BAYES COMP 2018
PROGRAM SUMMARY

All talks will be held at Roger de Llúria building, UPF Ciutadella Campus (Ramon Trias-Fargas 25-27). Rooms 40.S02 (yard) and 40.012 (ground floor).

Monday March 26, 2018

08:30-09:20 REGISTRATION

09:20-09:30 Welcome. Room 40.S02
Omiros Papaspiliopoulos (ICREA, Universitat Pompeu Fabra & Barcelona GSE) Chair Scientific Committee

Kerrie Mengersen (Queensland University of Technology and Past-president, International Society for Bayesian Analysis 2017)

09:30-10:20 Plenary Talk. Room 40.S02
Session chair: *David Rossell (Pompeu Fabra and Barcelona GSE, Spain)*

- Merlise Clyde (Duke University, USA)*: Computational Challenges in Bayesian Model Choice*

10:20-10:40 COFFEE BREAK. ROOM EXHIBITION HALL

10:40-12:10 Parallel sessions

Invited session I: Variable Selection. Room 40.S02
Session chair: *David Rossell* (Pompeu Fabra and Barcelona GSE, Spain)

- Sameer Deshpande (Wharton School, University of Pennsylvania, USA). *Simultaneous Variable and Covariance Selection with the Multivariate Spike-and-slab lasso*
- Christopher Hans (Ohio State University, USA). *Strategies for Differential Shrinkage in Regression with Non-Orthogonal Designs*
- Naveen Narisetti (Urbana-Champaign, USA). *Scalable Bayesian Variable Selection for High Dimensional Data*

Contributed session I: Stein's method in Computational Statistics. Room 40.012
Session chair: *Chris Oates* (Newcastle University, UK)

- Sebastian Vollmer (Warwick & Alan Turing Institute, UK). Measuring Sample Quality with Diffusions
- Christophe Ley (Ghent University, Belgium). A Probabilistic Measure of the Impact of the Prior in Bayesian statistic.
- François-Xavier Briol (Warwick & Imperial College London, UK). Stein Points.

12:10-14:00 LUNCH BREAK

14:00-15:30 Parallel sessions

Invited session II: Computational Machine Learning. Room 40.S02
Session chair: *Michalis Titsias* (AUEB, Greece)

- Michalis Titsias (AUEB, Greece). *Auxiliary gradient-based sampling algorithms*
- Yee W Teh (Oxford-Deep Mind, UK). *TBC*
- Chris Yau (Birmingham, UK). *TBC*

Invited session III: Sequential Monte Carlo. Room 40.012
Session chair: *Anthony Lee* (Bristol, UK)

- Kari Heine (Bath University, UK). *Parallelising particle filters with butterfly interactions*
- Kody Law (Oak Ridge National Laboratory, UK). *Some multilevel Monte Carlo methods*
- Jimmy Olsson (Lund University, Sweden). *Numerically stable online estimation of variance in particle filters*

15:30-15:50 COFFEE BREAK. ROOM EXHIBITION HALL

15:50-17:20 Parallel sessions

Contributed session II: Advances in Piecewise Deterministic Sampling Algorithms. Room 40.S02
Session chair: *Joris Bierkens* (TU Delft, The Netherlands).

- Joris Bierkens (TU Delft, The Netherlands). *Piecewise Deterministic Sampling Algorithms and Exact Subsampling.*
- George Deligiannidis (Oxford, UK) *Exponential Ergodicity of the Bouncy Particle Sampler* (joint work with Alexandre Bouchard-Côté, Arnaud Doucet)
- Pierre-André Zitt, (Université Paris-Est, France). *Irreducibility and Exponential Ergodicity of the Zig-Zag Process.*

Contributed session III: Scalable Monte Carlo: Strategies For Sub-Sampling And Distributed Inference. Room 40.012
Session chair: *Richard Everitt (University of Reading)*

- Hongsheng Dai (Essex, UK). *Monte Carlo Fusion for distributed analysis*
- Chris Nemeth (Lancaster, UK). Stochastic gradient MCMC: computationally efficient Monte Carlo via data subsampling
- Murray Pollock, (Warwick, UK). *TBC*

17:30-POSTER SESSION-DRINKS-SNACKS. ROOM EXHIBITION HALL

Tuesday March 27, 2018

09:30-10:20 Plenary Talk. Room 40.S02
Session chair: *Piotr Zwiernik* (Pompeu Fabra and BGSE, Spain)

- Gareth Roberts (University of Warwick, UK). *Complexity of MCMC*

10:20-10:40 COFFEE BREAK. ROOM EXHIBITION HALL

10:40-12:10 Parallel sessions

Invited session IV. Scalable inference for hierarchical models. Room 40.S02
Session chair: *Giacomo Zanella* (Bocconi, Italy)

- Art B. Owen (Stanford, USA). *Method of Moments for Large Crossed Linear Mixed Effects Models*
- Giacomo Zanella (Bocconi, Italy). *Scalable inferences for crossed random effects models*
- Tamara Broderick (MIT, USA). *Automated Scalable Bayesian Inference via Data Summarization*

Contributed session IV: Adaptive MCMC and High Performance Computing. Room 40.012
Session chair: *Louis Aslett,* (Durham, UK)

- Cyril Chimisov (Warwick, UK). AirMCMC: *Adapted Increasingly Rarely Markov Chain Monte Carlo samplers*
- Nick Tawn (Warwick, UK). *Adapting Towards Optimality of the Parallel Tempering Algorithm*
- Louis Aslett, (Durham, UK). *Contemporaneous MCMC*

12:10-14:30 LUNCH BREAK

14:00-15:30 Parallel sessions

Invited session V: Deterministic approximation methods and pseudo-likelihoods. Room 40.S02
Session chair: *Ioannis.Kosmidis* (Warwick, UK)

- Nancy Reid (Toronto, Canada). *Approximate Likelihood Functions.*
- Helen Ogden (Southampton, UK). *Statistical scalability of approximate likelihood inference*
- Erlis Ruli (Padova, Italy). *Model selection for (approximate) Bayesian inference via estimating functions*

Contributed session V: Computational Bayes for Cancer Genomics Applications. Room 40.012
Session chair: *Robert Castelo* (Pompeu Fabra, Spain)

- Gwenael Leday (Cambridge, UK). *Fast Bayesian inference in Gaussian graphical models and applications to Omics data.*
- Manuela Zucknick (Oslo, Norway). *A Bayesian Mallows rank model for joint analyses of cancer genomics data.*
- Mark van de Wiel, (VU University, The Netherlands). *Improving high-dimensional prediction by empirical Bayes learning from co-data*

15:30-15:50 COFFEE BREAK. ROOM EXHIBITION HALL

15:50-17:20 Parallel sessions

Invited session VI: Intractable Likelihood. Room 40.S02
Session chair: *Anthony Lee* (Bristol, UK)

- Sinan Yildirim (Sabanci, Turkey). *Utilising inference in state-space models with multiple paths from conditional sequential Monte Carlo*
- Jeremy Heng (Harvard, USA). *Controlled sequential Monte Carlo*
- Anthony Lee (Bristol, UK). *Unbiased approximations of products of expectations*

Invited session VII: Asymptotics of MCMC convergence rates. Room 40.012
Session chair: *James Johndow* (Stanford, USA)

- Daniel Rudolf (Goettingen, Germany). Approximation of geometrically ergodic Metropolis-Hastings algorithms
- James Johndow (Stanford, USA). Approximate MCMC: Theory and Practice
- Aaron Smith (Ottawa, Canada). Some Progress and Questions on the Mixing of Hamiltonian Monte *Carlo*

17:30-19:00 Parallel sessions

Contributed session VI: Recent Advances in Bayesian Computation for Characterizing Astrophysical Populations. Room 40.S02
Session chair: *Eric B. Ford* (Penn State, USA)

- Ruth Angus (Columbia, USA). *Planetary systems across time and space*
- Angie Wolfgang (Penn State, USA). *Characterizing the Mass-Radius Relationship for Exoplanet Populations*
- David Stenning (Imperial College London UK). *Using Bayesian Computing to Solve a Complex Problem in Astrophysics*

Fast communications session. Room 40.012
Session chair: *Vicens Gomez* (Pompeu Fabra, Spain)

- Josephine Merhi (Sorbonne Universités, France). *Simultaneous Bayesian Quantile Regression.*

- Alexander Bone (Inria Paris, France). *Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms.*
- Igor Koval (Inria Paria, France). *Statistical Learning of Spatiotemporal Patterns from Longitudinal Manifold-Valued Networks*.
- Ioannis Mitliagkas (Stanford, USA). *Improving Gibbs sampler scan quality with DoGS.*
- Matt Graham (National University of Singapore, Singapore). *Asymptotically exact inference in differentiable generative models.*
- Orhan Sonmez (Bogazici University, Turkey & Pompeu Fabra, Spain). *Employing Sequential Monte Carlo in model-based reinforcement learning.*

Wednesday, March 28, 2018

09:30-10:20 Plenary Talk. Room 40.S02
Session chair: *Omiros Papaspiliopoulos (ICREA, Pompeu Fabra and BGSE, Spain)*

Andrew Stuart (Caltech, USA). *Large Graph Limits of Learning Algorithms*

10:20-10:40 COFFEE BREAK. ROOM EXHIBITION HALL

10:40-12:10 Parallel sessions

Invited session VIII: Applied Math meets Bayes Comp. Room 40.S02
Session chair: *Andrew Duncan* (Sussex, UK)

- Jonathan Weare (University of Chicago, USA). *Stratification for Markov Chain Monte Carlo Simulation*
- Chris Oates (Newcastle, UK). *A Bayesian Conjugate-Gradient Method*
- Aretha Teckentrup (Edinburgh, UK). *Deep Gaussian process priors for Bayesian inverse problems*

Contributed session VII: Illustration of software for BayesComp: NIMBLE and Birch. Room 40.012
Session chair: *Robert Castelo* (Pompeu Fabra, Spain)

- Perry de Valpine (Berkeley, USA): *NIMBLE*
- Lawrence Murray (Upsalla, Sweden): *Birch*

12:10-13:30 LUNCH BREAK

13:30-15:00 Parallel sessions

Invited session IX: Hamiltonian Monte Carlo. Room 40.S02
Session chair: Alex Beskos (UCL, UK)

- Alex Beskos (UCL, UK). *Geometric MCMC for infinite-dimensional inverse problems*
- Jesus Sanz-Serna (Carlos III de Madrid, Spain). *Numerical integrators for the Hamiltonian Monte Carlo method*
- Sam Livingstone (UCL, UK). *Kinetic energy choice in Hamiltonian/Hybrid Monte Carlo*

Invited session X. Bayesian computing for Economics, Business and Finance. Room 40.012
Session chair: *Gregor Kastner* (WU Vienna, Austria)

- Nalan Basturk (Maastricht U, The Netherlands) *Time-varying learning combinations of Bayesian dynamic models and equity momentum strategies* (joint with A. Borowska, S. Grassi, L. Hoogerheide and H.K. van Dijk)
- Dimitris Korobilis (Essex U, UK). *Forecasting with many predictors using message passing algorithms*
- Gregor Kastner (WU Vienna, Austria). *Sparse Bayesian Vector Autoregressions in Huge Dimensions* (joint with F. Huber)

15:10-16:00 Plenary Talk. Room 40.S02
Session chair: *Sylvia Fruehwirth-Schnatter* (Wien, Austria)

- Antonietta Mira (Data Science Center, ICS, Università della Svizzera italiana and DISAT, Università dell'Insubria) *Approximate Bayesian Computation for inference in mechanistic network models (with Ritabrata Dutta (Data Science Center, ICS, Università della Svizzera italiana), JP Onnela and Sixing Chen (Department of Biostatistics, Harvard University))*

16:00-20:00 GUIDED VISITS/SPORT EVENTS

21:00-DINNER

BAYES COMP 2018
ABSTRACT SECTION

MONDAY MARCH 26, 2018

Plenary Talk
Merlise Clyde (Duke University, USA)*: Computational Challenges in Bayesian Model Choice*

Parallel sessions

Invited session I: Variable Selection.

Sameer Deshpande (Wharton School, University of Pennsylvania, USA). **Simultaneous Variable and Covariance Selection with the Multivariate Spike-and-slab lasso** (joint work with Veronika Rockova and Ed George)

We propose a Bayesian procedure for simultaneous variable and covariance selection using continuous spike-and-slab priors in multivariate linear regression models where q possibly correlated responses are regressed onto p predictors. Rather than relying on a stochastic search through the high-dimensional model space, we develop an ECM algorithm similar to the EMVS procedure of Rockova & George (2014) targeting modal estimates of the matrix of regression coefficients and residual precision matrix. Varying the scale of the continuous spike densities facilitates dynamic posterior exploration and allows us to filter out negligible regression coefficients and partial covariances gradually. Our method is seen to substantially outperform regularization competitors on simulated data. We demonstrate our method with a re-examination of data from a recent observational study of the effect of playing high school football on several later-life cognitions, psychological, and socio-economic outcomes.

Christopher Hans (Ohio State University, USA). **Strategies for Differential Shrinkage in Regression with Non-Orthogonal Designs** (joint work with Steven MacEachern and former PhD student Agniva Som)

Thick-tailed mixtures of g priors that mix over a single, common scale parameter have gained traction as a default prior in Bayesian regression settings. Such priors shrink all regression coefficients in the same manner and can negatively impact inference and model comparison in situations where differential shrinkage across regression coefficients is appropriate. In this talk I will review two known deficiencies of existing mixtures of g priors---"Essentially Least Squares (ELS) estimation" and a "Conditional Lindley's Paradox (CLP)"---that arise under a "conditional information asymptotic" regime that is motivated by the common data analytic setting where one regression coefficient is expected to be much larger than the others. The driver behind these undesirable behaviors is the use of a single, latent scale parameter that is common to all coefficients. Classes of block hyper-g priors that employ differential shrinkage across groups of coefficients have been proposed to avoid these behaviors, however the theory underlying these priors requires that the regression design matrix has a block orthogonal structure. This talk describes extensions to the theory underlying ELS and the CLP that are relevant for general, non-orthogonal designs, and introduces new prior distributions for imposing differential shrinkage in this setting. The priors rely on identifying blocks of related predictors that can be prioritized in terms of their relationship with the response. We discuss strategies for analysis that are robust to these modeling choices.

Naveen Narisetti (Urbana-Champaign). ***Scalable Bayesian Variable Selection for High Dimensional Data***

We consider the computational and statistical issues for high dimensional Bayesian model selection under the continuous spike and slab priors. To avoid large matrix computations needed in standard Gibbs sampling algorithms, we propose a novel Gibbs sampler called "Skinny Gibbs" which is much more scalable to high dimensional problems, both in memory and in computational efficiency. In particular, its computational complexity grows only linearly in p, the number of predictors, while retaining the property of strong model selection consistency even when p is much greater than the sample size n. We demonstrate its performance in several settings including logistic regression and quantile regression.

Contributed session I: Stein's method in Computational Statistics.

Sebastian Vollmer (University of Warwick & The Alan Turing Institute, UK). **Measuring Sample Quality with Diffusions**

Standard Markov chain Monte Carlo diagnostics, like effective sample size, are ineffective for biased sampling procedures that sacrifice asymptotic correct- ness for computational speed. Recent work addresses this issue for a class of strongly log-concave target distributions by constructing a computable discrepancy measure based on Stein's method that provably determines convergence to the target. We generalize this approach to cover any target with a fast-coupling Ito diffusion by bounding the derivatives of Stein equation solutions in terms of Markov process coupling times [Gorham et al., 2016]. As example applications, we develop computable and convergence-determining diffusion Stein discrepancies for log-concave, heavy-tailed, and multimodal targets and use these quality measures to select the hyperparameters of biased samplers, compare random and deterministic quadrature rules, and quantify bias-variance tradeoffs in approximate Markov chain Monte Carlo. Our explicit multivariate Stein factor bounds may be of independent interest.

Christophe Ley (Ghent University). **A Probabilistic Measure of the Impact of the Prior in Bayesian statistic.**

A key question in Bayesian analysis is the effect of the prior on the posterior, and how this effect could be assessed. As more and more data are collected, will the posterior distributions derived with different priors be very similar? This question has a long history; see for example Diaconis and Freedman [1986 a,b]. It is well-known that, asymptotically, the effect of the prior wanes as the sample size tends to infinity. Here we are interested, at fixed sample size, in explicit bounds on some measure of the distributional distance between posteriors based on a given prior and the no-prior data-only based posterior, allowing to detect at fixed sample size the effect of the prior. To reach this goal, we have developed in Ley et al. [2017] new aspects of the celebrated Stein Method for asymptotic approximations [Stein, 1972].

Francois-Xavier Briol (Warwick & Imperial College London). **Stein Points**

An important task in computational statistics is to approximate a posterior distribution with an empirical measure supported on a set of representative points. This paper focuses on methods where the selection of points is essentially deterministic, with an emphasis on achieving accurate approximation when the number of points is small. To this end, we present Stein Points. The idea is to exploit either a greedy or a conditional gradient method to iteratively minimise a kernel Stein discrepancy between the empirical measure and the posterior distribution. Our empirical results demonstrate that Stein Points enable accurate approximation of the posterior at modest computational cost. In addition, theoretical results are provided to establish convergence of the method.

14:00-15:30 Parallel sessions

Invited session II: Computational Machine Learning.

Michalis Titsias (AUEB, Greece). **Auxiliary gradient-based sampling algorithms**

We introduce a new family of MCMC samplers that combine auxiliary variables, Gibbs sampling and Taylor expansions of the target density. Our approach permits the marginalisation over the auxiliary variables yielding marginal samplers, or the augmentation of the auxiliary variables, yielding auxiliary samplers. The well-known Metropolis-adjusted Langevin algorithm (MALA) and preconditioned Crank-Nicolson Langevin (pCNL) algorithm are shown to be special cases. We prove that marginal samplers are superior in terms of asymptotic variance and demonstrate cases where they are slower in computing time compared to auxiliary samplers. In the context of latent Gaussian models, we propose new auxiliary and marginal samplers whose implementation requires a single tuning parameter, which can be found automatically during the transient phase. Extensive experimentation shows that the increase in efficiency (measured as effective sample size per unit of computing time) relative to (optimised implementations of) pCNL, elliptical slice sampling and MALA ranges from 10-fold in binary classification problems to 25-fold in log-Gaussian Cox processes to 100-fold in Gaussian process regression, and it is on par with Riemann manifold Hamiltonian Monte Carlo in an example where the latter has the same complexity as the aforementioned algorithms. We explain this remarkable improvement in terms of the way alternative samplers try to approximate the eigenvalues of the target. We introduce a novel MCMC sampling scheme for hyperparameter learning that builds upon the auxiliary samplers. The MATLAB code for reproducing the experiments in the article is publicly available and a Supplement to this article contains additional experiments and implementation details.

Invited session III: Sequential Monte Carlo.

Kari Heine (Bath University, UK). **Parallelising particle filters with butterfly interactions**
(joint work with Nick Whiteley (University of Bristol) and Ali Taylan Cemgil (Bogazici University))

In modern computing systems an increase in the computational power is primarily obtained by increasing the number of parallel processing elements (PE) rather than by increasing the speed of an individual PE. While in many cases such parallel computing systems have enabled the completion of increasingly complex computational tasks, they can only do so if the task in question admits parallel computations. This talk focuses on an important class of algorithms lacking such inherent parallelism, namely the sequential Monte Carlo (SMC) methods, or particle filters. We consider some new parallel particle filter algorithms whose interaction diagram coincides with the butterfly diagram best known from the context of fast Fourier transform. We present some new convergence results and consider the potential of these algorithms based on some numerical experiments.

Kody Law (Oak Ridge National Laboratory, UK). **Some multilevel Monte Carlo methods**

This talk will concern new algorithms for solving inference problems involving complex models which must be approximated. An example is Bayesian inverse problems where a PDE forward model typically appears inside the likelihood. 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, and can be smaller than i.i.d. sampling from the target at the finest approximation level. Recently several strategies have been employed to successfully apply this method to the problem of computationally intensive inference. Some of the resulting algorithms will be presented, such as multilevel sequential Monte Carlo samplers.

Jimmy Olsson (Lund University, Sweden). **Numerically stable online estimation of variance in particle filters** (joint work with Randal Douc, TELECOM SudParis)

This talk discusses variance estimation in sequential Monte Carlo methods, alternatively termed particle filters. The variance estimator that we propose is a natural modification of that suggested by H.P. Chan and T.L. Lai [A general theory of particle filters in hidden Markov models and some applications. Ann. Statist., 41(6):2877--2904, 2013], which allows the variance to be estimated in a single run of the particle filter by tracing the genealogical history of the particles. However, due particle lineage degeneracy, the estimator of the mentioned work becomes numerically unstable as the number of sequential particle updates increases. Thus, by tracing only a part of the particles' genealogy rather than the full one, our estimator gains long-term numerical stability at the cost of a bias. The scope of the genealogical tracing is regulated by a lag, and under mild, easily checked model assumptions, we prove that the bias tends to zero geometrically fast as the lag increases. As confirmed by our numerical results, this allows the bias to be tightly controlled also for moderate particle sample sizes.

15:50-17:20 Parallel sessions

Contributed session III: Advances in Piecewise Deterministic Sampling Algorithms

Joris Bierkens (TU Delft). **Piecewise Deterministic Sampling Algorithms and Exact Subsampling.**

In recent years PDMPs have gained substantial ground as a tool for Monte Carlo: drawing samples from a probability distribution. In this talk I will provide an introduction to this methodology, focusing on practical aspects such as how to implement these sampling methods, how can we use these processes for subsampling (resulting in certain cases in sampling methods whose efficiency does not depend on the size of the data), and how can we sample on restricted domains.

George Deligiannidis (University of Oxford). **Exponential Ergodicity of the Bouncy Particle Sampler** (joint work with Alexandre Bouchard-Côté, Arnaud Doucet)

Non-reversible Markov chain Monte Carlo schemes based on piecewise deterministic Markov processes have been recently introduced in applied probability, automatic control, physics and statistics. Although these algorithms demonstrate experimentally good performance and are accordingly increasingly used in a wide range of applications, geometric ergodicity results for such schemes have only been established so far under very restrictive assumptions. We give here verifiable conditions on the target distribution under which the Bouncy Particle Sampler algorithm introduced in (Peters and de With, 2012) is geometrically ergodic. This holds whenever the target satisfies a curvature condition and has tails decaying at least as fast as an exponential and at most as fast as a Gaussian distribution. This allows us to provide a central limit theorem for the associated ergodic averages. When the target has tails thinner than a Gaussian distribution, we propose an original modification of this scheme that is geometrically ergodic. For thick-tailed target distributions, such as t-distributions, we extend the idea pioneered in (Johnson and Geyer, 2012) in a random walk Metropolis context. We apply a change of variable to obtain a transformed target satisfying the tail conditions for geometric ergodicity. By sampling the transformed target using the Bouncy Particle Sampler and mapping back the Markov process to the original parameterization, we obtain a geometrically ergodic algorithm.

Pierre-André Zitt (Université Paris-Est). **Irreducibility and Exponential Ergodicity of the Zig-Zag Process.**

The zigzag process is a Piecewise Deterministic Markov Process which can be used in a MCMC framework to sample from a given target distribution. We prove the convergence of this process to its target under very weak assumptions, and establish a central limit theorem for empirical averages under stronger assumptions on the decay of the target measure. We use the classical "Meyn-Tweedie" approach. The main difficulty turns out to be the proof that the process can indeed reach all the points in the space, even if we consider the minimal switching rates.

<u>Contributed session: Scalable Monte Carlo: Strategies For Sub-Sampling And Distributed Inference.</u>

Hongsheng Dai (Essex, UK). **Monte Carlo Fusion for distributed analysis**

Unifying distributed analyses and inferences on shared parameters from multiple sources, into a single coherent inference, is surprisingly challenging. Such a problem arises in many settings (for instance, expert elicitation, multi-view learning, distributed `big data' problems etc.). We will present a framework and methodology which avoids any form of approximation error in obtaining the unified inference. The presentation will focus on the key theoretical underpinnings of this new methodology and simple (direct) Monte Carlo interpretations of the theory. There is considerable scope to tailor the theory introduced in this paper to particular application settings (such as the big data setting), construct efficient parallelised schemes, understand the approximation and computational efficiencies of other such unification paradigms, and explore new theoretical and methodological directions.

Chris Nemeth (Lancaster, UK). **Stochastic gradient MCMC: computationally efficient Monte Carlo via data subsampling**

Abstract: It is well known that Markov chain Monte Carlo (MCMC) methods scale poorly with dataset size. A popular class of methods for solving this issue is stochastic gradient MCMC (SGMCMC). These methods use a noisy estimate of the gradient of the log-posterior based on only a subset of the full data. This reduces the per iteration computational cost of the algorithm, but despite this, the computational cost is still proportional to the full dataset size. In this talk, I'll give an overview of popular SGMCMC algorithms and show how by reducing the variance of the gradient estimate with control variates, we can establish results under log-concavity assumptions on the target distribution and show that the computational cost required for a given level of accuracy is now independent of the dataset size.

TUESDAY MARCH 27, 2018

10:40-12:10 Parallel sessions

<u>Invited session. Scalable inference for hierarchical models</u>

Art B. Owen (Stanford, USA). **Method of Moments for Large Crossed Linear Mixed Effects Models** *(joint work with Katelyn Gao, Intel Inc.)*

Statistics faces two large crises: data sets are growing, and findings are not replicating. We might hope for the first crisis to ameliorate the second but it does not always happen. One place it fails to happen is in data sets with a tangle of cross-cutting correlations.
One of the simplest such problems is the linear crossed random effects model. For instance, in online shopping there is a random effect for the customer and another for the item. Simply computing the likelihood takes O(N^3/2) time. A Gibbs sampler takes O(N^1/2) time to mix with O(N) work per iteration. The problem is not amenable to sample splitting as any held out data are correlated with the held in data. This is a stark contrast to the ease with which hierarchical models can be handled. We propose to move away from likelihood and Bayes, returning to the method of moments. The costs are O(N). There is a mild loss of statistical efficiency. In return there are no tuning parameters, or convergence diagnostics or parametric distributional assumptions. Our motivating example comes from Stitch Fix which models customer satisfaction using such a crossed random effects model.

We thank Brad Klingenberg of Stitch Fix for data and helpful discussions.

Giacomo Zanella (Bocconi, Italy). **Scalable inferences for crossed random effects models** (joint with Omiros Papaspiliopolus and Gareth Roberts)

We analyze the complexity of Gibbs samplers for inference in crossed random effect models used in modern analysis of variance. Our theory is based on a multi-grid decomposition that allows to derive analytic expressions for the convergence rates of the algorithms. We demonstrate that for certain designs the plain vanilla Gibbs sampler is not scalable, in the sense that its complexity is worse than proportional to the number of parameters and data. We thus propose a simple modification leading to a collapsed Gibbs sampler that is provably scalable. Although our theory requires some balancedness assumptions on the data designs, we demonstrate in simulated and real datasets that the rates it predicts match remarkably the correct rates in cases where the assumptions are violated. We also show that the collapsed Gibbs sampler, extended to sample further unknown hyperparameters, outperforms significantly alternative state of the art algorithms.

Tamara Broderick (MIT, USA). **Automated Scalable Bayesian Inference via Data Summarization**

The use of Bayesian methods in large-scale data settings is attractive because of the rich hierarchical relationships, uncertainty quantification, and prior specification these methods provide. Standard Bayesian inference algorithms are often computationally expensive, however, making their direct application to large datasets difficult or infeasible. Other standard algorithms sacrifice accuracy in the pursuit of scalability. We take a new approach. Namely, we leverage the insight that data often exhibits redundancies to instead obtain a weighted subset of the data (called a "coreset") that is much smaller than the original dataset. We can then use this small coreset in any number of existing posterior inference algorithms without modification. We provide theoretical guarantees on the size and approximation quality of the coreset. In particular, we show that our method provides geometric decay in posterior approximation error as a function of coreset size. We validate on both synthetic and real datasets, demonstrating that our method reduces posterior approximation error by orders of magnitude relative to uniform random subsampling.

<u>Contributed session: Adaptive MCMC and High Performance Computing.</u>

Cyril Chimisov (Warwick, UK). ***AirMCMC: Adapted Increasingly Rarely Markov Chain Monte Carlo samplers*** (joint work with Krys Latuszynski and Gareth Roberts)

Adaptive MCMC algorithms are designed to self tune for optimal sampling performance, and hence they change the transition kernel of the underlying Markov chain as the simulation progresses, and characteristics of the target distribution are being revealed. This strategy breaks Markovianity, and while hugely successful in applications, these algorithms are notoriously difficult to analyse theoretically. We introduce a class of Adapted Increasingly Rarely Markov Chain Monte Carlo (AirMCMC) algorithms where the underlying Markov kernel is allowed to be changed based on the whole available chain output but only at specific time points separated by an increasing number of iterations. The main motivation is the ease of analysis of such algorithms. Under assumption of either simultaneous or (weaker) local simultaneous geometric drift condition, or simultaneous polynomial drift we prove the Mean Squared Error (MSE) convergence, Strong and Weak Laws of Large Numbers (SLLN, WLLN), Central Limit Theorem (CLT), and discuss how our approach extends the existing results. We argue that many of the known Adaptive MCMC may be transformed into an Air version and provide an empirical evidence that the performance of the Air version stays virtually the same.

References
[1] C. Chimisov, K. Latuszynski, and G. Roberts. Air Markov Chain Monte Carlo, arXiv:1801.09309, 2018

Nick Tawn (Warwick, UK). ***Adapting Towards Optimality of the Parallel Tempering Algorithm***

When a target distribution exhibits multi-modality, it is well known that MCMC schemes relying on localised proposals struggle to escape their local mode, becoming trapped in a subset of the state space leading to a bias sample. The parallel tempering algorithm is a popular solution but suffers from the curse of dimensionality. This talk introduces a new prototype population parallel tempering approach that is motivated by a non-centred reparametrisation that in certain settings exhibits vastly improved inter-modal mixing by mitigating the effects of high-dimensionality. Accompanied by theoretically derived optimal scaling results there is quantification of the accelerated mixing speed and an analysis of when this is applicable and indeed worth the expense of this new approach.

Louis Aslett (Durham, UK). **Contemporaneous MCMC**

Some population-based methods are amenable to highly parallel computation, but these methods typically don't help with the choice of transition kernel for the underlying MCMC scheme. We present a framework which allows development of adaptive methods for population schemes in a way that avoids the usual containment condition and diminishing adaptation requirements, and which allows fully asynchronous implementation across multiple GPUs. These methods therefore achieve iterative performance which scales nearly linearly and incorporates adaptation. Joint work with Murray Pollock and Gareth Roberts.

14:00-15:30 Parallel sessions

<u>Invited session: Deterministic approximation methods and pseudo-likelihoods</u>

Nancy Reid (Toronto, Canada). **Approximate Likelihood functions**

Approximate likelihood functions are widely used for inference in high dimensional models. Approaches that are used include Laplace approximation of integrals, approximation through simulation, or simplification using mis-specified, simpler, models. An overview of these approaches will be attempted, with reference to recent literature in high-dimensional inference.

Helen Ogden (Southampton, UK) **Statistical scalability of approximate likelihood inference**

In cases where evaluating the likelihood function exactly is not feasible, an alternative is to find some numerical approximation to the likelihood, then to use this approximate likelihood in place of the true likelihood to do inference about the model parameters. Such approximations are typically designed to be computationally scalable, but the statistical properties of these methods are often not well understood: fitting the model may be fast, but is the resulting inference any good? I will describe results which ensure that the approximate likelihood inference retains good asymptotic statistical properties, and discuss the statistical scalability of inference with an approximate likelihood, in terms of how the cost of conducting statistically valid inference scales as the amount of data increases. I will demonstrate the implications of these results for a particular family of approximations to the likelihood used for inference on an Ising model.

Erlis Ruli (Padova, Italy). **Model selection for (approximate) Bayesian inference via estimating functions**

The theory of estimating functions (EF) provides well-established inferential tools for modelling data by relaxing the stringent assumption that the data are generated from a parametric model. Classical applications of the EF theory are semi-parametric longitudinal data modelling and robust data analysis. However, the literature on EF theory has been mostly focused on parameter estimation and inference, while the problem of model selection remains relatively unexplored. Using the Approximate Bayesian Computation (ABC) machinery with EF as summary statistics, we show how to approximate the Bayes factor, which can be used to co

Contributed session: Computational Bayes for Cancer Genomics Applications

15:50-17:20 Parallel sessions

Invited session: Intractable Likelihood.

Sinan Yildirim (Sabanci, Turkey). **Utilising inference in state-space models with multiple paths from conditional sequential Monte Carlo**

Our work concerns static parameter inference in state-space models when the inference is utilised with sequential Monte Carlo (SMC) algorithms. Retaining one path from the samples in the cSMC algorithm involved in Metropolis within particle Gibbs may seem to be wasteful, and a natural question is whether it is possible to make use of multiple, even all possible, trajectories and average out the corresponding acceptance ratios. We show that this is indeed possible via the use of asymmetric acceptance ratios. The proposed schemes improve performance at a cost that can be parallellised.

Jeremy Heng (Harvard, USA). **Controlled sequential Monte Carlo**

Sequential Monte Carlo (SMC) methods are a set of simulation-based techniques used to approximate high- dimensional probability distributions and their normalizing constants. They have found numerous applications in statistics as they can be applied to perform state estimation for state-space models and inference for complex static models. Like many Monte Carlo sampling schemes, they rely on proposal distributions which have a crucial impact on their performance. In this talk, I will introduce a class of controlled SMC algorithms where the proposal distributions are determined by approximating the solution of an associated optimal control problem using an iterative scheme. Connections to existing work and some theoretical results on our proposed methodology will be discussed. Significant gains over state-of-the-art methods at a fixed computational complexity will also be illustrated on a variety of applications.

Anthony Lee (Bristol, UK). **Unbiased approximations of products of expectations**

I will describe recent work with Simone Tiberi (Zurich) and Giacomo Zanella (Bocconi), on the unbiased approximation of a product of n expectations. Such products arise, e.g., as values of the likelihood function in latent variable models, and unbiased approximations can be used in a pseudo-marginal Markov chain to facilitate inference. A straightforward, standard approach consists of approximating each term using an independent average of M i.i.d. random variables and taking the product of these approximations. While simple, this typically requires M to be O(n) so that the total number of random variables required is N = Mn = O(n^2) in order to control the relative variance of the approximation. Using all N random variables to approximate each expectation is less wasteful when producing them is costly, but produces a biased approximation. We propose an alternative to these two approximations that uses most of the N samples to approximate each expectation in such a way that the estimate of the product of expectations is unbiased. We analyze the variance of this approximation and show that it can result in N = O(n) being sufficient for the relative variance to be controlled as n increases. In situations where the cost of simulations dominates overall computational time, and fixing the relative variance, the proposed approximation is almost n times faster than the standard approach to compute.

Invited session: Asymptotics of MCMC convergence rates.

Daniel Rudolf (Goettingen, Germany). **Approximation of geometrically ergodic Metropolis-Hastings algorithms**

By using perturbation theory for Markov chains, we derive explicit estimates of the bias of an approximate 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.

James Johndrow (Stanford, USA). **Approximate MCMC: Theory and Practice**

We give some general results on approximations of uniformly and geometrically ergodic Markov chains. The results are motivated by applications to high-dimensional regression and Gaussian process models. We apply the insights from our general theory to construct a highly accurate approximate MCMC algorithm for global-local shrinkage priors. The power of these techniques is demonstrated in an application of the horseshoe prior to analysis of a GWAS dataset with 2,267 subjects and 98,385 predictors.

Aaron Smith (Ottawa, Canada). **Some Progress and Questions on the Mixing of Hamiltonian Monte Carlo**

Hamiltonian Monte Carlo (HMC) is a popular version of Markov chain Monte Carlo (MCMC), and variants of the algorithm drive the popular Bayesian computation software STAN. HMC is often able to make "longer jumps" than competitor algorithms such as Metropolis-Hastings and MALA, and this is widely believed to contribute to faster mixing on ill-posed and high-dimensional target distributions. In this talk, I will discuss recent progress on proving rigorous mixing bounds to justify these beliefs, including discussion of two recent papers with Oren Mangoubi. The first estimates the efficiency of HMC on "nice" high-dimensional unimodal targets, showing that HMC has much better performance than its competitor algorithms in this regime. The second gives precise estimates of the efficiency of HMC on multimodal targets; these imply that the standard HMC algorithm often has worse performance than its competitors in this regime. Finally, I will discuss several questions left open by our work and other recent progress.

17:30-19:00 Parallel sessions

Contributed session: Recent Advances in Bayesian Computation for Characterizing Astrophysical Populations

Ruth Angus (Columbia, USA). **Planetary systems across time and space**

The diverse population of planetary systems observed today reflects the history of planet formation and evolution across our galaxy. To learn about the processes behind planet formation and evolution from the observed distribution, we need to know the dates at which these planetary systems were formed. However, the ages of stars and planets are notoriously difficult to infer. In this talk I present a new hierarchical model designed to dramatically improve both the precision and accuracy of stellar (and planetary) ages. This model uses Gaussian process regression to infer stellar ages from time-series data, which are combined with other age indicators in a self-calibrated hierarchical probabilistic model. This new model will reveal time-dependent trends in the planet population, helping to validate planet formation theories and revealing the processes that sculpt planetary systems over time.

Angie Wolfgang (Penn State, USA). ***Characterizing the Mass-Radius Relationship for Exoplanet Populations***

David Stenning (Imperial College London UK). **Using Bayesian Computing to Solve a Complex Problem in Astrophysics**

Computer models are becoming increasingly prevalent in a variety of scientific settings; these models pose challenges because the resulting likelihood function cannot be directly evaluated. For example, astrophysicists develop computer models that predict the photometric magnitudes of a star as a function of input parameters such as age and chemical composition. A goal is to use such models to derive the physical properties of globular clusters—gravitationally bound collections of up to millions of stars. Recent observations from the Hubble Space Telescope provide evidence that globular clusters host multiple stellar populations, with stars belonging to the same population sharing certain physical properties. We embed physics-based computer models into a statistical likelihood function that assumes a hierarchical structuring of the parameters in which physical properties are common to (i) the cluster as a whole, or to (ii) individual populations within a cluster, or are unique to (iii) individual stars. A Bayesian approach is adopted for model fitting, and we implement an adaptive MCMC scheme that greatly improves convergence relative to its precursor, non-adaptive MCMC algorithm. Our method constitutes an advance over standard practice, which involves fitting single computer models by comparing their predictions with one or more two-dimensional projections of the data.

Fast communications session.

Josephine Merhi Bleik (Sorbonne universités, Université de Technologie de Compiègne, Laboratoire de Mathématiques Appliquées de Compiègne, Département de Génie Informatique, Centre de Recherches de Royallieu, France). **Simultaneous Bayesian quantile regression**

Quantile regression has received increasing attention both from theoretical and empirical viewpoint. It is used to explore the relationship between quantiles of the response distribution and available covariates. Since a set of quantiles often provides more complete description of the response distribution than the mean, quantile regression offers practically an important alternative to the classical mean regression. We are interested in estimating several quantiles simultaneously in a regression context via the Bayesian approach. Assuming that the error term has an asymmetric Laplace distribution ALD and using a relation that links two distinct quantiles of the ALD distribution, we propose a simple fully Bayesian method that satisfies the property of non-crossing quantiles. To evaluate the performance of our method, we use Monte Carlo Markov Chain methods to simulate in the full conditional distributions since they do not admit a closed form representation.

Bône, Alexander (Inria Paris, Aramis project-team, France). **Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms**

We propose a method to learn a distribution of shape trajectories from longitudinal data, i.e. the collection of individual objects repeatedly observed at multiple time-points. The method allows to compute an average spatiotemporal trajectory of shape changes at the group level, and the individual variations of this trajectory both in terms of geometry and time dynamics. First, we formulate a non-linear mixed-effects statistical model as the combination of a generic statistical model for manifold-valued longitudinal data, a deformation model defining shape trajectories via the action of a finite-dimensional set of diffeomorphisms with a manifold structure, and an efficient numerical scheme to compute parallel transport on this manifold. Second, we introduce a MCMC-SAEM algorithm with a specific approach to shape sampling, an adaptive scheme for proposal variances, and a log-likelihood tempering strategy to estimate our model. Third, we validate our algorithm on 2D simulated data, and then estimate a scenario of alteration of the shape of the hippocampus 3D brain structure during the course of Alzheimer's disease. The method shows for instance that hippocampal atrophy progresses more quickly in female subjects, and occurs earlier in APOE4 mutation carriers. We finally illustrate the potential of our method for classifying pathological trajectories versus normal ageing.

Igor Koval (Inria Paris, Aramis project-team, France). **Statistical Learning of Spatiotemporal Patterns from Longitudinal Manifold-Valued Networks**

We introduce a non-linear mixed-effects model to learn spatiotemporal patterns on a network by considering longitudinal measures distributed on a fixed graph. The data come from repeated observations of subjects at different time points which take the form of measurement maps distributed on a graph such as an image or a mesh. They are assumed to belong to a Riemmanian manifold such that it is possible to define spatiotemporal trajectories in the space of measurements. The model learns a typical group-average trajectory characterizing the propagation of measurement changes across the graph nodes. The subject-specific trajectories are defined via spatial and temporal transformations of the group-average scenario, thus estimating the variability of spatiotemporal patterns within the group. To estimate population and individual model parameters, we adapted a stochastic version of the Expectation-Maximization algorithm, namely the MCMC-SAEM, in a high dimensional setting with an adaptive sampling scheme. The model, described in a generic form as a product of dependent parametric monotonous one-dimensional functions, is used to describe the propagation of cortical atrophy during the course of Alzheimer's Disease. Model parameters show the variability of this average pattern of atrophy in terms of trajectories across brain regions, age at disease onset and pace of propagation. We show that the personalization of this model yields accurate prediction of maps of cortical thickness in patients.

Ioannis Miliagkasy (Department of Computer Science Stanford University, US) and **Lester Mackey** (Microsoft Research, New England, US). **Improving Gibbs Sampler Scan Quality with DoGS**

The pairwise influence matrix of Dobrushin has long been used as an analytical tool to bound the rate of convergence of Gibbs sampling. In this work, we use Dobrushin influence as the basis of a practical tool to certify and efficiently improve the quality of a discrete Gibbs sampler. Our Dobrushin-optimized Gibbs samplers (DoGS) offer customized variable selection orders for a given sampling budget and variable subset of interest, explicit bounds on total variation distance to stationarity, and certifiable improvements over the standard systematic and uniform random scan Gibbs samplers. In our experiments with joint image segmentation and object recognition, Markov chain Monte Carlo maximum likelihood estimation, and Ising model inference, DoGS consistently deliver higher-quality inferences with significantly smaller sampling budgets than standard Gibbs samplers.

Matt Graham (National University of Singapore -work completed while at University of Edinburgh). **Asymptotically exact inference in differentiable generative models**

Abstract: Many generative models can be expressed as a differentiable function applied to input variables sampled from a known probability distribution. This framework includes both procedurally defined simulator models involving only differentiable operations such as based on numerical integration of ordinary and stochastic differential equation systems, and the generative component of learned parametric models currently popular in the machine learning literature such as variational autoencoders and generative adversarial networks. Though the distribution on the input variables to such models is known, often the distribution on the output variables is only implicitly defined. We present a method for performing efficient Markov chain Monte Carlo inference in such models when conditioning on observations of the model output. For some models this offers an asymptotically exact inference method where approximate Bayesian computation might otherwise be employed. We use the intuition that computing conditional expectations is equivalent to integrating over a density defined on the manifold corresponding to the set of inputs consistent with the observed outputs. This motivates the use of a constrained variant of Hamiltonian Monte Carlo which leverages the smooth geometry of the manifold to move

between inputs exactly consistent with observations. We validate the method by performing inference experiments in a diverse set of models.

Sönmez, Orhan (Bogazici University, Pompeu Fabra University) and **Cemgil, A. Taylan,** **(**Bogazici University, Turkey)

Employing Sequential Monte Carlo in Model-Based Reinforcement Learning

In contrast to the standard approaches to the reinforcement learning problem, we formulated the model-based version of the problem into an equivalent likelihood maximization problem which allows us to employ approximate probabilistic inference tools. Then, we proposed an expectation-maximization algorithm that utilizes a MonteCarlo sampling scheme in its expectation step. We also derived different sequential Monte Carlo algorithms such as sequential Monte Carlo samplers and sequentially interacting Markov chain Markov for this setting. Finally, we applied our algorithms to some of the benchmark reinforcement learning problems in the OpenAI framework.

Akihiko Nishimura, (UCLA, USA**) Discontinuous Hamiltonian Monte Carlo for sampling discrete parameters** (joint work with David Dunson and Jianfeng Lu)

Hamiltonian Monte Carlo (HMC) is a powerful sampling algorithm employed by several probabilistic programming languages. Its fully automatic implementations have made HMC a standard tool for applied Bayesian modeling. While its performance is often superior to alternatives under a wide range of models, one weakness of HMC is its inability to handle discrete parameters. In this article, we present discontinuous HMC, an extension that can efficiently explore discrete parameter spaces as well as continuous ones. The proposed algorithm is based on two key ideas: embedding of discrete parameters into a continuous space and simulation of Hamiltonian dynamics on a piecewise smooth density function. The latter idea has been explored under special cases in the literature, but the extensions introduced here are critical in turning the idea into a general and practical sampling algorithm. Discontinuous HMC is guaranteed to outperform a Metropolis-within-Gibbs algorithm as the two algorithms coincide under a specific (and sub-optimal) implementation of discontinuous HMC. It is additionally shown that the dynamics underlying discontinuous HMC have a remarkable similarity to a zig-zag process, a continuous-time Markov process behind a state-of-the-art non-reversible rejection-free sampler. We apply our algorithm to challenging posterior inference problems to demonstrate its wide applicability and superior performance.

WEDNESDAY, MARCH 28, 2018

10:40-12:10 Parallel sessions

Invited session: Applied Math meets Bayes Comp

Jonathan Weare (University of Chicago, USA). **Stratification for Markov Chain Monte Carlo Simulation**

I will discuss a Monte Carlo approach to computing statistical averages that is based on a decomposition of the target average of interest into subproblems that are each individually easier to solve and can be solved in parallel. It is a close relative of the classical stratified sampling approach that has long been a cornerstone of experimental design in statistics. The most basic version of the scheme computes averages with respect to a given density and is a generalization of the umbrella sampling method for the calculation of free energies. For this scheme we have developed error bounds that reveal that the existing understanding of umbrella sampling is incomplete and potentially misleading. We demonstrate that the improvement from umbrella sampling over direct simulation can be dramatic in certain regimes. Our bounds are motivated by new, more detailed perturbation bounds for stochastic matrices. Finally, I will show an application of umbrella sampling and a tempering variant of umbrella sampling in the estimation of cosmological constraints from supernova data.

Chris Oates (Newcastle, UK). **A Bayesian Conjugate-Gradient Method**

A fundamental task in numerical computation is the solution of large linear systems. The conjugate gradient method is an iterative method which offers rapid convergence to the solution, particularly when an effective preconditioner is employed. However, for more challenging systems a substantial error can be present even after many iterations have been performed. Such estimates are of little value unless further information can be provided about the magnitude of the error. In this talk I will describe a novel statistical model for this error set in a Bayesian framework. Our approach is a strict generalisation of the conjugate gradient method, which is recovered as the posterior mean for a particular choice of prior. The estimates obtained are analysed with Krylov subspace methods and a contraction result for the posterior is presented. The method is illustrated on a challenging problem in medical imaging.

Aretha Teckentrup (Edinburgh, UK). **Deep Gaussian process priors for Bayesian inverse problems**

Deep Gaussian processes have received a great deal of attention in the last couple of years, due to their ability to model very complex behaviour. In this talk, we present a general framework for constructing deep Gaussian processes, and provide a mathematical argument for why the depth of the processes is in most cases finite. We also present some numerical experiments, where deep Gaussian processes have been employed as prior distributions in Bayesian inverse problems. This is joint work with Matt Dunlop, Mark Girolami and Andrew Stuart.

Contributed session: Illustration of software for BayesComp: NIMBLE and Birch.

Perry de Valpine (Berkeley, USA): *NIMBLE*
Lawrence Murray (Upsalla, Sweden): *Birch*

13:30-15:00 Parallel sessions

Invited session X: Hamiltonian Monte Carlo

Jesus Sanz-Serna (Carlos III de Madrid). **Numerical integrators for the Hamiltonian Monte Carlo method**

Most of the computational effort in the Hamiltonian Monte Carlo is spent in numerically integrating the Hamiltonian dynamics, a task where Verlet/leapfrog algorithm is the method of choice. In the talk I will first analyze the reasons for the success of the Verlet algorithm and discuss to what extent there is room for improvement. I will then present some ideas to construct algorithms more efficient than Verlet. Some of these ideas have already been incorporated to Monte Carlo software for molecular dynamics sampling.

Alexandros Beskos (UCL, UK). **Geometric MCMC for infinite-dimensional inverse problems**

Bayesian inverse problems often involve sampling posterior distributions on infinite-dimensional function spaces. Traditional Markov chain Monte Carlo (MCMC) algorithms are characterized by deteriorating mixing times upon mesh-refinement, when the finite-dimensional approximations become more accurate. Such methods are typically forced to reduce step-sizes as the discretization gets finer, and thus are expensive as a function of dimension. Recently, a new class of MCMC methods with mesh-independent convergence times has emerged. However, few of them take into account the geometry of the posterior informed by the data. At the same time, recently developed geometric MCMC algorithms have been found to be powerful in exploring complicated distributions that deviate significantly from elliptic Gaussian laws, but are in general computationally intractable for models defined in infinite dimensions. In this work, we combine geometric methods on a finite-dimensional subspace with mesh-independent infinite-dimensional approaches. Our objective is to speed up MCMC mixing times, without significantly increasing the computational cost per step (for instance, in comparison with the vanilla preconditioned Crank–Nicolson (pCN) method). This is achieved by using ideas from geometric MCMC to probe the complex structure of an intrinsic finite-dimensional subspace where most data information concentrates, while retaining robust mixing times as the dimension grows by using pCN-like methods in the complementary subspace. The resulting algorithms are demonstrated in the context of three challenging inverse problems arising in subsurface flow, heat conduction and incompressible flow control. The algorithms exhibit up to two orders of magnitude improvement in sampling efficiency when compared with the pCN method.

Sam Livingstone (UCL, UK). **Kinetic energy choice in Hamiltonian/Hybrid Monte Carlo**

We consider how different choices of kinetic energy in Hamiltonian Monte Carlo affect algorithm performance. To this end, we introduce two quantities which can be easily evaluated, the composite gradient and the implicit noise. Results are established on integrator stability and geometric convergence, and we show that choices of kinetic energy that result in heavy-tailed momentum distributions can exhibit an undesirable negligible moves property, which we define. Two numerical studies illustrate our theoretical findings, showing that the standard choice which results in a Gaussian momentum distribution may not always be optimal.

<u>Invited session. Bayesian computing for Economics, Business and Finance</u>

Nalan Basturk (Maastricht, The Netherlands) ***Time-varying learning combinations of***

Bayesian dynamic models and equity momentum strategies (joint with A. Borowska, S. Grassi, L. Hoogerheide and H.K. van Dijk)

A novel dynamic asset-allocation approach is proposed where portfolios as well as portfolio strategies are updated at every decision period based on learning about their past performance. A general class of models is specified, combining a dynamic factor model and a vector autoregressive model while allowing for stochastic volatility. A Bayesian strategy combination is introduced where several portfolio strategies are combined using feedback mechanisms. The approach extends the mixture of expert's analysis by allowing the strategy weights to be interdependent, time-dependent and allowing for model incompleteness. The proposed estimation method extends the forecast combination literature by relying on an implied state space structure of the joint model for time series and strategy weights. Given the complexity of the resulting non-linear and non-Gaussian structure a robust and efficient particle filter is introduced, based on mixtures of Student's t distributions. Using US industry portfolio returns over almost a century of monthly data, our empirical results indicate that time-varying combinations of flexible time series models combined with carefully selected portfolio strategies outperform competing standard alternative ones in the financial literature in terms of mean returns and risk features. This demonstrates the usefulness of the proposed methodology from a risk management perspective.

Geert Mesters (Pompeu Fabra and Barcelona GSE) **Nonlinear Dynamic Factor Models with Interacting Level and Volatility**

Volatility is an important ingredient in economic and financial decision making and yet the interaction between the levels and volatilities of macroeconomic and financial variables is not well understood. We propose a class of nonlinear dynamic factor models that has factor structures for both levels and volatilities. Both sets of latent factors are modeled jointly in an unrestricted vector autoregressive model. We develop a computationally convenient approximate filtering method for the estimation of all factors. The algorithm relies on numerical integration and can be implemented by augmenting the Kalman filter with weighted least squares regressions. Some theoretical bounds and a simulation study show that the methodology is highly accurate when compared to feasible alternative methods. The model is applied in two empirical studies. First, we consider euro area government bond yields between 2008 and 2012 and show that the volatility factor became an economically significant predictor of the yield levels in several countries. Bond purchases by the European Central Bank reduced yields but not the dispersion of pricing errors. Second, the model is applied for forecasting the levels of U.S. macroeconomic variables. We show that the inclusion of interacting volatility factors improves out-of-sample forecasts.

Gregor Kastner (WU Vienna, Austria). **Sparse Bayesian Vector Autoregressions in Huge Dimensions** (joint with F. Huber)

We develop a Bayesian vector autoregressive (VAR) model with multivariate stochastic volatility that is capable of handling vast dimensional information sets. Three features are introduced to permit reliable estimation of the model. First, we assume that the reduced-form errors in the VAR feature a factor stochastic volatility structure, allowing for conditional equation-by-equation estimation. Second, we apply recently developed global-local shrinkage priors to the VAR coefficients to cure the curse of dimensionality. Third, we utilize recent innovations to efficiently sample from high-dimensional multivariate Gaussian distributions. This makes simulation-based fully Bayesian inference feasible when the dimensionality is large but the time series length is moderate. We demonstrate the merits of our approach in an extensive simulation study and apply the model to US macroeconomic data to evaluate its forecasting capabilities.

POSTER SESSION

1. Aritz Adin - Universidad de Navarra, Spain
2. Omer Deniz Akyıldız - Universidad Carlos III, Spain
3. Angelos Alexopoulos - MRC Biostatistics Unit, Cambridge, UK
4. Alejandra Avalos-Pacheco - University of Warwick, UK
5. Arturo Avelino - Harvard University, USA
6. Jack Baker - Lancaster University, UK
7. Cecilia Balocchi - Wharton School of Business, University of Pennsylvania, USA
8. Marco Banterle - Brunel University London, UK
9. Alessandro Barp - Imperial College, UK
10. Mark Bell - Reading University, UK
11. Alexandre Bone - Inria Paris, France
12. Agnieszka Borowska - VU Amsterdam / Tinbergen Institute
13. Francois-Xavier Briol - University of Warwick and Imperial College London, UK
14. Rafael Cabañas - University of Almería, Spain
15. Annalisa Cadonna - WU Vienna, Austria
16. Zhanglong Cao - Otago, New Zealand
17. Hongsheng Dai - University of Essex, UK
18. Nikos Demiris - AUEB, Greece
19. Sameer K. Deshpande - Wharton School of Business, University of Pennsylvania, USA
20. Richard Everitt - Reading University, UK
21. Axel Finke - National University of Singapore, Singapore
22. Sebastian Funk - London School of Hygiene and Tropical Medicine, UK
23. Andrew Golightly - Newcastle University, UK
24. Matt Graham - National University of Singapore, Singapore
25. Beniamino Hadj-Amar - University of Warwick, UK
26. Jeremy Heng - Harvard University, USA
27. Darjus Hosszejni - WU Vienna, Austria
28. Jonathan Huggins - MIT, USA
29. Thanh Huy Nguyen - Telecom Paristech, France
30. Sofia Maria Karadimitriou - University of Sheffield, UK
31. Luke Kelly - University of Oxford, UK
32. Dimitrios Kiagias - University of Sheffield, UK
33. Guillaume Kon Kam King - Collegio Carlo Alberto and Universita di Torino, Italy
34. Jere Koskela - Technische Universität Berlin, Germany
35. Igor Koval - Inria Paris, France
36. Divakar Kumar - University of Warwick, UK
37. Brieuc Lehmann - MRC Biostatistics Unit, Cambridge, UK
38. Wentao Li - Newcastle University, UK
39. Oren Mangoubi - EPFL, Switzerland
40. Dean Markwick - University College London, UK

41. Raira Marotta - Universidade General do Rio de Janeiro, Brazil
42. Kaspar Martens - University of Oxford, UK
43. Isaac Matthews - Newcastle University, UK
44. Felipe Medina - Reading University, UK
45. Kerrie Mengersen - KUT, Australia
46. Josephine Merhi Bleik - Université de Téchnologie de Compiègne, France
47. Marcin Mider - University of Warwick, UK
48. Romain Mismer - Paris VI, France
49. Ioannis Mitliagkas - Stanford University, USA
50. Vincent Moens - Institute of Neuroscience, Belgium
51. Shariq Mohammed - University of Conneticut, USA
52. Giulio Morina - University of Warwick, UK
53. Brady Neal – MILA, Canada
54. Akihiko Nishimura, UCLA, USA
55. Brown University, USA
56. Brayan Ortiz - University of Washington, USA
57. Constantinos Perrakis - German Center for Neurodegenerative Diseases, Germany
58. Emilia Pompe - Oxford University, UK
59. Francesc Pons - Imperial College, UK
60. Dennis Prangle - Newcastle University, UK
61. Lewis Rendell - University of Warwick, UK
62. Lorentz Richter - BTU Berlin, Germany
63. Tom Ryder - Newcastle University, UK
64. Darius Savory - University College London, UK
65. Sebastian Schmon - University of Oxford, UK
66. Tobias Schwedes - Imperial College, UK
67. Zhe Sha - Bristol University, UK
68. Stephane Shao - Harvard University, USA
69. Umut Simsekli - Telecom Paristech, France
70. Michael T. Smith - University of Sheffield, UK
71. Orhan Sonmez - Bogazici University, Turkey
72. Josh Speagle - Harvard University, USA
73. Timothée Stumpf-Fétizon- Telefónica, Spain
74. Massimiliano Tamborrino - Johannes Kepler University Linz, Austria
75. Bram Thijssen - Netherlands Cancer Institute, The Netherlands
76. Miquel Torrens – Universitat Pompeu Fabra, Spain
77. Irene Tubikanec - JKU Linz, Austria
78. Kathryn Turnbull - Lancaster University, UK
79. Paul Vanetti - University of Oxford, UK
80. Dootika Vats - University of Warwick, UK
81. Gonzalo Vicente - Universidad de Navarra, Spain
82. Callum Vyner - Lancaster University, UK
83. Changye Wu - Paris Dauphine, France
84. Alexandre Zenon - INCIA, Bourdeaux, France
85. Rob Zinkov - University of Oxford

Workshop Organizers:

- Nicolas Chopin (ENSAE)
- Sylvia Fruehwirth-Schnatter (Wirtschaftsuniversitaet Wien)
- Mark Girolami (Imperial College London)
- Jim Hobert (University of Florida)
- Krzysztof Łatuszyński (University of Warwick)
- Hedibert Lopes (Insper)
- Ioanna Manolopoulou (University College London)
- Omiros Papaspiliopoulos (Chair, ICREA, Universitat Pompeu Fabra & Barcelona GSE)
- David Rossell (Universitat Pompeu Fabra & Barcelona GSE)

Local Committee

- Robert Castelo (Universitat Pompeu Fabra)
- Vicenç Gómez (Universitat Pompeu Fabra)
- Theo Kypraios (University of Nottingham)
- Geert Mesters (Universitat Pompeu Fabra & Barcelona GSE)
- David Rossell (Universitat Pompeu Fabra & Barcelona GSE)
- Piotr Zwiernik (Universitat Pompeu Fabra & Barcelona GSE)