-author of ten books entitled "The Linear Hypothesis" (2nd edition, 1970), "Elementary Statistics" (1974), "Linear Regression Analysis" (1977), "Estimation of Animal Abundance" (2nd edition, 1982, now reprinted as a paperback by Blackburn press, 2002), "Multivariate Analysis" (1984, reprinted in the Wiley paper back series, 2004), Nonlinear Regression Analysis (1989, with Chris Wild, reprinted in the Wiley paper back series, 2003), "Adaptive Sampling" (1996, with Steve Thompson), "Chance Encounters: Introduction to Data Analysis and Inference" (with Chris Wild, 2000), together with a Teachers' Manual and Web site, and "Linear Regression", second edition (with Alan Lee, 2003). A further book has been started entitled "A matrix handbook for Statisticians". His research interests have included: animal abundance, adaptive sampling, linear models, blood genetics, and large sample theory.

In 2004 George graduated with a Diploma in Counselling from a theological college and has been counseling part-time for the past three years. Among other activities he enjoys keeping fit, learning modern piano, and has played bass guitar in a church music team for the past 14 years. He will continue with the latter as long as they will still have him! Apart from book writing he has now retired from teaching and research in statistics.