AUSTRALIAN RESEARCH COUNCIL CENTRE OF EXCELLENCE
FOR MATHEMATICAL AND STATISTICAL FRONTIERS

2019 Enabling Algorithms Theme Symposium

Dates: Thursday 13th June 2019,
Friday 14th June 2019.

Venue: Room 430, Level 5, Building 4,
University of Technology Sydney,
South-west corner of Harris Street
and Thomas Street, Ultimo,
New South Wales, Australia.

Theme: Fast Approximate Inference.

SCHEDULE

Thursday 13th June 2019

9:00	David Nott (Condit...)
9:45	Andrew Zammit Mangion
10:30	Refreshments
11:00	Tutorial
11:30	Tutorial
12:00	Tutorial
12:30	Lunch
1:30	Sarah Romanes
2:00	Luca Maestrini
2:30	James Yu
3:00	Refreshments
3:30	Tui Nolan
4:00	Panel discussion

Friday 14th June 2019

9:00	Minh-Ngoc Tran
9:30	Weichang Yu
10:00	Edwin Bonilla
10:30	Refreshments
11:00	Tutorial
11:30	Tutorial
12:00	Tutorial
12:30	Lunch
1:30	John Ormerod
2:15	David Nott (High-...)
3:00	Refreshments
3:30	Discussion on possible collaborations

TITLES AND ABSTRACTS OF TALKS

ANDREW ZAMMIT MANGION, **University of Wollongong**

Approximate Inference with Deep Compositional Spatial Models

Non-stationary, anisotropic spatial processes are often used when modelling, analysing and predicting complex environmental phenomena. One such class of processes considers a stationary, isotropic process on a warped spatial domain. The warping function is generally difficult to fit and not constrained to be bijective, often resulting in 'space-folding.' In the first part of the talk I will propose modelling a bijective warping function through a composition of multiple elemental bijective functions in a deep-learning framework, which ensures that there is no space-folding by construction. In the second part I will discuss inference when the resulting deep spatial model is a non-stationary Gaussian process or a general non-Gaussian process composed from a set of processes that are conditionally Gaussian. In the latter case I show how stochastic variational inference can be used for parameter estimation and for approximating predictive distributions as Gaussian mixtures. Through experiments in one and two dimensions I show that the deep compositional spatial models are quick to fit using approximate inference algorithms, and are able to provide better predictions and uncertainty quantification than other deep stochastic models of similar complexity.

MINH-NGOC TRAN, **University of Sydney**

Doubly Geometry-Informed Variational Bayes

Increasingly complicated models in modern statistics have called for more efficient Bayesian estimation methods. This work develops a Variational Bayes algorithm that exploits both the information geometry of the manifold of probability distribution functions and the manifold structure of the variational parameters. The information geometry of the manifold of probability distributions results in the natural gradient which is the steepest ascent on this manifold. Utilising the manifold structure of the variational parameters leads to an efficient non-linear optimization technique that takes into account the constraint structure of the parameter space. The convergence of the proposed algorithm is theoretically guaranteed and its performance is tested on several challenging examples including deep neural networks.

DAVID NOTT, **National University of Singapore**

Conditionally Structured Variational Gaussian Approximation with Importance Weights

We develop flexible variational inference methods for models with complex latent variable structure, such as generalized linear mixed models and state space models. Our approach first splits the variables in these models into "global" parameters and "local" latent variables. Variational approximations are then defined sequentially, through a marginal density for the global parameters and a conditional density for local variables given global parameters. Each term in our approximation is Gaussian, but we allow the conditional covariance matrix for the local parameters to depend on the global parameters, which leads to an approximation that is not jointly Gaussian. The approximations are motivated by the fact that in many hierarchical models there are global variance and dependence parameters which determine the scale and dependence structure of local latent variables in their conditional posterior given the global parameters. We also consider parsimonious parametrizations by using conditional independence structure, and improved estimation of the log marginal likelihood and variational density using importance weights. These methods are shown to improve significantly on Gaussian variational approximation methods for a similar computational cost. (Joint work with Linda Tan and Aishwarya Bhaskaran.)

DAVID NOTT, **National University of Singapore**

High-dimensional Copula Variational Approximation Through Transformation

Variational methods are attractive for computing Bayesian inference for highly parametrized models and large datasets where exact inference is impractical. They approximate a target distribution - either the posterior or an augmented posterior - using a simpler distribution that is selected to balance accuracy with computational feasibility. Here we approximate an element-wise parametric transformation of the target distribution as multivariate Gaussian or skew-normal. Approximations of this kind are implicit copula models for the original parameters, with a Gaussian or skew-normal copula function and flexible parametric margins. A key observation is that their adoption can improve the accuracy of variational inference in high dimensions at limited or no additional computational cost. We consider the Yeo-Johnson and G&H transformations, along with sparse factor structures for the scale matrix of the Gaussian or skew-normal. We also show how to implement efficient reparametrization gradient methods with both Gaussian and skew-normal copula variational families and illustrate the effectiveness of the proposed methods in a range of examples. (Joint work with Michael Smith and Ruben Loaiza-Maya.)

SARAH ROMANES, **University of Sydney**

Using Variational Approximations to Efficiently Build a Generalised Discriminant Analysis (genDA) Algorithm

Discriminant Analysis (DA) methods, such as LDA and QDA, have long been used as effective classifiers for correlated, Gaussian data. However, the use of such classifiers is restricted when: a) The data is non-Gaussian, and/or b) The number of features is larger than the number of observations. Although diagonal discriminant analysis and Factor analysis based methods have been developed successfully to address problems arising from b), the challenge to develop DA methods for non-Gaussian data remains open. In this talk, we introduce our generalised DA method (genDA) as a novel attempt to extend DA for non-Gaussian responses. This method utilises Bayesian Generalised Linear Latent Variable Models (GLLVMs) to capture the correlation structure between features, and effectively using this information to classify new data points. Variational approximations are used to estimate such GLLVMs, with efficient optimisation routines implemented using Automatic Differentiation techniques provided by the 'TMB' package in 'R'. We will show performance results on simulated as well as real data, as well as address future directions for the development of this model.

JOHN ORMEROD, **University of Sydney**

Particle Based Collapsed Variational Approximation for Bayesian Linear Model Averaging

While Bayesian model averaging has several desirable properties, it is computationally expensive unless the number of models to be averaged over is small. Typically the number of models to be averaged grows exponentially in the number of covariates and some form of approximation is required. In this paper we explore a novel particle based collapsed variational approximation for Bayesian model averaging. The resulting objective function can be optimized in a highly parallel manner. We explore several different prior specifications which lead to Bayes factors with closed forms. We show empirically that our approach is fast and effective for moderately large problems on several simulated and publicly available datasets, particularly when parallel computing resources are available. An R package is available implementing our approach. (Joint work with Mark Greenaway).

JAMES YU, **University of Technology Sydney**

Fast and Accurate Frequentist Generalised Linear Mixed Model Analysis via Expectation Propagation

Generalised linear mixed models are a particularly powerful and well established statistical tool. Unlike linear mixed models, where the integrals arising in likelihood functions can be expressed in closed form, the likelihood functions expressed in generalised linear mixed models do not follow tractable solutions. Methods such as Gauss-Hermite quadrature and Laplace approximation are the standard approaches to overcome these integrals. While Gauss-Hermite quadrature is accurate, it is slow, rendering it unsuitable for analyses with more than two or three random effects. Laplace approximations are the most feasible solution, however, the approximate inference they provide in binary models is well known to be inaccurate. This talk aims to explain a new fast and accurate method of solving this integral for frequentist generalised linear mixed models called expectation propagation.

LUCA MAESTRINI, **University of Technology Sydney**

Double-loop Expectation Propagation for Statistical Models

Despite recent developments, issues and challenges involved in practical implementation of expectation propagation in Statistics have partially been explored. Expectation propagation is an intuitive variational approximation scheme which converges in many practical cases, but not always. We take advantage of a double-loop prescription for expectation propagation, guaranteed to converge to a local minimum of a likelihood lower bound, to derive an explicit form for simple statistical models. Our contributions show what is exactly involved in deriving and implementing double-loop expectation propagation and how to exploit connections with the standard single-loop version. (Joint work with Linda S.L. Tan and Matt P. Wand.)

WEICHANG YU, **University of Sydney**

Variational Discriminant Analysis with Variable Selection

A Bayesian method that seamlessly fuses classification via discriminant analysis and hypothesis testing is developed. Building upon the original discriminant analysis classifier, modelling components are added to identify discriminative variables. A combination of cake priors and a novel form of variational Bayes we call reverse collapsed variational Bayes gives rise to variable selection that can be directly posed as a multiple hypothesis testing approach using likelihood ratio statistics. Some theoretical arguments are presented showing that Chernoff consistency (asymptotically zero type I and type II error) is maintained across all hypotheses. We apply our method on some publicly available genomics datasets and show that our method performs well in practice. An R package VaDA has also been made available on Github.

TUI NOLAN, **University of Technology Sydney**

Streamlined Computing for Variational Bayesian Inference with Higher-Level Random Effects

Streamlined variational Bayesian inference for multilevel data analysis is hindered by the presence of sparse multilevel matrix problems. Existing methods are such that streamlined variational inference is restricted to mean-field variational Bayes algorithms for two-level random effects models. Streamlined solutions to sparse matrix problems will be presented, which arise in multilevel modeling and longitudinal data analysis. The solutions to two-level and three-level problems provide the blueprint for extensions to higher-level versions of the problem. These results can be applied to mean field variational Bayes algorithms by using a least squares representation of the updates. Whilst the linear system solutions are a concise recasting of existing results, the matrix inverse sub-block results are novel and facilitate streamlined mean field variational Bayesian inference for models containing higher-level random effects. In summary, the barriers for streamlining variational inference algorithms with higher level random effects are removed.

EDWIN BONILLA, **Data 61**

Automated Probabilistic Reasoning via Variational Inference

A large proportion of machine-learning research is devoted to probabilistic modelling and inference. One of the long-term goals in this area is that of developing flexible probabilistic models along with automated yet scalable inference algorithms for them. In this talk I will overview our recent advances in automated variational inference algorithms for non-parametric Bayesian models. These algorithms have desirable properties when using flexible approximate posteriors: (a) they are statistically efficient and (b) they can carry out hyper-parameter estimation straightforwardly. Besides enabling practitioners and researchers to investigate new models with minimal effort, such an approach can be generalised to address at least three major research challenges: intractable likelihoods, structured prediction and uncertainty quantification in deep-learning architectures.