ADVANCES IN SCALABLE BAYESIAN COMPUTATION

Luke Bornn (Harvard University),
Nando de Freitas (University of British Columbia),
Christian Robert (Ceremade - Université Paris-Dauphine),
Scott Schmidler (Duke University)

Sunday, March 2 to Friday March 7, 2014

1 Workshop Objectives

As the statistical models used to understand complex systems grow, the methods used to fit these models must scale accordingly. While advanced computational methods are being developed to fit these complex models, their speed and memory requirements often demand tremendous computational power via large clusters. This approach of relying on ever larger computational resources is quickly becoming unsustainable, particularly for practitioners for whom these resources are not available. As such, there is a large and growing need for statistically efficient methods which scale in terms of speed and memory while being straightforward to implement and communicate.

Several communities are currently working on methods for Bayesian computation which are scalable; however, interaction between these communities has been limited and as such significant cross-community learnings are being missed. As the communities grow and diverge, it is important to bridge the gap at this crucial stage in the development of scalable methods.

Currently, the exchanges between the Monte Carlo (including MCMC, SMC, and ABC), INLA, optimization, and other communities remain surprisingly limited. With each group having their own workshops and conferences, and little work being done to compare and educate between the various approaches, this workshop strived to:

(1) Gather people from different research communities and foster links between these communities. (2) Expose the various communities to the state of the art in scalable Bayesian computation methods, including MCMC, SMC, ABC, INLA, optimization, and other methods. (3) Classify the advantages, limitations, and possibilities from each class of methods, and equip all participants with the knowledge to bridge the existing gap in scalable Bayesian computation. (4) Create a venue to encourage innovation through the synthesis and development of new approaches by combining existing, currently disjoint, approaches to Bayesian computation.

In reaching out to statisticians, computer scientists, mathematicians, and others working in Bayesian computation as potential participants, the desire for such a workshop was overwhelming. To quote several of the contacted participants, a fantastic and very timely conference topic, sounds incredibly interesting and highly pertinent, a particularly timely topic, and a great opportunity to see what others are doing.

2 Overview of the Field

While linear models under the assumption of Gaussian noise were the hallmark of early 20th century statistics, the past several decades have seen an explosion in statistical models which produce complex and high-dimensional density functions for which simple, analytical integration is impossible. This growth was largely fueled by renewed interest in Bayesian statistics accompanying the Markov chain Monte Carlo (MCMC) revolution in the 1990s. With the computational power to explore the posterior distributions arising from Bayesian models, MCMC allowed practitioners to build models of increasing size and complexity.

Though advanced MCMC methods are being actively developed to fit the extreme scale data arising in fields ranging from sociology to climatology, their speed and memory requirements often demand tremendous computing power. This approach of relying on ever larger computational resources is quickly becoming unsustainable, particularly for practitioners for whom these resources are not available. As such, there is a persistent and growing need for statistically efficient methods which scale in terms of speed and memory while also allowing for straightforward implementation and communication.

In response to the computational cost associated with MCMC, the machine learning community has steadily moved towards variational inference, which approximates the original complex model with a simple, tractable model. While such approximations lead to the development of quick, simple solvers, they are known to systematically underestimate uncertainty, and it is notoriously difficult to judge the accuracy of the implicit approximation in general situations.

3 Presentation Highlights

3.1 Day 1

This was the first day of our workshop Advances in Scalable Bayesian Computation and the theme was probabilistic programming. Indeed, both Vikash Mansinghka and Frank Wood gave talks about this concept, Vikash detailing the specifics of a new programming language called Venture and Frank focussing on his state-space version of the above called Anglican. This is a version of the language Church, developed to handle probabilistic models and inference.

The other talks of Day 1 were of a more classical nature with Pierre Jacob explaining why non-negative unbiased estimators were impossible to provide in general, including an interesting objective Bayes example. Then Sumeet Singh presented a joint work with Nicolas Chopin on the uniform ergodicity of the particle Gibbs sampler with a nice coupling proof. And Maria Lomeli gave an introduction to the highly general Poisson-Kingman mixture models as random measures, which encompasses all of the previously studied non-parametric random measures, with an MCMC implementation that included a latent variable representation for the alpha-stable process behind the scene, representation that could be (and maybe is) also useful in parametric analyses of alpha-stable processes.

We also had an open discussion in the afternoon that ended up being quite exciting, with a few of us voicing out some problems or questions about existing methods and others making suggestions or contradictions.

Speaker: **Mansinghka, Vikash** (MIT)
Title: *Probabilistic computing for Bayesian inference*
Abstract: Although probabilistic modeling and Bayesian inference provide a unifying theoretical framework for uncertain reasoning, they can be difficult to apply in practice. Inference in simple models can seem intractable, while more realistic, flexible models can be difficult to specify, let alone implement correctly. My talk will describe three prototype probabilistic computing systems — including probabilistic programming languages, a Bayesian database system, and intentionally stochastic hardware — designed to mitigate these challenges.

I will focus on Venture, a new, open-source, Turing-complete probabilistic programming platform that aims to be sufficiently expressive, extensible and efficient for general-purpose use. Venture programmers specify models via executable code, where random choices correspond to latent variables; this approach can yield a 100x reduction in code size for state-of-the-art models. Multiple scalable, computationally universal inference algorithms are provided, based on MCMC, conditional SMC and mean field variational techniques. Unlike probabilistic programming tools like BUGS or Church, users can also reprogram the inference strategy for each application and easily implement novel approximation schemes. I will review applications of Venture to text modeling with millions of observations; 2D and 3D computer vision problems; and structured inverse problems in geophysics.

I will also touch on the ways the ideas behind Venture fit together into a mathematically coherent software and hardware stack for Bayesian inference.

This talk includes joint work with Daniel Roy, Eric Jonas, Daniel Selsam and Yura Perov.

Speaker: **Jacob, Pierre** (University of Oxford)
Title: *On non-negative unbiased estimators*
Abstract: I will talk about performing exact inference using Monte Carlo methods, that is, estimating an integral in such a way that the error goes to zero when the computational effort increases to infinity. This is in general possible when the target probability density function can be unbiasedly estimated pointwise, as in the pseudo-marginal approach. However it requires almost surely non-negative unbiased estimators which are not always available. In this talk I will talk about schemes providing unbiased estimators and about the sign problem, leading to a general result stating the non-existence of schemes yielding non-negative unbiased estimators. I will discuss the consequences of that result in some statistical settings.

Speaker: **Lomeli, Maria** (Gatsby Unit, University College London)
Title: *Marginal Sampler for σ-Stable Poisson-Kingman Mixture Models*
Abstract: Bayesian nonparametric mixture models reposed on random probability measures like the Dirichlet process allow for flexible modelling of densities and for clustering applications where the number of clusters is not fixed a priori. Our understanding of these models has grown significantly over the last decade: there is an increasing realisation that while these models are nonparametric in nature and allow an arbitrary number of components to be used, they do impose significant prior assumptions regarding the clustering structure which may or may not conform to actual prior beliefs.

In recent years, there is a growing interest in extending modelling flexibility beyond the classical Dirichlet process first proposed by Ferguson. Examples include Pitman-Yor processes, normalized inverse Gaussian processes, and normalized random measures. With each process, additional flexibility comes with additional developments in characterizing properties of each process. These characterizations are useful to develop tractable Markov chain Monte Carlo posterior simulation algorithms.

In this talk we explore the use of a very wide class of random probability measures, called σ-stable Poisson-Kingman processes or Gibbs-type priors with positive indices, for Bayesian nonparametric mixture modelling. This class of processes encompasses all known tractable random probability measures proposed in the literature so far. We argue that it forms a natural class in which to study. We propose a number of characterizations of the process which allow the development of tractable inference algorithms. Specifically, we develop an efficient marginal sampler which can be used for posterior simulation for the whole class of nonparametric priors.

(Joint work with Stefano Favaro and Yee Whye Teh)

Speaker: **Singh, Sumeetpal** (Unievrsity of Cambridge)
Title: *On Particle Gibbs Sampling*
Abstract: The particle Gibbs sampler is a Markov chain Monte Carlo (MCMC) algorithm to sample from the full posterior distribution of a state-space model. It does so by executing Gibbs sampling steps on an extended target distribution defined on the space of the auxiliary variables generated by an interacting particle system. This paper makes the following contributions to the theoretical study of this algorithm. Firstly, we present a coupling construction between two particle Gibbs updates from different starting points and we show that the coupling probability may be made arbitrarily close to one by increasing the number of particles. We obtain as a direct corollary that the Particle Gibbs kernel is uniformly ergodic. Secondly, we show how the inclusion of an additional Gibbs sampling step that reselects the ancestors of the Particle Gibbs extended target distribution, which is a popular approach in practice to improve mixing, does indeed yield a theoretically more efficient algorithm as measured by the asymptotic variance. Thirdly, we extend Particle Gibbs to work with lower variance resampling schemes. A detailed numerical study is provided to demonstrate the efficiency of Particle Gibbs and the proposed variants.

This is joint work with Nicolas Chopin.

Speaker: **Wood, Frank** (University of Oxford)
Title: *A new approach to probabilistic programming inference*
Abstract: Probabilistic programming languages hold the promise of dramatically accelerating the development of both new statistical models and inference strategies. In probabilistic programs, variables can take on random values at run time and inference is performed by calculating expectation values over all execution traces that are in agreement with a set of observed data. This allows statistical models to be represented in a concise and intuitive manner, enabling more rapid iteration over model variants. Inference schemes, when implemented as a backend to a programming framework, can easily be tested on a large collection of models, enabling a much more systematic comparison of the efficacy of inference strategies.

We introduce the use of particle Markov chain Monte Carlo to perform probabilistic programming inference. Our approach is simple to implement, easy to parallelize, and supports accurate inference in models that make use of complex control flow, including stochastic recursion. It also includes primitives from Bayesian nonparametric statistics. Our experiments show that this approach can be more efficient than previously introduced single-site Metropolis-Hastings methods.

Joint work with Brooks Paige, Jan Willem van de Meent, and Vikash Mansingkha

3.2 Day 2

The main theme of the second day was about brains In fact, Simon Barthelms research originated from neurosciences, while Dawn Woodard dissected a brain (via MRI) during her talk! Simons talk was quite inspiring, using Tibshirani et al.s trick of using logistic regression to estimate densities as a classification problem central to the method and suggesting a completely different vista for handling normalising constants. Then Raazesh Sainudiin gave a detailed explanation and validation of his approach to density estimation by multidimensional pavings/histograms, with a tree representation allowing for fast merging of different estimators. Raaz had given a preliminary version of the talk at CREST last Fall, which helped with focussing on the statistical aspects of the method. Chris Strickland then exposed an image analysis of flooded Northern Queensland landscapes, using a spatio-temporal model with changepoints and about 18,000 parameters, still managing to get an efficiency of O(np) thanks to two tricks. Then it was time for the group photograph outside in a balmy -18 and an open research time that was quite profitable.

In the afternoon sessions, Paul Fearnhead presented an auxiliary variable approach to particle Gibbs, which again opened new possibilities for handling state-space models, but also connecting to Xiao-Li Mengs reparameterisation devices. Questions were asked whether or not the SMC algorithm was that essential in a static setting, since the sequence could be explored in any possible order for a fixed time horizon. Then Emily Fox gave a 2-for-1 talk, mostly focussing on the first talk, where she introduced a new technique for approximating the gradient in Hamiltonian (or Hockey!) Monte Carlo, using second order Langevin. She did not have much time for the second talk, which intersected with the one she gave at BNPski in Chamonix, but focussed on a notion of sandwiched slice sampling where the target density only needs bounds that can get improved if needed. A cool trick! And the talks ended with Dawn Woodards analysis of time varying 3-D brain images towards lesion detection, through an efficient estimation of a spatial mixture of normals.

Speaker: **Simon Barthelmé** (University of Geneva)
Title: *LATKES in Space: flexible models for spatial sequences*
Abstract:

Most questions that arise in the analysis of spatial data are some variant of the following: “Why are there more X in area A than in other places?”. In epidemiology for example, the question pertains to disease cases, in ecology, to some particular plant or animal, and in astrostatististics, to galaxies. Point process models try to answer these questions by relating the intensity of a phenomenon in a certain region of space to a set of spatial covariates that might explain the data.

An interesting application of point process models is to eye movements, where the goal is to understand why people look in certain places more often than others. In applications of spatial statistics to eye movements the *temporal* dimension also plays an important role: where the eyes move to depends on where they were before. In this talk I will outline an extension of point process models to spatial sequences, based on *Log-Additive Transition Kernels* (LATKES).

I will show that the "logistic regression trick" used for non-parametric density modelling, can be extended to sequences of dependent data. A consequence is that inference in LATKES models can be transformed into logistic regression problem, at the cost of some Monte Carlo error. I will discuss links with the Poisson-multinomial transform, variational bounds on the likelihood, and inference in unnormalised statistical models, a set of topics which turn out to be closely related.

Speaker: **Sainudin, Raazesh** (University of Canterbury)
Title: *Statistical Regular Pavings for Bayesian Non-parametric Density Estimation*
Abstract: We present a novel method for averaging a sequence of histogram states visited by a Metropolis-Hastings Markov chain whose stationary distribution is the posterior distribution over a dense space of tree-based histograms. The computational efficiency of our posterior mean histogram estimate relies on a statistical data-structure that is sufficient for non-parametric density estimation of massive, multi-dimensional metric data. This data-structure is formalized as statistical regular paving (SRP). A regular paving (RP) is a binary tree obtained by selectively bisecting boxes along their first widest side. SRP augments RP by mutably caching the recursively computable sufficient statistics of the data. The base Markov chain used to propose moves for the Metropolis-Hastings chain is a random walk that data-adaptively prunes and grows the SRP histogram tree. We use a prior distribution based on Catalan numbers and detect convergence heuristically. The L1-consistency of the the initializing strategy over SRP histograms using a data-driven randomized priority queue based on a generalized statistically equivalent blocks principle is proved by bounding the Vapnik-Chervonenkis shatter coefficients of the class of SRP histogram partitions. The performance of our posterior mean SRP histogram is empirically assessed for large sample sizes simulated from several multivariate distributions that belong to the space of SRP histograms.

We also present arithmetical capabilities of the SRPs, including tree-based algorithms and structures for marginalization, conditional density extraction, fast look-up of product likelihood in validation, uniform approximation of other density estimates as SRP histograms and more general arithmetic operations. These operations are used in an SRP-based ABC method and in fast cross-validation for prior-selection.

This is joint work with Dominic Lee, Jennifer Harlow and Gloria Teng.

Speaker: **Christopher Strickland** (UNSW)
Title: *A scalable Bayesian changepoint methodology for large space-time data sets*
Abstract: A scalable Bayesian changepoint methodology is developed to analyse large space-time data sets, which allows for an unknown number of change points in the common components of a hierarchical space-time model. The computational cost scales linearly with the sample size, across both space and time, and the methodology is simulation efficient. The methodology is used to assess the impact of extended inundation on the ecosystem of the Gulf Plains bioregion in northern Australia. Our data set consists of nearly 5 million observations, and our methodology is sufficiently efficient to conduct a full Bayesian analysis in tens of minutes, despite the complexity of the proposed model.

Speaker: **Fearnhead, Paul** (Lancaster University)
Title: *Reparameterisations for Particle MCMC*
Abstract: Consider inference for a state-space model with unknown parameters. Standard implementation of Particle MCMC involves using an MCMC kernel to propose a new value for the parameters, and then a particle filter to propose a new value for the path of the state process conditional on this value for the parameters. However alternative approaches are possible. For example we could use the particle filter to update some of the parameters (as well as the state). This reduces the dependence in the MCMC update, but at the expense of greater Monte Carlo error when running the particle filter. We consider generalisations of this approach, based on introducing pseudo observations. This allows us to run the particle filter conditional on part of the information about the current set of parameters and state path.

We give examples of how this approach can improve the mixing in particle Gibbs, and help with the initialisation of the particle filter.

Speaker: **Fox, Emily** (University of Washington)
Title: *2-for-1: Stochastic Gradient Hamiltonian Monte Carlo and Bayesian Learning of DPP Kernels*
Abstract: In this talk, we will present two separate vignettes on Bayesian computation. One is stochastic

gradient Hamiltonian Monte Carlo (HMC). HMC methods provide a mechanism for defining distant proposals with high acceptance probabilities in a Metropolis-Hastings framework, enabling more efficient exploration of the state space than standard random-walk proposals. However, a limitation of HMC methods is the required gradient computation for simulating the Hamiltonian dynamical system—such a computation is infeasible in problems involving a large sample size or streaming data. We instead consider the impacts of a noisy gradient estimate computed from a subset of the data. Surprisingly, the natural implementation of the stochastic approximation can be arbitrarily bad. To address this problem we introduce a variant that uses second-order Langevin dynamics with a friction term that counteracts the effects of the noisy gradient, maintaining the desired target distribution as the invariant distribution.

In our second vignette, we consider learning the kernel parameters of determinantal point processes (DPPs), a class of repulsive point processes. While DPPs have many appealing properties, such as efficient sampling, learning the parameters of a DPP is still considered a difficult problem due to the non-convex nature of the likelihood function. We explore a set of Bayesian methods to learn the DPP kernel parameters even in large-scale and continuous DPP settings when the exact form of the eigendecomposition is unknown.

The HMC work is in collaboration with Tianqi Chen and Carlos Guestrin. The DPP work is joint with Raja Hafiz Affandi, Ryan Adams, and Ben Taskar.

Speaker: **Woodard, Dawn** (Cornell University)
Title: *Model-Based Image Segmentation / Efficiency of MCMC in Parametric Models*
Abstract: I will discuss two topics. First, I discuss an improved method for spatial model-based clustering, and apply it to segment three-dimensional Dynamic Contrast Enhanced Magnetic Resonance (DCE-MR) images. The approach extends an existing Monte Carlo Expectation-Maximization method for Markov random field mixture models, and is guaranteed to converge to a local maximum of the likelihood. The first extension is to show how to incorporate cluster weight parameters in a computationally tractable way; these parameters are needed to accurately capture small features in the image. Secondly, we incorporate a covariance decomposition to allow control over geometric characteristics of the segmentation. Thirdly, we give a consistent approximation to the observed-data likelihood.

In the second part of the talk I discuss work analyzing the efficiency of Markov chain Monte Carlo (MCMC) methods used in Bayesian computation. While convergence diagnosis is used to choose how long to run a Markov chain, it can be inaccurate and does not provide insight regarding how the efficiency scales with the number of parameters or other quantities of interest. We instead characterize the number of iterations of the Markov chain (the running time) sufficient to ensure that the approximate Bayes estimator obtained by MCMC preserves the property of asymptotic efficiency. We show that in many situations where the likelihood satisfies local asymptotic normality, the running time grows linearly in the number of observations n.

3.3 Day 3

The theme running through day three was industry, as the three speakers spoke of problems and solutions connected with Google, Facebook and similar companies. First, Russ Salakhutdinov presented some of the video hierarchical structures on multimedia data, like connecting images and text, with obvious applications at Google. The first part described Boltzman machines with impressive posterior simulations of characters and images. Then Steve Scott gave us a Google motivated entry to embarrassingly parallel algorithms. One of the novel things in the talk (for me) was the inclusion of BART in this framework, with the interesting feature that using the whole prior on each machine was way better than using a fraction of the prior, as predicted by the theory! And Joaquin Quinonero Candela provided examples of machine learning techniques used by Facebook to suggest friends and ads in a most efficient way.

Speaker: **Salakhutdinov, Ruslan** (University of Toronto)
Title: *Learning Structured, Robust, and Multimodal Models*
Abstract: Building intelligent systems that are capable of extracting meaningful representations from high-dimensional data lies at the core of solving many Artificial Intelligence tasks, including visual object recognition, information retrieval, speech perception, and language understanding. In this talk I will first introduce a broad class of hierarchical probabilistic models called Deep Boltzmann Machines (DBMs) and show that DBMs can learn useful hierarchical representations from large volumes of high-dimensional data with applications in information retrieval, object recognition, and speech perception. I will then describe a new class of

more complex models that combine Deep Boltzmann Machines with structured hierarchical Bayesian models and show how these models can learn a deep hierarchical structure for sharing knowledge across hundreds of visual categories, which allows accurate learning of novel visual concepts from few examples. Finally, I will introduce deep models that are capable of extracting a unified representation that fuses together multiple data modalities. I will show that on several tasks, including modelling images and text, video and sound, these models significantly improve upon many of the existing techniques.

Speaker: **Scott, Steve** (Google)
Title: *Bayes and Big Data: The Consensus Monte Carlo Algorithm*
Abstract: A useful definition of "big data" is data that is too big to comfortably process on a single machine, either because of processor, memory, or disk bottlenecks. Graphics processing units can alleviate the processor bottleneck, but memory or disk bottlenecks can only be eliminated by splitting data across multiple machines. Communication between large numbers of machines is expensive (regardless of the amount of data being communicated), so there is a need for algorithms that perform distributed approximate Bayesian analyses with minimal communication. Consensus Monte Carlo operates by running a separate Monte Carlo algorithm on each machine, and then averaging individual Monte Carlo draws across machines. Depending on the model, the resulting draws can be nearly indistinguishable from the draws that would have been obtained by running a single machine algorithm for a very long time. Examples of consensus Monte Carlo are shown for simple models where single-machine solutions are available, for large single-layer hierarchical models, and for Bayesian additive regression trees (BART).

Speaker: **Darren Wilkinson** (Newcastle University))
Title: *Parallelisation strategies for Monte Carlo algorithms*
Abstract: An overview will be presented of parallelisation strategies for Monte Carlo algorithms such as ABC, SMC, MCMC, and pMCMC, with particular emphasis on the problem of inference for intractable Markov process models. Different algorithms have differing degrees of amenability to parallelisation, leading to trade-offs with statistical efficiency. A hybrid approach will be discussed, which attempts to effectively utilise multicore systems. The potential benefits of functional programming languages, immutable data structures and "monadic" algorithm design for scalable Bayesian computation will also be examined.

3.4 Day 4

Still looking for a daily theme, parallelisation could be the right candidate, even though other talks this week went into parallelisation issues, incl. Steves talk Wednesday. Indeed, Anthony Lee gave a talk this morning on interactive sequential Monte Carlo, where he motivated the setting by a formal parallel structure. Then, Darren Wilkinson surveyed the parallelisation issues in Monte Carlo, MCMC, SMC and ABC settings, before arguing in favour of a functional language called Scala. In the afternoon session, Sylvia Frhwirth-Schnatter exposed her approach to the (embarrassingly) parallel problem, in the spirit of Steves, David Dunsons and Scotts. Marc Suchard mostly talked about flu and trees in a very pleasant and broad talk, he also had a slide on parallelisation to fit the theme! Although unrelated with parallelism, Nicolas Chopins talk was on sequential quasi-Monte Carlo algorithms, and was full of exciting stuff. Similarly, Alex Lenkoski spoke about extreme rain events in Norway with no trace of parallelism, but the general idea behind the examples was to question the notion of the calibrated Bayesian (with possible connections with the cut models).

Speaker: **Lee, Anthony** (University of Warwick)
Title: *On the role of interaction in sequential Monte Carlo algorithm*
Abstract: Motivated largely by issues surrounding parallel implementation of particle filters, we introduce a general form of sequential Monte Carlo (SMC) algorithm defined in terms of a parameterized resampling mechanism. We find that a suitably generalized notion of the Effective Sample Size (ESS), widely used to monitor algorithm degeneracy, appears naturally in a study of its convergence properties. We are then able to phrase sufficient conditions for time-uniform convergence in terms of algorithmic control of the ESS, in turn achievable by adaptively modulating the interaction between particles. This leads us to suggest novel algorithms which are, in senses to be made precise, provably stable and yet designed to avoid the degree of interaction which hinders parallelization of standard algorithms. As a by-product we prove time-uniform

convergence of the popular adaptive resampling particle filter. The resulting scheme provides some indication of how one can achieve scalable SMC algorithms in practice.

This is joint work with Nick Whiteley and Kari Heine.

Speaker: **Quinonero Candela, Joaquin** (Facebook)
Title: *Examples and Lessons from Machine Learning at Facebook*
Abstract: The problem of selecting what information to present users is prevalent at Facebook. In this talk I will give a couple of examples of machine learning applied to rank Facebook content. I will also describe a couple of simple tricks to operate at scale. Finally, I will share some practical lessons learnt from ranking ads, the area I have been working on the past year and a half.

Speaker: **Chopin, Nicolas** (ENSAE-CREST)
Title: *Sequential Quasi-Monte Carlo*
Abstract: We develop a new class of algorithms, SQMC (Sequential Quasi Monte Carlo), as a variant of SMC (Sequential Monte Carlo) based on low-discrepancy points. The complexity of SQMC is O(N log N) where N is the number of simulations at each iteration, and its error rate is smaller than the Monte Carlo rate $O(N^{-1/2})$. The only requirement to implement SQMC is the ability to write the simulation of particle x_t^n given x_{t-1}^n as a deterministic function of x_{t-1}^n and uniform variates. We show that SQMC is amenable to the same extensions as standard SMC, such as forward smoothing, backward smoothing, unbiased likelihood evaluation, and so on. In particular, SQMC may replace SMC within a PMCMC (particle Markov chain Monte Carlo) algorithm. We establish several convergence results. We provide numerical evidence in several difficult scenarios than SQMC significantly outperforms SMC in terms of approximation error.

(Joint work with Mathieu Gerber)

Speaker: **Frühwirth-Schnatter, Sylvia** (WU Vienna University of Economics and Business)
Title: *Merging parallel MCMC output for horizontally partitioned data*
Abstract: Horizontally partitioned data frequently occur when different entities cannot pool or share their data, either due to privacy protection, or if data are too large to be analyzed within a single analysis due to the computational burden. One way to handle this problem is to partition the data and perform independent manageable analyzes in parallel on each partition. Somewhat surprisingly, combining and merging the independently obtained results to obtain overall Bayesian inference for the whole sample is far from trivial.

The first part of the talks reviews and comments on several methods that have been proposed recently in this context, such as embarrassingly parallel MCMC, the consensus Monte Carlo algorithm, and parallel MCMC via Weierstrass sampler. However, none of the existing methods applies a fully Bayesian analysis based on MCMC sampling for each of the sub-samples and merging the draws into a single sample from the joint posterior distributions given all data.

The second part of the talk discussed two approaches to construct such a merged sample directly from the MCMC sub samples obtained for partitioned data. The first approach uses naive reweighting of the observations of one sample with the estimated posterior density of the other. The second approach applies an independence Metropolis-Hastings-algorithm to merge two samples from different partitions into one sample from the joint posterior. This algorithm is based on using one of the posteriors as proposal which leads to quite a simple acceptance rate for the draws from this posterior. Both procedure are applied sequentially over all sub samples.

We evaluate the performance of both approaches for various examples of univariate Bayesian inference and discuss how to extend these approaches to multivariate Bayesian inference and hierarchical models.

(Joint work with Alexandra Posekany.)

Speaker: **Lenkoski, Alex** (Norsk Regnesentral/Norwegian Computing Center)
Title: *Hierarchical Bayesian Methods in Modern Industrial Statistics*
Abstract: The reliance on computationally intensive hierarchical methods and the (re-)emergence of the Bayesian paradigm is now thoroughly established in academic statistical research. Further, through the parallel development of machine learning methods, innumerable techniques have been developed to handle a variety of problems faced by information-oriented technology companies. However, in the more staid domain of industrial statistics, uptake of computationally intensive methods has been somewhat more subdued.

This talk will present four applications across the industrial realm where advanced Bayesian methods either have been, or should be, used to good effect. These applications include: Mapping regional heterogeneity of extreme precipitation levels in Norway, issuing a calibrated prediction that an insurance claim is fraudulent, creating temporally coherent predictive distributions of wind power production in Germany and using bid-ask curve data to form distributional electricity price forecasts in the Nordic region. In the course of each project, a host of questions came up regarding the appropriate implementation and ultimate usefulness of modern, cutting-edge methods over more standard statistical routines. We will highlight these issues and hopefully discuss how to truly integrate current research into industrial solutions.

Speaker: **Suchard, Marc** (UCLA)
Title: *Scaling Bayesian models for large-scale infectious disease surveillance*
Abstract: Influenza viruses undergo continual antigenic evolution allowing mutant viruses to evade host immunity. Antigenic phenotype is often assessed through pairwise measurement of cross-reactivity between influenza strains using the hemagglutination inhibition (HI) assay. Large-scale experimentation is generating HI measurements between thousands of strains. Semi-parametric clustering of these measurements guides annual influenza vaccine selection, but current methods fail to account for uncertainty in cluster identification and the unobserved shared evolutionary history between strains. We propose a semi-parametric Bayesian model that uses multidimensional scaling to map pairwise measurements into a latent low-dimensional space. Viral coordinates in this space identify clusters and the viruses' shared history provides a prior distribution over coordinates through a Browian diffusion process along the history. Central to the success of this model lies a novel sum-product algorithm that transforms an $O(N^3K^3)$ into an $O(NK^2)$ computation for the Brownian diffusion, where N counts the number of viruses and K counts the dimension of the phenotypes in the database. The algorithm exploits the directed acyclic graph structure of the shared history and opens the door for joint inference of multiple, high-dimensional phenotypes. We show that seasonal influenza A/H3N2 evolves faster and in a more punctuated fashion than other influenza lineages. [Joint work with Trevor Bedford, Andrew Rambaut and Philippe Lemey]

4 Outcome of the Meeting

The meeting provided a great opportunity for many researchers from around the world to come together and discuss different approaches to making Bayesian computation scale to modern problems. It was a uniform success, with many excellent discussions between research groups which had previously not communicated, and a dissemination of ideas across fields which has opened many avenues for future research. To quote one participant, Simon Barthelmé (University of Geneva),

> Attending the workshop has been tremendously valuable. I was able to get an overview of the state-of-the-art in the field, and interact closely with some of its best researchers. What made it especially interesting for me is that I work in a neighbouring field and cannot easily keep myself abreast of all the latest developments. I've received excellent feedback on my own work, and I've come home with many new ideas as well as excellent memories of the workshop.

Current research in Bayesian computation seeks innovation through one of three mechanisms: adaptation, parallelization, or model simplification. The first, adaptation, strives to have the algorithm tune itself as it runs. The second, parallelization, seeks to cleverly distribute computation across clusters and modern graphical processing units. While these techniques generally improve performance, the $O(N^{-1/2})$ convergence of Monte Carlo methods remains.

In contrast, to exploit the idea of model simplification, the machine learning community is actively developing fast and efficient variational methods. Unfortunately, however, it can be difficult to gauge the quality of the approximation made by these methods. Efforts are being made to improve and better understand these approximations, though generally such improvements are model-specific, and as such there lacks a unified, yet efficient, computational framework for Bayesian modeling.

This workshop bridged these two communities, and has opened doors for building hybrid methods which avoid the inherent pitfalls of existing approaches.