Hierarchical Models and Markov Chain Monte Carlo

Conference in Honour of Adrian F.M. Smith

2 - 5 June 2011, Crete, Greece

Book of Poster Presentations

(Posters are sorted in the order submitted to the Organizing Committee)

Poster 1

Presenter's name:

Nuttanan Wichitaksorn (University of Sydney Business School)

Authors names:

Nuttanan Wichitaksorn and S.T. Boris Choy (University of Sydney Business School)

Title:

Bayesian Analysis of Seeming Unrelated Tobit Models through Copula and Scale-Mixture Margins

Abstract:

In this study, we extend our analysis on copula-based seemingly unrelated Tobit (SUR Tobit) model by expressing the marginal distributions of error terms as a scale mixture of normal. We model the dependence structure of our SUR Tobit model through Gaussian copula. To simplify the implementation of our model, we express the marginal Student-t errors as a scale mixture of normal and inverse-gamma distributions and Laplace errors as a scale mixture of normal and exponential distributions. The estimation results are presented through simulation and empirical studies.

Poster 2

Presenter's name:

Lizanne Raubenheimer (Rhodes University)

Authors names:

Lizanne Raubenheimer (Rhodes University) and Abrie J. van der Merwe (University of the Free State)

Title:

Bayesian Estimation of the Product of Binomial Proportions

Abstract:

In this paper we look at the procedure of deriving the probability matching prior for the product of different powers of *k* Binomial parameters. Datta and Ghosh (1995) derived the differential equation which a prior must satisfy if the posterior probability of a one sided credibility interval for a parametric function and its frequentist probability agree up to a certain point. In the case of two and three independently distributed Binomial variables, the Jeffreys', uniform and probability matching priors for the product of the parameters are compared. The weighted Monte Carlo method is used for the simulation from the posterior distribution in the case of the probability matching prior.

The product of different powers of *k* Binomial parameters can be used in applications to system reliability. The probability of system failure is also studied in cases where at least one of two types of components are required to be employed and where three components in parallel are needed.

Poster 3

Presenter's name:

Madeleine Cule (University of Oxford)

Authors names:

Madeleine Cule and Peter Donnelly (University of Oxford)

Title:

Modelling the transmission of Clostridium Difficile in hospitals

Abstract:

Clostridium difficile is a major cause of healthcare-associated diarrhoea, and controlling its spread within a healthcare setting is an ongoing focus of significant public health effort. However, the dynamics of the disease are still relatively poorly understood, and the complex, dynamic contact network of a large hospital system makes it difficult to apply standard techniques for analysing disease outbreaks. Focussing on data collected within

the Oxford Radcliffe Hospitals NHS Trust between 2007 and 2009 (containing a total of 250,000 hospital admissions and 1,500 confirmed cases of C. difficile), we describe a compartmental model for the transmission of C. difficile within the hospital system. A Bayesian approach allows us to estimate key epidemiological parameters, as well as to assess the impact of assumptions made about the dynamics of disease and the potential impact of interventions. An advantage of this framework is that it enables us to introduce additional genetic typing information about the bacteria as it becomes available, allowing us to refine our estimates of key parameters.

This is joint work with the UKCRC consortium "Modernising Medical Microbiology" and the Oxford Biomedical Research Centre.

Poster 4

Presenter's name:

Boris Choy (University of Sydney)

Authors names:

Boris Choy and C.P.Y. Lam (University of Sydney)

Title:

A new representation for skew normal distribution

Abstract:

This paper presents a new presentation for the skew normal distribution via scale mixtures. The scale mixtures form enables the skew normal distribution to be implemented more efficiently using Gibbs sampling algorithms. The technique can be easily extended to other well known skewed distributions such as skew *t*, skew variance gamma, skew Laplace and skew generalised error distributions. Empirical study on financial time series data will be performed**.**

Poster 5

Presenter's name:

Robert E. Weiss (UCLA School of Public Health)

Authors names:

Robert E. Weiss and Lei Qian (UCLA School of Public Health)

Title:

Bayesian Semiparametric Nonstationary Correlation Models for Longitudinal Data

Abstract:

We develop a matrix mixture correlation model for continuous longitudinal data with nonstationary correlations. We consider two situations, one where there is a known change point causing nonstationarity, and a second situation where the correlations change smoothly over time. The former is appropriate when a known change in the biological system has a substantial impact on outcomes. The second situation with continuously changing correlation happens for example in growth curve studies where growth in early life is very different from later growth. Our model allows both the correlation values and underlying structure to change over time and is capable of handling highly unbalanced data with a large number of repeated measurements. To illustrate the model, we fit our model to unbalanced CD4 cell counts from HIV seroconvertors and to a mice growth dataset.

Poster 6

Presenter's name:

Michail Papathomas (Conventry University and Imperial College London)

Authors names:

Michail Papathomas, John Molitor, Sylvia Richardson, Clive Hoggart and Paolo Vineis (Imperial College London)

Title:

Variable selection when modelling with the Dirichlet process, with an application in a genome-wide association study.

Abstract:

Standard regression analyses are often plagued with problems encountered when one tries to make meaningful inference going beyond main effects, using datasets that contain dozens of potentially correlated variables . We propose a method that addresses these problems by using, as its basic unit of inference, a profile, formed from a sequence of covariate values. These covariate profiles are clustered into groups using the Dirichlet process, and are associated via a regression model to a relevant outcome. Different variable selection approaches are introduced and compared. We apply these variable selection methods to a GWA study on lung cancer, in order to explore gene-gene and gene-environment interactions.

Poster 7

Presenter's name:

Sarah Michalak (Los Alamos National Laboratory)

Authors names:

Todd Graves, Sarah Michalak, and Lori Pritchett-Sheats (Los Alamos National Laboratory)

Title:

Hierarchical Models for High-Performance Computing Application-Failure and Hardware-Failure Data

Abstract:

Application-failure data and hardware-failure data can provide contrasting information about high-performance computing (HPC) platform health. For example, hardware-failure data may indicate that the system is in a typical state, while application users may be experiencing an unusually high number of failures. Using application-failure and hardware-failure data jointly to understand HPC platform health should provide a more accurate description of system health. This poster presents preliminary (non-joint) models for application-failure data and for hardware-failure data and considers on how the two might be combined.

Poster 8

Presenter's name:

Matti Pirinen (University of Oxford)

Authors names:

Matti Pirinen, Garrett Hellenthal, Colin Freeman, Zhan Su, Peter Donnelly and Chris Spencer (Wellcome Trust Centre for Human Genetics, University of Oxford)

Title:

Modelling Diverse Genetic Backgrounds within Individuals in Genome-wide Association Studies

Abstract:

Contemporary genome-wide association studies (GWAS) consider genotype data on thousands of individuals at up to millions of polymorphic loci. A central question is to quantify the statistical evidence about connections between the genetic variants and the phenotypes of interest. In the past few years GWAS have identified many loci that show convincing statistical association with complex human diseases, and the expectation is that these signals will lead us closer to the actual biological mechanisms behind the phenotypes.

Unless a causal variant is directly genotyped, we are only able to detect association signals because a typed marker is correlated with a causal variant (most likely) due to their spatial proximity on the genome. The strength of the correlation between the two depends on the historical genetic events on the region, such as mutations and recombinations, and consequently may be different in different populations. Thus, when studying for example African-Americans or Latinos, who have recent ancestry in several genetically different backgrounds, some association signals may be blurred if differences in local ancestries are not taken into account.

We introduce a method for association analysis in presence of diverse genetic backgrounds using hierarchical Bayesian modelling. In the first stage we phase diploid genotypes into chromosomes whose local ancestries we probabilistically estimate using Hidden Markov Model techniques. In the second stage we carry out the association analysis by modelling the effect sizes hierarchically and by accounting for the uncertainty in the local ancestry estimates. As special cases this framework covers common situations where the ancestries of individuals are assumed known from some external information (e.g. meta-analysis across several populations, modelling genotype-environment interactions).

We make use of an MCMC algorithm to estimate the effect size parameters and importance sampling to compute the marginal likelihoods. Despite of the use of sampling algorithms the method is readily applicable to the genome-wide data sets that are encountered in contemporary human genetics.

Poster 9

Presenter's name:

Tatiana Xifara (Lancaster University)

Authors names:

Tatiana Xifara and Chris Sherlock (Lancaster University)

Title:

Adaptive MCMC Using a Hidden Markov Model for Disease Interactions

Abstract:

In natural populations, animals are likely to be infected by different diseases either simultaneously or successively. Ecologists are interested in the interactions between parasite species. A longitudinal study was designed for that purpose by recording the sequences of infections events in a population of field voles. As the dataset contains many missing observations, a simple analysis is not preferred as it would entail discarding much of the dataset. We offer a more powerful alternative. For each disease *d*, we wish to evaluate the way in which the presence or absence of each of the other diseases affects the probability of contracting *d* and additionally, for some of the diseases, the time which infection with *d* lasts. We use a hidden Markov model for each disease and perform inference via a Gibbs sampler. One iteration of the Gibbs sampler cycles through each of the diseases, first sampling from the conditional posterior for the parameters for that particular disease using an adaptive Metropolis-Hastings step and then sampling from the hidden disease states using the Forward-Backward algorithm.

Poster 10

Presenter's name:

Dimitris Fouskakis (National Technical University of Athens)

Authors names:

Dimitris Fouskakis (National Technical University of Athens), Ioannis Ntzoufras (Athens University of Economics and Business) and David Draper (University of California, Santa Cruz)

Title:

Power Intrinsic Variable Selection for Normal Models based on Zellner's g-Prior

Abstract:

In order to express prior ignorance in Bayesian variable selection problems, proper prior distributions with large variances or non-informative improper distributions can be used. Bayes factors are well known for their sensitivity on prior variances, while, when using improper priors, Bayes factors cannot be determined because of the involvement of the unknown normalizing constants. This has urged the Bayesian community to develop various methodologies to overcome the problem of prior specification in model comparison and variable selection problems. An important part of this research is focused on the so-called objective model selection methods having their source on the intrinsic priors in order to provide an approximate proper Bayesian interpretation for intrinsic

Bayes factors. These intrinsic priors use improper priors as a starting point and overcome the problem of indeterminacy of the Bayes factor since the same constant is involved in all marginal likelihoods. In this paper we develop the methodology of intrinsic priors when using proper priors as a starting point. Specifically we focus on normal linear models and we use initially the Normal-inverse gamma Zellner's g-prior. We introduce the power intrinsic Zellner's g-prior where we use the intrinsic prior methodology in order to define the joint prior distribution of the model parameters and the error variance. Moreover, by borrowing ideas from the power prior approach we avoid the use of a minimal training sample and the sensitivity of posterior results on the selection (and size) of this training sample in our intrinsic prior methodology. The methodology is illustrated on both simulated and real examples and sensitivity analysis reveals broad stability of our conclusions.

Poster 11

Presenter's name:

Jukka Ranta (Risk Assessment Unit, Finnish Food Safety Authority Evira)

Authors names:

Jukka Ranta, Antti Mikkelä and Pirkko Tuominen (Risk Assessment Unit, Finnish Food Safety Authority Evira)

Title:

Bayesian methods in food safety risk assessment

Abstract:

The growing area of food safety risk assessments requires probabilistic methods to deal with problems concerning missing data, hidden variable processes, surveillance and biased or selective sampling, traceability, hierarchical dependency structures, expert knowledge elicitation, information synthesis, decision analysis, predictive modelling, source attribution methods, intervention and scenario simulations. New research for methods that are applicable for these, typically 'ill-posed' and 'nonstandard', problems is needed when available information and data are structured but always unbalanced and partial. Hence, there is scope for developing novel applications using the general framework of Bayesian methodology with hierarchical models and MCMC. Also, integration of decision analysis methods with models of food production chains under uncertainty would be fruitful for more comprehensive analysis of decision options, while accounting for uncertainties in a coherent manner. Some application examples are presented, over a decade 2001-2011 at Risk Assessment Unit, Finnish Food Safety Authority Evira.

Poster 12

Presenter's name:

Tony Pettitt (Queensland University of Technology)

Authors names:

Christopher Drovandi (Queensland University of Technology), Tony Pettitt (Queensland University of Technology), Robert Henderson (Royal Brisbane and Women's Hospital, Brisbane, Queensland) and Pamela McCombe (Royal Brisbane and Women's Hospital, Brisbane, Queensland and University of Queensland, Brisbane)

Title:

Hierarchical Bayesian modelling for motor unit number estimation: the role of the observed data likelihood.

Abstract:

Motor unit number estimation (MUNE) is a method which aims to provide a key quantitative indicator of disease progression for patients suffering from diseases of loss of motor units such as Motor Neuron disease. However the development of a reliable, repeatable and fast real-time MUNE method has proved elusive hitherto.

Ridall et al. (2007) implements a reversible jump Markov chain Monte Carlo (RJMCMC) algorithm to produce a posterior for the number of motor units using a Bayesian hierarchical model that takes into account biological information of motor unit activation. However we find that the approach in general fails the reliability objective on a wide range of clinical datasets as it generally suffers from poor cross-dimensional mixing.

The focus of this poster is to examine methods of inference that use the observed data likelihood. In particular we explore how the observed data likelihood can improve RJMCMC and investigate an alternative approach using the posterior distribution of the observed data likelihood, Aitkin, Liu and Chadwick (2009). The observed data likelihood for this model involves an intractable summation over all combinations of a set of latent binary variables whose joint sample space increases exponentially with the number of units. We provide a tractable and accurate approximation for this quantity. We compare the Aitkin, Liu and Chadwick (2009) method with the newly developed gold standard RJMCMC algorithm that requires very little user intervention. We find that the Aitkin, Liu and Chadwick (2009) method leads to biases for inference of the number of motor units.

References

Aitkin, M., Liu, C. C. and Chadwick, T. (2009). Bayesian model comparison and model averaging for small-area estimation. Annals of Applied Statistics 3 199-221.

Ridall, P., Pettitt, A., Friel, N., McCombe, P. and Henderson, R. (2007). Motor unit number estimation using reversible jump Markov chain Monte Carlo Methods (with Discussion). Journal of the Royal Statistical Society: Series C (Applied Statistics) 56 235-269.

Poster 13

Presenter's name:

Laura Ventura (University of Padova, Italy), Nicola Sartori (University of Padova, Italy) and Walter Racugno (University of Cagliari, Italy)

Authors names:

Laura Ventura (University of Padova, Italy), Nicola Sartori (University of Padova, Italy) and Walter Racugno (University of Cagliari, Italy)

Title:

A class of strong matching priors for higher-order Bayesian inference

Abstract:

For inference on a scalar parameter of interest ψ in the presence of a nuisance parameter λ, standard Bayesian techniques require a prior probability distribution $\pi(\psi,\lambda)$ on the entire parameter and are typically based on the marginal posterior distribution for ψ, integrating out λ. This procedure requires elicitation of $\pi(\psi,\lambda)$, which may be difficult both in the subjective and objective Bayesian contexts, and the computation of a multidimensional integral, which can be heavy when λ is high-dimensional. To avoid these drawbacks, recent Bayesian techniques use the modified profile likelihood $L_{mp}(\psi)$ as the basis for the posterior distribution $\pi_{mp}(\psi\,|y) \propto \pi\,(\psi)L_{mp}(\psi)$, where $\pi\,(\psi)$ is a prior distribution on ψ only. Although $\pi_{mp}(\psi\,|y)$ cannot always be considered as orthodox in a Bayesian setting, there is nowadays an extensive literature on the use of alternative likelihoods and on the associated default priors (see, e.g., Chang and Mukerjee, 2006, Chang et al., 2009, Ventura et al., 2009, Ventura and Racugno, 2011, and references therein).

In this contribution, starting from higher-order asymptotic expansions for $\pi_{mp}(\psi\,|y)$, we suggest a class of strong matching priors (Fraser and Reid, 2002) for ψ, i.e. priors for which there is a strong agreement of frequentist and Bayesian parametric inferential conclusions. These priors allow to derive a simple explicit higher-order approximation for $\pi_{mp}(\psi\,|y)$ and accurate asymptotic highest posterior density credible sets for ψ, even for

small sample sizes. Examples in the context of exponential families are discussed. We also show how accurate Bayesian inference can be performed in practice using likelihood asymptotics and the recent R package bundle hoa (Brazzale et al., 2007).

References

Brazzale, A.R., Davison, A.C., Reid, N. (2007). Applied Asymptotics. Cambridge University Press, Cambridge.

Chang, H., Mukerjee, R. (2006). Probability matching property of adjusted likelihoods. Statist.Probab.Lett., 76, 838–842.

Chang, H., Kim, B.H., Mukerjee, R. (2009). Bayesian and frequentist confidence intervals via adjusted likelihoods under prior specification on the interest parameter. Statistics, 43, 203-211.

Fraser, D.A.S., Reid, N. (2002). Strong matching of frequentist and Bayesian parametric inference. J.Statist.Plan.Inf., 103, 263–285.

Ventura, L., Cabras, S., Racugno, W. (2009). Prior distributions from pseudo-likelihoods in the presence of nuisance parameters. J.Amer.Stat.Assoc., 104, 768–774.

Ventura, L., Racugno, W. (2011). Recent advances on Bayesian inference for P(X < Y). Bayesian Analysis, to appear.

Poster 14

Presenter's name:

Fahimah Al-Awadhi (Kuwait University)

Authors names:

Merrilee Hurn (University of Bath), Peter J. Green (University of Bristol) and Fahimah Al-Awadhi (Kuwait University)

Title:

A Bayesian Hierarchical Model for Photometric Redshifts

Abstract:

The Sloan Digital Sky Survey (SDSS) is an extremely large astronomical survey conducted with the intention of mapping more than a quarter of the sky. Among the data it is generating are spectroscopic and photometric measurements, both containing

information about the redshift of galaxies. The former is precise and easy to interpret but expensive to gather, the latter is far cheaper but correspondingly harder to interpret. A recent paper by Csabai *et al.* (2003) describes various calibration techniques aiming to predict redshift from photometric measurements. In this paper, we investigate what a structured Bayesian approach to the problem can add. In particular, we are interested in providing uncertainty bounds associated with the underlying redshifts and the classifications of the galaxies. We find that a quite generic statistical modelling approach, using for the most part standard model ingredients, can compete with much more specific custom-made and highly-tuned techniques already available in the astronomical literature.

Poster 15

Presenter's name:

Ioannis Ntzoufras (Athens University of Economics and Business)

Authors names:

Ioannis Ntzoufras (Athens University of Economics and Business) and Anastasia Lykou (Lancaster University)

Title:

On Bayesian Lasso Variable Selection and the Specification of the Shrinkage Parameter

Abstract:

In this work, we propose a Bayesian implementation of the Lasso regression that accomplishes both shrinkage and variable selection. We focus on the appropriate specification for the shrinkage parameter λ through Bayes factors that evaluate the inclusion of each covariate in the model formulation. We associate this parameter with the values of Pearson and partial correlation at the limits between significance and insignificance as defined by Bayes factors. By this way, a meaningful interpretation of λ is achieved that leads to a simple specification of this parameter which is of prominent importance in Lasso literature. Finally, the usefulness of a hierarchical model using a hyperprior on the shrinkage parameter is also explored.

Poster 16

Presenter's name:

Paul Damien (University of Texas at Austin)

Authors names:

Paul Damien (University of Texas at Austin), Prakash Laud (Medical College of Wisconsin), Stephen Walker (University of Kent), and Daniel Zantedeschi (University of Texas at Austin)

Title:

Direct Sampling

Abstract:

In recent years Markov chain Monte Carlo (MCMC) methods have been used to provide a full Bayesian analysis, when the posterior joint distribution of interest is both analytically intractable, and it is not known how to draw independent samples. In this paper, a non-MCMC approach to sampling from posterior distributions in a Bayesian analysis is developed and illustrated. Some sampling problems, now thought to be best handled by MCMC methods alone, are tackled efficiently by independent samples.

Poster 17

Presenter's name:

Mu Niu (University of Sheffield)

Authors names:

Mu Niu, Paul Blackwell, Caitlin Buck, Tim Heaton (University of Sheffield)

Title:

Bayesian Modelling radiocarbon calibration curves with Wiener process

Abstract:

The level of radioactive carbon, ^{14}C, in the atmosphere has fluctuated through time and the mechanisms by which ^{14}C enters living organisms are not well understood. Given these complications, a data-based calibration curve is required in order for radiocarbon dating to be a viable absolute dating method. The data available for estimating this curve consist of historic materials which possess a radiocarbon determination together with an independently found estimate of calendar age, both of which are subject to error.

In this poster, we describe a tailor-made, Bayesian, non-parametric regression approach with observations at imprecise times which the authors have developed as part of their work with the wider research group who provide the internationally-agreed radiocarbon calibration curves. The underlying function which relates radiocarbon to calendar years is

modelled as a Wiener process with sample paths which are updated, in the light of the paired observations, through use of a Metropolis-within-Gibbs algorithm. The methodology is computationally efficient due to the independent increment property of the process and the use of a Brownian bridge.

The technique we describe was used to create the internationally agreed radiocarbon calibration curves known as IntCal09 and marine09, and will be used for the forthcoming (more major) up-dates to the curves (to be published in 2012). Such calibration curves are vital to all who use radiocarbon dating and contribute significantly to improving accuracy in archaeological and palaeoenvironmental (including palaeoclimate) chronologies.

Poster 18

Presenter's name:

Wolfgang Polasek (Institute for Advanced Studies, Vienna, Austria and University of Porto, Portugal)

Authors names:

Wolfgang Polasek (Institute for Advanced Studies, Vienna, Austria and University of Porto, Portugal)

Title:

MCMC estimation of extended Hodrick-Prescott (HP) filtering models

Abstract:

The Hodrick-Prescott (HP) method was originally developed to smooth time series, i.e. to get a smooth (long-term) component. We show that the HP smoother can be viewed as a Bayesian linear model with a strong prior for the smoothness component. Extending this Bayesian approach in a linear model set-up is possible by a conjugate and a non-conjugate model using MCMC. The Bayesian HP smoothing model is also extended to a spatial smoothing model. We have to define spatial neighbors for each observation and we can use in a similar way a smoothness prior as for the HP filter in time series. The new smoothing approaches are applied to the (textbook) airline passenger data for time series and to the problem of smoothing spatial regional data. This new approach can be used for a new class of model-based smoothers for time series and spatial models.

Poster 19

Presenter's name:

Oliver Ratmann (Duke University)

Authors names:

Oliver Ratmann, Christophe Fraser, Ge Donker, Katia Koelle

Title:

Bayesian inference for viral diseases based on epidemiological and genetic summary statistics

Abstract:

A key priority in infectious disease research is to understand the ecological and evolutionary dynamics of viral diseases, and the use of this knowledge in designing public health measures. Empirical data on disease incidence as well as viral genetic and antigenic variation is increasing and could ideally be used synergistically in furthering our understanding of these dynamics. Because these data are heterogeneous in form and highly interdependent, writing down a likelihood of all the data becomes extremely difficult.

We use Approximate Bayesian Computation (ABC) to fit phylodynamic models to epidemiological, genetic and antigenic data. This approach is sufficiently flexible to allow for phylodynamic models that account for non-linear disease dynamics, molecular selection and population structure. To aid Bayesian model choice in this context, we make extensive use of the computed ABC errors. Conflict between these errors typically arises irrespective of sufficiency and provides a useful diagnostic for model design.

We present results of this approach to elucidate epidemiological and evolutionary characteristics of influenza A/H3N2 dynamics, using epidemiological and antigenic data from the Netherlands and serially-sampled genetic data from northern continental Europe.

Poster 20

Presenter's name:

Maria Asuncion Beamonte (Universidad de Zaragoza)

Authors names:

M P Alonso (Universidad de Lleida), M. Asuncion Beamonte, P. Gargallo, M. Salvador (Universidad de Zaragoza),

Title:

Bayesian analysis of labour markets accesibility

Abstract:

In this work we use a Poisson interaction gravity model with spatial effects that allows us to study labour markets accessibility distinguishing between the attractiveness and impedance components. The estimation process of the model is carried out using Bayesian tools based on the use of auxiliary mixture sampling. For illustration purposes, the methodology is applied to a dataset containing information of commuters from the Spanish region of Aragón.

Poster 21

Presenter's name:

Qiming Lv (University of Sheffield)

Authors names:

Q. Lv, P. G. Blackwell, C. E. Buck, M. Charles (University of Sheffield), S. Colledge (University College London), A.Walker (University of Sheffield)

Title:

Out of Asia: A new framework for dating the spread of agriculture to Europe

Abstract:

The switch from hunting and gathering to agriculture is one of the most significant economic and social changes in human history. Motivated by our interests in the first arrival time of prehistorical agriculture across Europe, we are only dating the earliest cereals found on Neolithic sites and developing a tailored Bayesian statistical framework for their interpretation.

This new framework constructs a spatial-temporal process of the spread of agriculture on an irregularly sampled network (created via a Delaunay triangulation). The nodes in the network represent sites of statistical and geographic importance, augmented by points chosen to ensure that the network meets a minimum angle constraint. Edges indicate major corridors or geographical boundaries of unfavorable areas, plus additional edges to complete the triangulation. By formulating the first arrival time at each node as the sum of the travel time along the shortest path, we can break down the correlated first arrival times at each nodes into independent travel times along edges, thus explicitly disentangling the spatial dependence among neighboring sites and allowing landscape connectivity, geographic information and environmental conditions to enter as priors. Using an existing Bayesian radiocarbon calibration framework, we are then able to combine these spatial prior knowledge with observed radiocarbon dates of cereal grains as well as the large amount of uncertainties associated with them to make better inference. Preliminary results show that the new model can provide reduced uncertainty on estimates of arrival times for

nodes where data are available, make predictions (with clear uncertainty statements) about arrival times at the others, and provide a powerful tool for testing hypotheses concerning agricultural spread.

Poster 22

Presenter's name:

Linda Sharples (MRC Biostatistics Unit, Cambridge)

Authors names:

Linda Sharples, Nikolaos Demiris, Dave Lunn

Title:

Survival Extrapolation using the Poly-Weibull model

Abstract:

Recent studies of (cost-) effectiveness in cardiothoracic transplantation have required estimation of mean survival over the lifetime of the recipients. In order to calculate mean survival the complete survivor curve is required but is often not fully observed, so that survival extrapolation is necessary. After transplantation the hazard function is bathtub shaped, reflecting latent competing risks which operate additively in overlapping time periods. The Poly-Weibull distribution is a flexible parametric model that may be used to extrapolate survival and has a natural competing risks interpretation. In addition, treatment effects and subgroups can be modeled separately for each component of risk. We describe the model, develop inference procedures using freely available software and discuss problems with identifiability. The methods are applied to assessment of a new preservation device in cardiac transplantation.

Poster 23

Presenter's name:

Ioanna Manolopoulou (Duke University)

Authors names:

Ioanna Manolopoulou, Melanie P. Matheu, Michael D. Cahalan, Mike West and Thomas B Kepler

Title:

Semi-parametric Bayesian modelling of inhomogeneous tactic fields in single-cell motility

Abstract:

We develop dynamic models of single cell motion involving nonparametric representations of nonlinear spatial fields that guide cellular motility. Assuming a discretized diffusion model for the cell motion, the tactic field is flexibly modelled using radial basis kernel regression. Our methods are motivated by the temporal dynamics of lymphocytes in the lymph nodes, critical to the immune response. The primary goal is learning the structure of the tactic fields that fundamentally characterize the immune cell motion. We develop Bayesian analysis via customized Markov chain Monte Carlo methods for single cell models, and multi-cell hierarchical extensions for aggregating models and data across multiple cells. Our implementation explores data from multi-photon vital microscopy in murine lymph node experiments, and we use a number of visualization tools to summarize and compare posterior inferences on the 3−dimensional tactic fields.

Poster 24

Presenter's name:

Emma Jones (University of Sheffield)

Authors names:

E.M. Jones, C.E. Buck (University of Sheffield), C.D. Litton (University of Nottingham), C.Tyers (University of Sheffield) and A. Bayliss (Scientific Dating, English Heritage)

Title:

A Bayesian Approach to Dendrochronology

Abstract:

The width of tree-rings are determined by several factors including a local climatic signal apparent in that year, and the tree's growth trend. The climatic signal influences growth such that if the summer is warm and wet, the ring tends to be wider than if the summer is cold and dry. The growth trend describes the fast growth of the tree when it is young producing wide rings, followed by narrower rings as it ages. Other factors such as the soil conditions, presence of pests and diseases and competition for light and nutrients can also effect the ring width. The impact of these latter factors are collectively known as noise. It is assumed that trees within the same geographical region are exposed to the same climatic signal in each year, but that this differs from year to year.

Tree-ring dating involves matching sequences of tree-ring widths from timbers of unknown age to dated sequences known as `master' chronologies. Before matching takes place, all data are preprocessed to remove the growth trends. The timbers of unknown age (typically from a single building or woodland) are, firstly, sequentially matched against one another to identify the relative offsets with the `best' match. The sequence produced is known as a `site' chronology. The site chronology is then further matched to a local master chronology, to attempt to produce a date estimate for the site chronology.

Traditionally the quality of the matches (both within the site chronology and between the site chronology and the master chronology) are assessed via the classical statistical t-test. A match at a particular offset is only considered to be `best' if it produces the largest t-value of all of the possible offsets and is greater than (an arbitrary value of) 3.5. The success rate of dating varies within sites and across regions; the national average being approximately 60-70% but in some geographical areas the success rate can be much lower. One of the reasons for this is that the t-test does not utilise the wide range of information that could be used if a Bayesian model was used for tree-ring dating.

This poster highlights our recent work to develop a hierarchical Bayesian model for tree-ring dating which allows important prior information on parameters to be drawn into the inference process; this prior information can be taken from trees and can also be elicited from expert dendrochronologists. The model assumes that each ring width is composed of an overall climatic signal and some noise, and can be further extended to include climatic signals at varying geographic scales. Probabilities for a match at each offset can be produced conditional on the data and the prior specifications. The method removes the need to identify a single `best' match, but it does rely on careful prior specification of parameters. Consequently, we have collated ring width data from trees of known age from several woods in the UK and are using these to provide informative prior knowledge. Work has also just begun on ways to include expert knowledge from dendrochronologists.

Poster 25

Presenter's name:

George Kouvaras (National Technical University of Athens)

Authors names:

George Kouvaras and George Kokolakis (National Technical University of Athens)

Title:

On bivariate random probability measures

Abstract:

A Bayesian nonparametric model for the construction of bivariate random probability measures on a Polish space X is considered. The proposed model is built on a dependence structure between weighted gamma processes. Dependence is introduced by using a variant of Poisson/Gamma random field model. Results related to our approach are presented.

Poster 26

Presenter's name:

Fadlalla G. Elfadaly (The Open University, UK)

Authors names:

Fadlalla G. Elfadaly and Paul H. Garthwaite

Title:

On Eliciting Logistic Normal Priors for Multinomial Models

Abstract:

In Bayesian analysis of multinomial models, an important assessment task is to elicit an informative joint prior distribution for multinomial probabilities. It is well-known that the Dirichlet distribution is a conjugate prior for a multinomial model. However, due to its limited number of parameters and negative correlation structure, it is not very flexible for representing prior belief. A possible alternate prior with a larger number of parameters and a general dependence structure is the logistic normal distribution. We propose a method for quantifying opinion about the logistic normal prior of a multinomial model. Using an additive logistic transformation, multinomial probabilities are transformed into multivariate normal unconstrained variables. By assessing sets of probability quartiles, our method outputs a mean vector and a positive-definite covariance matrix for the logistic normal prior distribution. The proposed method requires a reasonable number of assessments that can be simply elicited using interactive user-friendly software written in Java.

Poster 27

Presenter's name:

Shahrul Mt-Isa (Imperial College London)

Authors names:

Shahrul Mt-Isa (Imperial College London), Nick Croft (Queen Mary University of London), Deborah Ashby (Imperial College London)

Title:

A closer look at the evidence of acceptable benefit-risk decision making for off-label medicines prescription in children

Abstract:

Over half of the medicines prescribed to children in the European Union have never been licensed for its use (off-label) (Conroy, 2000). The decision to prescribe these drugs off-label is a subjective judgment by medical professionals as a desperate attempt to treat very sick children. Realising this necessity, the Medicines Act 1968 neither prohibit nor promote the practice. Bayesian decision-making through evidence-based medicine can help improve and support such decisions (Ashby and Smith, 2000). Along the line of "Bayesian thinking" (Bernardo and Smith, 1994), we developed formal framework for the assessment of risks and benefits trade-off as a decision-making tool; and consequently examined how the evidence tantamount. A case study of domperidone as prescribed off-label to children for the treatment of gastro-oesophageal reflux disease (GORD) was used; formally quantified and examined under hierarchical Bayesian framework.

References

Ashby, D. and Smith, A. F. M. (2000) Evidence-based medicine as Bayesian decision-making. Stat Med 19(23), 3291-3305 .

Bernardo, J. M. and Smith, A. F. M. (1994) Bayesian Theory. John Wiley & Sons Ltd., Chicester, England.

Conroy S, Choonara I, Impicciatore P, Mohn A, Arnell H, Rane A, et al. (2000) Survey of unlicensed and off label drug use in paediatric wards in European countries. European Network for Drug Investigation in Children. BMJ Jan 8;320(7227):79-82.

Poster 28

Presenter's name:

Lawrence Pettit (Queen Mary University of London)

Authors names:

Lawrence Pettit, Nalaiyini Sothinathan, Karen Young

Title:

Outliers in Circular Data

Abstract:

Bayesian methods to analyse circular data are considered. The conditional predictive ordinate can be used to identify the most outlying observation. Two models are fitted, one which assumes that all the observations come from the same von Mises distribution, the other being an outlier model where spurious observations are assumed to come from this distribution but with a shifted mean direction. The models are fitted using Gibbs sampling with latent variables. A dataset on the homing instincts of frogs is analysed. A class of conjugate priors is illustrated graphically. Model comparison by Bayes factors shows strong support for the outlier model.

Poster 29

Presenter's name:

Joe L. Herman (University of Oxford)

Authors names:

J. L. Herman, R. Lyngs and J. Hein (University of Oxford)

Title:

Probabilistic models for protein structure alignment

Abstract:

Successful models for the evolution of protein sequences have been in existence since the 1960s, and are widely used in a large range of applications, ranging from inference of phylogeny to identification of binding site motifs. However, despite decades of research seeking to bridge the gap between sequence and structure, current methods for studying structural evolution are much less well developed.

We develop a rotation and translation invariant distribution to model a set of homologous structures, under the assumption that the structures represent perturbations from a common underlying equilibrium distribution. Simulating from this model via MCMC, we can generate samples of alignments for any set of structures. Using a recently developed method for maximum posterior decoding in the space of alignment columns, it is possible to generate a single summary alignment that minimises one of a family of loss functions. Working in the space of equivalence classes of alignment columns allows the convergence of the sampler to be monitored much more effectively than via a simple log likelihood trace, and enables efficient schemes to be constructed for representing the posterior.

The method is illustrated on a selection of protein families, and a test statistic is developed to assess the significance of any particular alignment. On a test set of homologues and decoys we are able to successfully cluster the data into the known groups.