Tuesday 26th June

Appleton Tower

Theme: Bayesian Computation

Agrawal: Raj

Minimal I-MAP MCMC for Scalable Structure Discovery in Causal DAG Models

Learning a Bayesian network (BN) from data can be useful for decision-making or discovering causal relationships, but traditional methods can fail in modern applications, which often exhibit a larger number of observed variables than data points. The resulting uncertainty about the underlying network as well as the ability to incorporate prior information recommend a Bayesian approach to learning the BN, but the highly combinatorial structure of BNs poses a striking challenge for inference. The current state-of-the-art method, order MCMC, is faster than previous methods but prevents the use of many natural structural priors and still has running time exponential in the maximum indegree of the true directed acyclic graph (DAG) of the BN. We here propose an alternative posterior approximation based on the observation that, if we incorporate empirical conditional independence tests, we can focus on a high-probability DAG associated with each permutation. We show that our method allows the desired flexibility in prior specification, removes timing dependence on the maximum indegree, yields provably good posterior approximations, and achieves superior accuracy, scalability, and sampler mixing on several datasets.

Asmann: Christian

Bayesian Estimation for Reduced Rank Regression Models using Post-Processing

Bayesian estimation of reduced rank regression models requires careful consideration of the well known identification problem via specification of appropriate prior distributions. We illustrate that efficient handling of the identification problem is possible via post-processing minimizing an appropriately specified loss function, when identification is either achieved via restriction on the Stiefel manifold or the second order moments of the latent factor. The resulting Bayesian estimation approach is illustrated within an simulation study and an empirical application and suggest a higher accuracy compared to alternative approaches.

Baker: Jack

Stochastic Sampling from the Probability Simplex and Dirichlet Processes

Traditional Markov chain Monte Carlo algorithms tend to perform prohibitively slowly on large datasets. Patterson and Teh (2013) developed a sampler for simplex spaces which can be applied in a large data setting. They apply this method to latent Dirichlet allocation and achieve state of the art performance. We revisit this problem and develop an alternative method to sample from simplex spaces that has no discretisation error and achieves a large improvement in performance. We use this method to develop a truncation free slice sampler for the Dirichlet Process which can be applied in a big data setting; we show it has a substantial speed up over the Gibbs sampler. We demonstrate our methods on a range of interesting problems, including recently developed sparse networks.

Banterle: Marco

Sparse covariance modeling in high-dimentional SUR models

For multivariate regression models with general residual covariance structure, the conjugacy property of a standard Normal-Inverse-Wishart prior distribution is lost when we select a different subset of predictors for each response. This type of model is often referred to as Seemingly Unrelated Regression (SUR) in the literature, and a standard Gibbs-sampling is available to sample from posterior full-conditionals of the parameters.

When dealing with high dimensional data, the naive Gibbs procedure is unfeasible for computational reasons. A particular reparametrisation that re-writes the joint likelihood by chain-conditioning all the univariate responses can alleviate the problem, but if the number of responses is high the computations are still gated by the cost of inverting and sampling the elements of the residual covariance matrix and a sparse solution is thus desirable.

Under usual Gaussian Graphical Model assumptions, where the covariance matrix is Hyper-Inverse-Wishart, we show that thanks to the Bartlett decomposition we can not only derive the correct posterior full conditional of its reparametrised elements, but also draw a connection between them and the zero-constrained elements of the sparse precision matrix when the underlying graph is decomposable.

Bell: Mark

Particle Gibbs Split-Merge sampler for the clustering of nodes in networks

A network describes the structure of relationships between a set of nodes. Three models for representing networks are stochastic block models (SBM), latent space models (LSM) and exponential random graph models (ERGM). For an observed network, the clustering is usually unknown. Dirichlet processes can be used as a prior on clusterings, however sampling from the

posterior distribution under this model presents a computational challenge. For SBMs and LSMs, the posterior normalising constant is intractable. For ERGMs, the problem is 'doubly intractable', since the likelihood density also possesses an intractable normalising constant. Conjugate priors and integration enable collapsing of the nuisance parameters, allowing samples to be drawn from the posterior. The Gibbs sampler can be used for this, but algorithms that split and merge existing clusters in one step have demonstrated greater efficiency. One such algorithm is the Particle Gibbs Split-Merge sampler (PGSMs).This decomposes the clustering problem into smaller sub-problems which are more suitable for applying Particle Gibbs. We use PGSMs to allocate the nodes of observed networks. For networks represented as ERGMs, we approximate marginal likelihoods with pseudo-likelihoods. Performance of PGSMs is compared for different network models; SBM, LSM and ERGM, on synthetic and historical networks.

Candela: Rosa

Insights into distributed variational inference for Bayesian Deep Learning

In the direction of equipping Deep Neural Networks with the ability to accurately quantify uncertainty, which is of huge interest in risk-sensitive applications, there has been a surge of interest in Bayesian DNNs and their nonparametric counterparts, namely Deep Gaussian Processes (DGPs).

Bayesian inference for DNNs and DGPs, however, leads to various forms of computational and mathematical intractabilities. A way to overcome these issues is to approximate the posterior over model parameters using variational inference, which gives also the possibility to adopt mini-batch-based stochastic gradient descent (SGD), making it possible to scale up to large datasets.

In this context, we assess the performance of synchronous and asynchronous distributed optimization of the variational posterior of Bayesian DNNs and DGPs approximated using random feature expansions. In synchronous SGD, a centralized server (aka the Parameter Server) waits for all the updates that computing nodes (aka Workers) calculate on individual batches, whereas in asynchronous SGD the Parameter Server applies the updates as they arrive.

We carry out an extensive analysis of the design choices of the two distributed implementations of stochastic variational inference and give insights into the trade-offs between design choices, such as number of Workers, learning rates, and mini-batch size.

Cristina Oliveira da Fonseca: Thais

Bayesian cross-validation of geostatistical models The problem of validating models for georeferenced data is challenging, since the conclusions can vary significantly depending on the locations of the validation set. This work proposes the use of cross-validation techniques to assess the goodness of fit of spatial models in different regions of the spatial domain to account for

uncertainty in the choice of the validation sets. An obvious problem with the basic cross-validation scheme is that it is based on selecting only a few out of sample locations to validate the model, possibily making the conclusions sensitive to which partition of the data into training and validation cases is utilized. A possible solution to this issue would be to consider all possible configurations of data divided into training and validation observations. From a Bayesian point of view, this could be computationally demanding, as estimation of parameters usually requires Monte Carlo Markov Chain methods. To deal with this problem, we propose the use of estimated discrepancy functions considering all configurations of data partition in a computationally efficient manner based on sampling importance resampling. In particular, we consider uncertainty in the locations by assigning a prior distribution to them. We illustrate the advantages of our proposal with simulated examples of homogeneous and inhomogeneous spatial processes. Furthermore, the methods are illustrated with an application to a rainfall dataset.

Dutta: Ritabrata

Well-Tempered Hamiltonian Monte Carlo on Active-Space

When the gradient of the log-target distribution is available, Hamiltonian Monte Carlo (HMC) has been proved to be an efficient simulation algorithm. However, HMC performs poorly when the target is high-dimensional and it has multiple isolated modes. To alleviate these problems we propose to perform HMC on a locally and continuously tempered target distribution. This tempering is based on an efficient approach to simulate molecular dynamics in high-dimensional space, known as well-tempered meta-dynamics. The tempering we suggest is performed locally and only along the directions of the maximum changes in the target which we identify as the active space of the target. The active space is the span of the eigenfunctions corresponding to the dominant eigenvalues of the expected Hessian matrix of the log-target. To capture the state dependent non-linearity of the target, we iteratively estimate the active space from the most recent batch of samples obtained from the target. Finally, we suggest a re-weighting scheme based on path-sampling to provide importance weights for the samples drawn from the continuously-tempered distribution. We illustrate the performance of this scheme for target distributions with complex geometry and multiple modes on high-dimensional spaces in comparison with traditional HMC with No-U-Turn-Sampler.

Ehlers: Ricardo

Zero variance and Hamiltonian Monte Carlo methods in GARCH models

In this paper, we develop Bayesian Hamiltonian Monte Carlo methods for inference in asymmetric GARCH models under different distributions for the error term. We implemented Zero-variance and Hamiltonian Monte Carlo schemes for parameter estimation to try and reduce the standard errors of the estimates thus obtaing more efficient results at the price of a small extra computational cost. We include a simulation study to investigate robustness of our methods and compare with other

algorithms. We also investigate how recently proposed model comparison techniques and detection of influential observations can be useful tools for this class of models. Finally, we illustrate with real time series data.

Elvira: Victor

Massively Recycled Importance Sampling

Importance sampling (IS) methods are broadly used to approximate posterior distributions and their moments in the context of Bayesian inference. In its standard approach, samples are drawn from a single proposal distribution and weighted properly. However, since the performance depends on the mismatch between the targeted and the proposal distribution, two strategies are often used. First, in multiple importance sampling (MIS), several proposal densities are employed. Second, in adaptive IS (AIS), the proposal densities are iteratively adapted in order to improve their performance. In both MIS and AIS, many different weighting schemes are possible and, as a consequence, for the same set of samples, several valid estimators can be built. In this work, we propose to build many different IS estimators and then combine them. This is done by massively reusing the same set of samples and applying different sets of weights. Note that no extra simulations are needed. Moreover, since all the weighting schemes use the same target evaluations, limited extra computations are required. More specifically, only extra proposal evaluations are needed, which are usually cheaper than the target evaluations. We provide algorithms for the optimal linear combination in terms of variance for both MIS and AIS.

Everitt: Richard

Variable dimension SMC for model comparison and streaming inference for the coalescent

Whole genome sequencing has had a big impact in studying the evolutionary history of pathogens. At the core of many studies is the need to infer the clonal ancestry of a sample from sequence data. This is usually performed using a Bayesian approach, with a coalescent prior on ancestries, and using Markov chain Monte Carlo (MCMC) for inference. However, MCMC can be computationally expensive, and is inflexible in that it needs to be run from scratch whenever new data is received. This talk describes work on a sequential Monte Carlo approach to inference in models with coalescent priors. The proposed approach is also applicable to Bayesian model comparison. An application to mixtures of Gaussians is presented, in which a deterministic transformation is used to improve the efficiency of the SMC.

This is joint work with Richard Culliford, Felipe Medina Aguayo and Daniel Wilson and a draft is available at arxiv.org/abs/1612.06468

Felipe Florez Rivera: Andres

Adaptive significance level test using weighted likelihoods for comparison of proportions

Statistical hypothesis testing is a widely used mathematical tool in various fields of knowledge. Nevertheless, it is a controversial topic due to the misuse of canonical (fixed) values of significance level which frequently leads to unreasonable decisions or “statistical paradoxes”, such as Lindley’s and Bartlett’s “paradoxes”.

In this work, we apply the procedure based on adaptive significance levels proposed by Pericchi & Pereira (2016), to the problem of equality test for proportions on contingence tables. Weighted likelihood methodology is used for such procedure, instead of usual maximum likelihood values over the hypotheses. The new procedure overcomes difficulties with fixed significance levels and it is in accordance with the likelihood principle.

Garbuno Inigo: Alfredo

Emulator-based history matching

Scientific understanding of real world processes has improved over the years through computer simulations. Such simulators represent complex mathematical models and are often implemented as expensive computer codes. The validity of using a particular simulator to draw accurate conclusions, relies in the assumption that the computer code is correctly calibrated. This calibration is often pursued under extensive experimentation and it requires comparison with the real world process. However, in some applications the collection of data is so expensive that only a handful of experiments is feasible. History matching is a calibration technique that iteratively discards regions of the input parameter space by using an implausibility measure. As such, history matching overcomes the limited experimental process and is able to provide regions of parameter space that are likely to replicate the observed phenomena, if any. This work uses the most of a Gaussian process as an emulator for computationally expensive computer codes. This results in an implausibility function that completely takes into account the uncertainty of the learned output. Three active learning criteria are extended to the history matching framework to improve the surrogate model. These learning criteria generate an educated guess on where to focus limited computational resources in a complete probabilistic setting. Examples cover real and simulated data experiments. The applicability of the procedure to different dimensional settings is shown in a proposed test bed of problems.

Giordano: Ryan

Automatic Measures of Hyperparameter Sensitivity for MCMC

High-quality, general-purpose Markov chain Monte Carlo (MCMC) tools like Stan have democratized Bayesian inference, allowing non-experts to easily calculate approximate posteriors for complex models. As with all mainstream Bayesian analysis, these models require the specification of priors and the likelihood, and it is practically important to know whether the posterior is robust to reasonable alternative choices. Complex models can be non-robust in counterintuitive ways, particularly for non-experts, and there is consequently a strong need for a fast, easy-to-use, automated tool to detect potential posterior non-robustness. To help fill this need, we provide an R package that allows Stan users to automatically calculate the local sensitivity of posterior means to user-defined hyperparameters. Using our tool, samples from a single MCMC run can be used, with minimal additional computation, to automatically calculate derivatives of posterior expectations with respect Stan variables that are indicated by the user as hyperparameters of interest. Our tool uses Stan's automatic differentiation capacity to automate classical results from Bayesian statistics. We discuss our methodology, illustrate the use and limitations of the package, and demonstrate its practical importance on real models taken from the Stan examples.

Giovannelli: Jean-François

Inverse problem and textures images

The paper deals with segmentation of textured images based on poor observation, including blur, noise and missing pixels. The inference then requires segmentation jointly with deconvolution, denoising and inpainting. The images are patchworks of textures that belong to a set of~K classes. Each class is described by a stationary Gaussian random field with a parametric power spectral density and the labels for the classes are modelled by a Potts field driven by a granularity coefficient. The method relies on a hierarchical model and a Bayesian strategy to jointly estimate the labels, the~K textured images in addition to hypermarameters: noise and image levels as well as texture and Potts parameters. The capability to estimate the latter is an important feature of the contribution. The estimation functions are designed in an optimal manner as a risk minimizer that yields the marginal posterior maximizer for the labels and the posterior mean for the other unknowns. The estimates are computed based on a convergent algorithm from samples of the posterior. They are obtained through an advanced MCMC algorithm: Perturbation-Optimization and Fisher Metropolis-Hastings steps within a Gibbs loop. Various numerical evaluations provide encouraging results despite the strong difficulty of the problem.

Graham: Matt

Scaling optimal-transport based ensemble filtering to high-dimensional spatiotemporal models

State inference in high-dimensional nonlinear dynamical models is a challenging task of significant interest in geophysical applications such as numerical weather prediction. The ensemble transport particle filter (ETPF) is an approach for approximating the state distribution at each timestep with a particle ensemble. Particles are propagated forward in time under the model dynamics before an optimal transport (OT) map is computed to update the ensemble for consistency with the latest observations. Compared to competing ensemble Kalman filter (EnKF) methods, the ETPF can improve inference quality in models with nonlinear dynamics or observation processes. Localisation approaches which exploit the often near independence of state variables at distant points are typically key to scaling EnKF methods to large spatiotemporal models. Previously proposed localised ETPF methods require computing OT maps for every spatial grid point, which as grid resolution is increased quickly becomes computationally intractable. We propose a new scalable localisation scheme which allows a fixed number of OT maps to be computed independent of the grid resolution and smoothly interpolates this sparse set of maps across space. We perform experiments in several challenging nonlinear spatiotemporal models and show our scheme allows the ETPF to significantly outperform EnKF approaches in such settings.

Graham: Matt

Continuously tempered Hamiltonian Monte Carlo

Hamiltonian Monte Carlo (HMC) is a powerful Markov chain Monte Carlo (MCMC) method for performing approximate inference in complex probabilistic models of continuous variables. In common with many MCMC methods, however, the standard HMC approach performs poorly in distributions with multiple isolated modes. We present a method for augmenting the Hamiltonian system with an extra continuous temperature control variable which allows the dynamic to bridge between sampling a complex target distribution and a simpler unimodal base distribution. This augmentation both helps improve mixing in multimodal targets and allows the normalisation constant of the target distribution to be estimated. The method is simple to implement within existing HMC code, requiring only a standard leapfrog integrator. We demonstrate experimentally that the method is competitive with annealed importance sampling and simulating tempering methods at sampling from challenging multimodal distributions and estimating their normalising constants.

Gullikstad Hem: Ingeborg

Penalised complexity priors for multilevel models: An outlook

Penalised complexity (PC) priors are proper prior distributions constructed to penalise increased complexity component-wise in a statistical model (Simpson et al. 2017). One aim is to reduce overfitting. We are working on a generalisation of this approach towards multilevel models. Instead of putting priors separately on each variance parameter in the model, we propose to put a prior on the total model uncertainty and propagate this uncertainty down to the individual components. We show that this approach is related to putting a prior on R^2. As of today, this prior is developed for a random intercept model with two variance components, which leads to one prior on the total variance, and one on the weight indicating how much of the variance belongs to the group error. We present results from a simulation study comparing the new PC prior approach to commonly used prior distributions. An outlook on how this approach might be generalised will be given.

Hartmann: Marcelo

Laplace approximation and the natural gradient for Gaussian process regression with the heteroscedastic Student-t model Student-t model has been widely used as a useful tool for robustifying data analysis. This presentation considers the Laplace method to derive approximate inference for the Gaussian process (GP) regression in the location and scale parameters of the Student-t probabilistic model. This allows both mean and variance of the data to vary as a function of covariates. The challenge in the approximate inference for the GP regression with the Student-t probabilistic model, lies in the analytical intractability of the posterior distribution and the lack of concavity of the log-likelihood function. We present the natural gradient adaptation for the estimation process which primarily relies on the property that the Student-t model naturally has orthogonal parametrization with respect to the location and scale paramaters. Due to this particular property of the model, we also introduce an alternative Laplace approximation by using the Fisher information matrix in place of the Hessian matrix of the negative log-likelihood function.

According to experiments this alternative approximation provides very similar posterior approximations and predictive performance when compared to the traditional Laplace approximation. We also compare both of these Laplace approximations with the Monte Carlo Markov Chain (MCMC) method. Moreover, we compare our heteroscedastic Student-t model and the GP regression with the heteroscedastic Gaussian model. We also discuss how our approach can improve the inference algorithm in cases where the probabilistic model assumed for the data is not log-concave.

Hosszejni: Darjus

Efficient Bayesian Inference for the Stochastic Volatility Model with Leverage

The sampling efficiency of MCMC methods in Bayesian inference for stochastic volatility (SV) models is known to highly depend on the actual parameter values, and the effectiveness of samplers based on different parameterizations differs significantly. We derive novel samplers for the centered and the non-centered parameterizations of the practically highly relevant SV model with leverage, where the return process and innovations of the volatility process are allowed to correlate. Moreover, based on the idea of ancillarity-sufficiency interweaving, we combine the resulting samplers in the hope of achieving superior sampling efficiency, irrespectively of the baseline parameterization. The method is implemented using R and C++, with the help of Rcpp for easy interfacing between the two languages. Finally, we carry out an extensive comparison to already existing sampling methods for this model.

Huan: Xun

Choosing Embedding for Capturing Model Misspecification Using Global Sensitivity Analysis and Bayes Factor Computation

Applications of Bayesian model calibration often assume that data are consistently generated from the model---that is, the model is correct. In reality, of course, all models are approximations to the truth. It is thus important to quantify the uncertainty due to model misspecification, and to propagate it to model predictions. We focus on an approach that embeds a stochastic representation of model error, or discrepancy, in select model parameters of the misspecified model. This strategy seeks to capture the data-model discrepancy through the internal stochastic representation while maintaining the original model overall-structure, which is useful when one wishes to retain desirable properties (such as physical constraints) in the predictive quantities through the original, albeit imperfect, model. Instituting embedding in all model parameters would be computationally intractable, and often unnecessary in practice. We thus study the selection problem of which parameters to augment with stochastic terms, with the aim of finding the fewest embeddings that capture the discrepancy sufficiently well. Numerical comparisons are performed using model-based global sensitivity analysis, as well as data-driven model selection with Bayes factor computation.

Jacobi: Liana

Automated Sensitivity Analysis for Bayesian Inference via Markov Chain Monte Carlo

Bayesian inference relies heavily on numerical Markov chain Monte carlo (MCMC) methods for the estimation of intractable high-dimensional posterior distributions and requires specific inputs. In this

paper we develop a new general and efficient numerical approach to address important robustness concerns of MCMC analysis with respect to prior input assumptions, a major obstacle to wider acceptance of Bayesian inference, and MCMC algorithm performance (convergence) reflected in dependence to chain starting values. Current input robustness analysis relies heavily on a restrictive and computationally very costly bumping-type approaches based on rerunning the algorithm with a small set of different inputs as well as convergence and efficiency diagnostics based on the autcorrelation of the draws. We introduce a comprehensive input sensitivity analysis based on first order derivatives of MCMC output with respect to the hyper-parameters and starting values to analyse prior robustness and algorithm convergence and efficiency. The approach builds on recent developments in sensitivity analysis of high-dimensional numerical integrals for classical simulation methods using automatic numerical differentiation methods. We introduce a range of new robustness measures to enable researchers to routinely undertake a comprehensive sensitivity analysis of their MCMC results. The methods are implemented for a range of Gibbs samplers and illustrated using both simulated and real data examples.

Jagath Senarathne: SG

Bayesian sequential design for Copula models: a comparison of designs selected under different Copula models

Copula models provide flexible structures to derive the joint distribution of multivariate responses. However, they are rarely considered in the experimental design context, particularly in a Bayesian framework where model and parameter uncertainty are considered. Here, we explore a variety of such models which explain dependence structures in experiments where bivariate discrete and mixed responses are observed. A sequential Monte Carlo algorithm is adopted to reduce the computational effort required in deriving efficient sequential designs. Moreover, the performance of the total entropy utility function is evaluated under different Copula models, which allows us to derive designs for the dual objectives of parameter estimation and model discrimination for Copula models. We illustrate our findings by considering an example with bivariate binary data and an application in pharmacology with mixed outcomes. These empirical findings provide a new understanding of parameter estimation and model discrimination for Copula models in the experimental design context.

Jo: Seongil

Stochastic variational inference for the Levy adaptive regression with multiple kernels

This paper presents a variational approach to a L{\'e}vy adaptive regression kernel (LARK) model that represents functions with an overcomplete system. In particular, we develop a stochastic variational inference method for a LARK model with multiple kernels (LARMuK) which estimates arbitrary functions that could have jump discontinuities. The algorithm is based on a variational EM method with probabilistic procedure. We compare the proposed algorithm to simulation-based Markov

chain Monte Carlo (MCMC) method using numerical experiments and discuss its potential and limitations.

Jones: David

Warp Bridge Sampling: The Next Generation

Bridge sampling (Meng and Wong, 1996) is an efficient Monte Carlo approach for estimating the normalizing constants that regularly appear in Bayesian statistics and science (indeed bridge sampling has a maximum likelihood interpretation, see Nicolae, Meng, and Kong, 2008). However, like related Monte Carlo methods, the precision of bridge sampling estimators is limited by the amount of overlap between the sampling and unnormalized densities in question. Warp transformations that increase the overlap between densities (without altering their normalizing constants) were proposed by Meng and Schilling (2002), but are focused on unimodal densities. In scientific applications, densities are often multi-modal and therefore it has so far been difficult to exploit the full power of bridge sampling. To overcome this challenge, in this presentation we introduce a stochastic transformation, the Warp-U transformation, that transforms multi-modal densities into unimodal ones. Constructing a Warp-U transformation begins with approximating the density of interest p with a Normal mixture phi_mix, but the stochastic nature of the Warp-U transformation means it goes beyond this simple setup. Indeed, the final overlap achieved after the Warp-U transformation is often substantially larger than that between p and phi_mix, and we have theoretical guarantees that it will never be smaller. We compare our approach to some competing methods and illustrate with an application to exoplanet hunting.

Keller: Merlin

Using Mixture Modeling to Perform Bayesian Model Selection and Averaging

A new method for Bayes factor computation and Bayesian model averaging is introduced, following the spirit of the mixture model approach for Bayesian model selection in the seminal work of Kamary, 2014. Specifically, we show that classical Bayesian model averaging corresponds to the special case of a single-datum mixture.

This result allows to redefine classical Bayesian model selection & averaging in such a way that improper priors can be chosen for parameters shared by all candidate models.

Based on this mixture-model formulation, we propose a generic algorithm which yields: model-specific as well as Bayesian model averaged posterior samples, together with estimated posterior probabilities for each candidate model.

We illustrate the advantages of the proposed approach on synthetic and real-life datasets.

Kelly: Luke

Accelerated inference in a complex phylogenetic model

Lateral transfer, a process whereby species acquire evolutionary traits outside of ancestral relationships, is a frequent source of model misspecification in phylogenetic inference. We describe an explicit generative model lateral transfer. As we only observe data at the leaves of the tree, we must integrate out the unobserved trait histories under our model when inferring the overall phylogeny. The computational cost of this massive integration over the latent histories is exponential in the number of taxa, in contrast to the linear cost for a model which ignores lateral transfer. We first describe how to efficiently construct a convergent sequence of approximations to the likelihood parameters, which we then accelerate using techniques from numerical analysis. The accelerated parameter estimates are sufficiently accurate that we can use them to construct an unbiased likelihood estimator and run a pseudo-marginal sampler to perform inference. The parameter effective sample sizes per unit time from our accelerated inference scheme are an order of magnitude higher compared to the corresponding Metropolis--Hastings algorithm using a brute-force method to compute the likelihood parameters.

Klein: Nadja

Implicit Copulas from Bayesian Regularized Regression Smoothers We show how to extract the implicit copula of a response vector from a Bayesian regularized regression smoother with Gaussian disturbances. The copula can be used to compare smoothers that employ different shrinkage priors and function bases. We illustrate with three popular choices of shrinkage priors — a pairwise prior, the horseshoe prior and a g prior augmented with a point mass as employed for Bayesian variable selection — and both univariate and multivariate function bases.

To evaluate the implicit copula we first construct a Gaussian copula by conditioning on the regularization parameters, and then mix over them using numerical or Monte Carlo methods. This greatly simplifies computation of the implicit copula compared to direct evaluation. The copulas are combined with non-parametric margins to extend the regularized smoothers to non-Gaussian data. Efficient Markov chain Monte Carlo schemes for evaluating the copula are given for this case. Using both simulated and real data, we show how such copula smoothing models can improve the quality of resulting function estimates and predictive distributions.

Lehmann: Brieuc

Extending the Bayesian exponential random graph model

While exponential random graph models (ERGMs) were originally developed in the context of social networks, recently there has been significant interest in applying ERGMs to neuroimaging data in both structural and functional connectivity studies. As opposed to social network studies, neuroimaging datasets generally consist of multiple networks across several individuals. This presents an additional layer of complexity - studies to date have circumvented this issue either by combining multiple networks into a single network, or by fitting an ERGM to each network separately. Ideally, however, one would like to pool information across networks to better estimate the model parameters. Further, in contrast to binary social networks, brain connectivity structure is typically weighted. Motivated by these facts, we develop two extensions for the Bayesian ERGM. Firstly, we establish a novel hierarchical framework for ERGMs. Secondly, we introduce an adaptation of the Bayesian ERGM for weighted networks. In this presentation, we will address the challenge posed by the intractability of the likelihood, illustrating our results with both real and simulated data.

Leriou: Ilias

Bayesian Survival Analysis under Model Uncertainty using the Generalized F Distribution

In this work, we consider inference for survival data using parametric survival models based on the Generalized F distribution. The proposed approach is very flexible in the sense that it includes a wide variety of popular distributions frequently used in survival analysis. Estimation of the parameters of this distribution is possible using MCMC methodology. Finally, we will account for model uncertainty by implementing MCMC based Bayesian Model and Variable Selection techniques by considering different choices of priors that take into account both the generality of this distribution and the very nature of survival data (censoring). The proposed methodology will be presented using real and simulated datasets.

Low-Choy: Samantha

informBCT: an R package for Bayesian Classification Trees with informative priors and decision theory

A new R package, informBCT, implements Bayesian classification trees with non-informative or informative priors, and will be made available through the Comprehensive R Archive Network (CRAN). The package has many features for data analysis including: variable selection, informative

priors, a Bayesian decision-theoretic layer to evaluate posterior distributions of tree performance from posteriors of tree parameters, as well as plotting the best trees, and saving trees. Informative priors can be placed on: the size of the tree, variable selection and/or the splitting rule. Moreover, informBCT can handle multi-category predictor variables, a difficulty for other tree packages—such as BART, bartMachine and tgp (Bayesian treed Gaussian process models) which all code categorical variables as dummy variables. We illustrate use of informBCT for two moderate-sized problems: predicting acute lymphoblastic leukaemia with 50 records on 55,000 genes; and confirming absence of cryptosporidium bacteria with 1,131 absences out of 1,332 records on 5 variables.

Ludkin: Matthew

Explicit, non-reversible, contour-hugging MCMC

Both the Bouncy Particle Sampler (BPS) and the Discrete Bouncy Particle Sampler (DBPS) are non-reversible Markov chain Monte Carlo algorithms whose action can be visualised in terms of a particle moving with a fixed-magnitude velocity. Both algorithms include an occasional step where the particle `bounces' off a hyperplane which is tangent to the gradient of the target density, making the BPS rejection-free and allowing the DBPS to propose relatively large jumps whilst maintaining a high acceptance rate. Analogously to the concatenation of leapfrog steps in HMC, we describe an algorithm which omits the straight-line movement of the BPS and DBPS and, instead, at each iteration concatenates several discrete `bounces' to provide a proposal which is on almost the same target contour as the starting point, producing a a large proposed move with a high acceptance probability. Combined with a separate kernel designed for moving between contours, an explicit bouncing scheme which takes account of the local Hessian at each bounce point ensures that the proposal respects the local geometry of the target, and leads to an efficient, skew-reversible MCMC algorithm.

M Schmon: Sebastian

Large Sample Asymptotics of the Pseudo-Marginal Method

Abstract The pseudo-marginal algorithm is a variant of the Metropolis–Hastings algorithm which allows us to sample asymptotically from a probability distribution when it is only possible to estimate unbiasedly an unnormalized version of its density with respect to a suitable dominating measure. It has become very popular in Bayesian statistics as there are many models for which the likelihood is intractable but can be estimated unbiasedly. Practically one has to trade off the number of Monte Carlo samples used to obtain this estimator against the asymptotic variances of the pseudo-marginal averages. The lower the variance of the likelihood estimator is the lower the variances of these averages typically are but increasing the number of samples also increases the computational complexity. Recent works optimizing this trade-off make the seemingly unrealistic

assumption that the distribution of the additive noise introduced by the log-likelihood estimator is Gaussian of mean and variance independent of the parameter value at which it is evaluated. Under weak regularity conditions we show here that, as the number of data points T tends to infinity, a space-rescaled version of the pseudo-marginal chain converges weakly towards another pseudo-marginal chain for which this assumption indeed holds. We provide numerical results on how to optimally scale the normal random walk proposal and the noise variance of this limiting Markov chain in both low and high-dimensional scenarios. This complements and validates currently available results.

Märtens: Kaspar

Scalable probabilistic non-linear dimensionality reduction for mixed variable types

The widespread availability of heterogeneous data sources in numerous industries requires the ability to integrate mixed data types. However, while there are numerous analytical techniques available for modeling continuous \emph{or} discrete data, few are available specifically for the integration of both. We propose an extension of the Gaussian Process Latent Variable Model (GP-LVM) to mixed variable types. We make use of the variational inducing point framework in combination with black-box variational inference to provide a scalable implementation. This additionally handles missing data and accounts for potentially non-linear covariate effects. The latter is achieved by introducing a kernel defined on the input space where some of the dimensions are treated as fixed, but the rest as random variables. The latter therefore allows us to combine nonlinear regression and latent variable modelling within a single GP-driven framework. We demonstrate the utility of our methods for estimating disease progression indices in cancer and acute conditions.

Mohammadi: Reza

Bayesian Structure Learning of High-dimensional Graphical Models with Application to Brain Connectivity

We consider the problem of Bayesian structure learning in high-dimensional graphical models, motivated by brain connectivity applications. In graphical models, Bayesian frameworks provide a straightforward tool, explicitly incorporating underlying graph uncertainty. In principle, the Bayesian approaches are based on averaging the posterior distributions of the quantity of interest, weighted by their posterior graph probabilities. However, Bayesian inference has not been used in practice for high-dimensional graphical models, because computing the posterior graph probabilities is hard and the number of possible graph models is very large. We discuss the computational problems related to Bayesian structure learning and we offer several solutions to cope the high-dimensionality problems. We apply our method to high-dimensional fMRI data from brain connectivity studies to

show its empirical usefulness. In addition, we have implemented our method in the R package BDgraph which is available online.

Muchmore: Patrick

Random Riemann sums for likelihood-free Markov chain Monte Carlo with stochastically scaled Gaussian vectors

A wide variety of probabilistic phenomena can be encapsulated in the model of a Gaussian vector stochastically scaled by a positive random variable, and in this work we consider the problem of density estimation for such random vectors. Although straightforward to define conceptually, in all but a few cases the resultant random vectors posses density functions which are computationally intractable, if not completely unknown. We first derive a standard Monte Carlo (MC) estimator which reduces the problem to a combination of sampling from the positive real numbers and evaluating the multivariate Gaussian density. We also illustrate how significantly higher accuracy can be achieved, at roughly the same computational cost, by employing the theory of random Riemann sums (RRS). Both the MC and RRS approaches produce unbiased estimates of the likelihood. Therefore, by appealing to the recently introduced theory of pseudo-marginal Markov chain Monte Carlo, they enable exact inference despite an intractable, or even unknown, likelihood function.

Nemeth: Christopher

Pseudo-Extended Markov Chain Monte Carlo

Sampling from the posterior distribution using Markov chain Monte Carlo (MCMC) methods can require an exhaustive number of iterations to fully explore the correct posterior. This is often the case when the posterior of interest is multi-modal, as the MCMC sampler can become trapped in a local mode for a large number of iterations. In this paper, we introduce the pseudo-extended MCMC method as an approach for improving the mixing of the MCMC sampler in complex posterior distributions. The pseudo-extended method augments the state-space of the posterior using pseudo-samples as auxiliary variables, where on the extended space, the MCMC sampler is able to easily move between the well-separated modes of the posterior.

We apply the pseudo-extended method within a Hamiltonian Monte Carlo sampler, and show that by using the No U-turn algorithm (Hoffman and Gelman, 2014), our proposed sampler is completely tuning free. We compare the pseudo-extended method against well known tempered MCMC algorithms and show the advantages of the new sampler on a number of challenging examples from the statistics literature.

Niemi: Jarad

Fully Bayesian analysis of hierarchical count regression models applied to RNAseq

Heterosis, or hybrid vigor, is the enhancement of the phenotype of hybrid progeny relative to their inbred parents. To identify genes displaying a heterosis pattern in their expression, we construct a gene-specific overdispersed count regression model. Since there are ~40,000 genes and ~10 samples, we build a hierarchical model for the gene-specific parameters to provide a data-based borrowing of information across genes. To implement a fully Bayesian analysis, we construct a novel parallelized Markov chain Monte Carlo algorithm that efficiently utilizes the architecture of a graphical processing unit through embarrassingly parallel computations and parallel reductions. We demonstrate the utility of the method to identify gene expression heterosis through a variety of simulation studies and analyze an RNAseq maize dataset to identify genes with 6 different types of heterosis.

Nishimura: Akihiko

Bayesian variable selection for "large n and large p"

In a modern observational study based on healthcare databases, the number of observations is typically in the order of 10^5 and that of the predictors in the order of 10^4 ~ 10^5. Although the sample size is large, it is rarely large enough to estimate the regression coefficients without some sparsity assumptions. Variable selection plays an important role in such situations, and there is a rich literature on Bayesian shrinkage priors and their desirable theoretical properties. The progress in the computational methods for Bayesian variable selection, however, has largely been limited to the “p >> n” case. The posteriors are amenable to Gibbs sampling, but the need to repeatedly sample from high-dimensional Gaussian distributions creates a major computational bottleneck. In this talk, we present a novel algorithm to speed up this bottleneck based on the following observation: we can generate a random vector b such that the solution of a linear system $\Phi \beta = b$ has the desired Gaussian distribution. An accurate solution of the linear system can then be found by the conjugate gradient algorithm with only a small number of the matrix-vector multiplications, without ever explicitly inverting Φ. We apply our algorithm to analyze a data set from the OHDSI project — in which the design matrix is of size 72,489 by 22,175 — and demonstrate over three-fold speed-up in the posterior computation.

Oyebamiji: Oluwole

Bayesian parameter inference for a dynamic simulator of microbial communities

Individual-based (IB) modelling has been widely used for studying the emergence of complex interactions of bacterial biofilms and their environment. We describe the emulation and calibration of an expensive dynamic simulator of an IB model of microbial communities. We used a combination of dynamic linear models and a Gaussian process to estimate the model parameters of our dynamic emulators. The emulators incorporated a smoothly varying and nonstationary trend that is modelled as a deterministic function of explanatory variables while the Gaussian process is allowed to capture the remaining intrinsic local variations. We then applied this emulation strategy for parameter calibration of a newly developed Newcastle University Frontiers in Biology (NUFEB) model for simulation of microbial communities. The simulation-based sensitivity analysis identified carbon substrate, oxygen concentration and maximum specific growth rate for heterotrophic bacteria as the most critical variables for predictions. The calibration results also indicated a general reduction the uncertainty levels in maximum specific growth rate for heterotrophic and ammonia-oxidizing bacteria. The study has helped to identify the trade-off in using different types of models for microbial simulation. The approach illustrated here provides a tractable and computationally efficient technique for calibrating the parameters of an expensive computer model.

Pao-Yen Wu: Paul

Understanding changes in physiological state during triathlon cycling using hidden markov models

Advances in instrumentation in cycling provide an unprecedented opportunity to better understand how the physiological state of athlete's change during a triathlon cycling race. Variability in power output is linked to physiological state and fatigue, which is a latent state whose dynamics are currently poorly understood for individual athletes. One approach to understanding changes in physiological state is the Hidden Markov Model (HMM). The data is challenging as there are only eight races with four athletes on two courses. At the same time, there are thousands of instrument readings per race relating to power, the main response of interest, and covariates of velocity, location, elevation and cadence. Such challenging data is reflected in an inability to find a solution for three or more states when fitting this HMM using expectation maximisation. We develop a HMM with athlete random effects sampled using MCMC to estimate fatiguing processes for individual athletes. The approach allowed us to achieve model convergence and explore the relationship between physiological states and both power and variability in power output. The model enables us to analyse race performance and predict fatigue development and recovery, ultimately enabling the customisation and optimisation of race strategy and training.

Polymeropoulos: Alexios

Bayesian Variable Selection for Multinomial Regression Models with Mixtures of g-Priors

Mixtures of g-priors are well established in linear regression models by Liang et al. (2007) and generalized linear models by Bove and Held (2011) and Li and Clyde (2013) for variable selection. This approach enabled us to overcome the problem of specifying the dispersion parameter by imposing a hyper-prior on it. By this way we allow for our model to "learn" about the shrinkage from the data. In this work, we implement Bayesian variable selection methods based on g-priors and their mixtures in multinomial regression models. More precisely, we follow two approaches: (a) the traditional implementation by extending the approach of Bove and Held (2011) for multinomial models, and (b) an augmented implementation by considering a g-prior conditional on the latent structure. We will study and compare the two approaches. Furthermore, we will focus on handling sparsity issues appearing when the number of covariates and categories is large, the need of specifying different covariate selection across different pairwise logit structures, and the impact of the selection of the reference category on the final results. All proposed methods will be presented in simulated datasets.

Pompe: Emilia

Multicore adaptive MCMC for multimodal distributions

Poor performance of standard Monte Carlo techniques on target distributions with isolated modes is a well-described problem in the MCMC literature. Multimodal distributions play a significant role in applications, such as Bayesian linear regression. The posterior distribution for the parameter beta is often multimodal under high collinearity between explanatory variables and a shrinkage prior..

We propose a new Monte Carlo method, Multicore Adaptive MCMC for Multimodal Distributions. The idea of this technique is based on two algorithms running in parallel. The first one is responsible for finding the modes of the target distribution, whereas the second algorithm is an MCMC sampler, using the knowledge it has about the locations of the modes. The MCMC sampler relies on steps of two types: local ones, preserving the mode and jumps to a region associated with a different mode. The method we propose is adaptive, so it learns the optimal parameters while it runs.

Our work motivated introducing a new class of algorithms, Auxiliary Variable Adaptive MCMC. We prove some general ergodic results for the whole class before specialising to the case of Multicore Adaptive MCMC for Multimodal Distributions.

The performance of the algorithm is tested on a Bayesian linear regression example.

Prangle: Dennis

Black-box Variational Inference for Stochastic Differential Equations

Parameter inference for stochastic differential equations is challenging due to the presence of a latent diffusion process. Working with an Euler-Maruyama discretisation for the diffusion, we use variational inference to jointly learn the parameters and the diffusion paths. We use a standard mean-field variational approximation of the parameter posterior, and introduce a recurrent neural network to approximate the posterior for the diffusion paths conditional on the parameters. This neural network learns how to provide Gaussian state transitions which bridge between observations in a very similar way to the conditioned diffusion process. The resulting black-box inference method can be applied to any SDE system with light tuning requirements. We illustrate the method on a Lotka-Volterra system and an epidemic model, producing accurate parameter estimates in a few hours.

Rastelli: Riccardo

Noisy inference for latent position network models

Latent position models are nowadays widely used for the analysis of networks in a variety of research fields. In fact, these models possess a number of desirable theoretical properties, and are particularly easy to interpret. However, statistical methodologies that estimate latent position models generally require a computational cost which grows with the square of the number of nodes in the graph. This makes the analysis of large social networks impractical. In this paper, we propose a new method characterised by a linear computational complexity, which may be used to fit latent position models on networks with several tens of thousands of nodes. Our approach relies on an approximation of the model likelihood, where the amount of noise introduced can be arbitrarily reduced at the expense of computational efficiency. We show several theoretical results that study how the likelihood error propagates to the invariant distribution of the Markov chain Monte Carlo sampler. In particular, we illustrate that one can achieve a substantial reduction in computing time and still recover the unknown latent structure correctly. Finally, we propose applications of our method to simulated networks and to a large coauthorships network, demonstrating the usefulness of our approach.

Rigon: Tommaso

A note on quadratic approximations of logistic log-likelihoods

Quadratic approximations of logistic log-likelihoods are fundamental to ease estimation and inference for binary data. Although the expansions underlying Newton-Raphson and Fisher scoring methods have attracted much of the interest, there has been also an increasing focus on quadratic

bounds that uniformly minorize the logistic log-likelihood. A relevant contribution, within this class of approximations, relies on a convex duality argument to derive a tractable family of tangent quadratic expansions indexed by a location parameter. Although this approximation is still being successfully implemented to facilitate computation in several models, less attempts have been made to understand the formal reasons underlying its excellent performance. To address this gap we provide a novel connection between the above bound and a recent Polya-gamma data augmentation for logistic regression. This result places the computational methods associated with the aforementioned bound within more general frameworks, i.e. expectation-maximization for frequentist estimation and global mean-field variational Bayes for approximate Bayesian inference, characterized by several optimality properties.

Riva Palacio: Alan

Bayesian Nonparametric Survival Regression Models

We present a broad class of Bayesian non-parametric regression models for survival analysis data. The construction of the model relies in a vector of completely random measures which allows for dependence among its entries resulting on the possibility of the borrowing of information between different co-variates. As a particular case, the Cox regression model in a neutral to the right setting (Kim and Lee, 2003) can be recovered; however the more general case of non-proportional hazards can be included in the model as well. We provide an explicit characterization of the posterior distribution which is shown to fall in the same class of distributions given by the model. We outline numerical methods for the inference via posterior mean estimates with a special focus when using a compound random measure (Griffin and Leisen, 2017) for the vector of completely random measures.

Safta: Cosmin

Bayesian Estimation of Spatio-Temporal Parameters for Atmospheric Transport Models

We present a hierarchical Bayesian framework to estimate parameters of an atmospheric pollutant transport model based on point concentration measurements. Methane is a powerful greenhouse gas and understanding the relative importance of anthropogenic and biogenic sources is crucial for developing and improving emission models. Reconciling atmospheric measurements with inventory-based estimates for various emissions sectors remains a great challenge due to large discrepancies between predictions and observations. Current approaches for measuring regional emissions yield highly uncertain estimates because of the sparsity of measurement sites and the presence of multiple simultaneous sources.

The set of parameters considered here corresponds to boundary conditions gleaned from methane emission databases. Transport model parameters are also considered in the inference process. To

tackle the high-dimensionality that results from parameters distributed spatially and temporally, we adopt a low-dimensional representation of these parameters as spectral random fields. The stochastic components in these representations are further parameterized as polynomial chaos expansions.

Next, we adopt a probabilistic model for spatio-temporal correlations to account for noise and inaccuracies due to limited sample size. We sample the posterior distributions of these parameters using transitional Markov Chain Monte Carlo techniques and we present inference results pertaining to California air basins.

Shao: Stephane

Bayesian model comparison with the Hyvärinen score

The Bayes factor is a widely used criterion in Bayesian model comparison and can be regarded as a ratio of out-of-sample predictive scores under the logarithmic scoring rule. However, when some of the candidate models involve vague or improper priors on their parameters, the Bayes factor features an arbitrary multiplicative constant that hinders its interpretation. As an alternative, we consider model comparison using the Hyvärinen score of Dawid & Musio (2015). We provide a method to consistently estimate this score for parametric models, using sequential Monte Carlo (SMC). In particular, we show that it can be estimated for models with tractable likelihoods via SMC samplers, and for nonlinear non-Gaussian state-space models by using SMC^2. We prove the asymptotic consistency of this new model selection criterion under strong regularity assumptions. We illustrate the method on diffusion models for population dynamics and Lévy-driven stochastic volatility models.

Shestopaloff: Alexander

On efficient Markov chain Monte Carlo sampling in discrete spaces

Hamiltonian Monte Carlo (HMC) is a state-of-the-art Markov chain Monte Carlo (MCMC) approach to sample high-dimensional target distributions in continuous spaces. This method generates “informed” trajectories whose endpoints are used as proposals within a Metropolis-Hastings scheme. By “informed”, we mean that the paths evolve according to the Hamiltonian dynamics which exploit the local structure of the target (i.e. gradient information), allowing the chain to take large steps while not ending up in areas with negligible densities. Unfortunately, there is no direct way of generalising HMC to discrete spaces. To address this problem, various authors have proposed to transform the discrete sampling problem as sampling from an auxiliary continuous distribution which is either piecewise constant (Pakman et al., 2013, arXiv:1311.2166v2; Nishimura et al., 2017, arXiv:1705.08510v2) or multimodal (Zhang et al., 2012) and then use HMC on the resulting auxiliary target. We present an alternative approach which bypasses the introduction of such an auxiliary

continuous target, yet is able to generate "informed" trajectories of high target probability mass. We demonstrate its correctness and shows that it compares favourably to alternative methods on a variety of examples arising in Bayesian statistics.

Zhang, Y., Ghahramani, Z., Storkey, A.J. and Sutton, C.A. (2012).

Sidén: Per

Efficient covariance approximations for large sparse precision matrices

Abstract. The use of sparse precision (inverse covariance) matrices has become popular because they allow for efficient algorithms for joint inference in high-dimensional models. Many applications require the computation of certain elements of the covariance matrix, such as the marginal variances, which may be non-trivial to obtain when the dimension is large. This paper introduces a fast Rao-Blackwellized Monte Carlo sampling-based method for efficiently approximating selected elements of the covariance matrix. The variance and confidence bounds of the approximations can be precisely estimated without additional computational costs. Furthermore, a method that iterates over subdomains is introduced, and is shown to additionally reduce the approximation errors to practically negligible levels in an application on functional magnetic resonance imaging data. Both methods have low memory requirements, which is typically the bottleneck for competing direct methods.

Souris: Allyson

A new blocked Gibbs sampler for the truncated multivariate normal distribution

We propose a blocked Gibbs sampler to approximate simulations from a multivariate truncated normal distribution with linear constraints. Existing Gibbs samplers for the truncated multivariate normal distribution typically proceed by sampling the coordinates one-at-a-time, leading to slow mixing and convergence. Our main idea is to approximate the multivariate truncated normal distribution by another distribution, which is achieved by replacing the indicator function by a smooth sigmoidal approximation. We provide theoretical support to validate our approximation and develop a data-augmentation Gibbs sampler to sample from this approximate distribution, utilizing properties of the sigmoidal function. This sampler uses the Polya--Gamma sampler from Polson and Scott (2013) and a multivariate normal sampler from Bhattacharya, Chakraborty, and Malllick (2016), allowing efficient block updating of all parameters. The proposed sampler is particularly efficient when the ambient dimension far exceeds the number of constraints. Comparisons to existing Gibbs samplers are conducted and we apply the sampler to high-dimensional classification problems.

South: Leah

Advances in zero-variance control variates for SMC

Sequential Monte Carlo (SMC) is a powerful method for sampling from the posterior distribution of static Bayesian models and estimating the evidence for model choice. SMC applies reweighting, resampling and mutation steps to transition a population of particles through a sequence of distributions. Derivative-based proposals have been considered in the mutation step [Sim et al., 2012] as a way to improve efficiency when derivatives of the log power posteriors are available in closed form or can be unbiasedly estimated. We propose to take further advantage of these derivatives by applying zero-variance control variates (ZV-CVs) to lower the variance of posterior expectations and the standard SMC and path sampling estimates of the evidence. However, one issue with ZV-CVs is that the number of control variates grows rapidly with the number of so-called polynomials and the dimension of the model parameter. We use higher order polynomials to improve fit and consider regularisation and dimension-reduction techniques to reduce the variance of these estimators in higher dimensions. We demonstrate that our approach can lead to more precise estimates of the evidence and improved posterior expectations and we give some intuition on why the SMC and path sampling evidence estimators produce remarkably similar results.

Sutton: Matthew

Bayesian Variable Selection Regression Of Multivariate Responses For Group Data

We propose two multivariate extensions of the Bayesian group lasso for variable selection and estimation for data with high dimensional predictors and multi-dimensional response variables. The methods utilize spike and slab priors to yield solutions which are sparse at either a group level or both a group and individual feature level. The incorporation of group structure in a predictor matrix is a key factor in obtaining better estimators and identifying associations between multiple responses and predictors. The approach is suited to biological studies where the response is multivariate, and each predictor is embedded in some biological grouping structure such as gene pathways. We are currently adapting our method in a meta-analysis approach to incorporate the biological grouping structure for multiple studies to detect pleiotropy. The performance of the proposed approaches is compared to state-of-the-art variable selection strategies on simulated datasets.

Tait: Daniel

Multiplicative Latent Force Models Using Neumann Series Expansions

Bayesian modelling of dynamic systems must achieve a compromise between providing a complete mechanistic specification of the process while retaining the flexibility to handle those situations in

which data is sparse relative to the model complexity, or a full specification is hard to motivate. Latent force models achieve this dual aim by specifying a parsimonious linear evolution equation and an additive latent Gaussian process forcing term. The resulting model is equivalent to a GP regression model with mean and kernel function depending on the linear dynamics. In this work we extend the latent force framework to allow for multiplicative interactions between the latent GPs and the state variables. This enables the embedding of additional geometric structure in the model, at the expense of a tractable posterior distribution. To proceed with inference we introduce a sampling method derived from a truncated Neumann series expansion of the trajectories, and further provide a variational approximation to this method. We use real and simulated datasets to demonstrate the ability of our model to recover the geometric structure of the trajectories and the latent GPs. We also demonstrate the improved performance of our method compared to adaptive gradient matching methods for lower sample densities.

Tardella: Luca

Algorithms and diagnostics for the analysis of preference rankings with the extended Plackett-Luce model

Choice behavior and preferences typically involve numerous and subjective aspects that are difficult to be identified and quantified. For this reason, their exploration is frequently conducted through the collection of ordinal evidence in the form of ranking data. A ranking is an ordered sequence resulting from the comparative evaluation of a given set of items according to a specific criterion. Multistage ranking models, including the popular Plackett-Luce distribution (PL), rely on the assumption that the ranking process is performed sequentially, by assigning the positions from the top to the bottom one (forward order). A recent contribution to the ranking literature relaxed this assumption with the addition of the discrete reference order parameter, yielding the novel Extended Plackett-Luce model (EPL). Inference on the EPL and its generalization into a finite mixture framework was originally addressed from the frequentist perspective. In this work, we propose the Bayesian estimation of the EPL without and with order constraints on the reference order parameter.

We additionally propose a novel model diagnostic to assess the adequacy of the EPL parametric specification. The usefulness of the proposal is illustrated with applications to simulated and real datasets.

Joint work with Cristina Mollica

Touloupou: Panayiota

Scalable inference for epidemic models with individual level data

As individual level epidemiological and pathogen genetic data become available in ever increasing quantities, the task of analysing such data becomes more and more challenging. Inferences for this type of data are complicated by the fact that the data are usually incomplete, in the sense that the times of acquiring and clearing infection are not directly observed, making the evaluation of the model likelihood intractable. A solution to this problem can be given in the Bayesian framework with unobserved data being imputed within Markov chain Monte Carlo (MCMC) algorithms, at the cost of considerable extra computational effort.

In this work we describe a novel method for updating individual level infection states within MCMC algorithms that respects the dependence structure inherent within epidemic data. We apply our new methodology to an epidemic of Escherichia coli O157:H7 in feedlot cattle in which 8 competing strains were identified using genetic typing methods. We show that surprisingly little genetic data is needed to produce a probabilistic reconstruction of the epidemic trajectories, despite some possibility of misclassification in the genetic typing. We believe that this complex model, capturing the interactions between strains, would not have been able to be fitted using existing methodologies. Joint work with Simon Spencer and Bärbel Finkenstädt Rand.

As individual level epidemiological and pathogen genetic data become available in ever increasing quantities, the task of analysing such data becomes more and more challenging. Inferences for this type of data are complicated by the fact that the data are usually incomplete, in the sense that the times of acquiring and clearing infection are not directly observed, making the evaluation of the model likelihood intractable. A solution to this problem can be given in the Bayesian framework with unobserved data being imputed within Markov chain Monte Carlo (MCMC) algorithms, at the cost of considerable extra computational effort.

In this work we describe a novel method for updating individual level infection states within MCMC algorithms that respects the dependence structure inherent within epidemic data. We apply our new methodology to an epidemic of Escherichia coli O157:H7 in feedlot cattle in which 8 competing strains were identified using genetic typing methods. We show that surprisingly little genetic data is needed to produce a probabilistic reconstruction of the epidemic trajectories, despite some possibility of misclassification in the genetic typing. We believe that this complex model, capturing the interactions between strains, would not have been able to be fitted using existing methodologies. Joint work with Simon Spencer and Bärbel Finkenstädt Rand.

van der Merwe: AJ

A Bayesian-Frequentists Approach for detecting Outliers in a One-way Variance Components Model

The most common Bayesian approach for detecting outliers is to assume that outliers are observations which have been generated by contaminating models. An alternative idea was used by

Zellner (1975) and Chaloner (1994). They studied the properties of realized regression error terms. Posterior distributions for individual realized errors and for linear and quadratic combinations of them were derived. In this note the theory and results derived by Chaloner (1994) are extended. Since it is not clear to us what the frequentist properties of the Bayesian procedures of Chaloner and Zellner are (i.e. what the size of the type I error or the power of their tests are) a Bayesian-frequentist approach is used for detecting outliers in a one-way variance components model. For illustration purposes, the Sharples (1990) contaminated data are used. It is concluded that the Bayesian frequentist approach seems to be more conservative than Chaloner's method.

Vanetti: Paul

Piecewise Deterministic Markov Chain Monte Carlo

A novel class of non-reversible Markov chain Monte Carlo schemes relying on continuous-time piecewise deterministic Markov Processes has recently emerged. In these algorithms, the state of the Markov process evolves according to a deterministic dynamics which is modified using a Markov transition kernel at random event times. These methods enjoy remarkable features including the ability to update only a subset of the state components while other components implicitly keep evolving. However, several important problems

remain open. The deterministic dynamics used so far do not exploit the structure of the target. Moreover, exact simulation of the event times is feasible for an important yet restricted class of problems and, even when

it is, it is application specific. This limits the applicability of these methods and prevents the development of a generic software implementation. In this paper, we introduce novel MCMC methods addressing these limitations by bringing together piecewise deterministic Markov processes, Hamiltonian dynamics and slice

sampling. We propose novel continuous-time algorithms relying on exact Hamiltonian flows and novel discrete-time algorithms which can exploit complex dynamics such as approximate Hamiltonian dynamics arising from symplectic integrators. We demonstrate the performance of these schemes on a variety of

applications.

Vasco: Daniela

Rain-drop plots: Graphical diagnostics for visualizing morphology of multiple plausible tree models

Here we propose new graphical diagnostics for Bayesian Classification Trees, which will enable visualization of multiple trees from the posterior distribution.

Classification Trees are a popular method for classification, especially for the easy interpretation of a single model. However, in order to improve its predictive performance and make it more robust for application in different data sets many models are combined using techniques such as Bagging and Boosting. This improvement in predictive performance comes at a cost: losing interpretability of the model. However, Bayesian Classification Trees provide interpretability whilst allowing fine-tuning of prediction performance.

However, similar to data mining algorithms for Classification Trees and Random Forests, there are very few diagnostics available to consider multiple possible trees, and evaluate or compare all models, aside from considering predictive performance. We consider ways of making the morphology of single trees visible en masse, namely, the shape and location of splits relevant to variable selection and prediction.

This approach will be illustrated for a published case study, using R and will show why we have called them Rain-drop plots.

Wiqvist: Samuel

Accelerating delayed-acceptance Markov chain Monte Carlo algorithms

Delayed-acceptance (DA) Markov chain Monte Carlo (Christen and Fox 2005) is a general method that utilizes a surrogate model of the target distribution to accelerate the ordinary Metropolis-Hastings algorithm, while still targeting the correct distribution. We introduce an accelerated DA algorithm, which targets an approximation of the target distribution.

We consider the problem of parameters inference in a Bayesian setting where a surrogate model of the likelihood function is introduced in the delayed-acceptance scheme. We employ a Gaussian process as a surrogate model of the likelihood function, but other options are possible. In our accelerated DA algorithm the calculations in the ``second stage'' of the delayed-acceptance scheme are reordered in such as way that we can obtain a significant speed-up, when the evaluation of the likelihood function is computationally intensive.

Simulations studies, considering both a simple toy model and a more involved stochastic differential equation model for protein-folding data, show that we obtain a speed-up using our accelerated version, and that for the more complex application our algorithm is twice as fast as a standard DA MCMC. Inference results show that the standard delayed-acceptance algorithm and our accelerated and approximated version are similar.

Keywords: Markov chain Monte Carlo, delayed-acceptance, particle Markov chain Monte Carlo, Gaussian processes

Christen, J. and C. Fox. Markov chain Monte Carlo using an approximation. Journal of Computational

and Graphical statistics, 14(4):795–810, 2005.

Wu: Yuefeng

Convergence of Adaptive Rejection Metropolis Sampling with Hit and Run Random Direction

Hit and run adaptive rejection Metropolis sampling (HARARMS) was proposed in 2015 for sampling from non-log-concave multivariate distributions, which improves the adaptive rejection Metropolis sampling (ARMS) algorithm by incorporating the hit and run sampling. It is not rare that the ARMS is trapped away from some subspace with significant probability in the support of the multivariate distribution. While ARMS updates samples only in the directions that are parallel to dimensions, HARARMS updates samples in arbitrary directions determined by the hit and run algorithm, which makes it almost not possible to be trapped in any isolated subspaces. Despite the difficulty to show the convergence of adaptive scheme in general, ARMS preserves stationary distribution of the Gibbs sampler, which can be shown as an application of the auxiliary variables method. HARARMS replaces the Gibbs sampler used for ARMS in multi-dimension by hit and run scheme, which did not get enough attention in statistics area. Combining the convergence of hit and run and ARMS, our main result shows the convergence of HARARMS.

Yang: Jun

Complexity Results for MCMC Derived from Quantitative Bounds

This paper considers how to obtain MCMC quantitative convergence bounds which can be translated into tight complexity bounds in high-dimensional setting. We propose a modified drift-and-minorization approach, which establishes a generalized drift condition defined in a subset of the state space. The subset is called the ``large set'', and is chosen to rule out some ``bad'' states which have poor drift property when the dimension gets large. Using the ``large set'' together with a ``centered'' drift function, a quantitative bound can be obtained which can be translated into a tight complexity bound.

As a demonstration, we analyze a certain realistic Gibbs sampler algorithm and obtain a complexity upper bound for the mixing time, which shows that the number of iterations required for the Gibbs sampler to converge is constant. It is our hope that this modified drift-and-minorization approach can be employed in many other specific examples to obtain complexity bounds for high-dimensional Markov chains.

YE: LIFENG

MC-CAVI: A Monte Carlo Implementation of CAVI algorithm and its application in Metabolomics

The coordinate ascent variational inference (CAVI) is one of the most commonly used algorithms in variational inference (VI) for approximating difficult-to-compute probability densities, such as posterior distribution in Bayesian statistics. In CAVI, the variational distribution is often assumed to factorise over some partition of the latent variables. In order to update variational distribution of each partition, the expectation of the logarithm of the joint probability of the data and latent variables, taken over all variables not in the partition needs to be derived analytically, which can be challenging in complex models. We discuss a Monte Carlo CAVI (MCCAVI) algorithm that uses Markov chain Monte Carlo (MCMC) methods in the calculation of the expectation required in CAVI. The performance of MCCAVI algorithm is compared with the CAVI algorithm and MCMC methods through in simulations. We also show that under suitable regularity conditions, an MCCAVI algorithm will, with high probability, converge to a maximiser of the target distribution. We further illustrate the MCCAVI algorithm through an application in NMR spectroscopy.

Zhe Liu: Jeremiah

Posterior Inference Algorithms for Exponential-Family Bayesian Factor Models: How fast do they converge?

Posterior Inference Algorithms for Exponential-Family Bayesian Factor Models: How fast do they converge?

Bayesian latent factor model with outcomes from exponential family has shown promising performance in ecology and computational biology applications. Many algorithms are proposed for posterior inference of this model in the literature, but the relative merit of these algorithms, in terms of their convergence rates and goodness-of-fit, under different model specifications remains to be explored. This paper answers this question and provides practitioners a principled way to choose sampling algorithms when applying Bayesian latent factor model. We start with a review on existing popular MCMC and Variational Inference (VI) algorithms for Bayesian latent factor model, including Metropolis-Hasting, No-U-Turn Sampler, Mean-field VI and Normalizing-flow VI. We then study how to adapt a kernel-based metric, Kernel Stein's Discrepancy (KSD), to quantify the discrepancy between posterior samples generated and the target posterior distributions under the high-dimensional latent factor model. Using this metric, we conduct simulation studies to examine the convergence rates and goodness-of-fit of each algorithm, under different assumptions on the distribution of the outcome and the rank of the model matrix. Finally, we illustrate the usefulness our methods on two applications in air pollution source apportionment and microbial community ordination.