Abstracts

Strategies for Multilevel Monte Carlo

Ajay Jasra[a] and Kody Law[b]

[a] *National University of Singapore*
[b] *The University of Manchester, UK*

ABSTRACT

This talk will concern the problem of inference when the posterior measure involves continuous models which require approximation before inference can be performed. Typically one cannot sample from the posterior distribution directly, but can at best only evaluate it, up to a normalizing constant. Therfore one must resort to computationally-intensive inference algorithms in order to construct estimators. These algorithms are typically of Monte Carlo type, and include for example Markov chain Monte Carlo, importance samplers, and sequential Monte Carlo samplers. The multilevel Monte Carlo method provides a way of optimally balancing discretization and sampling error on a hierarchy of approximation levels, such that cost is optimized. Recently this method has been applied to computationally intensive inference. This non-trivial task can be achieved in a variety of ways. This talk will review 3 primary strategies which have been successfully employed to achieve optimal convergence rates – in other words faster convergence than i.i.d. sampling at the finest discretization level. Some of the specific resulting algorithms, and applications, will also be presented.

Variable Selection with Big Data based on Zero Norm and via Sequential Monte Carlo

Jin-Chuan Duan

National University of Singapore

ABSTRACT

Selecting a subset from many potential explanatory variables in linear regressions has long been the subject of research interest, and the matter is made more important in the era of big data when many more variables become available/accessible. Of late, the l_1-norm penalty based techniques such as Lasso of Tibshirani (1996) have become very popular. However, the variable selection problem in its natural setting is a zero-norm penalty problem, i.e., a penalty on the number of variables as opposed to the l_1-norm of the regression coefficients. The popularity of the l_1-norm penalty or its variants has more to do with computational considerations, because selection with the zero-norm penalty is a highly demanding combinatory optimization problem when the number of potential variables becomes large. In this paper, we devise a sequential Monte Carlo (SMC) method as a practical tool for zero-norm variable selection problems, and the selection task can, for example, be completed under half-an-hour using a typical desktop computer for problems with 1,000 potential variables and 100,000 observations. The essence of our SMC method is to view the selection problem as a discrete probability function defined over all possible combinations comprising, say, k regressors out of m $\geq$ k potential variables, and the peak of this function corresponds to the optimal combination. The solution technique sets out to sequentially generate samples, and after a while the final sample is capable of representing this probability function. We demonstrate through a simulation study the method's superiority in selecting right variables.

Limit Theorems for Sequential MCMC Methods

Arnaud Doucet[a], Axel Finke[b], and Adam M. Johansen[c,d]

[a] *Oxford University, UK*
[b] *National University of Singapore*
[c] *The University of Warwick, UK*
[d] *The Alan Turing Institute, UK*

ABSTRACT

Both SMC methods (a.k.a. "particle filters") as well as sequential MCMC methods constitute classes of algorithms which can be used to approximate (a sequence of) probability distributions and their normalising constants. While SMC methods sample particles conditionally independently at each time step, sequential MCMC methods sample particles according to an MCMC kernel. Introduced over twenty years ago in Berzuini et al. (1997), sequential MCMC methods have attracted renewed interest recently as they empirically outperform SMC methods in some applications (Septier & Peters, 2016). We establish a strong law of large numbers and a central limit theorem for sequential MCMC methods and provide conditions under which errors can be controlled uniformly in time. In the context of state-space models, we also provide conditions under which sequential MCMC methods can indeed outperform standard SMC methods in terms of asymptotic variance of the corresponding Monte Carlo estimators.

Dynamic Degree-corrected Blockmodels for Social Networks: A Nonparametric Approach

Linda Tan

National University of Singapore

ABSTRACT

We propose a nonparametric approach to model social networks using degree-corrected stochastic blockmodels. The static model consists of a stochastic blockmodel formulated using a probit regression and popularity parameters are incorporated to account for degree heterogeneity. Dirichlet processes are used to detect community structure and induce clustering in the popularity parameters. We further extend the static model to dynamic networks. We derive Gibbs samplers for posterior inference under a Bayesian approach. The models are illustrated using real social networks.

Unbiased Estimators and Multilevel Monte Carlo

Matti Vihola

University of Jyväskylä, Finland

ABSTRACT

Multilevel Monte Carlo (MLMC) and recently proposed debiasing schemes are closely related methods which can be applied in scenarios where exact simulation methods are difficult to implement, but biased estimators are easily available. An important example of such a scenario is the inference with continuous-time diffusion processes, where the process is difficult to simulate exactly but time-discretized approximations are available. I will present a new general class of unbiased estimators which admits earlier debiasing schemes as special cases, and new lower variance estimators which behave asymptotically like MLMC, both in terms of variance and cost, under general conditions. This suggests that bias can often be eliminated entirely with arbitrarily small extra cost.

Stable Approximation Schemes for Optimal Filters

Joaquín Míguez

Universidad Carlos III de Madrid, Spain

ABSTRACT

We explore truncation schemes for the approximation of (possibly unstable) optimal filters. In particular, let $S = (\pi_0, \kappa_t, g_t)$ be a state space model defined by a prior distribution π_0, Markov kernels $\{\kappa_t\}$ and potential functions $\{g_t\}$, $t \geq 1$, and let $\{C_t\}$ be a sequence of compact subsets of the state space. In the first part of the talk, we describe a simple and systematic procedure to construct a system $S' = (\pi_0, \kappa'_t, g'_t)$, with potentials truncated within the sets $\{C_t\}$, such that the optimal filters generated by S and S' can be made arbitrarily close, with approximation errors independent of time t. Then, in the second part, we investigate the stability of the approximate optimal filters. Specifically, given a system S with a prescribed prior π_0, we seek sufficient conditions to guarantee that the truncated system S' (with *the same* prior π_0) generates a sequence of optimal filters which are stable and, at the same time, can attain arbitrarily small approximation errors.

Subspace Acceleration for Large-Scale Bayesian Inverse Problems

Tiangang Cui

Monash University, Australia

ABSTRACT

Algorithmic scalability to high-dimensional models is one of the central challenges in solving large-scale Bayesian inverse problems. By exploiting the interaction among various information sources and model structures, we will present a set of certified dimension reduction methods for identifying the intrinsic dimensionality of inverse problems. The resulting reduced dimensional subspaces offer new insights into the acceleration of classical Bayesian inference algorithms for solving inverse problems. We will discuss some old and new algorithms that can be significantly accelerated by the reduced dimensional subspaces.

Connections Between Optimization and Sampling

Daniel Paulin

University of Oxford, UK

ABSTRACT

Suppose that U is a real valued convex function in a d dimensional Euclidean space.

The problems of

1. finding the minimum of U
2. sampling from a distribution whose negative log-likelihood is U (up to an additive normalising constant)

look quite different at first sight. However, it turns out that there are important connections between these problems. In this short course, we will review some recent results in the literature exploring these connections, and highlight some open problems.

Bayesian Static Parameter Estimation for Partially Observed Diffusions using Multilevel Monte Carlo

Ajay Jasra[a], Kengo Kamatani[b], Kody Law[c], and Yan Zhou[a]

[a] *National University of Singapore*
[b] *Osaka University, Japan*
[c] *The University of Manchester, UK*

ABSTRACT

This talk will consider estimating parameters for partially observed SDE, which is known to be a challenging problem. A popular class of methods for solving this problem are particle MCMC (pMCMC) methods, such as particle marginal Metropolis-Hastings. Such methods leverage non-negative unbiased estimators of the marginal likelihood from a particle filter conditioned on a given parameter value within a pseudo-marginal MCMC algorithm in order to obtain an asymptotically exact algorithm without ever computing the marginal likelihood exactly. A further remarkable innovation of the pMCMC beyond standard pseudo-marginal methods is the ability to sample consistently from the joint distribution on the parameters and the hidden state path. Here we assume furthermore that the SDE giving rise to the hidden process cannot be solved exactly, and must be approximated at finite resolution. It is well-known that in such contexts the multilevel Monte Carlo (MLMC) method can be used to substantially reduce the cost to achieve a given level of error. The idea is to represent the target expectation as a telescopic sum of increments of increasing cost, and estimate the increments using targets which are coupled in such a way that the increments have decreasing variance. A schedule of decreasing sample numbers can then be carefully constructed based upon the relationship between the variance and the cost, resulting in a substantial speedup. In the context of interest here it is not clear how to construct an exact coupling, and we instead appeal to a carefully constructed approximate coupling of the pairs of particle filters. It will be shown how to construct a consistent estimator with optimal speedup via the approximate coupling.

Uniform Estimates for Particle Filters

Adrian Bishop[a], Pierre Del Moral[b], Alice Guionnet[c], Kengo Kamatani[d], Aline Kurtzmann[e], Laurent Miclo[f], Bruno Rémillard[g], and Julian Tugaut[h]

[a] *Data61/CSIRO and University of Technology Sydney, Australia*
[b] *Institut National de Recherche en Informatique et en Automatique, France*
[c] *École normale supérieure de Lyon, France*
[d] *Osaka University, Japan*
[e] *Université de Lorraine, France*
[f] *Université Paul Sabatier, France*
[g] *HEC Montréal, Canada*
[h] *Télécom Saint-Étienne, France*

ABSTRACT

This talk is concerned with the long time behavior of particle filters and Ensemble Kalman filters. These filters can be interpreted as mean field type particle interpretation of the filtering equation and the Kalman recursion. We present a series of old and new results on the stability properties of these filters. We initiate a comparison between these particle samplers and discuss some open research questions.

A Short Introduction to Approximate Bayesian Computation (ABC)

David Nott

National University of Singapore

ABSTRACT

Bayesian inference is an attractive framework for combining information and uncertainty quantification. However, for some statistical models of interest it may be difficult to compute the likelihood, and this complicates the application of usual Bayesian computational methods. If it is possible to simulate data from the model, Bayesian inference can sometimes be performed without evaluating the likelihood, using so-called likelihood-free inference methods. This talk is a short introduction to these methods focusing mostly on approximate Bayesian computation (ABC) approaches. After considering first the earliest ABC algorithms, more advanced topics will be discussed such as Markov chain Monte Carlo and Sequential Monte Carlo ABC, regression post-processing adjustments, variational methods and expectation propagation.

Approximating High-dimensional Posteriors with Nuisance Parameters via Integrated Rotated Gaussian Approximation

Willem van den Boom[a], Galen Reeves[b], and David B. Dunson[b]

[a]*National University of Singapore*
[b]*Duke University, USA*

ABSTRACT

Posterior computation for high-dimensional data with many parameters can be challenging. This talk lays out a new method for approximating posterior distributions of a low- to moderate-dimensional parameter in the presence of a high-dimensional nuisance parameter. The focus is on regression models and the key idea is to separate the likelihood into two components through a rotation. One component involves only the nuisance parameters, which can then be integrated out using a novel type of Gaussian approximation. We provide theory on approximation accuracy that holds for a broad class of forms of the nuisance component and priors. Simulations show that the proposed method can outperform state-of-the-art posterior approximation approaches.

On the Benefits of Correlating Particles within SMC

Arnaud Doucet[a], Axel Finke[b], Adam M. Johansen[c,d], and Alexandre H. Thiery[b]

[a] *Oxford University, UK*
[b] *National University of Singapore*
[c] *The University of Warwick, UK*
[d] *The Alan Turing Institute, UK*

ABSTRACT

We introduce a general particle MCMC framework which admits most known (reversible) MCMC kernels as special cases and also establish useful connections to various multiple-proposal MCMC approaches such those introduced in Tjelmeland (2004) and Neal (2003; 2011). One of the main benefits of our general particle MCMC framework is that it permits additional dependence between the auxiliary variables generated in the underlying sequential Monte Carlo (SMC)-type algorithm. We examine the benefits of such an approach in two settings:

1. For "unconditional" SMC algorithms, several non-standard approaches such as discrete particle filters (Fearnhead, 1998), sequential Quasi-Monte Carlo methods (Gerber & Chopin, 2015), antithetic particle filters (Bizjajeva & Olsson, 2016) and sequential MCMC methods (Berzuini et al., 1997; Septier & Peters, 2016) can be interpreted as special cases of the general framework. This immediately implies, e.g. that these methods yield unbiased estimates of normalising constants so that they can be used within pseudo-marginal approaches.

2. For "conditional" SMC algorithms, correlating the particles sampled at each time step via MCMC kernels leads to the "embedded hidden Markov models" method from Shestopaloff & Neal (2018). This is a "local" conditional SMC algorithm because it places the particles locally around the reference path. We construct an alternative local conditional SMC algorithm using a random-walk proposal which is suitably scaled with the model dimension. This yields a conditional SMC algorithm which is stable in high dimensions for a fixed number of particles (in contrast, standard conditional SMC algorithms require the number of particles to grow exponentially with dimension).

A Divide-and-Conquer Bayesian Approach to Large-Scale Kriging

Cheng Li

National University of Singapore

ABSTRACT

We propose a three-step divide-and-conquer strategy within the Bayesian paradigm that delivers massive scalability for any spatial process model. We partition the data into a large number of subsets, apply a readily available Bayesian spatial process model on every subset, in parallel, and optimally combine the posterior distributions estimated across all the subsets into a pseudo posterior distribution that conditions on the entire data. The combined pseudo posterior distribution replaces the full data posterior distribution for predicting the responses at arbitrary locations and for inference on the model parameters and spatial surface. Based on distributed Bayesian inference, our approach is called "Distributed Kriging" (DISK) and offers significant advantages in massive data applications where the full data are stored across multiple machines. Our theoretical results show that the DISK posterior distribution achieves the near optimal Bayes L2-risk in estimating the true spatial surface in a variety of applications and provide upper bounds for the number of subsets in these applications as a function of the sample size. The model-free feature of DISK is demonstrated by scaling posterior computations in spatial process models with a stationary full-rank and a nonstationary low-rank Gaussian process (GP) prior. A variety of simulations and a geostatistical analysis of the Pacific Ocean sea surface temperature data validate our theoretical results.

Advancements of the EnKF: inverse problems and optimal transport

Neil Chada

National University of Singapore

ABSTRACT

The ensemble Kalman filter (EnKF) is a powerful but simple algorithm aimed at merging models with observational data. Since its formulation in 1994 it has seen a wide applicability to various mathematical disciplines. In this talk we provide two emerging fields connected to Kalman filtering: inverse problems and optimal transport. The main focus in this talk will be on Bayesian inverse problems such as hierarchical learning, and box-constraint optimisation. We will also specify various limit analysis. At the end we present some recent findings on optimal transport such as scaling limits and its use for the EnKF.

Importance sampling correction versus standard averages of reversible MCMCs in terms of the asymptotic variance

Jordan Franks and Matti Vihola

University of Jyväskylä, Finland

ABSTRACT

We will discuss in this talk an ordering criterion for the asymptotic variances of two consistent Markov chain Monte Carlo (MCMC) estimators: an importance sampling (IS) estimator, based on an approximate reversible chain and subsequent IS weighting, and a standard MCMC estimator, based on an exact reversible chain. Essentially, we relax the criterion of the Peskun-Tierney ordering, and obtain, in place of a strict ordering, a bound of the asymptotic variance of IS by that of the direct MCMC. Although simple examples show that IS can have arbitrarily better or worse asymptotic variance than Metropolis-Hastings and delayed acceptance MCMC, our ordering implies that IS is guaranteed to be competitive up to a factor depending on the supremum of the (marginal) IS weight. We elaborate upon the criterion in case of unbiased estimators as part of an auxiliary latent variable framework. We show how the criterion involving the marginal (parameter space) IS weight implies asymptotic variance guarantees for IS in terms of direct MCMC versions of pseudomarginal, particle marginal Metropolis-Hastings MCMC, and delayed acceptance chains. These guarantees help justify the usage of IS, especially with massive parallelisation, as the IS correction phase can be fully parallelised. Besides the asymptotic variance, we also discuss convergence properties of the IS chain, which can show greater robustness to unbiased estimator noise than do direct versions of the aforementioned chains. This talk is based on [1].

References

[1] J. Franks and M. Vihola. Importance sampling correction versus standard averages of reversible MCMCs in terms of the asymptotic variance. Preprint arXiv:1706.09873v3, 2017.

Inference for Multi-object Dynamical Systems: Methods and Analysis

Jeremie Houssineau

National University of Singapore

ABSTRACT

The notion of single-object dynamical system and the associated Bayesian inference methodology are the subject of a vast literature and have been appropriately understood for the practical purpose of most application areas. However, the theoretical understanding of the problem as well as the availability of efficient solutions is more limited in the situation where i) multiple dynamical systems are simultaneously present, ii) the number of constituent single-object systems changes over time according to a birth-death process, and iii) the object-wise observations, in addition to being likely to fail, are corrupted by the presence of false positives. In this talk, I will give an overview of the techniques currently available to model and solve this class of problems. I will also cover the few attempts that have been made toward the analysis of multi-object systems as well as some of the many remaining open questions.

The Birch Probabilistic Programming Language

Lawrence Murray

Uppsala University, Sweden

ABSTRACT

I will introduce a new universal probabilistic programming language called Birch (www.birch-lang.org). Probabilistic programming is an emerging research area across machine learning, statistics, artificial intelligence, and programming languages. It is a programming paradigm that aims to accelerate workflow with new programming languages and software tools tailored for probabilistic modeling and inference. Birch, in particular, is an imperative language geared toward object-oriented and generic programming paradigms. It aims to support complex models and complex inference methods in an intuitive way. Both are implemented in the same language, with specific support for probabilistic constructs. The language draws inspiration from several sources, notably from LibBi—for which it is something of a successor—but in moving from model specification language to universal probabilistic programming language it draws ideas from modern object-oriented programming languages such as Swift, too. I will present some of the design philosophy, some of the language features that support the probabilistic programming paradigm, and several practical examples.

Controlled sequential Monte Carlo

Jeremy Heng

Harvard University, USA

ABSTRACT

Sequential Monte Carlo methods, also known as particle methods, are a popular set of techniques to approximate high-dimensional probability distributions and their normalizing constants. They have found numerous applications in statistics and related fields as they can be applied to perform state estimation for non-linear non-Gaussian state space models and Bayesian inference for complex static models. Like many Monte Carlo sampling schemes, they rely on proposal distributions which have a crucial impact on their performance. We introduce here a class of controlled sequential Monte Carlo algorithms, where the proposal distributions are determined by approximating the solution to an associated optimal control problem using an iterative scheme. We provide theoretical analysis of our proposed methodology and demonstrate significant gains over state-of-the-art methods at a fixed computational complexity on a variety of applications.

Monte Carlo Computational Methods for Stochastic Nonlinear Optimal Control

Adrian Bishop

Data61/CSIRO and University of Technology Sydney, Australia

ABSTRACT

This talk will cover a computational framework for nonlinear stochastic optimal control based on Monte Carlo methods. Roughly speaking, the optimal controller in the nonlinear stochastic setting is related to the solution of a deterministic partial differential equation (PDE): the Hamilton-Jacobi-Bellman (HJB) equation. We will discuss how the solution to the HJB equation can be formulated in terms of an expectation over a stochastic trajectory defined by an uncontrolled stochastic differential equation. Indeed, this relationship is a simple consequence of the Feynman-Kac formula (and more generally its nonlinear variants). It then follows that this expectation (or path-integral) can be approximated via Monte Carlo simulation and we show specifically how this leads to a Monte Carlo approximation of the optimal controller. We discuss a number of details involved in carrying out this Monte Carlo simulation; some of which are particular to the control framework. We discuss the stability of the controlled system given the approximated controller. A number of other recent extensions and (open) discussion points will also be touched upon.

Sidestepping intractability by augmentation: auxiliary variable inference methods

Matthew Graham

National University of Singapore

ABSTRACT

This three-part session will review approaches for performing approximate inference in 'intractable' models - specifically where a density function of the target distribution cannot be evaluated with a practical amount of computational resource. This can arise for example when performing Bayesian inference in undirected graphical models or complex simulators specified in code. Even if we cannot evaluate a density exactly we can often however compute an unbiased density estimate and a range of algorithms have been proposed for performing approximate inference in this case. I will give an introductory overview of these approaches and show that they can be considered as particular choices of 'standard' approximate inference methods applied to a distribution on a space augmented with additional auxiliary variables generated when computing the density estimate. Although these auxiliary variables are often hidden in the algorithms used in practice, a series of recent papers have shown that there can be substantial computational gains to explicitly considering both the auxiliary and target variables when performing inference. I will review a selection of these auxiliary variable methods, attempting to give some intuitions as to why they can improve performance as well illustrating these benefits with various numerical examples.

Ensemble Kalman Filter and its performance in high dimension settings

Xin Tong

National University of Singapore

ABSTRACT

Ensemble Kalman filter (EnKF) is an important data assimilation method for high-dimensional geophysical systems. Efficient implementation of EnKF in practice often involves various inflation and localization techniques. We will first discuss these formulation and how are they applied.

Then we rigorously analyze the filter error of EnKF for linear systems, and show that the filter error can be dominated by the ensemble covariance, as long as either 1) there is a low effective dimension, or 2) the forecast covariance matrix admits a stable localized structure. In particular, this indicates that with small system and observation noises, the filter error will be accurate in long time even if the initialization is not.

Posterior convergence analysis of α-stable processes

Neil Chada

National University of Singapore

ABSTRACT

This work is concerned with the theoretical understanding of α-stable processes $\tilde{X}$ in $\mathbb{R}^d$. Our motivation for this is in the context of Bayesian inverse problems where we consider the treatment of these processes as prior forms. Specifically we aim to tackle the issue of edge-preserving inversion, which has remained a challenge in this field. We derive numerous convergence results referring to finite-dimensional convergence of random variables. In doing so we introduce a number of variants which these processes can take, such as a linear representation and through Poisson processes. Well-posedeness for the inverse problem is also considered for these set of priors. Our results and methodology are specific to the case of both finite dimensional data and for $d = 1$ dimensional random fields.