Bayesian Statistics and Algorithms
February 29 - March 4, 2016

MINI-COURSES:

Simon Barthelmé: Expectation-Propagation for Approximate Inference.

The Expectation-Propagation algorithm was introduced by Minka in 2001, and is today still one of the most effective algorithms for approximate inference. It is relatively difficult to implement well but in certain cases it can give results that are almost exact, while being much faster than MCMC. In this course I will review EP and classical applications to Generalised Linear Models and Gaussian Process models. I will also introduce some recent developments, including applications of EP to ABC problems, and discuss how to parallelise EP effectively.

Christine Keribin: Variational Bayes methods and algorithms.

Bayesian posterior distributions can be numerically intractable, even by the means of Markov Chain Monte Carlo methods. Bayesian variational methods can then be used to compute directly (and fast) a deterministic approximation of these posterior distributions. In this course, I describe the principles of the variational methods and their application in Bayesian inference, review main theoretical results and discuss their use on examples.

Jean-Michel Marin & Christian P. Robert: Variational Bayes methods and algorithms.

In this short course, we recall the basics of Markov chain Monte Carlo (Gibbs & Metropolis sampelrs) along with the most recent developments like Hamiltonian Monte Carlo, Rao-Blackwellisation, divide & conquer strategies, pseudo-marginal and other noisy versions. We also cover the specific approximate method of ABC that is currently used in many fields to handle complex models in manageable conditions, from the original motivation in population genetics to the several reinterpretations of the approach found in the recent literature. Time allowing, we will also comment on the programming developments like BUGS, STAN and Anglican that stemmed from those specific algorithms.

TALKS:

Leonardo Bottolo: Bayesian hierarchical mixture model for financial time series

In finance and econometrics, there has been an explosion of papers analysing time series by adopting AR-GARCH-type models. Of crucial interest in this context is the ability of these models to forecast the return and volatility at time t+1 conditional on the available information at time t. The nature of the applications drives the choice of these models to be based on forecasting power rather than the ability to explain the data generation mechanism.

In this work we present a Bayesian modelling framework for multivariate AR-GARCH-type models that utilizes a prior density placed on each ARMA-GARCH parameter in a way that it is possible to compute posterior probabilities that any subset of parameters across time series belongs to the same cluster. Within each cluster, the parameters are assumed to be exchangeable so the borrowing strength effect pools information for parameter estimation. This can be viewed as a model averaging or model mixing strategy that offers forecasts that take into account model uncertainty.

The benefits of the proposed model over existing methods are illustrated in a real case scenario when the prediction of returns and volatilities of 600 stocks of the European Stock Market in day t + 1 are required.

Joint work with Petros Dellaportas

Nicolas Chopin: Leave Pima Indians alone: binary regression as a benchmark for Bayesian computation.

Whenever a new approach to perform Bayesian computation is introduced, a common practice is to showcase this approach on a binary regression model and datasets of moderate size. This paper discusses to which extent this practice is sound. It also reviews the current state of the art of Bayesian computation, using binary regression as a running example. Both sampling-based algorithms (importance sampling, MCMC and SMC) and fast approximations (Laplace and EP) are covered. Extensive numerical results are provided, some of which might go against conventional wisdom regarding the effectiveness of certain algorithms. Implications for other problems (variable selection) and other models are also discussed.

Joint work with James Ridgway.

Guillaume Dehaene: Expectation Propagation is exact in the large-data limit.

Expectation Propagation (EP, Minka 2001) is a popular algorithm for approximating posterior distributions. While it is known empirically to give good approximations at a low computational cost (Nickisch and Rasmussen, 2008), it's also very poorly understood. In this talk, I will present some new results on EP in the large-data limit. I will show that Gaussian-EP iterations asymptotically behave like the Newton algorithm, from which it follows that Gaussian-EP is asymptotically exact. I will then give an upper bound on the errors of the EP approximation and show that their decay speed compares favorably to the errors of the Laplace approximation of the posterior.

Pierre Druilhet: Convergence modes for prior distributions.

Improper priors are often criticized for leading to paradoxes or inconsistency. However, a posterior distribution obtained from an improper prior is most often the limit of a sequence of posteriors obtained from proper priors. In this talk, we compare two types of convergence modes for a sequence of prior distributions that aims to approximate an improper prior. The first one is derived from finitely additive probabilities whereas the second one is obtained by considering a projective space of measures. Then, we discuss their properties and the implication on the Bayesian paradigm. At last, we revisit some paradoxes.

Richard Everitt: An overview of noisy MCMC and SMC.

The development of exact approximate Monte Carlo methods, in which unbiased estimates of densities are used within Markov chain Monte Carlo (MCMC) or sequential Monte Carlo (SMC) algorithms without loss of exactness, is one of the most important recent innovations in the field. This talk gives an overview of work on inexact approximations or noisy methods, where (often low variance) alternatives to unbiased approximations are used instead. In all cases the exactness of the algorithm is lost, but in some cases this proves to be insignificant compared to computational savings or improved variance of estimates produced by finite runs. In the context of many applied researchers accepting the use of other approximate methods (such as approximate Bayesian computation) further investigation of the use of noisy methods is warrented.

Nial Friel: Accelerating Bayesian inference for intractable likelihood models using noisy MCMC.

Statistical models with intractable likelihoods abound. There have been many approaches to overcome this computational impasse. One such is noisy MCMC, where the central idea is to replace an exact transition kernel (which leave the target posterior distribution invariant) with an approximation of the transition kernel. It is possible, under certain conditions, to prove that such a noisy MCMC algorithm yields a stationary distribution which converges to the target distribution. However, it can be quite slow to run, since it relies on repeatedly simulating from the likelihood function. Here we suggest an approach where the computationally intensive likelihood simulation is carried out off-line. Parameter inference is then carried out on- line using the off-line likelihood simulations. We show that our algorithm can be placed in the framework of noisy MCMC and can prove convergence to the target distribution. We illustrate that this approach gives impressive speed up compared to existing approaches.

Pierre Latouche: Goodness of fit of logistic models for random graphs.

We consider binary networks along with covariate information on the edges. In order to take these covariates into account, logistic-type models for random graphs are often considered. One of the main questions which arises in practice is to assess the goodness of fit of a model. To address this problem, we add a general term, related to the graphon function of W- graph models, to the logistic models. Such an extra term can be approximated from a blockwise constant function obtained using stochastic block models with increasing number of clusters. If the given network is fully explained by the covariates, then a sole block should be estimated from data. This framework allows to derive a testing procedure from a model based selection context. Bayes factors or posterior odds can then be used for decision making. Overall, the logistic model considered necessitates two types of variational approximations to derive the model selection approach.

Denys Pommeret: Combining ridge parameter with the g-prior of Zellner.

In a Bayesian regression model a common prior is the g-prior of Zellner. Its use necessitates an inversion of the covariance matrix. This is problematic when the variables are highly correlated or when the number of parameters is too big. We then introduce a ridge parameter. We illustrate this approach through a variable selection in a probit mixed model.

Dennis Prangle: A data augmentation approach to high dimensional ABC.

A well known limitation of ABC methods is that they perform poorly for high dimensional data, with the use of summary statistics providing only a partial solution. This talk outlines a new approach to high-dimensional ABC: a more efficient method to estimate the likelihood given some parameter value .

This method begins performing several simulations to produce a rough estimate of the likelihood. The most successful simulations are then perturbed in an attempt to improve their matches to the observation. These can be used to produce a more accurate estimate of the likelihood. These steps are iterated. This process can be formalised as a sequential Monte Carlo algorithm, employing data augmentation MCMC moves to perform the perturbations. The talk will describe the method and present simple examples.

Pierre Pudlo: Adaptive multiple important sampling.

Among Monte Carlo techniques, the importance sampling requires fine tuning of a proposal distribution, which can be done through iterative schemes. Sequential adaptive algorithms have been proposed to calibrate the sampling distribution, e.g. Douc et al. (2007, Annals of Statis). The Adaptive Multiple Importance Sampling of Cornuet et al. (2012, Scandinavian) provides a significant improvement in stability and effective sample size by introducing a recycling procedure of past particles. After presenting some sequential algorithms, we will show consistency results of the adaptive multiple importance sampling.

Sylvia Richardson: Exploring the presence of complex dependence structures in epidemiological and genomic data

Faced with data containing a large number of inter-related explanatory variables, finding ways to investigate complex multi-factorial effects is an important statistical task. This is particularly relevant for epidemiological study designs where large numbers of covariates are typically collected in an attempt to capture complex interactions between host characteristics and risk factors. A related task, which is of great interest in stratified medicine, is to use multi-omics data to discover subgroups of patients with distinct molecular phenotypes and clinical outcomes, thus providing the potential to target treatments more precisely. Flexible clustering is a natural way to tackle such problems. It can be used in an unsupervised or a semi-supervised manner by adding a link between the clustering structure and outcomes and performing joint modelling. In this case, the clustering structure is used to help predict the outcome. This latter approach, known as âĂúprofile regressionâĂù, has been implemented recently using a Bayesian non parametric DP modelling framework, which specifies a joint clustering model for covariates and outcome, with an additional variable selection step to uncover the variables driving the clustering (Papathomas et al, 2012). In this talk,

two related issues will be discussed. Firstly, we will focus on categorical covariates, a common situation in epidemiological studies, and examine the relation between: (i) dependence structures highlighted by Bayesian partitioning of the covariate space incorporating variable selection; and (ii) log linear modelling with interaction terms, a traditional approach to model dependence. We will show how the clustering approach can be employed to assist log-linear model determination, a challenging task as the model space becomes quickly very large (Papathomas and Richardson, 2015). Secondly, we will discuss clustering as a tool for integrating information from multiple datasets, with a view to discover âÄúusefulâÄù structure for prediction. In this context several related issues arise. It is clear that each dataset may carry a different amount of information for the predictive task. Methods for learning how to reweight each data type for this task will therefore be presented. In the context of multi-omics datasets, the efficiency of different methods for performing integrative clustering will also be discussed, contrasting joint modelling and stepwise approaches. This will be illustrated by analysis of genomics cancer datasets.

M Papathomas, J Molitor, C Hoggart, D Hastie & S Richardson (2012). Exploring Data From Genetic Association Studies Using Bayesian Variable Selection and the Dirichlet Process: Application to Searching for Gene &ó Gene Patterns. Genetic Epidemiology. 36:663-674.

M Papathomas & S Richardson (2015). On the utility of the Dirichlet process for linear model determination: application to graphical log-linear model determination. To appear in Journal of Statistical Planning and Inference. Available from: http://arxiv.org/abs/1401.7214. Joint work with Michael Papathomas and Paul Kirk.

James Ridgway: On the properties of variational approximations of Gibbs posteriors.

The PAC-Bayesian approach is a powerful set of techniques to derive non- asymptotic risk bounds for random estimators. The corresponding optimal distribution of estimators, usually called the Gibbs posterior, is unfortunately intractable. One may sample from it using Markov chain Monte Carlo, but this is often too slow for big datasets. In this talk we consider instead variational approx- imations of the Gibbs posterior, which are fast to compute. We undertake a general study of the properties of such approximations. Our main finding is that such a variational approximation has often the same rate of convergence as the original PAC-Bayesian procedure it approximates. We illustrate our results on several learning tasks (classi cation, ranking, matrix comple- tion), discuss how to implement a variational approximation in each case, and illustrate the good properties of said approximation on real datasets.

Stéphane Robin: Exact Bayesian inference for some models with discrete parameters.

Bayesian inference often requires the integration over the parameter space. In the case of continuous parameters, the use of conjugate priors may allow to achieve this integration in an exact manner. For models involving discrete parameters the integration over the parameter space is often intractable due to the combinatorial complexity induced by a systematic exploration. Two models involving a discrete parameter will be considered: (i) change-point detection, in which the set of locations of the change-points is discrete, (ii) the inference

of the structure of a graphical model, for which the topology of the underlying graph is discrete as well. In both cases we will show that exact Bayesian inference can be achieved using some specific algebraic properties (up to some restriction in the latter case). These properties allow to compute key quantities such as the posterior distribution of the location of a change-point or the posterior probability for a given edge to be part of the graphical model. Both approaches will be illustrated with biological applications.

Judith Rousseau: Nonparametric mixture models and HMMS.

In the recent years some results have been obtained about the identifiability of mixture models - possibly dynamical - when the emission distributions are not specified. In particular, in the case of independent and identically distributed hidden states leaving on a finite state space, the parameters (emission distributions and weights of the mixture) are identifiable when each individual is associated to three independent observations. In the case of non independent hidden states, then as soon as the transition matrix is invertible, then the parameters are identifiable. In this work, we investigate estimation in these models and discuss some aspects of semi- parametric Bayesian estimation, including Bersntein von Mises theorems for the weights (or transition matrices) and estimation of the number of hidden states.
This is a joint work with Elisabeth Gassiat, Elodie Vernet and Kerrie Mengersen

Daniel Rudolf: Approximations of geometrically ergodic Markov chains.

By using perturbation theory for Markov chains we derive explicit estimates of the bias of an approximate or noisy version of a geometrically ergodic Markov chain. We apply this result to a noisy Metropolis-Hastings algorithm and discuss also some consequences for the integration error of such Markov chain Monte Carlo methods.

Adeline Samson: Computational methods for stochastic differential equations.

In this talk, I will focus on parametric estimation for stochastic differential equations. Because the likelihood is generally not available, we resort to computational methods (MCMC, Sequential Monte Carlo, Particle MCMC, Stochastic EM) to estimate the parameters either from a Bayesian point of view or with a frequentist one. I will review and compare some of these methods and their advantages and drawbacks depending on the Bayesian or frequentist approaches.
(joint work with M. Delattre, AgroParisTech, Ditlevsen, Copenhagen, S. Donnet, INRA)

Darren Wilkinson: Bayesian Hierarchical Modelling of Genetic Interaction in Yeast .

Saccharomyces cerevisiae (often known as budding yeast, or brewers yeast) is a single-celled micro-organism that is easy to grow and genetically manipulate. As it has a cellular organisation that has much in common with the cells of humans, it is often used as a model organism for studying genetics. High-throughput robotic genetic technologies can be used to study the fitness of many thousands of genetic mutant strains of yeast, and the resulting

data can be used to identify novel genetic interactions relevant to a target area of biology. The processed data consists of tens of thousands of growth curves with a complex hierarchical structure requiring sophisticated statistical modelling of genetic independence, genetic interaction (epistasis), and variation at multiple levels of the hierarchy. Starting from simple stochastic differential equation (SDE) modelling of individual growth curves, a Bayesian hierarchical model can be built with variable selection indicators for inferring genetic interaction. The methods will be applied to data from experiments designed to highlight genetic interactions relevant to telomere biology.