48th Gregynog Statistical Conference
Programme

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

Friday 13th April	16.00	*Tea*	
	17.15	Dr Peter Thwaites *Graphical representation of asymmetric problems*	Leeds
	19.00	*Dinner*	
	20.15	*Workshop on graphical representations*	
Saturday 14th April	08.00	*Breakfast*	
	09.30	Prof Russell Cheng *Optimization by Random Search: Statistical Aspects*	Southampton
	11.00	*Coffee*	
	11.30	Prof Marion Scott *The role of Statistics in environmental science, through to policy, regulation and management*	Glasgow
	13.00	*Lunch*	
		Afternoon free	
	16.00	*Tea*	
	17.30	Jacky Civil *Modelling the risk of oceanic aircraft collisions*	NATS
	19.00	*Dinner*	
	20.15	Dr Huw Llewellyn *Probabilistic reasoning by elimination and its medical applications*	Aberystwyth
Sunday 15th April	8.00	*Breakfast*	
	9.15	Dr Dan Jackson *Handling missing data in smoking cessation trials*	MRC, Cambridge
	10.45	*Coffee*	
	11.15	Dr Martin Kolb *Quasi-stationary distributions for diffusion processes*	Warwick
	12.30	*Lunch and finish*	

Speakers

Dr Peter Thwaites	Leeds
Prof Marion Scott	Glasgow
Jacky Civil	NATS
Dr Martin Kolb	Warwick
Prof Russell Cheng	Southampton
Dr Dan Jackson	MRC, Cambridge
Dr Huw Llewellyn	Aberystwyth

Staff

Students

Aberystwyth

John Lane
Alan Jones & Diane Jones

Jeff Smith
Martin Hathaway
Charlotte Jones-Todd
Luke Williams
Maldwyn Francis
Amy Hodgkins
Cerys Rand

Bangor

Chris Whitaker
Rhiannon Whitaker

Yongzhong Sun
Lu Zou

Birmingham

Jen Marsh (non-resident)

Warwick

Jane Hutton
John Fenlon
Tony Lawrance
John Copas

Monia Ranalli
Francisco J. Rubio
Axel Finke
Ionna Nteka
Amogh Deshpande
Markus Elze
Tomasz Lapinski
Murray Pollock
Yufan Zhao
Dialid Ramirez
Manuele Leonelli
Kirsty Hey
Tom Honnor

Abstracts

Peter Thwaites Leeds

The graphical representation of asymmetric problems

Graphs (or networks) are a convenient means of representing statistical models where there are many variables. If some sort of temporal or causal ordering is known then directed graphs such as Bayesian Networks (BNs) can be used. These graphs illustrate the (conditional) independence structure of a model, if this is simple.

In many cases however, this structure is not simple, and BNs require some sort of modification in order to represent this. As a result the independence structure is no longer described completely by the topology of the graph.

In this talk I describe a class of graphical models which allow the analyst to represent the complete independence structure of a problem through the topology of a simple graph.

Russell Cheng Southampton

Optimization by Random Search; Statistical Aspects

We consider the use of computer simulation in real time decision making to choose, using random search, from a large number of alternative ways of operating a system, but where there is only limited time to carry out the search. A balance has to be struck between making long simulation runs where system performance is accurately measured but only a few alternatives can be considered, and examining a large number of alternatives but using short runs where system performance is poorly estimated. We discuss how to choose simulation run length for optimal balance.

Numerical results involving a real example arising in the provision of fire service emergency cover are presented. This stemmed from work initially carried out for the Fire and Rescue Service in the (then) Home Office Department of the Deputy Prime Minister which investigated how real-time computer simulation models of fire brigade operations might help in handling large incidents. (One of the largest of which occurred at Southampton University shortly after this work started!)

Marian Scott Glasgow

The role of Statistics in environmental science, through to policy, regulation and management

Sensing the Natural Environment: “Sensor networks will produce a revolution in our understanding of the environment by providing observations at temporal and spatial scales that are not currently possible. Expanding observational scales will enable a deeper and broader understanding of environmental variability and change that will, in turn, improve public awareness, enabling better informed public policies and addressing the intrinsic interdependence of human society and the natural environment.” (NSF, 2004). National and international environmental policy setting and evaluation requires a strong and robust evidence base. The key to the delivery of this deeper and broader understanding is the development of spatio-temporal modelling able to handle uncertainty, be computationally robust (and able to deal with massive datasets), to accommodate time-varying (non stationary) spatial processes where the data come from multiple sources, to have an appropriate inferential framework and which can deliver visualisation tools. In addition, such capability will ensure risk informed decision making, in emerging areas such as monitoring of impacts with regard to marine renewable energy developments, security (in an urban environment), water resources (floods, droughts, quality and quantity), and carbon budgets.

Regulation: Within the European Union, there are a number of regulatory frameworks dealing with the aquatic environment, of which the Water Framework Directive (WFD, 2003), transposed in Scotland to protect, improve and promote sustainable sustainable use of

Scotland's water environment, is perhaps one of the most significant. Two others that will be considered are the Bathing Waters Directive (BWD, 2006), for predicting microbiological health risk and the Floods Directive (FD, 2009), subsequently the Flood Risk Management (Scotland) Act which requires a national assessment of flood risk by 2011, and flood risk and hazard maps by 2013.
The directives frequently define in a generic way the sampling that is required, eg (WFD, Monitoring for the water framework directive 2000/60/EC): "Member states must ensure that enough individual water bodies of each water type are monitored and determine how many stations are required to determine the ecological and chemical status of the water body". In the BWD, data (and hence sampling) requirements are also defined for any bathing water assessment. Monitoring networks (ideally long term and stable) provide the key evidence base for change, yet IPCC (2008) commented that "observational data and data access are pre-requisites for adaptive management, yet many observational networks are shrinking."

In terms of reporting on a policy statement, there are three basic questions:
a) What is happening?
b) Why is it happening?
c) Are the changes significant?
Making sense of the Natural Environment The ability to visualise complex data, such as those which might be generated from an array of environmental sensors is an important skill, enhancing understanding of the system being studied and facilitating communication of the results to both technical and non-technical audiences. Within the legislative requirement that EU governments must report on the state of the environment, until recently this has taken the form of a published report, but it is intended that future reports should be immediate and interactive so that citizens and other users might interrogate the reports.

Making sense of the Natural Environment The ability to visualise complex data, such as those which might be generated from an array of environmental sensors is an important skill, enhancing understanding of the system being studied and facilitating communication of the results to both technical and non-technical audiences. Within the legislative requirement that EU governments must report on the state of the environment, until recently this has taken the form of a published report, but it is intended that future reports should be immediate and interactive so that citizens and other users might interrogate the reports.
Natural visualisation techniques include time series and spatial surfaces plots, and multivariate representations of the patterns of relationships between the different determinands. However these displays typically do not show the dynamic nature of the system being studied, nor the landscape on which the system operates so that more recent developments have included introducing animation to demonstrate the spatio-temporal development of the system under study and the changing relationships.

Statistical challenges: There remain however a number of challenges in terms of design of sensor network systems, handling potential large volumes of data, pre-processing, including temporal and spatial co-location issues over the array and then ultimately modelling and visualisation. These challenges are of particular importance in the environment reporting developments mentioned in the opening paragraph but also have wider relevance and applicability.

Jacky Civil, Operational Analysis Department, NATS

Modelling the Risk of Oceanic Aircraft Collisions

The assurance of aircraft and passenger safety is the first priority in air traffic management. As such, extensive risk modelling and simulation work is required before any proposed changes to procedure, airspace design or separation rules can be trialled and implemented. This work explores the risk models used to assess a proposal to reduce by half the minimum separation between two in-trail aircraft flying the same route over the North Atlantic Ocean. A distribution is derived for the gain/loss of separation between two in-trail aircraft over a fixed period of time, with components due to weather forecast errors, aircraft speed-keeping errors and navigational equipment inaccuracy.

Huw Llewelyn, Aberystwyth

Probabilistic reasoning by elimination and its medical applications

Reasoning by elimination involves listing possibilities and then showing that all but one is not possible. For example, if there are a number people on the Orient Express and one of them must have committed a murder and if all but one has an alibi, then the one without an alibi must be the murderer!

In practice of course, it is not that simple. For example, if there is tenderness in the right lower abdomen (t), it is probably appendicitis (A) or 'nothing important' (N) and probably not something else (S). If there is localised rigidity (r), then as this occurs rarely in 'nothing important' and commonly in appendicitis, so it is unlikely to be 'nothing important' and is probably appendicitis.

(1) The probability of appendicitis during such reasoning can be calculated as follows by using only one finding at a time to 'eliminate' one diagnosis at a time (i.e. without using the independence assumption):

$p(A|t\&r) \geq 1/\{1 + [p(N)*p(r|N) + p(t)*p(S|t)] / [p(A)*\{p(t|A) + p(r/A) + 2 - 1\}]\}$

(eg) $1/\{1 + [\ 0.5 * 0.04 + 0.67*0.01] / [\ 0.33*\{\ 0.75 + 0.8 + 2 - 1\}]\} = 0.873$

(2) By making a dependence assumption we can produce an approximation:

$p(A|t\&r) \approx 1/\{1 + [p(N)*p(r|N)]/[p(A)*p(r|A)] + [p(t)*p(S|t)]/[p(A)*p(t|A)]\}$

$1/\{1 + [\ 0.5 * 0.04\ ] / [0.33 * 0.80] + [0.67 * 0.01]/[\ 0.33* 0.75\ ]\} = 0.908$

(3) By applying Bayes theorem and rearranging the above expression (2) we can dispense with all the likelihoods:

$p(A|t\&r) \approx 1/\{1+p(N|r)/p(A|r) + p(S|t)/p(A|t)]\}$

$1/\{1+ 0.05 / 0.67 + 0.01 / 0.375\ \} = 0.908$

These expressions allow the validity of probabilistic reasoning by elimination to be tested by examining the relevant frequencies or distributions used in the reasoning process. The first step in the diagnostic process is to consider the possibility of treatable disease. This happens by the patient complaining of a symptom or by performing a population screening test. This first step can be analysed using Bayes theorem.

The next step is to form a differential diagnosis – a list of possibilities, the shorter the better. We then choose a diagnosis from this list and to try to confirm it by looking for a finding that occurs commonly in the chosen diagnosis and rarely in another in the list. (A finding will never occur in another diagnosis only if it represents the absence of one of its necessary criteria or if it is a sufficient criterion of the 'chosen' diagnosis). This step uses ratios of sensitivities, which may include ratios of likelihood densities.

The reasoning process can also be based on combining diagnostic probabilities based on single findings at a time by using expression (3), which dispenses with the need to use likelihoods in the calculation. In addition to reasoning with differential diagnoses, probabilistic reasoning by elimination can be used to assess the reliability of data gathered from individual patients and from groups of patients e.g. in clinical trials.