The Open University

Statistics Department

1999 Report

1. Membership

Dr K.J. McConway (Head of Department)
Professor F. Critchley (from September)
Professor J.C. Gower
Professor D.J. Hand (to March)
Dr N. Adams (to March)
Dr C.P. Farrington
Dr G.E. Iossif
Dr M.C. Jones
Miss M.N. Kanaan (from October)
Dr M.G. Kelly (to March)
Dr A.D. Lunn
Dr P. Sebastiani
Dr S.K. Vines
Mrs J.E. Williams

Visitors

Dr G.M. Constantine, (CSIRO Adelaide, Australia, retired)
Professor M.J. Faddy (University of Canterbury, New Zealand)
Mr M. Joseph (AGC Ltd, Australia)
Dr I. Koch (University of Newcastle, Australia)
Mr A. Neri (Italy)
Dr P.C. Taylor (University of Hertfordshire)

Full-time research students

Mr Y. Abou-Salama (from October)
Mr T.C. Benton (to March)
Miss M.N. Kanaan (to September)
Mr R.J. Till (to March)
Mr B.E.T.H. Twala
Mrs J. Warwick

Part-time research students

Mr G. Blunt
Ms J. Burrett
Mr S. Cresswell
Mr A. Garrett

Dr M. King
Mr M. Lunt

Support staff

Ms E. Ostrowski, Departmental Secretary

2. Introduction

This year was another very productive and successful one for the Department. We hold a grade 4 rating from the 1996 Research Assessment Exercise, and research activity continued strongly in 1999. During 1999, Department members published a book, 21 papers in refereed journals and 12 book chapters etc. In addition, our teaching continued to flourish.

An undergraduate Diploma in Statistics was introduced. Work started on replacements for the current pair of second level courses. A case was successfully made for the introduction of a new course in quantitative methods for business.

It was the Department's great loss that Professor David Hand left to take the Chair in Statistics at Imperial College, London. Huge credit is due to David for his major contributions to increasing the reputation and activity of the Department in the 11 years he was here. Those that remain would like to take this opportunity to publicly thank him for his magnificent efforts. Niall Adams, Mark Kelly, Tim Benton and Rob Till went with David to Imperial, and Paola Sebastiani will join them too in 2000.

But this has created the opportunity for renewal and, indeed, expansion. Professor Frank Critchley has joined us from the University of Birmingham as Research Professor. We will also be advertising the Established Chair and three lectureships in early 2000.

Information on the Department is available through our Web pages at http://Statistics.open.ac.uk/index.html

3. Teaching

Presentation

The Department presented four courses (30 CATS points each), MDST242 *Statistics in Society*, M246 *Elements of Statistics*, M343 *Applications of Probability* and M346 *Linear Statistical Modelling*. Altogether the Department's courses attracted a total of about 2,100 students.

A revised version of M346 was prepared for commercial publication, and was published in September by Edward Arnold.

Each of the Faculty of Mathematics and Computing's first level courses, MU120 *Open Mathematics* and MST121 *Using Mathematics*, contains substantial amounts of statistical material (roughly one quarter of each) with which Gillian Iossif is heavily involved.

MST121, M246 and, newly, M346 are also presented to students of the Singapore Institute of Management.

Production

The department began work on preparing a revitalised and more compatible pair of second level courses, MU247 *Understanding and Interpreting Data* and M248 *Analysing Data* to replace MDST242 and M246, respectively. M248 is slated to be a revision and reorganisation of M246; MU247 is a completely new rewrite. M248 is due to appear in 2003, MU247 in 2005. Gillian Iossif is Course Team Chair of both courses.

The department also began the task of planning and outlining a new course in quantitative methods for business, BM240. This 30-point second level course is intended to be an important contribution to the Open University's new undergraduate degree in Business Studies. Kevin McConway and Chris Jones (Course Team Chair) were most heavily involved in 1999.

Other

The new undergraduate *Diploma in Statistics* came to fruition in 1999. This qualification is awarded to students who successfully study M246, M343 and M346 (or their predecessors) plus one other course from a list of electives.

Kevin McConway also contributed to the presentation of U205 *Health and Disease* and D300 *Professional Judgement and Decision Making*.

4. Research interests

At the start of the year, the Department was the home of two research groups, in *Statistics in Finance* (Hand, Adams, Kelly, McConway) and *Classification Methods* (Hand, Adams, Gower, Ngouenet, Taylor). With changes in personnel, two new research groups have emerged: *Multivariate Statistics* (Critchley, Abou-Salama, Gower, Jones, Lunn, Taylor, Twala, Vines, Constantine (CSIRO, retired), Cook (University of Minnesota), Marriott (National University of Singapore)) and *Medical Statistics* (Vines, Farrington, Kanaan, Lunn, Warwick, M. Lunn (University of Oxford)). In addition, Dr McConway is a member of the University's *Ecology and Conservation* research group.

Members of the Department are involved in a variety of research projects, including:

• *Influence analysis* (Critchley)

• *Regression graphics* (Critchley, Cook (University of Minnesota))

• *Multivariate analysis* (Critchley, Gower)

• *Differential geometry in statistics* (Critchley, Marriott (National University of Singapore))

• *Statistics of infectious diseases (estimation of epidemic parameters and of vaccine efficacy; evaluation of vaccine safety; epidemic detection systems)* (Farrington, Kanaan)

• *Interval-censored survival data* (Farrington)

• *Capture-recapture methods in epidemiology* (Farrington)

• *Goodness of fit of GLMs with sparse data* (Farrington)

• *Case crossover and case series studies* (Farrington, Vines)

• *Multidimensional scaling, Procrustes methods* (Gower)

• *Biplots* (Gower, Hand)

• *Measurement theory and metadata* (Hand)

• *Assessing classification rules* (Hand, Adams)

• *Credit risk scoring* (Hand, Adams, Kelly, McConway)

• *Discrimination and classification* (Hand, Adams, Taylor, Webb (DERA))

• *Distribution theory* (Jones)

• *Skew t distributions* (Jones, Faddy (University of Canterbury))

• *Semiparametric density estimation* (Jones, Hjort (University of Oslo), Storvik (Norwegian Computing Centre))

• *Local dependence* (Jones, Koch (University of Newcastle, Australia))

• *Miscellaneous smoothing topics* (Jones, Gijbels (Universite Catholique de Louvain), Hall (Australian National University), Hjort (University of Oslo), Lloyd (University of Hong Kong), Park (Hankuk University of Foreign

Studies), Signorini (Quantiles Ltd), Wright (University of Plymouth), Yao (London School of Economics and Political Science))

• *Spatial statistics* (Kanaan, Taylor, Mugglestone (Rothamsted Experimental Station), Alexander (Rothamsted Experimental Station))

• *Bayesian methods in meta-analysis* (Lunn)

• *Medical statistics, clinical trials (categorical data, repeated measures, generalised estimating equations, Bayesian methods)* (Lunn)

• *Performance-based handicapping systems in sport* (Lunn, Cresswell)

• *Bayesian statistics* (Lunn, McConway, Sebastiani, Vines)

• *Statistical methods in plant ecology and evolution* (McConway)

• *Comparative methods using phylogenies* (McConway, Silvertown (Biology, OU))

• *The determinants of species richness in flowering plants* (McConway, Silvertown (Biology, OU), Chase (Royal Botanical Gardens, Kew))

• *Methods for comparing cladogenesis rates in phylogenies* (McConway, Silvertown (Biology, OU), Dodd (Biology, OU))

• *Ecological and genetic correlates of long-term population trends in plants* (McConway, Silvertown (Biology, OU), Miller (Rothamsted Experimental Station), Lupton (Rothamsted Experimental Station), Macnair (University of Exeter))

• *Reasoning under uncertainty* (Sebastiani)

• *Decision theoretic foundations of statistical methods* (Sebastiani)

• *Statistical methods in machine learning, artificial intelligence and robotics* (Sebastiani, Ramoni (KMI, OU))

• *Graphical modelling* (Sebastiani, Ramoni (KMI, OU))

• *Statistical experimental design* (Sebastiani, Wynn (University of Warwick))

• *Bayesian analysis of case-control studies* (Vines)

• *Principal components analysis* (Vines)

• *Minimum distance estimation* (Warwick, Jones, Basu (Indian Statistical Institute), Harris (University of Northern Arizona), Hjort (University of Oslo))

5. Publications

5.1 Books

McConway, K.J., Jones, M.C. and Taylor, P.C. (1999) *Statistical Modelling Using Genstat*, Arnold, London.

5.2 Publications in refereed academic journals

Dawid, A.P. and Sebastiani, P. (1999) Coherent dispersion criteria for optimal experimental design. *Annals of Statistics*, 27, 65-81.

Delaney, G., Rus, M., Gebski, V., Lunn, A.D. and Lunn, M. (1999) An Australasian assessment of the Basic Treatment Equivalent model derived from New South Wales data. *Australasian Radiology*, 43, 500-6.

Delaney, G., Rus, M., Gebski, V., Lunn, A.D. and Lunn, M. (1999) Refinement of the Basic Treatment Equivalent model to reflect radiotherapy treatment throughput using Australasian data. *Australasian Radiology*, 43, 507-13.

Faddy, M.J. and Jones, M.C. (1999) Modelling and analysis of data that exhibit temporal decay. *Applied Statistics*, 48, 229-37.

Farrington, C.P. and Gay, N.J. (1999) Interval censored survival data with informative examination times: parametric models and approximate inference. *Statistics in Medicine*, 18, 1235-48.

Farrington, C.P. and Grant, A.D. (1999) The distribution of time to extinction in subcritical branching processes: applications to outbreaks of infectious diseases. *Journal of Applied Probability*, 36, 771-9.

Gay, N.J., Hesketh, L.M., Osborne, K.P, Farrington, C.P., Morgan-Capner, P. and Miller, E. (1999) The prevalence of hepatitis B infection in adults in England and Wales. *Epidemiology and Infection*, 122, 133-8.

Gower, J.C. (1999) Geometry, algebra and data analysis. *The Mathematical Scientist*, 24, 75-89.

Gower, J.C. (1999) Rank-one and rank-two departures from symmetry. *Computational Statistics and Data Analysis*.

Gower, J.C. and Krzanowski, W.J. (1999) Analysis of distance for structured multivariate data. *Applied Statistics*, 48, 505-19.

Gower, J.C., Meulman, J.J., and Arnold, G.M. (1999) Non-metric linear biplots. *Journal of Classification*, 16, 181-96.

Hand, D.J. (1998) Data mining ¾ reaching beyond statistics. *Research in Official Statistics*, 2, 5-17.

Hawker, J.I., Bakhshi, S.S., Ali, S. and Farrington, C.P. (1999) Ecological analysis of ethnic differences in relation between tuberculosis and poverty. *British Medical Journal*, 319, 1031-4.

Iossif, G. (1999) The graphics calculator as a teaching aid in statistics. *Teaching Statistics*, 21, 45-8.

Jones, M.C. (1999) Distributional relationships arising from simple trigonometric formulas. *The American Statistician*, 53, 99-102. Supplement, 53, 393.

Jones, M.C., Signorini, D.F. and Hjort, N.L. (1999) On multiplicative bias correction in kernel density estimation. *Sankhya*, Series A, 61, 422-30.

Stanghellini, E., McConway, K.J. and Hand, D.J. (1999) A discrete variable chain graph for applicants for credit. *Applied Statistics*, 48, 239-52.

Taylor, B., Miller, E., Farrington, C.P., Petropoulos, M.-C., Favot-Mayaud, I., Li, J. and Waight, P.A. (1999) Autism and measles, mumps and rubella vaccine: no epidemiological evidence for a causal association. *Lancet*, 353, 2026-9.

Taylor, P.C. and Hand, D.J. (1999) Finding superclassifications with acceptable misclassification rates. *Journal of Applied Statistics*, 26, 579-90.

Will, R.J., Cousens, S.N., Farrington, C.P., Smith, P.G., Knight, R.S.G. and Ironside, J.W. (1999) Deaths from variant Creutzfeldt-Jakob disease. *Lancet*, 353, 979.

Zhang, S., Karunamuni, R.J. and Jones, M.C. (1999) An improved estimator of the density at the boundary. *Journal of the American Statistical Association*, 94, 1231-41.

5.3 Refereed conference contributions and book chapters

Alexander, C.J., Mugglestone, M.A., Taylor, P.C., Thompson, R. and Morgan, G.W. (1999) Spectral analysis of spatial point patterns in incompletely observed regions. In *Spatial Temporal Modelling and its Applications*, eds K.V. Mardia, R.G. Aykroyd and I.L. Dryden, Leeds University Press, 113-6.

Farrington C.P. (1999) Interval-censoring. In *Encyclopedia of Statistical Science Update*, vol. 3, ed S. Kotz, Wiley, 371-6.

Gower, J.C. (1999) Discussion of contrast between psychometric and statistical approaches to multiway data analysis. *Bulletin of the International Statistical Institute*, 58, 101-2.

Hand, D.J. (1998) Data mining ¾ reaching beyond statistics. In *Proceedings of the KESDA98 Conference*.

Ramoni, M. and Sebastiani, P. (1999) Bayesian methods in intelligent data analysis. In *Intelligent Data Analysis; an Introduction*, eds M. Berthold and D.J. Hand, Springer, 128-66.

Ramoni, M. and Sebastiani, P. (1999) Learning conditional probabilities from incomplete data: an experimental comparison. In *Proceedings of Uncertainty 99: The 7th International Workshop on Artificial Intelligence and Statistics*, Morgan Kaufman, 260-5.

Ramoni, M., Sebastiani, P. and Dybowsky, R. (1999) Robust outcome prediction for intensive-care unit patients. In *Proceedings of the AIMDM-99 Workshop on Prognostic Models in Medicine: Artificial Intelligence and Decision Analytic Approaches*, 45-9.

Sebastiani, P. and Ramoni, M. (1999) Model folding for data subject to nonresponse. In *Proceedings of Uncertainty 99: The 7th International Workshop on Artificial Intelligence and Statistics*, Morgan Kaufman, 287-92.

Sebastiani, P., Ramoni, M. and Cohen, P. (1999) Unsupervised classification of sensory input in a mobile robot. In *Proceedings of IJCAI Workshop on Neural, Symbolic and Reinforcement Methods for Sequence Learning*, AAAI Press, 23-8.

Sebastiani, P., Ramoni, M., Cohen, P., Warwick, J. and Davis, J. (1999) Discovering dynamics using Bayesian clustering. In *Proceedings of the Third International Symposium on Intelligent Data Analysis*, Lecture Notes in Computer Science, Springer, 199-210.

Taylor, P.C. (1999) Statistical methods. In *Intelligent Data Analysis; an Introduction*, eds M. Berthold and D.J. Hand, Springer Verlag, 67-127.

5.4 Letters to the editor of refereed journals

Taylor, B., Miller, E and Farrington, P. (1999) MMR vaccination and autism. *The Lancet*, 354, 950-1.

6. Seminars and conferences at the Open University

Both the 1998 and 1999 departmental statistics conferences were held in 1999.

The 1998 conference was a two-day affair held off campus at St Hugh's College, Oxford at the beginning of January 1999. The conference was a special one in honour of the brilliant career of Professor John Gower. The conference was entitled *The Geometry of Data Analysis*, generally known as *Gowerfest*, and those of John's friends who spoke were:

Dr G.M. Constantine (CSIRO, Adelaide, Australia) *The analysis of asymmetry*

Professor F. Critchley (University of Birmingham) *On distance in statistics. Differential geometry in statistics? Whatever next?*

Dr J.-B. Denis (BIA-INRA, France) *Biadditivity for square tables*

Dr G Dijksterhuis (ID-DLO, The Netherlands) *Procrustes: where is he now?*

Professor A. de Falguerolles (Université Paul Sabatier, Toulouse, France) *Bilinear models, square tables and biplots*

Professor K.R. Gabriel (University of Rochester, U.S.A.) *Examples of generalized bilinear modelling*

Professor M. Greenacre (Pompeu Fabra University, Barcelona, Spain) *Geometry of block matrices*

Professor W.J. Krzanowski (University of Exeter) *Analysis of distance for structured multivariate data: an extension of MANOVA*

Professor P Legendre (Université de Montréal, Canada) *Corrections for negative eigenvalues: which one to use?*

Professor J.J. Meulman (Leiden University, The Netherlands) *Relational data analysis and graphical display*

Professor J.A. Nelder (Imperial College London) *The analysis of designed experiments; past and present*

Professor R.W. Payne (IACR-Rothamsted) *Diagnostic key construction for allowing for compromise between time and cost of identification*

In December, the Department reverted to its regular series of one-day statistics conferences. The theme this year was *Dynamic Models*. The conference was organised by Paola Sebastiani and Liz Ostrowski and comprised an excellent set of talks. These were:

Professor J.K. Lindsey (Limburgs University, Belgium) *An introduction to hidden Markov models*

Dr W.R. Gilks (MRC Cambridge) *Simulation methodology for dynamic models*

Dr M. Ramoni (Open University) *Bots, battles and Bach: Bayesian clustering by dynamics*

Dr S.J. Godsill (University of Cambridge) *Sequential simulation methods for Bayesian filtering and smoothing*

Professor V.S. Isham (University College London) *Dynamical models for hydrology*

Professor J.Q. Smith (University of Warwick) *Building large scale dynamic Bayes nets*

The Department also runs an active seminar programme with invited external speakers. During the year we heard the following:

Professor J.Q. Smith (University of Warwick) *Discrete Bayesian trees, hidden variables and pretty pictures*

Professor A. O'Hagan (University of Sheffield) *Bayesian uncertainty analysis*

Professor F. Critchley (University of Birmingham) *An invariant look at Euclidean distance matrices*

Dr P. Sasieni (Imperial Cancer Research Fund, London) *From genotypes to genes: doubling the sample size*

Professor G. Dunn (University of Manchester) *Modelling compliance-response relationships in clinical trials*

Professor G.O. Roberts (University of Lancaster) *Bayesian inference for discretely observed diffusions*

Finally, in September, we ran an internal Research Afternoon at which the speakers were Frank Critchley, Paddy Farrington, John Gower and Chris Jones.

7. Conference papers presented

C.P. Farrington

Twentieth Meeting of the International Society for Clinical Biostatistics (ISCB), Heidelberg, Germany: *Case series designs for vaccine safety evaluation*

RSS Theme Conference on Risk (RSS99), Coventry: *Estimating vaccine risks: experience with a new method*

Workshop on Design and Analysis of Infectious Disease Studies, Oberwolfach, Germany: *Estimation of the reproduction number for infectious diseases*

J.C. Gower

Large Scale Data Analysis Conference, Cologne, Germany: *The quality of graphical representations of categorical variables in multidimensional scaling*

D.J. Hand

Henry Stewart Conference, London: *Understanding, applying, and not misusing datamining techniques used in marketing*

M.C. Jones

Recent Advances in Probability and Statistics, Athens, Greece: *A skew t distribution*

Recent Advances in Nonparametric Statistics, Lausanne, Switzerland: *Kernel smoothing: introduction, review and something else*

P. Sebastiani

5th Workshop on Case Studies in Bayesian Statistics, Pittsburgh, USA: *Profiling customers from in-house data*

Highly Structured Stochastic Systems Conference, Pavia, Italy: *Iterative approximation of the marginal likelihood from incomplete data*

Prognostic Models in Medicine: Artificial Intelligence and Decision Analytic Approaches, Aalborg, Denmark: *Robust outcome prediction for intensive-care unit patients*

Uncertainty99, Ft Lauderdale, U.S.A.: *Model folding for data subject to nonresponse*

S.K. Vines

13th Annual Meeting of ASSESS: SPSS users' group: *Generating simple principal components in SPSS*

J. Warwick

22nd Research Students Conference in Probability and Statistics, Bristol: *Minimum distance estimation*

8. Other seminars and talks given

C.P. Farrington

RSS Manchester Local Group: *Analysis of interval-censored survival data*

J.C. Gower

University of Aachen, Germany: *Applications of the Leverrier-Faddeev algorithm*

D.J. Hand

Zeneca Pharmaceuticals: *Synergy*

Zeneca Pharmaceuticals: *Data mining*

RSS East Midlands Local Group, Nottingham: *Data mining: puff or potential?*

University of Odense, Denmark: *Prediction in medical statistics*

M.C. Jones

University of Leeds*: Two separate topics: a skew t distribution and minimum distance estimation*

University of Edinburgh*: Two separate topics: a skew t distribution and minimum distance estimation*

University of St Andrews*: Two separate topics: minimum distance estimation and local dependence*

University of Aberdeen: *Two separate topics: a skew t distribution and local dependence*

RSS Exeter Local Group, Plymouth*: Kernel smoothing: introduction, review and something else*

Imperial College London: *A skew t distribution*

P. Sebastiani

University of Southampton: *Robust Bayesian classification*

University College London: *Robust Bayesian classification*

University of Massachusetts, Amherst, USA.: *Finding episodes boundaries*

9. Editorships

C.P. Farrington

• Editorial board member: *Statistics in Medicine*

• Editorial board member: *Epidemiology and Infection*

J.C. Gower

• Associate editor: *Psychometrika*

• Associate editor: *Journal of Classification*

• Consultant editor: *The Mathematical Scientist*

• Statistical editor: *Journal of Agricultural Science*

D.J. Hand

• Editor-in-Chief (Founding Editor): *Statistics and Computing*

• Editorial board member: *Journal of Applied Statistics*

• Editorial board member: *Statistical Methods in Medical Research*

• Editorial board member: *Data Mining and Knowledge Discovery*

• Editorial board member: *Studies in Classification, Data Analysis, and Knowledge Acquisition* (book series, Springer-Verlag)

• Editorial advisory committee member: *Research in Official Statistics*

• Editorial board member: *Evaluation of Intelligent Systems* (web-site)

M.C. Jones

• Joint Editor: *Journal of the Royal Statistical Society*, Series B

• Associate editor: *Statistics and Computing*

P. Sebastiani

• Editorial board member: *Evaluation of Intelligent Systems* (web-site)

10. Grants awarded

C.P. Farrington

• Estimation of the reproduction number and critical immunisation threshold for infectious diseases. £20,311 from Open University Research Committee.

P. Sebastiani

• Dynamic maps. US$ 60,000 from USA's Air Force Rome Lab.

• Bayesian knowledge extractor. ECU 200,210 from European Community.

11. Other activities

F. Critchley

• is external examiner for the undergraduate and masters statistics programme at the University of Aberdeen

• examined a PhD for the University of St Andrews

C.P. Farrington

• is a member of the nvCJD Epidemiology subcommittee of SEAC (Spongiform Encephalopathy Advisory Committee).

• is an honorary scientist at the Communicable Disease Surveillance Centre, London.

• is an honorary member of the Faculty of Public Health Medicine and an honorary lecturer at the Department of Epidemiology and Population Sciences, London School of Hygiene and Tropical Medicine.

J.C. Gower

• is Honorary Secretary of the Sir Ronald Aylmer Fisher Memorial Committee of Great Britain.

• is a Foreign Member of the Advisory Board of IOPS (Interuniversitaire Onderzoekschool voor Psychometrie en Sociometrie).

D.J. Hand

• is external examiner for the BSc and MSc in Statistics at University College London, and for the MSc in Statistics at the University of Kent.

• examined PhDs for the University of Cambridge and the University of Amsterdam, The Netherlands.

M.C. Jones

• was elected a Fellow of the Institute of Mathematical Statistics

• examined PhDs for the University of Aberdeen, Swiss Federal Institute of Technology, Lausanne and the University of Oslo, Norway.

• is a member of the Editorial Policy Board, the Programme Committee and the Research Section Committee of the Royal Statistical Society.

• spent a short time as Acting Faculty of Mathematics and Computing Sub-Dean (Research).

P.C. Taylor

• is a member of Statistical Computing Section Committee and Programme Committee of the Royal Statistical Society.

S.K. Vines

• is a member of the Anglia and Oxford Multicentre Research Ethics Committee.

• continues to act as advisor on statistical matters to all research students associated with the Open University, and to give appropriate courses and talks.

• is a member of the General Applications Section Committee and Programme Committee of the Royal Statistical Society.

J.E. Williams

• continued in her role as Faculty of Mathematics and Computing Sub-Dean (Courses in Presentation).