APRIL 2004

Fortieth Gregynog Statistical Conference
Programme

The talks will take place in Seminar Room 1 (2nd Floor, far end).

Friday
16 April

16.00 *Tea*
17.00 Professor Mark Steel University of Warwick
Non-Gaussian Bayesian Geostatistical Modelling
19.00 *Dinner*

Saturday
17 April

08.00 *Breakfast*
09.30 Professor Frank Dunstan UW College of Medicine
The UK Biobank study - some epidemiological and statistical issues
11.00 *Coffee*
11.30 Dr Richard Boys University of Newcastle
A Bayesian approach to DNA sequence segmentation using hidden Markov models.
13.00 *Lunch*

Afternoon free

15.30 *Tea*
18.00 *Dinner*
19.45 John Fenlon HRI and University of Warwick
Topics in quantal assay

Sunday
18 April

08.00 *Breakfast*
09.30 Dr Nick Fieller University of Sheffield
Using NMR images as the primary outcome measure in clinical trials
11.00 *Coffee*
11.30 Professor John Gittins University of Oxford
Stochastic Models for the Planning of Pharmaceutical Research
13.00 *Lunch and finish*

Speakers

Dr Richard Boys	University of Newcastle
Professor Frank Dunstan	University of Wales College of Medicine
John Fenlon	Horticulture Research International and University of Warwick
Dr Nick Fieller	University of Sheffield
Professor John Gittins	University of Oxford
Professor Mark Steel	University of Warwick

Staff

Aberystwyth
Alan Jones
Dr John Lane
Sylvia Lutkins

Bangor
Chris Whitaker
Rhiannon Whitaker

Birmingham
Dr Richard Atkinson
Prof Tony Lawrance

Cardiff
Terry Iles

Keele
Prof Peter Jones

Swansea
Dr Mark Kelbert
Dr Alan Watkins

University of Wales College of Medicine
Dr Barry Nix

Warwick
Dr Jim Griffin
Dr Jane Hutton
Dr Jen Marsh
Abdel Berkaoui
Yuan Shen
Wenjuan Zhang

Students

Owen Bodger
Adam Shore
Hannah Finselbach

Jessica Read
Mark Kelly
Venkat Timmaraju

Stephen Conner
Kun Li
Beatriz Penaloza
Judith Anzures Cabrera
Claudia Lozada-Can
Katherine Boyd
Erick Lekone
Hugo Maruri Aguilar
Jionglong Su
Christopher Howitt
Ying Yin
Maria Costa
Yu-Chen Wen
Chih Min Lei
Elina Vladimirou

Abstracts

Non-Gaussian Bayesian Geostatistical Modelling

M.B. Palaciosa *Universidad Publica de Navarra, Spain*
M.F.J. Steel *University of Warwick*

Geostatistical data are considered a partial realization of an underlying random field the index set of which are locations that vary continuously through a fixed region in space. Sampling models are usually based on Gaussian processes. This assumption has an advantage in prediction, since it provides linear predictors justifying the use of the spatial prediction method called kriging. However, distributions of some real data deviate from the Gaussian distribution presenting e.g. heavy tails.

In this paper we propose a more flexible class of sampling models than those based on Gaussian processes. We start from the spatial linear model which is expressed as a spatial trend plus a stationary Gaussian error process. We extend the sampling model to non-Gaussianity by including a scale parameter at each location. We make sure that we obtain a stochastic process satisfying Kolmogorov's conditions. In addition, the scale parameters are spatially correlated to ensure the process is mean square continuous. From the analysis of the sampling properties, we obtain the expressions for the moments and show the relation between one of the hyperparameters of the scale parameter distribution and the kurtosis of the process. In addition, this more general stochastic process allows us to accommodate and identify observations that would be outliers under a Gaussian sampling process, by means of the distribution of the scale parameters. For the spatial correlation structure of the error component, we adopt the flexible Matern class where the smoothness parameter is treated as unknown. Furthermore, the nugget effect is included in the spatial correlation to capture measurement errors and/or microscale variations.

Bayesian inference is performing using a Markov chain Monte Carlo algorithm. Inference on the nugget effect, the smoothness parameter and the range of the correlation function are of special interest. We use a carefully elicited proper prior, and perform a sensitivity analysis in terms of the effects on the posterior distribution. We also perform predictive inference at isolated points and over a regular grid of points. Our methods are illustrated with a variety of data sets. We compare our model with the Gaussian model through Bayes factors.

The UK Biobank study - some epidemiological and statistical issues

Prof Frank Dunstan *University of Wales College of Medicine*

The UK Biobank Study is a huge new longitudinal study, funded by the Department of Health, Wellcome and the MRC, which aims to investigate genetic causes of major diseases and to examine gene-environment interactions. It is controversial for a variety of scientific, financial and ethical reasons. The talk will describe the aims and planned design of the study, outline some of the controversies and discuss some of the statistical considerations that will arise in both planning and analysis stages of the project.

A Bayesian approach to DNA sequence segmentation using hidden Markov models

Dr Richard Boys *University of Newcastle*

Many deoxyribonucleic acid (DNA) sequences display compositional heterogeneity in the form of segments of similar structure. This talk describes a Bayesian method which identifies such segments by using a Markov chain governed by a hidden Markov model. Markov chain Monte Carlo (MCMC) techniques are employed to compute all posterior quantities of interest and, in particular, allow inferences to be made regarding the number of segment types and the order of Markov dependence in the DNA sequence. The method is applied to the segmentation of the bacteriophage lambda genome, a common benchmark sequence used for the comparison of statistical segmentation algorithms.

Topics in Quantal Assay

John Fenlon *HRI and University of Warwick*

Over the past twenty years I have been involved in the design and analysis of bioassay experiments with particular reference to bio-insecticides. In this talk I will focus on various practical and theoretical issues associated with design and analysis of such experiments drawing on a wide variety of examples. After a brief 'historical' introduction I will show why particulate pesticides may need to be considered differently from chemical agents, and discuss several alternative models. With biological control agents time to response and chronic response are also important.

Other important topics relate to variable dose, non-specific mortality and competing risks, together with the treatment of overdispersion.

Using NMR Images as the Primary Outcome in Clinical Trials

Dr Nick Fieller *University of Sheffield*

Medical imaging provides a non-destructive method of direct investigation of effects of treatments on target tissues. This allows tissue to be examined on several occasions during the course of treatment, thus avoiding inter-individual variability. This presentation investigates methodology for assessing the statistical differences between images and hence the effectiveness (or otherwise) of the treatment. The particular study described here involved collection of three-dimensional images by MRI before and after treatment on individuals receiving one of a range of doses. Data extracted from the images were the separate voxel values of a parameter of interest. Statistical analysis focuses on the frequency distributions of these voxel values. The main analysis is based upon functional principal component analysis of the kernel density estimates obtained from each distribution. Statistical assessment of the effects of treatment is based upon a randomization test involving the differences in PC scores between images before and after treatment. The talk outlines the various stages in the development of the analysis, starting with exploratory analysis and then making many simplifying assumptions followed by successively tying up these loose ends.

Stochastic Models for the Planning of Pharmaceutical Research

Prof John Gittins *Oxford University*

Recent work on two aspects of this large topic will be described. Both are concerned with the 'research' or 'discovery' phase of pharmaceutical R&D, during which there is a search for new chemical entities which are sufficiently promising to be used in clinical trials.

First, a stochastic optimisation model for the allocation of resources over the successive stages of a discovery-phase project. Secondly, a statistical procedure for the selection of compounds to be submitted to the screening tests which characterise this search process, and based on the Gittins index.

If time permits I will also say something about work on a Bayesian methodology for determining the size of a clinical trial.