One-day workshop on Modern Statistical Methods in Health and Environment

Starts: Friday 9June 2017 10:00
Ends: Friday 9 June 2017 17:30

The workshop is sponsored by Brunel University and jointly organized with the Emerging Application Section (EAS) of the RSS

Workshop programme

10.00 Registration and Coffee

10.10 Welcome

10.20 Mark Girolami (Imperial College London)

A stochastic formulation of a dynamical singly constrained spatial interaction model

11.10 Richard Chandler (University College of London)

Natural hazards, risk and uncertainty – some hot topics

12:00 Clifford Lam (London School of Economics)

Nonlinear Shrinkage Estimation in Quadratic Inference Function Analysis for Correlated Data

12:50 Lunch

13.50 Nicky Best (GSK)

Constructing and using informative Bayesian priors in drug development

14.40 Marta Blangiardo (Imperial College London)

Investigating the health effect of multi-pollutant exposure using a time series approach

15.30 Coffee break

15.50 Bianca de Stavola (University College of London)

Multiple questions for multiple mediators

16.40 Richard Samworth (University of Cambridge)

High-dimensional changepoint estimation via sparse projection

17:30 End

Appendix:

Mark Girolami (Imperial College London)

Title: A stochastic formulation of a dynamical singly constrained spatial interaction model

Abstract: One of the challenges of 21st-century science is to model the evolution of complex systems. One example of practical importance is urban structure, for which the dynamics may be described by a series of non-linear first-order ordinary differential equations. Whilst this approach provides a reasonable model of spatial interaction as are relevant in areas diverse as public health and urban retail structure, it is somewhat restrictive owing to uncertainties arising in the modelling process.

We address these shortcomings by developing a dynamical singly constrained spatial interaction model, based on a system of stochastic differential equations. Our model is ergodic and the invariant distribution encodes our prior knowledge of spatio-temporal interactions. We proceed by performing inference and prediction in a Bayesian setting, and explore the resulting probability distributions with a position-specific metropolis-adjusted Langevin algorithm. Insights from studies of interactions within the city of London from retail structure are used as illustration.

Richard Samworth (University of Cambridge)

Title: High-dimensional changepoint estimation via sparse projection

Abstract: Changepoints are a very common feature of Big Data that arrive in the form of a data stream. We study high-dimensional time series in which, at certain time points, the mean structure changes in a sparse subset of the coordinates. The challenge is to borrow strength across the coordinates in order to detect smaller changes than could be observed in any individual component series. We propose a two-stage procedure called 'inspect' for estimation of the changepoints: first, we argue that a good projection direction can be obtained as the leading left singular vector of the matrix that solves a convex optimisation problem derived from the CUSUM transformation of the time series. We then apply an existing univariate changepoint detection algorithm to the projected series. Our theory provides strong guarantees on both the number of estimated changepoints and the rates of convergence of their locations, and our numerical studies validate its highly competitive empirical performance for a wide range of data generating mechanisms.

Richard Chandler (UCL)

Title: Natural hazards, risk and uncertainty – some hot topics

Abstract: Society is, and has always been, vulnerable to damage resulting from natural disasters: earthquakes, floods, wind storms, tsunami and so on. In recent decades, the global human and economic cost of such disasters has increased rapidly: reasons for this include an increasing global population, infrastructure construction in regions that were previously undeveloped, and the effects of climate and other environmental change. The 2015 United Nations Global Assessment Report on Disaster Risk Reduction estimated the future cost of natural disasters at 314 billion US dollars per year, in terms of impacts on the built environment alone. There is considerable incentive, therefore, to develop a better understanding of natural hazards and the associated risks, in order to inform strategies for improving societal resilience.

For all natural hazards, our understanding of the relevant processes comes from a combination of data and models. But data are often sparse, incomplete and prone to errors or inhomogeneities; and, as a well-known statistician once remarked, all models are wrong. Our understanding of natural hazards and their consequences is inevitably incomplete, therefore. Hazard scientists, planners and policymakers must acknowledge the underlying uncertainties, and account for them appropriately. However, scientists often lack training in the kinds of statistical techniques that are required to characterise and communicate uncertainty appropriately in problems that are often highly complex; and planners and policymakers often lack training in how to make rational and robust decisions in the presence of uncertainty. At the same time, the last 20 years have seen enormous progress by the statistical and other communities in tackling the relevant issues. This creates tremendously exciting opportunities for collaboration, with potential to reshape the way in which risk from natural disasters is handled: in the UK this has been recognised by the Natural Environment Research Council in their funding of a four-year research programme entitled Probability, Uncertainty and Risk in the Environment (PURE). In this talk I will present examples to illustrate some of these opportunities, partly drawing on my own experience from the PURE programme, with application areas including the construction of seismic hazard maps, earthquake engineering and the assessment of climate change impacts.

Marta Blangiardo (Imperial College London)

Title: Investigating the health effect of multi-pollutant exposure using a time series approach

Abstract: Airborne particles are a complex mix of organic and inorganic compounds, with a range of physical and chemical properties. The analysis of air pollution exposure as made up of multiple metrics of pollutants and/or sources, as well as the quantification of the magnitude of their simultaneous health effect, represents a new challenging aspect of epidemiological research.

In this talk I will present two approaches to analyse the simultaneous effect of multiple air pollutants on short-term health outcomes within a time series framework. The first consists of a clustering technique to reduce the exposure dimensionality while the second is based on a joint hierarchical model of the exposure and health outcome. I will then show how the proposed approaches work on two examples to evaluate the effect of several air pollutants on mortality in Greater London.

Clifford Lam (London School of Economics)

Title: Nonlinear Shrinkage Estimation in Quadratic Inference Function Analysis for Correlated Data

Abstract: Quadratic inference function (QIF) analysis is more efficient than the generalized estimating equations (GEE) approach when the working covariance matrices for the data are misspecified. Since QIF naturally requires a weighting matrix which is the inverse of a sample covariance matrix of non-identically distributed data, finite sample performance can be greatly affected when the number of independent data points is not large enough, which is usually the case in cluster randomized trials or many longitudinal studies. While nonlinear shrinkage is very successful in regularizing the extreme eigenvalues of a sample covariance matrix, the method is only restricted to independent and identically distributed data. We propose a novel nonlinear shrinkage approach for a sample covariance matrix of non-identically distributed data, which improves finite sample performance of QIF, and gives room for increasing the potential number of working correlation structures for even better performance. Together with a nonlinearly shrunk weighting matrix, we derive the asymptotic normality of the parameter estimators, and propose another nonlinear shrinkage approach to estimate the asymptotic covariance matrix more accurately. We demonstrate the performance of the proposed methods through simulation experiments and a real data analysis.

Bianca de Stavola (UCL Great Ormond Street Institute of Child Health)

Title: Multiple questions for multiple mediators

Investigating the mechanisms that may explain the causal links between an exposure and a temporally distal outcome often involves multiple interdependent mediators. Until recently, dealing with multiple mediators was restricted to settings where mediators relate to exposure and outcome only linearly. Extensions proposed in the causal inference literature to allow for interactions and non-linearities in the presence of multiple mediators initially concentrated on natural direct and indirect effects. These however are not all identifiable. More recent developments have focussed on interventional direct and indirect effects (Vansteelandt & Daniel, 2017) which can be identified under less restrictive assumptions, with generalizations dealing with time-varying exposures, mediators and confounders also possible (VanderWeele & Tchetgen Tchetgen, 2017).
The questions that can be addressed when estimating interventional effects differ from those asked by natural effects in subtle ways. In this talk I will review them, discuss their differences in emphasis, assumptions, and interpretation, and propose ways of exploiting these differences to assess the robustness of our conclusions using an epidemiological investigation of the mechanisms linking maternal pre-pregnancy weight status and offspring eating disorders behaviours to illustrate these points.

References
VanderWeele TJ, Tchetgen Tchetgen EJ. Mediation analysis with time-varying exposures and mediators. J R Stat Soc B (in press).
Vansteelandt S; Daniel RM. Causal mediation analysis with multiple mediators Epidemiology, 2017; 28 (2): 258–265.

Nicky Best (Statistical Innovation Group, GSK)

Title: Constructing and using informative Bayesian priors in drug development

In recent years, there have been several drivers that have led to a growing interest in use of Bayesian methods at all stages of pharmaceutical product development. Drug development is increasingly costly, risky and inefficient, prompting the search for innovative clinical trial designs to increase flexibility, improve decision making and maximize the use of accumulated knowledge. The status quo of null hypothesis significance testing and $p<0.05$ as the evidential standard for drawing conclusions about a scientific hypothesis has also been challenged by scientists across many disciplines in recent years. This led to the American Statistical Association issuing a statement on the subject last year, and to the organisation of a special workshop involving leading US regulatory, industry and academic statisticians to debate the meaning of “substantial evidence” and explore opportunities to advance the use of Bayesian methods for drug development and regulatory decision making.

In this talk, I will discuss two specific settings in which Bayesian methods and informative priors have played a critical role in a drug development setting, focusing particularly on the rationale and methods used to construct the prior. The first setting relates to paediatric clinical trials, where there is often relevant historical evidence of efficacy available from confirmatory trials of the same drug in the adult population. Several methods are available for constructing a ‘dynamic borrowing’ prior, in which the amount of historical information included in the posterior distribution for the paediatric trial depends on the extent of agreement between the historical and current data sources. The second setting relates to internal company decision making based on predicted probability of success of a future trial design. At GSK, we routinely use expert elicitation methods to construct prior distributions that represent our current beliefs about the efficacy of a drug. These priors are then used to compute the assurance (probability of success or expected power) for different trial designs at the next stage of development.