BayesComp 2023 abstract booklet

3rd biennial meeting of the ISBA Section on Bayesian Computation

Levi Summit, Finland

15–17 March, 2023 *(with satellite meetings March 12–14)*

Keynote Talks

Tamara Broderick

Massachusetts Institute of Technology, USA

Title: An Automatic Finite-Sample Robustness Check for Bayes and Beyond: Can Dropping a Little Data Change Conclusions?

Commonly researchers will run a statistical analysis on a data sample, with the goal of applying any conclusions to a new population. For instance, if economists conclude microcredit is effective at alleviating poverty based on observed data, policymakers might decide to distribute microcredit in other locations or future years.

Typically, the original data is not a perfect random sample of the population where policy is applied – but researchers might feel comfortable generalizing anyway so long as deviations from random sampling are small, and the corresponding impact on conclusions is small as well. Conversely, researchers might worry if a very small proportion of the data sample was instrumental to the original conclusion. So we propose a method to assess the sensitivity of statistical conclusions to the removal of a very small fraction of the data set. Manually checking all small data subsets is computationally infeasible, so we propose an approximation based on the classical influence function. Our method is automatically computable for MAP, variational Bayes, MLE, and other common estimators – and we discuss extensions to MCMC. We provide finite-sample error bounds on approximation performance and a low-cost exact lower bound on sensitivity. We find that sensitivity is driven by a signal-to-noise ratio in the inference problem, does not disappear asymptotically, is not decided by misspecification, and is not eliminated by taking a Bayesian approach. While some empirical applications are robust,

conclusions of several influential economics papers can be changed by removing (much) less than 1% of the data.

Anthony Lee

University of Bristol, UK

Title: How many steps are needed for random walk Metropolis? Explicit convergence bounds for Metropolis Markov chains

One of the simplest and enduringly popular general-purpose Monte Carlo Markov chains evolving on $\mathbb{R}^d$ is the random walk Metropolis (RWM) Markov chain. Despite its relative simplicity, explicit convergence bounds that scale suitably with dimension have proved elusive for decades. In recent years, progress has been made to show that for distributions with strongly convex and gradient-Lipschitz potentials there exists a specific proposal variance giving an explicit bound on the $\mathcal{L}^2$-mixing time. We refine these results and obtain explicit spectral gap and $\mathcal{L}^2$-mixing time bounds for RWM with arbitrary proposal variances in any dimension, demonstrating the robustness of the algorithm. We obtain the correct scaling with dimension of the spectral gap for sufficiently regular target distributions, and the mixing time bounds are of reasonable (not astronomical) order. Our positive results are quite generally applicable in principle. Essentially the same analysis can be performed for the preconditioned Crank–Nicolson Markov chain, obtaining dimension-independent bounds under suitable assumptions.

This is joint work with C. Andrieu, S. Power and A. Wang.

Veronika Rockova

The University of Chicago Booth School of Business, USA

Title: Adversarial Bayesian Simulation

In the absence of explicit or tractable likelihoods, Bayesians often resort to approximate Bayesian computation (ABC) for inference. Going beyond ABC, my talk surveys recent optimization approaches to simulation-based Bayesian inference in likelihood-free situations. In particular, I will focus on deep neural samplers based on generative adversarial networks (GANs) and on adversarial variational Bayes. Both ABC and GANs compare aspects of observed and fake data to simulate from posteriors and likelihoods, respectively. I will discuss the Bayesian GAN (B-GAN) sampler that directly targets the posterior by solving an adversarial optimization problem. B-GAN is driven by a deterministic mapping learned on the ABC reference by conditional GANs. Once the mapping has been trained, iid posterior samples are obtained by filtering noise at a negligible additional cost. My talk also mentions more traditional posterior sampling approaches (ABC and Metropolis-Hastings) based on classification.

Invited Sessions

Parallel I

10:30 — 12:00, Wednesday 15 March

State-space modelling and particle filtering

Location: Auditorium and Online - *Chair* : Nicolas Chopin

Speaker: Adrien Corenflos (Aalto University)

Title: Auxiliary MCMC and particle Gibbs samplers for parallelisable inference in latent dynamical systems

Abstract: In this talk, I present two classes of auxiliary MCMC samplers for state and parameter inference in Feynman-Kac (state-space) models. The first one builds on the work "Auxiliary gradient-based sampling algorithms" of Titsias and Papaspiliopoulos (2018, JRSSB). It consists in forming local linear Gaussian state-space model approximations to the Feynman-Kac model for use in a Hastings-within-Gibbs scheme via Kalman filtering/sampling. The second one uses particle Gibbs instead to sample from the auxiliary target distribution and can be seen as a generalisation of the method introduced in "Conditional sequential Monte Carlo in high dimensions" by Finke and Thiery (2023, Annals of Stats). I will also be discussing parallelisation of the two methods along the time dimension, by which we mean that the computational complexity of the method essentially scales as the logarithm of the number of time steps when using a GPU, at essentially no statistical cost. The presentation is based on the article "Auxiliary MCMC and particle Gibbs samplers for parallelisable inference in latent dynamical systems" available at https://arxiv.org/abs/2303.00301.

Speaker: Hai-Dang Dau (University of Oxford)

Title: On backward smoothing algorithms

Abstract: In the context of state-space models, skeleton-based smoothing algorithms rely on a backward sampling step which by default has a $O(N^2)$ complexity (where N is the number of particles). Existing improvements in the literature are unsatisfactory: a popular rejection sampling– based approach, as we shall show, might lead to badly behaved execution time; another rejection sampler with stopping lacks complexity analysis; yet another MCMC-inspired algorithm comes with no stability guarantee. We provide several results that close these gaps. In particular, we prove a novel non-asymptotic stability theorem, thus enabling smoothing with truly linear complexity and adequate theoretical justification. We propose a general framework which unites most skeleton-based smoothing algorithms in the literature and allows to simultaneously prove their convergence and stability, both in online and offline contexts. Furthermore, we derive, as a special case of that framework, a new coupling-based smoothing

algorithm applicable to models with intractable transition densities. We elaborate practical recommendations and confirm those with numerical experiments.

Speaker: Axel Finke (Loughborough University)

Title: Conditional sequential Monte Carlo in high dimensions

Abstract: The iterated conditional sequential Monte Carlo (i-CSMC) algorithm from Andrieu, Doucet and Holenstein (2010) is an MCMC approach for efficiently sampling from the joint posterior distribution of the T latent states in challenging time-series models, e.g. in non-linear or non-Gaussian state-space models. It is also the main ingredient in particle Gibbs samplers which infer unknown model parameters alongside the latent states. In this work, we first prove that the i-CSMC algorithm suffers from a curse of dimension in the dimension of the states, D: it breaks down unless the number of samples ("particles"), N, proposed by the algorithm grows exponentially with D. Then, we present a novel "local" version of the algorithm which proposes particles using Gaussian random-walk moves that are suitably scaled with D. We prove that this iterated random-walk conditional sequential Monte Carlo (i-RW-CSMC) algorithm avoids the curse of dimension: for arbitrary N, its acceptance rates and expected squared jumping distance converge to non-trivial limits as $D \to \infty$. If $T = N = 1$, our proposed algorithm reduces to a Metropolis–Hastings or Barker's algorithm with Gaussian random-walk moves and we recover the well known scaling limits for such algorithms.

Stein Discrepancies

Location: Room 1 - *Chair* : Jeremias Knoblauch

Speaker: Marina Riabiz (King's College London)

Title: Optimal Thinning of MCMC Output

Abstract: The use of heuristics to assess the convergence and compress the output of Markov chain Monte Carlo can be sub-optimal in terms of the empirical approximations that are produced. Typically a number of the initial states are attributed to "burn in" and removed, whilst the remainder of the chain is "thinned" if compression is also required. In this talk we consider the problem of retrospectively selecting a subset of states, of fixed cardinality, from the sample path such that the approximation provided by their empirical distribution is close to optimal. A novel method is proposed, based on greedy minimisation of a kernel Stein discrepancy, suitable for the case in which the target distribution has an intractable normalizing constant. Theoretical results guarantee consistency of the proposed approach and its effectiveness is demonstrated in the challenging context of parameter inference for ordinary differential equations. Software is available in the Stein Thinning package in Python, R and MATLAB.

Speaker: Chris. J. Oates (Newcastle University)

Title: Gradient-Free Kernel Stein Discrepancy

Abstract: Stein discrepancies have emerged as a powerful statistical tool, being applied to fundamental statistical problems including parameter inference, goodness-of-fit testing, and sampling. The canonical Stein discrepancies require the derivatives of a statistical model to be computed, and in return provide theoretical guarantees of convergence detection and control. However, for complex statistical models, the stable numerical computation of derivatives can require bespoke algorithmic development and render Stein discrepancies impractical. This talk focuses on posterior approximation using Stein discrepancies, and introduces a collection of non-canonical Stein discrepancies that are gradient free, meaning that derivatives of the statistical model are not required. Sufficient conditions for convergence detection and control are established, and applications to sampling and variational inference are presented.

Speaker: Heishiro Kanagawa (Newcastle University)

Title: Controlling Moments with Kernel Stein Discrepancies

Abstract: Quantifying the deviation of a probability distribution is challenging when the target distribution is defined by a density with an intractable normalizing constant. The kernel Stein discrepancy (KSD) was proposed to address this problem and has been applied to various tasks including diagnosing approximate MCMC samplers and goodness-of-fit testing for unnormalized statistical models. In this talk, I will discuss a convergence control property of the diffusion kernel Stein discrepancy (DKSD), an instance of the KSD proposed by Barp et al. (2019). This talk will present a result extending that of Gorham and Mackey (2017), which showed that the KSD controls the bounded-Lipschitz metric, to functions of polynomial growth. Specifically, the DKSD controls the integral probability metric defined by a class of pseudo-Lipschitz functions, a polynomial generalization of Lipschitz functions. I will also demonstrate practical sufficient conditions on the reproducing kernel for the stated property to hold. In particular, we will see that the DKSD detects non-convergence in moments with an appropriate kernel.

Measuring Quality of MCMC Samples

Location: Room 2 - *Chair* : Dootika Vats

Speaker: Hyebin Song (Pennsylvania State University)

Title: Efficient shape constrained inference with applications in autocovariance sequence estimation

Abstract: I will present a novel shape-constrained estimator of the autocovariance sequence resulting from a reversible Markov chain. A motivating application for studying this problem

is the estimation of the asymptotic variance in central limit theorems for Markov chains. Our approach is based on the key observation that the representability of the autocovariance sequence as a moment sequence imposes certain shape constraints, which we can exploit in the estimation procedure. I will discuss the theoretical properties of the proposed estimator and provide strong consistency guarantees for the proposed estimator. Finally, I will empirically demonstrate the effectiveness of our estimator in comparison with other current state-of-the-art methods for Markov chain Monte Carlo variance estimation, including batch means, spectral variance estimators, and the initial convex sequence estimator.

Speaker: Medha Agarwal (University of Washington)

Title: Globally-centered autocovariances in MCMC

Abstract: Autocovariances are a fundamental quantity of interest in Markov chain Monte Carlo (MCMC) simulations with autocorrelation function (ACF) plots being an integral visualization tool for performance assessment. Unfortunately, for slow- mixing Markov chains, the empirical autocovariance can highly underestimate the truth. We propose a globally-centered estimator of the autocovariance function (G-ACvF) that combines information from parallel Markov chains and exhibits significant theoretical and empirical improvements. I will discuss the impact of this improved estimator in three critical MCMC output analysis applications: (a) ACF plots, (b) estimates of the Monte Carlo asymptotic covariance matrix, and (c) estimates of the effective sample size. Further, I will discuss statistical properties like strong consistency, large-sample bias, and variance of our improved Monte Carlo asymptotic covariance estimator. This talk is based on joint work with Dootika Vats (Indian Institute of Technology Kanpur).

Speaker: James Flegal (University of California Riverside)

Title: Assessing and Visualizing Simultaneous Simulation Error

Abstract: Monte Carlo experiments produce samples in order to estimate features such as means and quantiles of a given distribution. However, simultaneous estimation of means and quantiles has received little attention. In this setting we establish a multivariate central limit theorem for any finite combination of sample means and quantiles under the assumption of a strongly mixing process, which includes the standard Monte Carlo and Markov chain Monte Carlo settings. We build on this to provide a fast algorithm for constructing hyperrectangular confidence regions having the desired simultaneous coverage probability and a convenient marginal interpretation. The methods are in- corporated into standard ways of visualizing the results of Monte Carlo experiments enabling the practitioner to more easily assess the reliability of the results. We demonstrate the utility of this approach in various Monte Carlo settings including simulation studies based on independent and identically distributed samples and Bayesian analyses using Markov chain Monte Carlo sampling.

What lies beneath - Some recent advances in Bayesian nonparametrics

Location: Room 3 and Online - *Chair* : Sirio Legramanti

Speaker: Jeff Miller (Harvard University)

Title: Truth-agnostic diagnostics for calibration under misspecification

Abstract: Under model misspecification, it is known that usual Bayes posteriors often do not properly quantify uncertainty about the true or pseudo-true parameters. However, this lack of correct calibration is always defined with respect to a somewhat arbitrary notion of what is "true", making it conceptually unappealing and difficult to check empirically. We introduce a notion of internally coherent uncertainty quantification that is completely agnostic to the truth. Specifically, we consider the probability that two confidence sets constructed from independent data sets have nonempty overlap, and we establish a lower bound on this overlap probability that holds for any valid confidence sets. We show that, under misspecification, credible sets from the usual Bayes posterior can strongly violate this bound, indicating that it is not internally coherent. Meanwhile, we find that bagged posteriors typically satisfy the bound, indicating that they provide improved reproducibility.

Speaker: Raffaele Argiento (Università degli Studi di Bergamo)

Title: Clustering grouped data via hierarchical mixture models

Abstract: In this work we propose a new approach to cluster grouped data by consider- ing a mixture model with a hierarchical structure. In particular, the weights of the mixing measures are obtained by normalization of independent finite point processes, and the dependence across groups is ensured by assuming that the random probability measures share the same support. Our model induces both a local (within groups) and a global (between groups) clusters. We characterize the law and study the properties of this group-dependent clustering structure by providing a closed form expression for the partially exchangeable partition probability function. Furthermore, we provide a posterior characterization of the vector of random mixing measures and design a conditional Gibbs sampler to perform full Bayesian inference. The performances of the proposed approach are illustrated via simulated data and a real application.

Speaker: Botond Szabo (Bocconi University)

Title: Contraction rates and uncertainty quantification for sparse variational approximations in Gaussian process regression

Abstract: We study the theoretical properties of a variational Bayes method in the Gaussian Process regression model. We consider the inducing variables method introduced by Titsias (2009b) and derive sufficient conditions for obtaining contraction rates for the corresponding variational Bayes (VB) posterior. As examples we show that for three particular covariance kernels (Matérn, squared exponential, random series prior) the VB approach can achieve optimal,

minimax contraction rates for a sufficiently large number of appropriately chosen inducing variables. We also derive guarantees for the frequentist coverage of the corresponding variational Bayes credible sets. The theoretical findings are demonstrated by numerical experiments.

Parallel II

15:00 — 16:30, Wednesday 15 March

Advances in twisted models for sequential Monte Carlo

Location: Auditorium and Online - *Chair* : Anthony Lee

Speaker: Adam Johansen (Warwick University)

Title: Twisted/Controlled Particle Filters for Online Inference

Abstract: The use of twisted proposals to enhance the performance of particle algorithms used in online contexts to approximate filtering and smoothing distributions is explored. A novel algorithm which allows the use of the type of twisting strategies which have proved powerful in offline settings for parameter inference in hidden Markov Models and more general static parameter estimation via sequential Monte Caro samplers for online inference is developed. Some proof-of-concept numerical results will be presented. This is joint work with Liwen Xue.

Speaker: Nicolas Chopin (ENSAE Paris)

Title: Nested cubing integration: how to get a O(N^{-5}) error when you compute your favourite integral

Abstract: We propose two novel unbiased estimators of the integral $\int_{[0,1]^s} f(u)du$ for a function f, which depend on a smoothness parameter r N. The first estimator integrates exactly the polynomials of degrees p < r and achieves the optimal error n^{−1/2−r/s} (where n is the number of evaluations of f) when f is r-times continuously differentiable. The second estimator is computationally cheaper but it is restricted to functions that vanish on the boundary of [0, 1]^s. The construction of the two estimators relies on a combination of cubic stratification and control variates based on numerical derivatives. We provide numerical evidence that they show good performance even for moderate values of n. This is joint work work with Mathieu Gerber.

Speaker: Joshua Bon (Queensland University of Technology)

Title: Monte Carlo twisting for particle filters

Abstract: We consider the problem of designing efficient particle filters for twisted Feynman–Kac models. Particle filters using twisted models can deliver low error approximations of statistical quantities and such twisting functions can be learnt iteratively. Practical implementations of these algorithms are complicated by the need to (i) sample from the twisted transition dynamics, and (ii) calculate the twisted potential functions. We expand the class of applicable models using rejection sampling for (i) and unbiased approximations for (ii) using a random weight particle filter. We characterise the average acceptance rates within the particle filter in order to control the computational cost, and analyse the asymptotic variance. Empirical results show the mean squared error of the normalising constant estimate in our method is smaller than a memory-equivalent particle filter but not a computation-equivalent filter. Both comparisons are improved when more efficient sampling is possible which we demonstrate on a stochastic volatility model.

Robust innovations in gradient-based MCMC

Location: Room 1 - *Chair* : Samuel Livingstone

Speaker: Mauro Camara Escudero (University of Bristol)

Title: Approximate Manifold Sampling

Abstract: Sampling from a probability density concentrated around a lower-dimensional submanifold is of importance in numerous applications arising in machine learning, statistics and statistical physics. This task is particularly challenging due to the extreme anisotropy and high-dimensionality of the problem, and the correlation between the variables. We propose a new family of bespoke algorithms to sample efficiently from these densities and show their computational superiority to general purpose and specialized samplers. In particular, we show that in some cases it is able to efficiently sample from densities that are several orders of magnitude tighter around the submanifold.

Speaker: Chris Sherlock (University of Lancaster)

Title: The apogee to apogee path sampler

Abstract: Amongst Markov chain Monte Carlo algorithms, Hamiltonian Monte Carlo (HMC) is often the algorithm of choice for complex, high-dimensional target distributions; however, its efficiency is notoriously sensitive to the choice of the integration-time tuning parameter, T. When integrating both forward and backward in time using the same leapfrog integration step as HMC, the set of local maxima in the potential along a path, or apogees, is the same whatever point (position and momentum) along the path is chosen to initialise the integration. We present the Apogee to Apogee Path Sampler (AAPS), which utilises this invariance to create a simple yet generic methodology for constructing a path, proposing a point from it and accepting or rejecting that proposal so as to target the intended distribution. We demonstrate

empirically that AAPS has a similar efficiency to HMC but is much more robust to the setting of its equivalent tuning parameter, the number of apogees that the path crosses.

Speaker: Lionel Riou-Durand (University of Warwick)

Title: Adaptive Tuning for Metropolis Adjusted Langevin Trajectories

Abstract: Hamiltonian Monte Carlo (HMC) is a widely used sampler for continuous probability distributions. In many cases, the underlying Hamiltonian dynamics exhibit a phenomenon of resonance which decreases the efficiency of the algorithm and makes it very sensitive to hyperparameter values. This issue can be tackled efficiently, either via the use of trajectory length randomization (RHMC) or via partial momentum refreshment. The second approach is connected to the kinetic Langevin diffusion, and has been mostly investigated through the use of Generalized HMC (GHMC). However, GHMC induces momentum flips upon rejections causing the sampler to backtrack and waste computational resources. In this work we focus on a recent algorithm bypassing this issue, named Metropolis Adjusted Langevin Trajectories (MALT). We build upon recent strategies for tuning the hyperparameters of RHMC which target a bound on the Effective Sample Size (ESS) and adapt it to MALT, thereby enabling the first user-friendly deployment of this algorithm. We construct a method to optimize a sharper bound on the ESS and reduce the estimator variance. Easily compatible with parallel implementation, the resultant Adaptive MALT algorithm is competitive in terms of ESS rate and hits useful tradeoffs in memory usage when compared to GHMC, RHMC and NUTS.

Bayesian computation to track the pandemic

Location: Room 2 - *Chair* : Theodore Kypraios

Speaker: Daniela DeAngelis (University of Cambridge, United Kingdom)

Title: The challenges of Nowcasting and Forecasting the SARS-CoV2 pandemic

Abstract: In England policy decisions during the SARS-COV-2 pandemic have relied on prompt scientific evidence on the state of the pandemic. Through participation in advisory groups to the government we have contributed to the real time pandemic assessment from the early phase to the current times, providing estimates of key epidemic quantities and short-to-medium-term projections of severe disease.

In this talk I describe the statistical transmission modelling framework we have used throughout and the model developments introduced at various stages of the pandemic to tackle the challenges posed by a changing epidemiology.

The discussion will focus on the computational efforts to ensure that model developments would not jeopardise the ability to provide relevant outputs to policy makers in real time.

Speaker: Alfonso Diz-Lois Palomares (Norwegian Institute of Public Health, Norway)

Title: Monitoring the COVID-19 pandemic in Norway with a stochastic compartmental model and an Individual Based Model (IBM): From Sequential Monte Carlo (SMC) to Bayesian History Matching.

Abstract: Stochastic compartmental metapopulation models have proven to be a critical tool for monitoring the Covid-19 pandemic. In this work, I will showcase two different models that have been instrumental in Norway for decision-making related to interventions and resource management in the health sector, while also adhering to tight time and computational-budget constraints.

The first model is a stochastic compartmental model that assumes a daily-varying reproduction number to quantify the viral transmission in addition to other latent variables. We proposed a state-space formalization of the model and employed a Sequential Monte Carlo approach for inference, based on daily hospitalisations and positive test incidences. This flexible setup allows for the easy integration of multiple data sources in the likelihood while also enabling for the run of updates.

The second model is an Individual Based Model (IBM) that provides a highly detailed representation of the entire Norwegian population, from households to schools and workplaces. Even though the application of IBMs has been usually limited to scenario modelling as they tend to be computationally demanding and heavily parametrized, its use has not been exempt from time and resources limitations. In particular, this model has been used to generate hypothetical scenarios and evaluations that have been crucial at certain stages of the pandemic (such as simulation of incoming variants or vaccination prioritization). In this regard, I will focus on how the application of Gaussian Process (GP) emulators within the History Matching framework enabled inference in useful time despite a highly dimensional parameter space. And how this might be paramount to boost and expand the usefulness of this type of models.

This is joint work of the Oslo covid-19 modelling team at the National Institute of Public Health, Norsk Regnesentral, Telenor and the University of Oslo.

Speaker: Christopher Jewell (Lancaster University, United Kingdom)

Title: Inference for spatial epidemic models of SARS-CoV-2 in the UK: linking Bayesian statistics and policy

Abstract: During outbreaks such as Covid19, accurate situational awareness is crucial for informing adaptive control policy and for basing reliable quantitative forecasts of ongoing disease spread. Detailed epidemic models which incorporate population heterogeneity, such as spatial relationship and individual-level covariates, provide an opportunity to do this. However, their parameters must be estimated in a rigorous and unbiased way by formally fitting the model to the past history of the epidemic. Doing this is difficult due to censored epidemiologically-relevant events, such as infection times or the onset of infectiousness. For medium-sized populations – such as a national population of farms – data-augmentation MCMC provides an inferential technology for unbiased parameter estimation. For much larger populations,

particle filtering promises much, but fails rapidly as model complexity increases. To address this issue, we develop a constrained Metropolis-Hastings proposal which honours the topology of the state-transition model to improve algorithmic performance. For the UK SARS-CoV-2 outbreak, a discrete-time metapopulation epidemic model is constructed in which spatial administrative units are coupled using human-mobility data taken from UK Census 2011 data. A MCMC algorithm is then proposed which respects the ordering of epidemiological events demanded by the underlying epidemic model. The model is written in Python using `gemlib`, a new computational library for epidemics based on TensorFlow and TensorFlow Probability for GPU performance, modularity, and reuse. Our work shows the potential of toplogically-aware MCMC methods for fitting epidemic models to large populations of millions of individuals. We also demonstrate the power of modern machine learning tooling to modularise and ultimately translate cutting edge statistical developments into applied epidemiological use rapidly in emergency situations.

Bayesian statistics for environmental data

Location: Room 3 and Online - *Chair* : Mari Myllymäki

Speaker: Jeffrey W. Doser (Michigan State University)

Title: Hierarchical Bayesian models for high-dimensional ecological data across large spatial domains

Abstract: Quantifying distributions of wildlife and plant species across space and time is a critical task for understanding global change effects and developing effective conservation and management actions. However, many modern ecological data sets present a variety of statistical and computational complexities when modeling ecological processes for hundreds of species across thousands of spatial locations. In this talk, I will discuss a flexible hierarchical Bayesian modeling framework designed to model geostatistical data in which both the number of locations and number of outcomes at each location is large. The proposed framework addresses these source of high dimensionality via spatial latent factors and highly scalable Nearest Neighbor Gaussian Process (NNGP) models. The approach accounts for autocorrelation within each response across spatial locations as well as among responses at any individual location. I will present two motivating case studies in which I use the proposed framework to predict forest biomass at the species-level across the continental US and quantify nonstationary effects of climate and land use change on a forest bird community in the eastern US.

Speaker: Jarno Vanhatalo (University of Helsinki)

Title: Joint Species Distribution Modeling of Percentage Cover Data with Exclusive Competition for Space

Abstract: Joint species distribution models (JSDM) are among the most important statistical tools in community ecology. They are routinely used for inference and various prediction tasks, such as to build species distribution maps or biomass estimation over spatial areas. Existing JSDM's cannot, however, model mutual exclusion between species, which may happen in some species groups, such as mosses in the bottom layer of a peatland site. We tackle this deficiency in the context of modeling plant percentage cover data, where mutual exclusion arises from limited growing space and competition for light. We propose a hierarchical JSDM where multivariate latent Gaussian variable model describes species' niche preferences and Dirichlet-Multinomial distribution models the observation process and exclusive competition for space between species. We use both stationary and non-stationary multivariate Gaussian processes to model residual phenomena. We also propose a decision theoretic model comparison and validation approach to assess the goodness of JSDMs in four different types of predictive tasks. We apply our models and methods to a case study on modeling vegetation cover in a boreal peatland. Our results show that ignoring the interspecific interactions and competition for space significantly reduces models' predictive performance and leads to biased estimates for total percentage cover both for individual species and over all species combined. A model's relative predictive performance also depends on the model comparison methods highlighting that model comparison and assessment should resemble the true predictive task. Our results also demonstrate that the proposed joint species distribution model can be used to simultaneously infer interspecific correlations in niche preference as well as mutual exclusive competition for space and through that provide novel insight into ecological research.

Speaker: Janine Illian (University of Glasgow)

Title: Accounting for observation processes in spatio-temporal ecological data - developing methodology that is relevant and usable

Abstract: These days more and more data are being collected, analysed and interpreted to inform decisions - as we have clearly seen, for example, in the recent Covid-19 pandemic. Here, statisticians have a responsibility to support users who are interpreting the results of a statistical analysis so that appropriate conclusions can be drawn from it. At the same time, increasingly complex analysis tools are being developed, and are now easily accessible to non-statisticians through a large number of different R packages. We as developers of complex statistical methods have a related responsibility to support the adequate use of the methods we develop - in particular when users are quantitatively trained, yet non-specialist, scientists.

When introducing these users to our methodology we need to strike the right balance between treating methodology as a mere black box and explaining every single technical detail, while providing an adequate understanding of the methodology that allows users to independently decide on appropriate model choices. This is needed to encourage the use of our methods as well as to establish a fruitful dialogue with the users to improve and successfully build on exciting methods.

In addition, when we develop statistical modelling approaches, it is important to ensure that these are relevant to the users and that they take into account the specific needs of the

user community. This involves exploring and engaging with the specific data structures and associated scientific questions typically addressed within a field. We will focus here on the context of ecology and discuss specific data structures and questions arising within that field.

In general, is a strong interest within ecological research in understanding how individuals - plants, animals or other organisms - interact with each other and with the environment. The spatial pattern formed by the locations of individuals in space can reflect both the inter-actions among individuals as well preferences of different species for specific environmental conditions or habitats. A statistical analysis based on spatial or spatio-temporal point process methodology can analyse these patterns and - as a result - reveal, e.g. specific habitat preferences in a changing environment. Due to the development of computationally efficient model fitting methodology such as integrated nested Laplace approximation (INLA), realistically complex spatial and spatio-temporal models may be formulated and fitted to spatial point pattern data within feasible time.

In many cases, however, it is not necessarily straight forward to observe ecological data since the environment a species lives in might be hard to access or the general area of interest might be very large. To account for this ecologists have to develop data collection methods that are feasible for a specific system and species with it. As a result, the observation processes vary with the nature of general system a study is interested in (e.g. is it terrestrial or aquatic?) and the specific behavioural patterns of a species data (e.g. are there any detection issues or are we likely to have seen every individual in the areas we surveyed?). In order to provide practically relevant modelling methodology - and software - these different observation processes have to be taken into account. With this in mind, the R package inlabru has been developed and combined with the computational efficiency of INLA. Here the observation process is seen as an operation on the ecological process of interest, e.g. detection issues may be modelled as a thinning of a point process. The software allows us to estimate the parameters of the detection process as well as those of the process of interest simultaneously. More generally, inlabru is not only relevant to spatial point processes and ecological observation processes. It provides a general and flexible, computationally efficient fitting tool for complex spatial and spatio-temporal models. It is a wrapper around the well-known package R-INLA with a streamlined interface to R-INLA with the aim of simplifying the user's code required for fitting spatial and other complex models. It also provides additional features that are particularly relevant in applied contexts. In particular, it facilitates point processes modelling for log Gaussian Cox processes, accounts for complex observation processes and allows for non-linear functional relationships.

In the talk we will briefly discuss the capabilities of inlabru, but we will also put a strong emphasis on exploring the need for - as well as approaches to - communicating the methodology well to potential users. This is picking up on the responsibility of model developers to support scientists who use the methodology. In other word, this talk uses the example of the software package inlabru and the associated broad range of statistical methodology to outline an approach to addressing the issue of juggling the right balance between treating an approach as a black box and explaining the every single mathematical detail of a modelling

approach. In particular, we will discuss our recent thoughts on and attempts to finding a level of explanation that summarises what the methodology does through focusing on the role of the different model components and how they are reflected in the syntax of the package.

Parallel III

16:30 — 18:00, Wednesday 15 March

Optimisation meets sampling

Location: Room 1 - *Chair* : Chris Nemeth

Speaker: Nikolas Nusken (King's College London)

Title: Stein transport for Bayesian inference

Abstract: This talk is about Stein transport, a novel methodology for Bayesian inference that pushes an ensemble of particles along a predefined curve of tempered probability distributions. The driving vector field is chosen from a reproducing kernel Hilbert space and can equivalently be obtained from either a suitable kernel ridge regression formulation or as an infinitesimal optimal transport map. These viewpoints connect sampling with optimisation and the problem of approximating the solutions to high-dimensional PDEs. The update equations of Stein transport resemble those of Stein variational gradient descent (SVGD) but introduce a time-varying score function as well as specific weights attached to the particles. I will also discuss the geometric underpinnings that connect Stein transport and SVGD.

Speaker: Louis Sharrock (University of Bristol)

Title: Coin Sampling: Gradient-Based Bayesian Inference without Learning Rates

Abstract: In recent years, particle-based variational inference (ParVI) methods such as Stein variational gradient descent (SVGD) have grown in popularity as scalable methods for Bayesian inference. Unfortunately, the properties of such methods invariably depend on hyperparameters such as the learning rate, which must be carefully tuned by the practitioner in order to ensure convergence to the target measure at a suitable rate. In this talk, we introduce a suite of new particle-based methods for scalable Bayesian inference based on coin betting, which are entirely learning-rate free. We illustrate the performance of our approach on a range of numerical examples, including several high-dimensional models and datasets, demonstrating comparable performance to other ParVI algorithms.

Speaker: Alberto Cabezas-Gonzalez (Lancaster University)

Title: Preconditioning of posterior densities with approximate transport maps

Abstract: We discuss a novel framework for efficiently sampling from complex probability distributions using a combination of normalizing flows and Monte Carlo methods. The central idea is to learn a diffeomorphism that "pushes forward" the non-Gaussian structure of the target distribution to an approximately Gaussian distribution. We then use Markov Chain Monte Carlo to sample from the transformed distribution. The samples are finally "pulled back" using the inverse normalizing flow, yielding samples that approximate the stationary target distribution of interest. The diffeomorphism is implemented as a normalizing flow, which looks to capture the global structure of the target, thus easing the sampling problem. Designing efficient MCMC algorithms usually relies on using local gradient information from the target distribution and optimally tuning the free parameters of the algorithm. The method of preconditioning hopes to exploit the global structure of the target to allow for simple, tuning-free algorithms for sampling. As an illustration of this method, we present Transport Elliptical Slice Sampling, a combination of normalizing flows and elliptical slice sampling optimized for modern computer architectures, where its adaptation mechanism utilizes parallel cores to rapidly run multiple Markov chains for a few iterations.

Robustness to model misspecification

Location: Room 2 - *Chair* : Jeff Miller

Speaker: Jonathan Huggins (Boston University)

Title: Robust, Structurally-Aware Inference with Mixture Models

Abstract: Learning latent structures with real-world interpretations is a common task in statistics. Of particular importance are mixture models for discovering unobserved groups (e.g., cell types) that generate observed data (e.g., single-cell gene expression levels). However, mixture model inference is usually ill-defined a priori because the assumed observation model is only an approximation to the true data-generating process and the number of groups is unknown. Thus, as the number of observations increases, rather than obtaining better inferences, the opposite occurs: the data is explained by hallucinating spurious groups that compensates for the shortcomings of the observation model. However, there are two important sources of prior knowledge that we can exploit to obtain well-defined results no matter the dataset size: causal information (e.g., the latent groups cause the observed signal but not vice-versa) and a domain knowledge about roughly how wrong the observation model is. We propose a new model selection criteria that, while model-based, uses this available knowledge to obtain mixture model inferences that are robust to misspecification of the observation model and account for (known) causal structure. We prove a consistency result for our approach and validate it in applications to flow cytometry and single-cell RNA-sequencing data.

Speaker: Catherine Xue (Harvard TH Chan School of Public Health)

Title: Robust discovery of mutational signatures using power posteriors

Abstract: Mutational signatures are distinctive patterns of mutations resulting from carcinogenic molecular processes, such as UV radiation, molecular effects of chemical agents, and defective DNA repair mechanisms. Non-negative matrix factorization (NMF) models have been used to discover mutational signatures and deconvolve their respective contributions in individual tumors from sequencing data. However, any assumed model is only a rough approximation to reality; as a consequence, the results are sometimes misleading and irreproducible. We propose an alternative approach to mutational signature inference that, by leveraging the power posterior, is robust to using an approximate model and, by using a novel sparsity-inducing prior, automatically infers the number of signatures. We demonstrate the robustness and accuracy of our approach on simulated data and real data from the ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium.

Speaker: Ryan Giordano (MIT EECS)

Title: Approximate data deletion and replication with the Bayesian influence function

Abstract: Many model-agnostic statistical diagnostics are based on repeatedly re-fitting a model with some observations deleted or replicated. Cross-validation, the non-parametric bootstrap, and outlier detection via case deletion are examples of this technique. However, for Bayesian statistical procedures based on Markov Chain Monte Carlo (MCMC), re-computing posteriors for many slightly different datasets can be computationally prohibitive. Instead of exactly re-fitting, one might use the entire dataset and a single MCMC run to form a linear approximation to the effect of re-weighting observations. In the robust statistics literature, the leading term of this linear approximation is known as the influence function. We show that, for Bayesian posteriors, the influence function takes the form of a set of easily-estimated posterior covariances, and that the error of the linear approximation vanishes asymptotically for finite-dimensional posteriors under standard regularity conditions. However, in models for which the number of parameters grows with the size of the data, N, we show that the error of the linear approximation based on the influence function does not vanish, even for finite-dimensional subsets of the parameters whose posterior does concentrate at a $\sqrt{N}$ rate. We discuss the implications for infinitesimal jackknife covariances, the bootstrap, and approximate cross-validation, as well what is implicitly meant by "exchangeability' ' when using the influence function in this way.

Bayesian Regression on Networks

Location: Room 3 and Online - *Chair* : Sameer Deshpande

Speaker: Anna Menacher (Oxford)

Title: Bayesian Lesion Estimation with a Structured Spike-and-Slab Prior

Abstract: Neural demyelination and brain damage accumulated in white matter appear as hyperintense areas on T2-weighted MRI scans in the form of lesions. Modeling binary images at the population level, where each voxel represents the existence of a lesion, plays an important role in understanding aging and inflammatory diseases. We propose a scalable hierarchical Bayesian spatial model, called BLESS, capable of handling binary responses by placing continuous spike-and-slab mixture priors on spatially-varying parameters and enforcing spatial dependency on the parameter dictating the amount of sparsity within the probability of inclusion. The use of mean-field variational inference with dynamic posterior exploration, which is an annealing-like strategy that improves optimization, allows our method to scale to large sample sizes. Our method also accounts for underestimation of posterior variance due to variational inference by providing an approximate posterior sampling approach based on Bayesian bootstrap ideas and spike-and-slab priors with random shrinkage targets. Besides accurate uncertainty quantification, this approach is capable of producing novel cluster size based imaging statistics, such as credible intervals of cluster size, and measures of reliability of cluster occurrence. Lastly, we validate our results via simulation studies and an application to the UK Biobank, a large-scale lesion mapping study with a sample size of 40,000 subjects.

Speaker: Alexander Nikitin (Aalto University)

Title: Spatio-temporal graph kernel via SPDEs

Abstract: Gaussian processes (GPs) provide a principled and direct approach for inference and learning on graphs. In order to apply GPs to graph problems, it is crucial to choose proper kernels. In this talk, I will show how graph kernels were developed, and extend them to spatio-temporal problems on graphs. In particular, I will introduce a framework that connects stochastic partial differential equations (SPDEs) and GPs on graphs and use it for deriving graph kernels via SPDEs. We will derive spatio-temporal graph kernels from stochastic heat and wave equations and discuss applications of this approach to practical problems.

Speaker: Sameer Deshpande (University of Wisconsin–Madison)

Title: BART for network-linked data

Abstract: We consider regression with network-linked data in which (1) covariate-response pairs are observed at the vertices of a given network but (2) the regression relationship might different vertex-to-vertex. We describe how to use the popular Bayesian Additive Regression Trees (BART) model for this problem in a way that does not require pre-specifying the functional form of the regression function or how the regression function varies across the network. Key to our proposal are (1) a more general class of decision trees that can assign multiple levels of a categorical predictor to the left or right child of an internal node and (2) several stochastic processes that randomly partition a network into two, possibly connected, components. We will describe our software implementation and demonstrate its use on synthetic data and a real-world example involving neighborhood-level crime in the city of Philadelphia. A preprint describing some of this work is available at https://arxiv.org/abs/2211.04459.

Parallel IV

10:30 — 12:00, Thursday 16 March

Scalable Monte Carlo

Location: Auditorium and Online - *Chair* : Paul Fearnhead

Speaker: Francesca Crucinio (ENSAE, Paris)

Title: A divide and conquer sequential Monte Carlo approach to high dimensional filtering

Abstract: Abstract: We propose a divide-and-conquer approach to filtering which decomposes the state variable into low-dimensional components to which standard particle filtering tools can be successfully applied and recursively merges them to recover the full filtering distribution. It is less dependent upon factorization of transition densities and observation likelihoods than competing approaches and can be applied to a broader class of models. Performance is compared with state-of-the-art methods on a benchmark problem and it is demonstrated that the proposed method is broadly comparable in settings in which those methods are applicable, and that it can be applied in settings in which they cannot.

Speaker: Giacomo Zanella (Bocconi University)

Title: Complexity of Gibbs Samplers through Bayesian Asymptotics

Abstract: Gibbs samplers are popular algorithms to approximate posterior distributions arising from Bayesian hierarchical models. Despite their popularity and good empirical performances, however, there are still relatively few quantitative theoretical results on their scalability or lack thereof, e.g. much less than for gradient-based sampling methods. We introduce a novel technique to analyse the asymptotic behaviour of mixing times of Gibbs Samplers, based on tools from Bayesian asymptotics. Our methodology applies to high-dimensional regimes where both the number of datapoints and parameters increase, and allows us to study the complexity of Gibbs samplers fitting Bayesian hierarchical models under random data-generating assumptions. The methodology is illustrated on two-level hierarchical models with generic likelihoods. For this class, we are able to provide dimension-free convergence results under mild conditions. Specific examples with Gaussian, Binomial and Categorical likelihoods are also discussed.

Speaker: Nikola Surjanovic (University of British Columbia)

Title: Parallel tempering with a variational reference

Abstract: Sampling from complex target distributions is a challenging task fundamental to Bayesian inference. Parallel tempering (PT) addresses this problem by constructing a Markov chain on the expanded state space of a sequence of distributions interpolating between the posterior distribution and a fixed reference distribution, which is typically chosen to be the

prior. However, in the typical case where the prior and posterior are nearly mutually singular, PT methods are computationally prohibitive. In this work we address this challenge by constructing a generalized annealing path connecting the posterior to an adaptively tuned variational reference. The reference distribution is tuned to minimize the forward (inclusive) KL divergence to the posterior distribution using a simple, gradient-free moment-matching procedure. We show that our adaptive procedure converges to the forward KL minimizer, and that the forward KL divergence serves as a good proxy to a previously developed measure of PT performance. We also show that in the large-data limit in typical Bayesian models, the proposed method improves in performance, while traditional PT deteriorates arbitrarily. Finally, we introduce PT with two references—one fixed, one variational—with a novel split annealing path that ensures stable variational reference adaptation. The paper concludes with experiments that demonstrate the large empirical gains achieved by our method in a wide range of realistic Bayesian inference scenarios.

Computational challenges in modeling complex data

Location: Room 1 - *Chair* : Raffaele Argiento

Speaker: Amy Herring (Duke University, USA)

Title: Bayesian functional principal component analysis using relaxed nearly orthogonal processes

Abstract: Functional Principal Component Analysis (FPCA) is a prominent tool to characterize variability and reduce dimension of longitudinal and functional datasets. Bayesian implementations of FPCA are advantageous because of their ability to propagate uncertainty in subsequent modeling. To ease computation, many modeling approaches rely on the restrictive assumption that functional principal components can be represented through a pre-specified basis. Under this assumption, inference is sensitive to the basis, and misspecification can lead to erroneous results. Alternatively, we develop a flexible Bayesian model using Nearly Mutually Orthogonal (NeMO) processes. We define NeMO processes to enforce mutual orthogonality between principal components to ensure identifiability of model parameters. The joint distribution of NeMO processes is governed by a penalty parameter that determines the degree to which the processes are mutually orthogonal and is related to ease of posterior computation. In comparison to other methods, analysis using NeMO processes provides a more flexible, computationally convenient approach that facilitates accurate propagation of uncertainty. We demonstrate our proposed model in an application to study the associations of breastfeeding status, illness, and demographic factors with weight dynamics in early childhood in an at risk population. (Joint work with James Matuk and David Dunson)

Speaker: Sirio Legramanti (Università degli Studi di Bergamo, Italy)

Title: Concentration of discrepancy-based ABC via Rademacher complexity

Abstract: There has been increasing interest on summary-free versions of approximate Bayesian computation (ABC), which replace distances among summaries with discrepancies between the whole empirical distributions of the observed data and the synthetic samples generated under the proposed parameter values. The success of these solutions has motivated theoretical studies on the concentration properties of the induced posteriors. However, current results are often specific to the selected discrepancy, and mostly rely on existence arguments which are typically difficult to verify and provide bounds not readily interpretable. We address these issues via a novel bridge between the concept of Rademacher complexity and recent concentration theory for discrepancy-based ABC. This perspective yields a unified and interpretable theoretical framework that relates the concentration of ABC posteriors to the behavior of the Rademacher complexity associated to the chosen discrepancy in the broad class of integral probability semimetrics. This class extends summary-based ABC, and includes the widely-implemented Wasserstein distance and maximum mean discrepancy (MMD), which admit interpretable bounds for the corresponding Rademacher complexity along with constructive sufficient conditions for the existence of such bounds. Therefore, this unique bridge crucially contributes towards an improved understanding of ABC, as further clarified through a focus of this theory on the MMD setting and via illustrative simulations. (Joint work with Daniele Durante and Pierre Alquier)

Speaker: Gregor Kastner (University of Klagenfurt, Austria)

Title: Posterior predictive model assessment using formal methods in a spatio-temporal model

Abstract: We propose an interdisciplinary framework, Bayesian formal predictive model assessment. It combines Bayesian predictive inference, a well established tool in statistics, with formal verification methods rooting in the computer science community. Bayesian predictive inference allows for coherently incorporating uncertainty about unknown quantities by making use of methods or models that produce predictive distributions, which in turn inform decision problems. By formalizing these problems and the corresponding properties, we can use spatio-temporal reach and escape logic to formulate and probabilistically assess their satisfaction. This way, competing models can directly be compared based on their ability to predict the property satisfaction a posteriori. The approach is illustrated on an urban mobility application, where the crowdedness in the center of Milan is proxied by aggregated mobile phone traffic data. We specify several desirable spatio-temporal properties related to city crowdedness such as a fault-tolerant network or the reachability of hospitals. After verifying these properties on draws from the posterior predictive distributions, we compare several spatio-temporal Bayesian models based on their overall and property-based predictive performance. (Joint work with Laura Vana, Ennio Visconti, Laura Nenzi and Annalisa Cadonna)

Evidence synthesis: conflicts, splits and cuts

Location: Room 2 - *Chair* : Anne Presanis

Speaker: Noa Kallioinen (Aalto University)

Title: Detecting and diagnosing prior and likelihood sensitivity with power-scaling

Abstract: Determining the sensitivity of a posterior to perturbations of the prior and likelihood is an important part of the Bayesian workflow. We present a practical and computationally efficient sensitivity analysis approach using importance sampling to estimate properties of posteriors that result from power-scaling perturbations of the prior or likelihood. We suggest a diagnostic based on this that can indicate the presence of prior-data conflict or likelihood noninformativity. The approach can be readily included in Bayesian workflows with minimal work by the model builder. We present the approach and implementation (the priorsense R package), discuss limitations, and demonstrate the workflow on case studies of real data.

Speaker: Xuejun Yu (National University of Singapore)

Title: Variational inference for cutting feedback in misspecified models

Abstract: Bayesian analyses combine information represented by different terms in a joint Bayesian model. When one or more of the terms is misspecified, it can be helpful to restrict the use of information from suspect model components to modify posterior inference. This is called "cutting feedback", and both the specification and computation of the posterior for such "cut models" is challenging. In this paper, we define cut posterior distributions as solutions to constrained optimization problems, and propose variational methods for their computation. These methods are faster than existing Markov chain Monte Carlo (MCMC) approaches by an order of magnitude. It is also shown that variational methods allow for the evaluation of computationally intensive conflict checks that can be used to decide whether or not feedback should be cut. Our methods are illustrated in a number of simulated and real examples, including an application where recent methodological advances that combine variational inference and MCMC within the variational optimization are used.

Speaker: Robert Goudie (MRC Biostatistics Unit, University of Cambridge)

Title: Joining Bayesian submodels and translating predictive information into priors

Abstract: Synthesising multiple sources of prior information and data into a joint analysis provides more precise estimates and reduces the risk of biases introduced by using only partial data, but can be challenging to do in practice. I will discuss two aspects. First, constructing a joint model involving many data sources can be difficult. To simplify, each data source is often modelled separately, but this results in uncertainty not being fully propagated. I will describe a simple, general method that requires only small changes to existing models and software. We first form a joint Bayesian model based upon the original submodels using a generic approach we call "Markov melding". We show that this model can be fitted in submodel-specific stages, rather than as a single, monolithic model. We also show the concept can be extended to "chains

of submodels", in which submodels relate to neighbouring submodels via common quantities. Second, apriori information is often easier to obtain on observable quantities in a model (or other low dimensional marginals). However, the appropriate informative prior that matches this information is not always obvious, particularly for complex models. I will discuss an approach (and algorithm) for "translating" such information into priors on parameters.

Trust and adding new algorithms to probabilistic programming frameworks

Location: Room 3 and Online - *Chair* : Aki Vehtari

Speaker: Måns Magnusson (Uppsala University, Sweden)

Title: posteriordb: Testing, Benchmarking and Developing Bayesian Inference Algorithms

Abstract: Probabilistic programming frameworks (PPF), such as Stan, Tensorflow probability, and Pyro are becoming increasingly popular for data analysis and predictive modelling. Many of these frameworks rely on general inference algorithms (GIA) such as Hamiltonian Monte Carlo or Black-Box Variational inference. The recent development of new inference algorithms, adaptation methods and implementations has sparked an interest in the evaluation of PPFs and GIAs. We propose posteriordb, a database with posteriors with a large set of different models and data to evaluate and test PPFs and GIA efficiently.

Speaker: Lu Zhang (University of Southern California, USA)

Title: Pathfinder: Parallel quasi-Newton variational inference - and how it was tested

Abstract: This talk introduces Pathfinder, a parallel quasi-Newton algorithm that efficiently reaches regions of high probability mass. Pathfinder locates normal approximations to the target density along a quasi-Newton optimization path and returns draws from the approximation with the lowest estimated Kullback-Leibler divergence to the target distribution. The Monte Carlo KL divergence estimates are embarrassingly parallelizable in the core Pathfinder algorithm, which, along with multiple runs in the resampling version, further increases its speed advantage with multiple cores. Simulation studies have demonstrated that Pathfinder's approximate draws are superior to those generated by automatic differentiation variational inference (ADVI) and comparable to short chains of dynamic Hamiltonian Monte Carlo (HMC), as measured by 1-Wasserstein distance. Despite its efficiency, studies have revealed that the performance of Pathfinder is sensitive to the pathological geometry of the posterior. The talk will also present case studies to shed light on the reasons behind Pathfinder's sensitive behavior, along with a brief summary of ongoing research conducted to address this issue.

Speaker: Mitzi Morris (Columbia University, USA)

Title: From a your algorithm to a Stan model: BridgeStan

Abstract: BridgeStan provides efficient in-memory access through Python, Julia, and R to the methods of a Stan model, including log densities, gradients, Hessians, and constraining and unconstraining transforms, thus providing a much-needed tool for inference algorithm development.

I will present the BridgeStan interface and the specific hooks provided into the C++ Stan model object, then I will discuss the design of Stan model class and its implications for new algorithms.

Parallel V

15:00 — 16:30, Thursday 16 March

Piecewise deterministic Monte Carlo: Recent Advances

Location: Auditorium and Online - *Chair* : Gareth O Roberts

Speaker: Paul Fearnhead (Lancaster University)

Title: Recent Advances for PDMP Samplers

Abstract: Alternatives to standard MCMC algorithms that are based on simulating a continuous-time Markov process called a piecewise deterministic Markov process have shown promise. They are non-reversible and rejection free samplers and can scale well to big data scenarios. However they can be difficult to implement.

This talk will cover some recent work on extending the class of models that these PDMP samplers can be used for, covering efficient simulation of the PDMP using convex-concave decomposition of the rate, and algorithms that can deal with discontinuous target distributions, and that allow sampling from spaces of different dimension.

This is joint work with Augustin Chevallier and Matt Sutton.

Speaker: Sebastiano Grazzi (University of Warwick)

Title: Methods and applications of PDMP samplers with boundary conditions

Abstract: In this talk, I will formally introduce piecewise deterministic Markov processes (PDMPs) endowed with ""sticky floors"","soft/hard walls"" and ""teleportation portals"" which can be used for Monte Carlo simulation and allow to target efficiently a rich class of measures arising in Bayesian inference.

I will motivate and illustrate the framework with three challenging statistical applications: high-dimensional Bayesian variable selection, for sampling the latent space of infection times

with unknown infected population size in the SIR model with notifications and for sampling efficiently the invariant measure in hard-sphere models.

The class of processes presented here extends [1] and is joint work with J. Bierkens, G. Roberts, and M. Schauer.

[1] Sticky PDMP samplers for sparse and local inference problem; Bierkens J., Grazzi S., van der Maulen F., Schauer M.

keywords: PDMP samplers, Bayesian variable selection, SIR models, Hard-sphere models

Speaker: Gareth O Roberts (University of Warwick)

Title: Subsampling and the efficiency of PDMPs

Abstract: PDMPs provide a relatively new approach to MCMC with considerable promise and a few notable successes. There has been substantial and impressive advances in understanding these algorithms from a theoretical viewpoint. However one of the most exciting features they possess is that of principled subsampling. Specifically where we can decompose a target density multiplicatively (for example for independent data), they can be implemented using unbiased non-negative estimates of log-density gradients without leading to any bias. This leads to substantial reduction to the computational cost of implementing the algorithm. For instance for a data set of size N which obeys an $N^{-1/2}$ posterior contraction rate, control variate subsampling methods can lead to efficiency gain of efficiency of $O(N)$ leading to the potential for super-efficiency (the property that the entire algorithms running time complexity is less than that of calculating the likelihood once). On the other hand unlike other principled subsampling algorithms (such as SCALE), subsampling generally leads to a deterioration of mixing for the PDMP process. This talk will give results describing the extent of this deterioration. It will consider both transient and stationary phases, and will describe implications for the choice of subsample size.

Likelihood-free inference with kernel distances

Location: Room 1 - *Chair* : Francois-Xavier Briol

Speaker: Charita Dellaporta (University of Warwick)

Title: Robust Bayesian Inference for Simulator-based Models via the MMD Posterior Bootstrap

Abstract: Simulator-based models are models for which the likelihood is intractable but simulation of synthetic data is possible. They are often used to describe complex real-world phenomena, and as such can often be misspecified in practice. In this talk, I will present a novel

algorithm based on the posterior bootstrap and Maximum Mean Discrepancy (MMD) estimators. This leads to a highly-parallelisable Bayesian inference algorithm with strong robustness properties. This is demonstrated through an in-depth theoretical study which includes generalisation bounds, frequentist consistency and robustness of our posterior guarantees. The approach is also illustrated on a range of examples including a g-and-k distribution and a toggle-switch model.

Speaker: Lorenzo Pacchiardi (University of Oxford)

Title: Generalized Bayesian Likelihood-Free Inference Using Scoring Rules and Stochastic-Gradient MCMC

Abstract: We propose a framework for Bayesian Likelihood-Free Inference (LFI) based on Generalized Bayesian Inference. To define the generalized posterior, we use Scoring Rules (SRs), which evaluate probabilistic models given an observation. In LFI, we can sample from the model but not evaluate the likelihood; for this reason, we employ SRs which admit unbiased empirical estimates. We use the Energy and the Kernel SRs, for which our posterior enjoys consistency in a well-specified setting and outlier robustness, but our general framework applies to other SRs. The straightforward way to perform posterior inference relies on pseudo-marginal Markov Chain Monte Carlo (MCMC). While this works satisfactorily for simple setups, it mixes poorly, which makes inference impossible when many observations are present. Hence, we employ stochastic-gradient (SG) MCMC methods, which are rejection-free and have thus no mixing issues. The targets of both sampling schemes only approximate our posterior, but the error vanishes as the number of model simulations at each MCMC step increases. In practice, SG-MCMC performs better than pseudo-marginal at a lower computational cost when both are applicable and scales to higher-dimensional setups. In our simulation studies, we compare with related approaches on standard benchmarks and a chaotic dynamical system from meteorology; for the latter, SG-MCMC allows us to infer the parameters of a neural network used to parametrize a part of the update equations of the dynamical system.

Speaker: Ayush Bharti (Aalto University)

Title: Optimally-Weighted Estimators of the Maximum Mean Discrepancy for Likelihood-Free Inference

Abstract: Likelihood-free inference methods typically make use of a distance between simulated and real data. A common example is the maximum mean discrepancy (MMD), which has previously been used for approximate Bayesian computation, minimum distance estimation, generalised Bayesian inference, and within the nonparametric learning framework. The MMD is commonly estimated at a root-m rate, where m is the number of simulated samples. This can lead to significant computational challenges since a large m is required to obtain an accurate estimate, which is crucial for parameter estimation. In this paper, we propose a novel estimator for the MMD with significantly improved sample complexity. The estimator is particularly well suited for computationally expensive smooth simulators with low- to mid-dimensional inputs. This claim is supported through both theoretical results and an extensive simulation study on benchmark simulators.

New tools for high-dimensional Bayesian inference from physics and ML

Location: Room 2 - *Chair* : Antonietta Mira

Speaker: Aldo Glielmo (Bank of Italy and International School for Advanced Studies (SISSA))

Title: The information imbalance: a tool to measure the relative information content of distance measures

Abstract: Real-world data typically contain a large number of features that are often heterogeneous in nature, relevance, and also units of measure. When assessing the similarity between data points, one can build various distance measures using subsets of these features; and using the fewest features while still retaining sufficient information is crucial in many statistical learning schemes, particularly when data are sparse. In my talk I will describe the "information imbalance": a novel statistical concept that quantifies the relative information retained when using two different distance measures, and determine if they are equivalent, independent, or if one is more informative than the other [Glielmo et al, PNAS Nexus (2022)]. I will then show how the information imbalance can be used to find the most informative distance measure out of a pool of candidates, and present applications of this idea for the analysis of the Covid-19 epidemic spreading as well as for the construction of optimally informative descriptors of physical systems. The information imbalance is readily available in the open-source package "DADApy" (https://github.com/sissa-data-science/DADApy), along with a set of other recently introduced methods to characterise data-manifolds using solely distances between data points [Glielmo et al, Patterns (2022)]. I will conclude my talk with an illustration of the package and how to use it.

Speaker: Simone Ulzega (: Zurich University of Applied Sciences (ZHAW))

Title: Boosting Bayesian parameter inference of SDE models by Hamiltonian scale separation. A real-world case study in urban hydrology

Abstract: In essentially all applied sciences, data-driven modeling heavily relies on a sound calibration of model parameters to measured data for making probabilistic predictions. Bayesian statistics is a consistent framework for parameter inference where knowledge about model parameters is expressed through probability distributions. However, Bayesian inference with stochastic models can become computationally extremely expensive and it is therefore hardly ever applied. We propose a very efficient approach for boosting Bayesian parameter inference of stochastic differential equation (SDE) models calibrated to measured time-series, using a Hamiltonian Monte Carlo (HMC) approach combined with a multiple time-scale integration. We present the first application of this HMC algorithm to a real-world case study from urban hydrology.

Speaker: Dhiman Ray (Atomistic Simulations Research Line, Italian Institute of Technology, Genoa, Italy)

Title: Enhanced sampling of Rare Events

Abstract: A key challenge in computer simulation of molecular processes, such as chemical reactions and drug-receptor binding, is that the relevant timescales are beyond the reach of currently available computational resources. To address this issue, enhanced sampling methods were introduced, where an external biasing force is applied in a well-directed manner to accelerate the process in question for extracting meaningful quantitative information. Metadynamics is one such approach where gaussian hills are deposited along a carefully chosen reduced dimensional collective variable (CV) space to accelerate rare transitions. It allows for the exploration of high-energy regions of the configurational space and computes a free-energy landscape that is otherwise not possible with standard simulations. We recently improved upon this idea to develop the On-the-fly Probability Enhanced Sampling (OPES) which allows for a better convergence of the predicted marginal probability distribution in the CV space. The performance of these methods depends critically on the choice of the CV. To address this problem we introduced machine learning (ML) and data-driven approaches to develop a generalizable protocol for constructing optimal CVs. These techniques not only provide a more in-depth understanding of these complex molecular phenomena such as protein folding and drug binding but also show promise to be applied in other fields of science such as parameter optimization in data science and ML.

MCMC for Multi-Modal Distributions

Location: Room 3 and Online - *Chair* : Raiha Browning

Speaker: Saifuddin Syed (University of Oxford)

Title: Scalable MCMC with non-reversible parallel tempering

Abstract: Markov chain Monte Carlo (MCMC) methods are the most widely used tools in Bayesian statistics for making inferences from complex posterior distributions. For challenging problems where the posterior is high-dimensional with well-separated modes, MCMC algorithms can get trapped exploring local regions of high probability. Parallel tempering (PT) tackles this problem by delegating the task of global exploration to a tractable reference distribution (e.g. prior) which communicates to the target (e.g. posterior) through a sequence of parallel MCMC algorithms targeting distributions of increasing complexity to the target.

The classical approach to designing PT algorithms relied on a reversibility assumption, making PT challenging to tune and even deteriorating performance when introducing too many parallel chains. This talk will introduce a new non-reversible paradigm for PT (NRPT) that dominates its reversible counterpart while avoiding the performance collapse endemic to reversible

PT methods. We will then establish near-optimal tuning guidelines, an efficient black-box methodology scalable to GPUs.

Researchers have used NRPT at scale to study the evolutionary structure of cancer, model plasma dynamics in nuclear fusion reactors, discover magnetic polarization in supermassive black hole M87 and generate the first image Sagittarius A*, the supermassive black hole at the center of the Milky Way.

[1] Syed, S., Bouchard-Côté, A., Deligiannidis, G., and Doucet, A. (2021), "Non-Reversible Parallel Tempering: a Scalable Highly Parallel MCMC Scheme," Journal of Royal Statistical Society, Series B, 84, 321–350.

[2] Syed, S., Romaniello, V., Campbell, T., and Bouchard-Côté, A. (2021), "Parallel tempering on optimized paths," in International Conference on Machine Learning.

[3] Surjanovic, N., Syed, S., Bouchard-Côté, A., and Campbell, T. (2022), "Parallel Tempering With a Variational Reference," in Advances in Neural Information Processing Systems.

Speaker: Matt Moores (University of Wollongong)

Title: The Annealed Leap-Point MCMC Sampler (ALPS) for multi-modal posterior distributions

Abstract: Multi-modal distributions pose major challenges for the usual algorithms that are employed in statistical inference. These problems are exacerbated in high-dimensional settings, where techniques such as Markov Chain Monte Carlo (MCMC) and Expectation Maximisation (EM) typically rely upon localised update mechanisms. Such localised algorithms can effectively become trapped in one of the local modes, leading to biased inference and underestimation of uncertainty.

In this talk, I will introduce the Annealed Leap-Point Sampler (ALPS), an MCMC algorithm that augments the state space of the target distribution with a sequence of modified, annealed (cooled) distributions. The temperature of the coldest state is chosen such that the corresponding annealed target density of each individual mode can be closely fitted by a Laplace approximation. As a result, independent MCMC proposals based on a mixture of Gaussians can jump between modes even in high-dimensional problems. The ability of this method to "mode hop" at the super-cold state is then filtered through to the target state by swapping information between neighbours, in a similar manner to parallel tempering. ALPS also incorporates the best aspects of current gold-standard approaches to multi-modal sampling in high-dimensional contexts.

We have implemented ALPS as an open-source R package. Our method is demonstrated using examples of multi-modal distributions that arise in econometrics and chemistry. These applications include a seemingly unrelated regression (SUR) model of longitudinal data from U.S. manufacturing firms, as well as a model of line shape and broadening for curve fitting in Raman spectroscopy.

This is joint work with Nick Tawn and Gareth Roberts at the University of Warwick: Tawn, Moores & Roberts (2021) arXiv preprint https://arxiv.org/abs/2112.12908

Speaker: Krzysztof Latuszynski (University of Warwick)

Title: A framework for adaptive MCMC targeting multimodal distributions

Abstract: We propose a new Monte Carlo method for sampling from multimodal distributions. The idea of this technique is based on splitting the task into two: finding the modes of a target distribution and sampling, given the knowledge of the locations of the modes. The sampling algorithm relies on steps of two types: local ones, preserving the mode; and jumps to regions associated with different modes. Besides, the method learns the optimal parameters of the algorithm while it runs, without requiring user intervention. Our technique should be considered as a flexible framework, in which the design of moves can follow various strategies known from the broad MCMC literature. In order to design an adaptive scheme that facilitates both local and jump moves, we introduce an auxiliary variable representing each mode and we define a new target distribution on an augmented state space X $\times$ I, where X is the original state space of and I is the set of the modes. As the algorithm runs and updates its parameters, the resulting new target distribution also keeps being modified. This motivates a new class of algorithms, Auxiliary Variable Adaptive MCMC. We prove general ergodic results for the whole class before specialising to the case of our algorithm. This is joint work with Emilia Pope and Chris Holmes.

Parallel VI

10:30 — 12:00, Friday 17 March

Advances in theory and methodology of MCMC

Location: Auditorium and Online - *Chair* : Matti Vihola

Speaker: Christophe Andrieu (University of Bristol)

Title: Comparison of Markov Chains Monte Carlo via Weak Poincaré Inequalities with Applications

Abstract: Being able to compare performance of Markov chain Monte Carlo algorithms can provide some insights into their efficient implementation. The Peskun-Tierney order provides us with simple and efficient tools to compare MCMC. However the results are usually not quantitative, except in very specific scenarios.

In this work we show that such quantitative bounds can be derived in scenarios previously not covered, in particular scenarios involving sub-geometric convergence using particular functional inequalities.

We show that this provides new insights into the practical use of pseudo-marginal algorithms, analyse the effect of averaging in Approximate Bayesian Computation (ABC) and the use of products of independent averages, and also to study the case of lognormal weights relevant to particle marginal Metropolis–Hastings (PMMH).

Joint with Anthony Lee, Sam Power and Andi Q. Wang, available as arXiv:2112.05605 or Ann. Statist. 50 (6) 3592 - 3618, 2022.

Speaker: Błażej Miasojedow (University of Warsaw)

Title: Solidarity of Gibbs samplers: the spectral gap

Abstract: I will show the solidarity principle of the preeminent classes of the Markov chain Monte Carlo algorithms, namely the Gibbs samplers. It is proven that if any of the random scans or d! deterministic scans have a spectral gap then all of them have. Our methods rely on the geometric interpretation of the Gibbs samplers as alternating projection algorithms and analysis of the rate of convergence in the von Neumann–Halperin method of cyclic alternating projections. I will show that despite being not reversible the deterministic scan admits a property typical for reversible Markov chains, that is the Central Limit Theorem for functions merely in L2( ). Additionally, relations between the spectral gaps and the geometric ergodicity will be presented.

The talk is based on joint work with Iwona Chlebicka and Krzysztof Łatuszyński

Speaker: Jimmy Olsson (KTH)

Title: Effortless bias reduction in self-normalized importance sampling

Abstract: Self-normalized importance sampling (SNIS) is a classical method for approximating expectations under a given target distribution, known only up to a normalization constant, on the basis of independent samples from some proposal distribution and associated importance weights. While the use of self-normalization can have a positive effect on the dispersion of the estimator, it introduces bias. In this talk, I propose a new estimator whose complexity and variance are essentially the same as that of SNIS, but whose bias is significantly reduced. The estimator is a wrapper in the sense that it includes the same proposal samples and importance weights as SNIS, but makes clever use of iterated sampling–importance resampling (i-SIR) and Gibbs sampling to form a bias-reduced version of the SNIS estimator. The proposed algorithm is furnished with rigorous theoretical results, including new bias, variance and high-probability bounds, which are illustrated numerically. Finally, if time allows, I will also discuss the extension of the proposed sampling technology to Feynman–Kac particle models and the so-called particle-based, rapid incremental smoother (PaRIS).

Piecewise deterministic Monte Carlo: Theory

Location: Room 1 - *Chair* : Sebastiano Grazzi

Speaker: Ardjen Pengel (TU Delft)

Title: Convergence Diagnostics for High-Dimensional Piecewise Determinstic Monte Carlo

Abstract: We present strong invariance principles for PDMPs and demonstrate how these can be used to analyse convergence diagnostics for simulation output of Piecewise Deterministic Monte Carlo samplers for high-dimensional problems. Strong invariance principles describe the error term of a Brownian approximation of the partial sums of a stochastic process. While these strong approximation results have many applications, the results for continuous-time settings have been limited. Strong invariance principles provide a unified framework for analysing commonly used estimators of the asymptotic variance in settings with a dependence structure. Estimation of the MCMC standard error is in turn a necessary building block for many convergence diagnostics.

Speaker: Marc Corstanje (VU Amsterdam)

Title: A geometric extension to PDMPs

Abstract: In recent years, there has been a trend in MCMC that involves the utilisation of a geometric structure in the state space. Method such as Riemannian MALA or Riemannain HMC are shown to resolve certain shortcomings of other Monte Carlo methods particularly for high dimensional target distributions with strong correlations. The structure of Riemannian manifolds also provides a straightforward way to extend some well-known PDMP samplers to Riemannian PDMPs. We present a method to simulate the trajectory of a bouncy particle sample and a zig-zag sampler in a geometric setting and show results for this method to sample from distributions on the special orthogonal group.

Speaker: Joris Bierkens (TU Delft)

Title: Scaling of Piecewise Deterministic Monte Carlo for Anisotropic Targets

Abstract: Piecewise deterministic Markov processes (PDMPs) are a type of continuous-time Markov process that combine deterministic flows with jumps. Recently, PDMPs have garnered attention within the Monte Carlo community as a potential alternative to traditional Markov chain Monte Carlo (MCMC) methods. The Zig-Zag sampler and the Bouncy particle sampler are commonly used examples of the PDMP methodology which have also yielded impressive theoretical properties, but little is known about their robustness to extreme dependence or isotropy of the target density. In this talk we investigate this effect in detail in the stylised but important Gaussian case.

Approximate Bayesian Computation

Location: Room 2 - *Chair* : Christian Robert

Speaker: Umberto Picchini (Chalmers University of Technology and University of Gothenburg)

Title: Guided sequential ABC schemes for intractable Bayesian models

Abstract: Sequential algorithms such as sequential importance sampling (SIS) and sequential Monte Carlo (SMC) have proven fundamental in Bayesian inference. However, probabilistic models often do not admit a readily available likelihood function or one that is computationally cheap to approximate. In the last 20 years, simulation-based approaches have flourished to bypass the likelihood intractability by implicitly making use of it via model simulations. The most studied class of simulation-based inference methods is arguably approximate Bayesian computation (ABC). For ABC, sequential Monte Carlo (SMC-ABC) is the state-of-art sampler. However, since the ABC paradigm is intrinsically wasteful, sequential ABC schemes can benefit from well-targeted proposal samplers that efficiently avoid improbable parameter regions. We construct novel proposal samplers that are conditional to summary statistics of the data. In a sense, the proposed parameters are "guided" to rapidly reach regions of the posterior surface that are compatible with the observed data. This speeds up the convergence of these sequential samplers, thus reducing the computational effort, while preserving the accuracy in the inference. We provide a variety of guided samplers easing inference for challenging case-studies, including multimodal posteriors, highly correlated posteriors, hierarchical models with high-dimensional summary statistics.

Joint work with Massimiliano Tamborrino available at https://arxiv.org/abs/2206.12235

Speaker: Giorgos Vasdekis (University College London)

Title: Pseudo-marginal Piecewise Deterministic Monte Carlo.

Abstract: Piecewise Deterministic Markov Processes (PDMPs) have recently caught the attention of the MCMC community for having a non-diffusive behavior, potentially allowing them to explore the state space efficiently. This makes them good candidates to generate MCMC algorithms. One important problem in Bayesian computation is inference for models where pointwise evaluation of the posterior is not available, but one has access to an unbiased estimator of the posterior. A technique to deal with this problem is the Pseudo-marginal Metropolis Hastings algorithm. In this talk we describe a PDMP algorithm that can be used in the same posterior free setting and can be seen as the analogue of Pseudo-marginal for Piecewise Deterministic Monte Carlo. We show that the algorithm targets the posterior of interest. We also provide some numerical examples, focusing on the case of Approximate Bayesian Computation (ABC), a popular method to deal with problems in the setting of likelihood free inference.

This is joint work with Richard Everitt (University of Warwick).

Speaker: Jeremias Knoblauch (University College London)

Title: An Optimization-centric View on Bayes' Rule

Abstract: I summarize a recent line of research and advocate for an optimization-centric generalisation of Bayesian inference. The main thrust of this argument relies on identifying the tension between the assumptions motivating the Bayesian posterior and the realities of modern large-scale Bayesian Machine Learning. Our generalisation is a useful conceptual device, but also has methodological merit: it can address various challenges that arise when the standard Bayesian paradigm is deployed in a Machine Learning context—including robustness to model misspecification, robustness to poorly chosen priors, or inference in intractable likelihood models.

Normalising Flows to Enhance Bayesian Sampling

Location: Room 3 and Online - *Chair* : Joshua Bon

Speaker: Marylou Gabrié (École Polytechnique (CMAP))

Title: Approximate transport with flows for sampling: training without data and MC(MC) corrections

Abstract: Deep generative models parametrize very flexible families of distributions able to fit complicated datasets of images or text. These models provide independent samples from complex high-distributions at negligible costs. On the other hand, sampling exactly a target distribution, such a Bayesian posterior, is typically challenging: either because of dimensionality, multi-modality, ill-conditioning or a combination of the previous. In this talk, I will discuss recent works trying to enhance traditional inference and sampling algorithms with learning. I will present in particular flowMC, an adaptive MCMC with Normalizing Flows along with first applications and remaining challenges.

Speaker: Laurence Davies (Queensland University of Technology)

Title: Transport Reversible Jump Proposals

Abstract: Reversible jump Markov chain Monte Carlo (RJMCMC) proposals that achieve reasonable acceptance rates and mixing are notoriously difficult to design in most applications. Inspired by recent advances in deep neural network-based normalizing flows and density estimation, we demonstrate an approach to enhance the efficiency of RJMCMC sampling by performing transdimensional jumps involving reference distributions. In contrast to other RJMCMC proposals, the proposed method is the first to apply a non-linear transport-based approach to construct efficient proposals between models with complicated dependency structures. It is shown that, in the setting where exact transports are used, our RJMCMC proposals have the desirable property that the acceptance probability depends only on the model probabilities. Numerical experiments demonstrate the efficacy of the approach.

Speaker: Michael Arbel (INRIA Grenoble - Rhône-Alpes)

Title: Annealed Flow Transport Monte Carlo

Abstract: Annealed Importance Sampling (AIS) and its Sequential Monte Carlo (SMC) extensions are state-of-the-art methods for estimating normalizing constants of probability distributions. We propose here a novel Monte Carlo algorithm, Annealed Flow Transport (AFT), that builds upon AIS and SMC and combines them with normalizing flows (NFs) for improved performance. This method transports a set of particles using not only importance sampling (IS), Markov chain Monte Carlo (MCMC) and resampling steps-as in SMC, but also relies on NFs which are learned sequentially to push particles towards the successive annealed targets. We provide limit theorems for the resulting Monte Carlo estimates of the normalizing constant and expectations with respect to the target distribution. Additionally, we show that a continuous-time scaling limit of the population version of AFT is given by a Feynman–Kac measure which simplifies to the law of a controlled diffusion for expressive NFs. We demonstrate experimentally the benefits and limitations of our methodology on a variety of applications.

Parallel VII

13:00 — 14:30, Friday 17 March

Machine Learning meets Adaptive MCMC

Location: Auditorium and Online - *Chair* : Maxim Panov

Speaker: Alain Durmus (Ecole Polytechnique)

Title: The Kick-Kac teleportation algorithm: boost your favorite Markov Chain Monte Carlo using Kac formula

Abstract: In this work, we propose to target a given probability measure by combining two Markov kernels with different invariant probability measures. In its basic form, the mechanism consists in picking up the current position and moving it according to a -invariant Markov kernel as soon as the proposed move does not fall into a predefined region. If this is the case, then we resort to the last position in this region and move it according to another auxiliary Markov kernel before starting another excursion outside the region with the first kernel. These state dependent interactions allow to combine smoothly different dynamics that can be tailored to each region while the resulting process still targets the probability measure π thanks to an argument based on the Kac formula. Under weak conditions, we obtain the Law of Large numbers starting from any point of the state space, as a byproduct of the same property for the different implied kernels. Geometric ergodicity and Central Limit theorem are also established.

Generalisations where the indicator function on the region target is replaced by an arbitrary acceptance probability are also given and allow to consider any Metropolis Hastings algorithm as a particular case of this general framework. Numerical examples, including mixture of Gaussian distributions are also provided and discussed.

Speaker: Eric Moulines (Ecole Polytechnique)

Title: Ex2MCMC: Sampling through Exploration Exploitation

Abstract: We develop an explore-exploit Markov chain Monte Carlo algorithm (Ex2MCMC) that combines multiple global proposals and mobile moves. The proposed method is massively parallelizable and extremely computationally efficient. We prove the V-uniform geometric ergodicity of Ex2MCMC under realistic conditions and compute explicit bounds on the mixing rate showing the improvement due to multiple global moves. We show that Ex2MCMC allows fine-tuning of exploitation (local moves) and exploration (global moves) via a novel approach to propose dependent global moves. Finally, we develop an adaptive scheme, FlEx2MCMC, that learns the distribution of global trains through normalizing flows. We illustrate the efficiency of Ex2MCMC and its adaptive versions in many classical sampling benchmarks. We also show that these algorithms improve the quality of sampling GANs as energy-based models.

Speaker: Achille Thin (AgroParisTech)

Title: NEO: Non equilibrium sampling on the orbits of a deterministic transform or Optimization meets Importance Sampling

Abstract: Sampling from a complex distribution and approximating its intractable normalizing constant Z are challenging problems classically addressed in MCMC. We present here novel schemes. Given an invertible map T, these schemes combine (with weights) elements from the forward and backward Orbits through points sampled from a proposal distribution . The map T does not leave the target π invariant, hence the name NEO, standing for Non-Equilibrium Orbits. This map can be interpreted as some optimizing operator. we provide from this transformation unbiased estimators of the normalizing constant (NEO-IS) as well as multiple estimates of the normalizing constant to build an iterated sampling-importance resampling mechanism to sample from (NEO-MCMC). We also provide detailed theoretical results for both methods. In particular, we show that NEO-MCMC is uniformly geometrically ergodic and establish explicit mixing time estimates under mild conditions. Finally, we show that we can extend these schemes starting from an initial distribution that can be adapted along the procedure.

Junior advances in scalable Bayesian methods

Location: Room 1 - *Chair* : Alejandra Avalos-Pacheco

Speaker: Andrea Bertazzi (TU Delft)

Title: Piecewise deterministic Monte Carlo with splitting schemes

Abstract: Piecewise deterministic Markov processes (PDMPs) received substantial interest in recent years as an alternative to classical Markov chain Monte Carlo algorithms. While theoretical properties of PDMPs have been studied extensively, their practical implementation remains limited to specific applications in which bounds on the gradient of the negative log-target can be derived. In order to address this problem, we propose to approximate PDMPs using splitting schemes, that means simulating the deterministic dynamics and the random jumps in separate stages. As expected, basic symmetric splittings of PDMPs are approximations of second order. Focusing on the Zig-Zag sampler and the Bouncy Particle sampler, we discuss the properties of approximations obtained with splitting schemes. Finally, we show how to design a skew-reversible Metropolis-Hastings algorithm based on the Zig-Zag sampler. Numerical simulations suggest that our algorithms can outperform existing methods.

Speaker: François-Xavier Briol (University College London)

Title: Multilevel Bayesian quadrature

Abstract: Multilevel Monte Carlo is a key tool for approximating integrals involving expensive scientific models. The idea is to use approximations of the integrand to construct an estimator with improved accuracy over classical Monte Carlo. We propose to further enhance multilevel Monte Carlo through Bayesian surrogate models of the integrand, focusing on Gaussian process models and the associated Bayesian quadrature estimators. We show, using both theory and numerical experiments, that our approach can lead to significant improvements in accuracy when the integrand is expensive and smooth, and when the dimensionality is small or moderate. We conclude the paper with a case study illustrating the potential impact of our method in landslide-generated tsunami modelling, where the cost of each integrand evaluation is typically too large for operational settings.

Speaker: Karla Monterrubio-Gomez (The University of Edinburgh)

Title: On MCMC for Variationally Sparse Gaussian Process: A Pseudo-Marginal Approach

Abstract: Gaussian processes (GPs) are frequently used in machine learning and statistics to construct powerful models. However, when employing GPs in practice, important considerations must be made, regarding the high computational burden, approximation of the posterior, choice of the covariance function and inference of its hyperparmeters. To address these issues, Hensman et al. [2015] combine variationally sparse GPs with Markov chain Monte Carlo (MCMC) to derive a scalable, flexible and general framework for GP models. Nevertheless, the resulting approach requires intractable likelihood evaluations for many observation models. To bypass this problem, we propose a pseudo-marginal (PM) scheme that offers asymptotically exact inference as well as computational gains through doubly stochastic estimators for the intractable likelihood and large datasets. In complex models, the advantages of the PM scheme are particularly evident, and we demonstrate this on a two-level GP regression model with a nonparametric covariance function to capture non-stationarity.

Statistical Computing for Phylogenetics

Location: Room 2 - *Chair* : Luiz Max Carvalho

Speaker: Luke Kelly (University College Cork)

Title: Lagged couplings diagnose Markov chain Monte Carlo phylogenetic inference

Abstract: Phylogenetic inference attempts to reconstruct the ancestry of a set of observed taxa and is an intractable statistical problem on a complex, high-dimensional space. The likelihood function is an integral over unobserved evolutionary events on a tree and is frequently multimodal. MCMC methods are the primary tool for Bayesian phylogenetic inference, but constructing sampling schemes to efficiently explore the associated posterior distributions and diagnosing their performance are difficult tasks.

Couplings have recently been used to construct unbiased MCMC estimators and estimate useful bounds on the convergence of chains to their stationary distribution. We describe a procedure to couple a pair of Markov chains targeting a posterior distribution over a space of phylogenetic tree topologies, branch lengths, scalar parameters and latent variables such that the chains meet exactly after a random, finite number of iterations and remain coupled. We use the meeting times to diagnose convergence in total variation distance jointly across all components of the model on trees with up to 200 leaves.

Speaker: Jason Xu (Duke University)

Title: Likelihood-based Inference for Stochastic Epidemic Models on Dynamic Contact Networks

Abstract: Stochastic epidemic models such as the Susceptible-Infectious-Removed (SIR) model are widely used to model the spread of disease at the population level, but fitting these models present significant challenges when missing data or latent variables are present. In particular, the likelihood function of the partially observed data is typically considered intractable. We will discuss recent advances that enable likelihood computations without model simplifications in the presence of missing infection and recovery times via efficient data-augmented samplers. Our methods target the exact posterior without relying on model-based forward simulation, and apply to several classic stochastic compartmental models and allow for disease-dependent contact networks to evolve dynamically. We apply our methods to high-resolution mobile contact tracking data from the eX-FLU study of influenza on a college campus and observational data from the COVD-19 pandemic

Speaker: Ian Roberts (University of Warwick)

Title: Bayesian Inference for the Structured Coalescent

Abstract: The structured coalescent models the common ancestry of organisms sampled from a spatially structured population. A realisation of this process consists of a phylogenetic tree relating the samples superimposed with a migration history containing the geographic location of each ancestor. Current inference methods either attempt to simultaneously infer the phylogenetic tree and plausible migration histories, which is computationally expensive, or rely on approximations to the structured coalescent. I will present a Markov Chain Monte Carlo (MCMC) scheme which strikes a balance between these extremes by sampling migration histories (along with governing static parameters) for a fixed phylogenetic tree under the structured coalescent model.

Recent Advances in Variational Inference

Location: Room 3 and Online - *Chair* : Anna Menacher

Speaker: Debdeep Pati (Texas A&M University, USA)

Title: Approximate Bayes inference with soft distributional constraint

Abstract: Flexible Bayesian models are typically constructed using limits of large parametric models with a multitude of parameters that are often un-interpretable. In this article, we offer a novel get-around by proposing an exponentially tilted empirical likelihood carefully designed to concentrate near a simpler parametric family of distributions of choice with respect to a novel variant of the Wasserstein metric. It finds applications in a wide variety of robust inference problems, where we intend to make inference on the parameters associated with the centering distribution in presence of outliers. Importantly, our novel formulation of the modified Wasserstein metric enjoys great computational simplicity, exploiting the Sinkhorn regularization of discrete optimal transport problem, and being inherently parallelizable. We demonstrate superior performance of our methodology when compared against the state-of-the-art robust Bayesian inference methods. Interestingly, we demonstrate equivalence of our approach with a nonparametric Bayesian formulation under a suitable asymptotic framework, testifying to its flexibility. The constrained entropy maximization setup that sits at the heart of our likelihood formulation finds its utility beyond robust Bayesian inference; illustrations are provided in the field of demographic parity in machine learning.

Speaker: Yingzhen Li (Imperial College London, United Kingdom)

Title: Variational Bayes for deep learning: towards function-space uncertainty quantification

Abstract: The Bayesian deep learning community concerns about Bayesian methods on quantifying uncertainty for deep neural networks. To date, approximate inference methods (incl. variational Bayes) have been researched extensively for neural network weight-space posterior inference. However, ultimately we want to measure uncertainty for neural network outputs, i.e.,

function-space uncertainty. In this talk, I will review existing variational Bayes approaches applied to function-space posterior inference for neural networks. Many of these approaches have connections to Gaussian processes (GPs), and I will also discuss our latest work on quantifying uncertainty for Transformer models with sparse GP inspired techniques.

Speaker: Trevor Campbell (University of British Columbia, Canada)

Title: MixFlows: Principled Variational Bayesian Inference via Approximately Measure-Preserving Maps

Abstract: This talk will introduce mixed variational flows (MixFlows), a new variational family for Bayesian inference that consists of a mixture of pushforwards of an initial reference distribution under repeated applications of a map. Like most variational families, MixFlows enable efficient i.i.d. sampling, density evaluation, and unbiased ELBO estimation. But unlike other families, MixFlows enable MCMC-like convergence guarantees; and crucially, these guarantees hold without the need to solve any nonconvex optimization problem. In particular, we show that when the flow map is ergodic and measure-preserving, MixFlow distributions converge to the target distribution. We also provide bounds on the accumulation of error in practical implementations where the flow map is approximated. Finally, we provide an implementation of MixFlows based on uncorrected discretized Hamiltonian dynamics combined with deterministic momentum refreshment. Simulated and real data experiments show that MixFlows can provide more reliable posterior approximations than several black-box normalizing flows, as well as samples of comparable quality to those obtained from state-of-the-art MCMC methods.

Contributed Sessions

Parallel VIII

14.30 — 16.00, Friday 17 March

note: each contributed talk will be allocated 20 minutes, including time for questions.

Manifolds & Preconditioning

Location: Auditorium and Online - *Chair* : Heikki Haario

Speaker: Alexandros Beskos (University College London)

Title: Manifold Markov chain Monte Carlo methods for Bayesian inference in diffusion models

Abstract: Bayesian inference for nonlinear diffusions, observed at discrete times, is a challenging task that has prompted the development of a number of algorithms, mainly within the computational statistics community. We propose a new direction, and accompanying methodology, borrowing ideas from statistical physics and computational chemistry, for inferring the posterior distribution of latent diffusion paths and model parameters, given observations of the process. Joint configurations of the underlying process noise and of parameters, mapping onto diffusion paths consistent with observations, form an implicitly defined manifold. Then, by making use of a constrained Hamiltonian Monte Carlo algorithm on the embedded manifold, we are able to perform computationally efficient inference for a class of discretely observed diffusion models. Critically, in contrast with other approaches proposed in the literature, our methodology is highly automated, requiring minimal user intervention and applying alike in a range of settings, including: elliptic or hypo-elliptic systems; observations with or without noise; linear or non-linear observation operators. Exploiting Markovianity, we propose a variant of the method with complexity that scales linearly in the resolution of path discretisation and the number of observation times.

Speaker: Max Hird (University College London)

Title: Preconditioning for MCMC

Abstract: Linear transformation of the state variable (linear preconditioning) is a common technique that often drastically improves the practical performance of a Markov chain Monte Carlo algorithm. Despite this, however, quantifying the benefits of linear preconditioning is not well-studied theoretically, and rigorous guidelines for choosing preconditioners are not always readily available. Mixing time bounds for various samplers (HMC, MALA, Unadjusted HMC, Unadjusted Langevin) have been produced in recent works for the class of strongly log-concave and Lipschitz target distributions and depend strongly on a quantity known as the condition number. We study linear preconditioning for this class of distributions, and

under appropriate assumptions we provide bounds on the condition number after using a given linear preconditioner. The bounds are easy to interpret and can be used to quantify the mixing properties before and after preconditioning, as well as helping the practitioner choose a good preconditioner to use. We also present counterintuitive examples in which common preconditioning strategies that are used in popular software packages can actually increase the condition number, and therefore lead to a worse-performing algorithm. This is joint work with Samuel Livingstone.

Speaker: Elena Bortolato (University of Padova)

Title: Convergence of MCMC algorithms on manifolds through coupling techniques

Abstract: Some problems in statistics and machine learninig require sampling distributions on submanifolds embedded in R^D. To target such distributions, in the last twenty years constrained Markov Chain Monte Carlo methods have been developed (Brubaker et al. 2012, Zappa et al. 2018, Lelievre et al. 2019). Assessing the convergence of such algorithms still remains an open problem. We propose to apply coupling techniques (Heng and Jacob, 2019, Jacob et al. 2020) that help monitoring the practical convergence of the chains. In particular, we derive couplings of Metropolis-Rosenbluth-Teller-Hastings-type and Hamiltonian Monte Carlo-type algorithms on smooth manifolds and present some applications in the domain of likelihood-free inference. Joint work with Pierre E. Jacob and Robin J. Ryder.

Sampling Algorithms

Location: Room 1 - *Chair* : Dootika Vats

Speaker: Filippo Ascolani (Bocconi University)

Title: Complexity of Gibbs Samplers through Bayesian asymptotics

Abstract: Gibbs samplers are popular algorithms to approximate posterior distributions arising from Bayesian models. Despite their popularity and good empirical performances, however, there are still relatively few quantitative theoretical results on their scalability or lack thereof, e.g. much less than for gradient-based sampling methods.

We introduce a novel technique to analyse the asymptotic behaviour of mixing times of Gibbs Samplers, based on tools of Bayesian asymptotics. Our methodology applies to high-dimensional regimes where both number of datapoints and parameters increase, under random data-generating assumptions. This allows us to provide a fairly general framework to study the complexity of Gibbs samplers fitting complex hierarchical Bayesian models.

The methodology is applied to two-level hierarchical models with likelihoods belonging to a general class (e.g. Binomial or Normal with unknown variances) and exponential family priors. In this framework we are able to provide dimension-free convergence results for Gibbs Samplers

under mild conditions. Moreover, we provide appropriate bounds on the rate of convergence using spectral theory.

Speaker: Alberto Cabezas Gonzalez (Lancaster University)

Title: Transport Elliptical Slice Sampling

Abstract: We introduce a new framework for efficient sampling from complex probability distributions, using a combination of normalizing flows and elliptical slice sampling (Murray et al., 2010). The core idea is to learn a diffeomorphism, via normalizing flows, that maps the non-Gaussian structure of our target distribution to an approximately Gaussian distribution. We can then sample from our transformed distribution using the elliptical slice sampler, which is an efficient and tuning-free Markov chain Monte Carlo (MCMC) algorithm. The samples are then pulled back using an inverse normalizing flow to yield samples which approximate the stationary target distribution of interest. Our transformed elliptical slice sampler (TESS) is efficiently designed for modern computer architectures, where its adaptation mechanism utilizes parallel cores to rapidly run multiple Markov chains for only a few iterations. Numerical demonstrations show that TESS produce Monte Carlo samples from the target distribution with lower autocorrelation compared to non-transformed samplers. Additionally, assuming a sufficiently flexible diffeomorphism, TESS demonstrates significant improvements in efficiency when compared to gradient-based proposals designed to run on parallel computer architectures.

Speaker: Miika Kailas (University of Jyväskylä)

Title: Online mass matrix adaptation for Hamiltonian Monte Carlo

Abstract: We consider adaptive Markov Chain Monte Carlo methods within the Hamiltonian Monte Carlo (HMC) sampler and its dynamic variant, the No U-Turn Sampler (NUTS). In particular we study strategies for full-rank mass matrix adaptation and make two primary contributions. First, we study regularization strategies for online estimates relating to full-rank mass matrix adaptation in HMC and variants. Second and more importantly, we propose a novel adaptation target for the mass matrix. Contrasting with the usual choice of choosing the mass matrix as the inverse of (an estimate of) the covariance matrix of the target distribution, a global quantity, our alternative proposal is instead an average over local geometric quantities relating to the stability of discretized Hamiltonian dynamics. The proposed target and its estimators are computationally cheap and simple to implement, and our empirical studies show that the proposed adaptation strategies are applicable to challenging problems in hundreds of dimensions.

(joint work with M. Vihola)

Speaker: Federica Milinanni (KTH Royal Institute of Technology)

Title: Large Deviation Principle for the Metropolis-Hastings algorithm

Abstract: For MCMC methods, good performance measures for the convergence of the underlying Markov chains are essential. For instance, such performance measures can be used to

compare different MCMC methods, or to tune parameters within a given method. Examples of common tools for investigating convergence properties include the spectral gap, mixing times and functional inequalities (Poincaré, log-Sobolev). In recent years there has been an interest in studying the performance of MCMC methods using tools from large deviation theory, specifically the rate function associated with the empirical measure of an MCMC method. In this talk we analyze the standard Metropolis-Hastings (MH) algorithm from this perspective. We consider the MH algorithm for a target measure defined on a Polish space. We state a large deviation principle for the corresponding empirical measure, generalising previous results for the MH algorithm on a finite state space, and we illustrate in some examples how the rate function depends on parameters of the method (in particular, parameters in the proposal distribution).

Composable, Graphical & Network Models

Location: Room 2 - *Chair* : Antonietta Mira

Speaker: Junpeng Lao (Google)

Title: A Functional Programming Approach to Composable Bayesian Workflow

Abstract: Bayesian modeling in practice is an iterative process, in which a practitioner implicitly or explicitly follows the Bayesian workflow (Gelman et al 2020) to build models and inferences that are closest to the "reality" within the computational constraints. A composable model building capability is often desired as it makes developing bigger and more complex Bayesian models easier: for example, changing the priors of a collection of random variables. Moreover, a composable approach could enable more flexibility in constructing inferences that optimize for local model structure, thus have the opportunity to improve inference quality overall, as compared to the general inference methods a statistical package usually offers (e.g., NUTS with different schemes of adaptation). In this talk, I will explain how adopting a functional programming perspective benefits the development of composable Bayesian modeling and programmable inference, with example using TensorFlow Probability on JAX (for the modeling part) and Blackjax (for the inference part).

Speaker: Jack Jewson (Universitat Pompeu Fabra)

Title: Graphical model inference with external network data

Abstract: A frequent challenge when using graphical models in applications is that the sample size is limited relative to the number of parameters to be learned. Our motivation stems from applications where one has external data, in the form of networks between variables, that provides valuable information to help improve inference. Specifically, we depict the relation between COVID19 and social and geographical network data, and between stock market and economic and policy networks extracted from text data. We propose a graphical LASSO

framework where likelihood penalties are guided by the external network data. We also propose a spike-and-slab prior framework that depicts how partial correlations depend on the networks, which helps interpret the fitted graphical model. We develop computational schemes and software implementations in R and probabilistic programming languages. Our applications show how one may significantly improve interpretation, statistical accuracy, and out-of-sample prediction, in some instances using significantly sparser graphical models than would otherwise be necessary.

Speaker: Alejandra Avalos Pacheco (Vienna University of Technology (TU Wien))

Title: Bayesian Inference of Multiple Ising Models for Heterogeneous Data

Abstract: Multiple Ising models can be used to model the heterogeneity induced in a set of binary variables by external factors. These factors may influence the joint dependence relationships represented by a set of graphs across different groups. This talk presents the inference for this class of models and proposes a Bayesian methodology based on a Markov Random Field prior for the multiple graph setting. Such prior enables the borrowing of strength across the different groups to encourage common edges when supported by the data. Sparse inducing spike-and-slab priors are employed on the parameters that measure graph similarities to learn which subgroups have a shared graph structure. Two Bayesian approaches are developed for the inference of multiple Ising models with special focus on model selection: (i) a Fully Bayesian method for low-dimensional graphs based on conjugate priors specified with respect to the exact likelihood, and (ii) an Approximate Bayesian method based on a quasi-likelihood approach for high-dimensional graphs where the normalization constant required in the exact method is computationally intractable. The performance of the proposed methods are studied and compared with competing approaches through an extensive simulation study. Both inferential strategies are employed for the analysis of data resulting from two public opinion studies in US. The first one analyzes the confidence in political institutions in different groups divided by the time users spent on web pages. The second one studies the opinion on public spending in diverse inter-generational groups.

Speaker: Olivier Zahm (Inria)

Title: Gradient-based data and parameter dimension reduction for Bayesian models

Abstract: We consider the problem of reducing jointly the dimensions of parameters and data in non-Gaussian Bayesian inference problems. Our goal is to identify an informed subspace of the parameters and an informative subspace of the data so that a high-dimensional inference problem can be approximately reformulated in low-to-moderate dimensions, thereby improving the computational efficiency of many inference techniques. To do so, we exploit gradient evaluations of the log-likelihood function. Furthermore, we use an information-theoretic analysis to derive a bound on the posterior error due to parameter and data dimension reduction. This bound relies on logarithmic Sobolev inequalities, and it reveals the appropriate dimensions of the reduced variables. We compare our method with classical dimension reduction techniques, such as principal component analysis and canonical correlation analysis, on applications ranging from mechanics to image processing.

Particle Filters, Importance Sampling & GIbbs

Location: Room 3 and Online - *Chair* : Christophe Andrieu

Speaker: Matti Vihola (University of Jyväskylä)

Title: Conditional particle filters with bridge backward sampling

Abstract: The performance of the conditional particle filter (CPF) with backward sampling is often impressive even with long data records. Two known exceptions are when the observations are weakly informative and when the dynamic model is slowly mixing. These are both present when sampling from finely time-discretised continuous-time path integral models, but can occur with hidden Markov models too. Multinomial resampling, which is commonly employed in the (backward sampling) CPF, resamples excessively for weakly informative observations and thereby introduces extra variance. A slowly mixing dynamic model renders the backward sampling step ineffective. We detail two conditional resampling strategies suitable for the weakly informative regime: the so-called killing' resampling and the systematic resampling with mean partial order. To avoid the degeneracy issue of backward sampling, we introduce a generalisation that involves backward sampling with an auxiliarybridging' CPF step, which is parameterised by a blocking sequence. We present practical tuning strategies for choosing an appropriate blocking. Our experiments demonstrate that the CPF with a suitable resampling and the developed 'bridge backward sampling' can lead to substantial efficiency gains in the weakly informative regime.

Based on arXiv:2205.13898, joint work with S. Karppinen (Jyväskylä) and S. S. Singh (Cambridge)

Speaker: Victor Elvira (University of Edinburgh)

Title: Massively Recycled Importance Sampling

Abstract: In the context of Bayesian inference, importance sampling (IS) methods are broadly used to approximate posterior distributions and related moments. In its standard approach, samples are simulated from a single-proposal distribution and weighted properly. However, since the IS performance depends on the mismatch between the targeted and the proposal densities, two strategies are often used. First, in multiple importance sampling (MIS), several proposals are employed. Second, in adaptive IS (AIS), the proposals 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.

Speaker: Raiha Browning (University of Warwick)

Title: A Gibbs sampler for flexible estimation of the temporal excitation pattern of discrete-time self-exciting processes

Abstract: Hawkes processes are a self-exciting stochastic process, whereby past events increase the probability of future events occurring. A key feature of these processes is the conditional intensity function, (t|H(t−1)), where H(t−1) is the history of the process up to time t − 1. (t|H(t−1)) is comprised of two components: a baseline rate, representing independent events, and a self-exciting term that describes the self-excitation. Most standard models of Hawkes processes rely on a parametric form for the self-exciting term of the intensity function, referred to as the triggering kernel, which describes the influence of past events. This is likely to be insufficient to capture the true excitation pattern, particularly for complex data. In this work we present a Bayesian nonparametric approach to modelling the triggering kernel for discrete-time Hawkes processes, such that it takes the form of any step function since the location and heights of each step are unknown. This allows for significantly more flexibility than a parametric form. The proposed model exploits the branching structure representation of the Hawkes process, which describes the parent and offspring relationship between the events in the process through time. This allows the likelihood to be re-expressed to exploit conjugacy in the resulting posterior distribution. To estimate the step function, we propose a novel clustering algorithm based on the Chinese restaurant process, thus determining the locations of each change point in the step function.

Lightning Talks

Parallel III

16.30 — 18.00, Wednesday 15 March

note: each lightning talk will be allocated 3 minutes (a single slide)

Location: Auditorium and Online - *Chair* : Christian Robert

Posters

18.00 — 20.00, Wednesday 15 March

Presenter: **Adam Howes** (Imperial College London)

Title: Integrated nested Laplace approximations for extended latent Gaussian models, with application to the Naomi HIV model

Abstract: Naomi (Eaton et al, 2021) is a spatial evidence synthesis model used to produce district-level HIV epidemic indicators in sub-Saharan Africa. Multiple outcomes of interest, including HIV prevalence, HIV incidence and treatment coverage are jointly modelled using both household survey data and routinely reported health system data. The model is provided as a tool for countries to input their data to and generate estimates. In this setting, computationally intensive inference methods like MCMC are impractical. We propose a new inference method which combines the simplified integrated nested Laplace approximation approach of Wood (2020) with adaptive Gauss-Hermite quadrature to enable fast and accurate inference for Naomi and other extended latent Gaussian models. Using data from Malawi, our method provides substantially more accurate inferences than the empirical Bayes Gaussian approximation approach used currently, and is comparable to Hamiltonian Monte Carlo with the No-U-Turn sampler. By extending the aghq package (Stringer, 2021) we facilitate flexible and easy use of our method when provided a TMB C++ template for the log-posterior.

Presenter: **Alberto Bassi** (ETH Zürich and Eawag)

Title: Metadynamics for enhanced high-dimensional Bayesian inference

Abstract: Bayesian inference for models with a large number of unobserved (latent) variables poses a formidable numerical challenge. One of the most effective solutions is to use Hamiltonian Monte Carlo on the product space of model parameters and latent variables. However, posteriors of large numbers of coupled variables tend to be multi-modal. This is akin to the problem of multi-modal free energy landscapes encountered in chemical physics applications, which is solved by means of biased Molecular Dynamics simulations such as Metadynamics, see [Bussi and Laio, 2020]. Here, we show the usefulness of Metadynamics for Bayesan inference with hidden Markov models.

Presenter: **Alessandro Mastrototaro** (KTH Royal Institute of Technology)

Title: Adaptive online variance estimation in particle filters: the ALVar estimator

Abstract: We present a new approach—the ALVar estimator—to estimation of asymptotic variance in sequential Monte Carlo methods, or, particle filters. The method, which adjusts adaptively the lag of the estimator proposed in [Olsson, J. and Douc, R. (2019). Numerically stable online estimation of variance in particle filters. Bernoulli, 25(2), pp. 1504–1535] applies to very general distribution flows and particle filters, including auxiliary particle filters with adaptive resampling. The algorithm operates entirely online, in the sense that it is able to monitor the variance of the particle filter in real time and with, on the average, constant computational complexity and memory requirements per iteration. Crucially, it does not require the calibration of any algorithmic parameter. Estimating the variance only on the basis of the genealogy of the propagated particle cloud, without additional simulations, the routine requires only minor code additions to the underlying particle algorithm. Finally, we prove that the ALVar estimator is consistent for the true asymptotic variance as the number of particles tends to infinity and illustrate numerically its superiority to existing approaches. (This is a joint work with Jimmy Olsson, KTH. https://arxiv.org/abs/2207.09590)

Presenter: **Alicia Gill** (University of Warwick)

Title: Bayesian Inference of Reproduction Number from Genomic and Epidemic Data using MCMC Methods

Abstract: Typically, reproduction number R(t) is inferred using only epidemic data, such as prevalence per day. However, prevalence data is often noisy and partially observed, and it can be difficult to identify whether you have observed many cases of a small epidemic or few cases of a large epidemic. Genomic data is therefore increasingly being used to understand infectious disease epidemiology, and inference methods incorporating both genomic and epidemiological information are an active area of research.

We use Markov chain Monte Carlo methods to infer parameters of the epidemic using both a dated phylogeny and partial prevalence data to improve inference compared with using only one source of information. To do this, we have implemented a sequential Monte Carlo algorithm to infer the latent unobserved epidemic, which is then used to infer the reproduction number as it varies through time. We then analyse the performance of this approach using simulated data.

Presenter: **Andrew Chin** (Johns Hopkins University)

Title: Hamiltonianizing a piecewise deterministic Markov process: a bouncy particle sampler with "inertia"

Abstract: The bouncy particle sampler is among the most prominent examples of piecewise deterministic Markov process samplers, a state-of-the-art paradigm in Bayesian computation. Inspired by recent connections to the Hamiltonian Monte Carlo paradigm, we present a Monte Carlo algorithm intimately related to the bouncy particle sampler but relying on Hamiltonian-like deterministic dynamics. The dynamics generate a piecewise linear trajectory similar to the bouncy particle sampler's. However, changes in its velocity occur deterministically in the manner of Hamiltonian dynamics, dictated by the auxiliary "inertia" parameter we introduce. We show that the proposed dynamics, while technically non-Hamiltonian, are reversible and volume-preserving and thus constitute a valid Metropolis proposal mechanism. We further establish that the dynamics, when combined with periodic refreshment of the inertia parameter, converge to the bouncy particle sampler in the limit of increasingly frequent refreshment. The dynamics can be simulated exactly on log-concave target distributions, easily accommodate parameter constraints, and require minimal tuning, yielding an efficient rejection-free sampling algorithm on a range of target distributions. We call this algorithm the Hamiltonian bouncy particle sampler. We demonstrate the algorithm's competitive performance on a variety of synthetic and real-data target distributions in high-dimensional spaces.

Presenter: **Andrew Johnson** (Aalto University)

Title: Practical Bayesian Modelling of Correlated Non-Normal Outcomes with Copulas

Abstract: The benefits of Bayesian partial pooling with hierarchical models, especially with small data sizes, are well-established. These benefits are also applicable in the multivariate case (i.e., observations clustered within individuals). However, modelling outcomes as correlated is often a non-trivial task if they do not all follow a continuous multivariate distribution. This can be seen in clinical trial settings where participants are assessed on a battery of measures. Commonly assessed outcomes would include frequency of symptoms, presence of a diagnosis, and self-rated symptom severity.

An accessible and powerful approach for the Bayesian modelling of these mixed-type multivariate outcomes is the use of discrete-data augmentation with copula models. Copula models allow for the univariate distributions of outcomes to modelled separately from the dependence them. This enables multivariate models with arbitrary outcome distributions.

We demonstrate that the use of Gaussian and Student-T copulas with correlated Gaussian, Poisson, and Binomial outcomes results in reduced bias in estimated parameters as well as increased predictive performance, especially with smaller sample sizes. Additionally, we demonstrate that these copula models are highly accessible and easy to implement using our ex-

tensions to the brms R package – allowing this dependence to be modelled using only one additional line of code.

Presenter: **Anna Elisabeth Riha** (Aalto University)

Title: Safe iterative model building and a multiverse of Bayesian models

Abstract: When using statistical models and data to analyse a phenomenon of interest, it is common to arrive at an explicit or implicit set of candidate models. This pre-selection of models can easily lead to flawed conclusions. Multiverse analysis creates multiple models at once based on combinations of sensible modelling choices with the aim of increasing transparency and uncertainty communication in model building. Models in a multiverse can appear equally plausible without additional assessment of the validity and relevance of models. This can overwhelm and lead to misinterpretation, especially in scenarios with large numbers of models. Starting from multiverse analysis and advances in Bayesian workflow, this work provides tools to support and evaluate iterative model building in scenarios where the true data generating process is unknown. These tools allow us to identify smaller sets of useful and well-specified models that lead to robust conclusions and enable post hoc evaluation of selected models. Specifically, we introduce filtering and clustering to multiverse analysis to assess the validity and connections of models and ultimately arrive at a curated set of models. To further illustrate the suggested procedures, we summarise different scenarios encountered when analysing treatment effect data and show how our tools support model evaluation.

Presenter: **Anton Beloconi** (Swiss TPH)

Title: Malaria, climate variability and the effect of interventions: modelling transmission dynamics

Abstract: Assessment of the relative impact of climate change on malaria dynamics is a complex problem. Climate is a well-known factor that plays a crucial role in driving malaria outbreaks in epidemic transmission areas. However, its influence in endemic environments with intensive malaria control interventions is not fully understood, mainly due to the scarcity of high-quality, long-term malaria data. The demographic surveillance systems in Africa offer unique platforms for quantifying the relative effects of weather variability on the burden of malaria. Here, using a process-based stochastic transmission model, we show that in the lowlands of malaria endemic western Kenya, variations in climatic factors played a key role in driving malaria incidence during 2008–2019, despite high bed net coverage and use among the population. Bayesian computation and inference based on pMCMC were compared to iterated filtering. The model accounts for the main mechanisms related to malaria dynamics, including immunity, infectivity,

and human migration, and opens the possibility to forecast malaria in endemic regions, taking into account the interaction between future climatic conditions and intervention scenarios.

Presenter: **Anton Stratmann** (Forschungszentrum Jülich)

Title: Approximate Bayesian Inference on Polytopes

Abstract: Network models arguably make up the most important model class in systems biology. To calibrate these models, up to hundreds of parameters need to be inferred from noisy data using non-linear algebraic or ordinary differential equation (AE/ODE) systems with thousands of dimensions. Peculiar to network models is that their parameters are constrained to a convex polytope. For parameter inference, Markov chain Monte Carlo (MCMC) approaches, tailored to handle polytopes efficiently, are the state-of-the-art. In the omics age, mechanistic knowledge of biochemical systems continually increases and network models grow in size. Therewith, the solution of the underlying AE/ODE systems become even more costly and the number of parameters increases, driving MCMC to the limit of computational feasibility. As MCMC alternative, approximate Bayesian methods promise a lower computational cost, at the expense of precision. Prominent approximate methods are, however, rarely studied in the context of linearly constrained parameter spaces. We here compare the performance of MCMC with two approximate methods - a previously developed analytical approximation approach, using expectation propagation, and a new variational inference method, using expectation maximization - from the perspective of approximation quality versus runtime. With models of different complexity, we report computational advances, discuss lessons learned, and point to open questions.

Presenter: **Antonio Di Noia** (ETH Zurich and USI Lugano)

Title: Detection of model identifiability issues testing the Intrinsic Dimension of MCMC posterior samples

Abstract: We propose a test methodology to assess model identifiability. To this aim, after running a MCMC simulation targeting the parameter posterior distribution, we test if the path of the Markov chain has Intrinsic Dimension equal to the number of model parameters. If the null hypothesis is not rejected we conclude that the model is well identified, meaning that the posterior distribution has full support. As a first step we introduce convenient Intrinsic Dimension estimators, a well-known methodological field in literature. Every estimation approach relies on the hypothesis that the underlying generating process is governed by a probability law whose support is a low-dimensional sub-manifold. Under reasonably weak assumptions on the underlying stochastic process, it is possible to derive distributional results on ratios of

nearest-neighbor distances among points. Most importantly, one can show that these distributions are parametrized by the Intrinsic Dimension leading to the introduction of a broad class of estimators which allow a straightforward uncertainty quantification and ensure available asymptotic properties. As a natural extension an asymptotic test is also proposed and explicitly adopted to detect identifiability issues in MCMC posterior approximations, although its range of applicability goes far beyond model identifiability assessment.

Presenter: **Antonio Peruzzi** (Ca'Foscari University of Venice)

Title: Media Bias and Polarization through the Lens of a Markov Switching Latent Space Model

Abstract: The news consumption landscape has changed quickly in recent years. News outlets are now more than ever incentivized to provide their audience with slanted news, while the intrinsic homophilic nature of online social media may exacerbate polarized opinions. In this work, we propose a new dynamic Latent Space (LS) model for time-varying online audience duplication networks exploiting social media content to provide a way of measuring news outlets' political leaning and polarization regimes. Our model, estimated within a Bayesian framework, endows latent coordinates with a proper interpretation in terms of political ideology via an observable proxy. The work has the twofold aim of making advancements concerning LS models and the estimation of media bias and social-media polarization. The developed model is applied to a dataset on the online activity of national and local news outlets from four European countries (France, Germany, Italy, and Spain). We find evidence of a strong positive correlation between our media slant measure and a well-grounded external source of media bias (72% correlation with the PEW Research index). We further provide insights on the in-platform polarization regimes across the four countries considered.

Presenter: **Antti Kero** (Sodankylä Geophysical Observatory, University of Oulu)

Title: HPC-approach to Ionospheric Situational Awareness (HISSA)

Abstract: The ionised part of the upper atmosphere – the ionosphere – is a region where the neutral atmosphere and near-Earth space meet. Its dynamics are controlled both by magnetospheric forcing from above and by neutral atmospheric processes from below. In HISSA project, we develop a HPC approach to the bayesian inference data-analysis for the upcoming EISCAT_3D incoherent scatter radar, also utilising an independent remote station, KAIRA, allowing for unprecedented 3D volumetric imaging of the ionosphere.

Presenter: **Augusto Magalhães** (Aalto University)

Title: Estimation of pattern formation in stochastic reaction-diffusion systems with the Block Particle Filter

Abstract: Reaction-diffusion systems have attracted great interest as they form a basis for understanding pattern-formation dynamics in theoretical biology. In this work, we consider a differential description of the evolution of the state of a reaction-diffusion system under environmental fluctuations. We are interested in estimating the state of the system when only partial observations are available. To describe how observations and states are related, we combine multiplicative noise-driven dynamics with an observation model. More specifically, we assume that the observations are subjected to error in the form of additive noise. We focus on the task of reconstructing the state of a Belousov-Zhabotinskii chemical reaction. We simulate a reaction conducted in a quasi-two-dimensional physical domain, such as on the surface of a Petri dish. We aim at reconstructing the emerging chemical patterns based on noisy spectral observations. For this task, we consider a finite difference representation on the physical domain, where nodes are chosen according to observation sites. We approximate the solution to this state estimation problem with the Block Particle Filter, a Sequential Monte Carlo method capable to address the associated high-dimensionality of this state-space representation. Keywords: state estimation, stochastic partial-differential equations, high-dimensional particle filtering, pattern formation, Belousov-Zhabotinskii reaction model.

Presenter: **Ayush Bharti** (Aalto University)

Title: Approximate Bayesian Computation with Domain Expert in the Loop

Abstract: Approximate Bayesian computation (ABC) is a popular likelihood-free inference method for models with intractable likelihood functions. As ABC methods usually rely on comparing summary statistics of observed and simulated data, the choice of the statistics is crucial. This choice involves a trade-off between loss of information and dimensionality reduction, and is often determined based on domain knowledge. However, handcrafting and selecting suitable statistics is a laborious task involving multiple trial-and-error steps. In this work, we introduce an active learning method for ABC statistics selection which reduces the domain expert's work considerably. By involving the experts, we are able to handle misspecified models, unlike the existing dimension reduction methods. Moreover, empirical results show better posterior estimates than with existing methods, when the simulation budget is limited.

Presenter: **Bernardo Williams Moreno Sánchez** (University of Helsinki)

Title: Geometric No-U-Turn-Sampler

Abstract: Geometric Markov Chain Monte Carlo algorithms utilize the underlying geometry of the distribution for improved exploration and convergence in sampling, by using a metric tensor that changes at each position. Unfortunately, the practical algorithms are slow due to the high computational cost of the metric operations. Recently, Hartmann et al. (2022) introduced an efficient metric that removes the need for explicit metric tensor inversion and showed that it works well with Lagrangian Monte Carlo (LMC). However, their sampler still depends on user-given parameters for step lengths and integration time.

Building on this new metric, we present the geometric No-U-Turn Sampler (gNUTS) that extends the concept of the NUTS algorithm to Riemannian manifolds. We introduce a geodesic stopping criterion that tests for U-turns along trajectories on the manifold, and hence can automatically select the user-given parameters.

We also consider further improvements for computational speed by approximating Hessian-vector products with linear cost. The resulting algorithm is a significant step towards more generally usable geometric MCMC algorithms, combining ease of use of NUTS as general inference engine for probabilistic programming with efficient use of Riemannian geometry.

Presenter: **Breeshey Roskams-Hieter** (University of Edinburgh)

Title: Leveraging variational autoencoders for multiple data imputation

Abstract: Missing data persists as a major barrier to data analysis across numerous applications. Recently, deep generative models have been used for imputation of missing data, motivated by their ability to capture highly non-linear and complex rela- tionships in the data. In this work, we investigate the ability of deep models, namely variational autoencoders (VAEs), to account for uncertainty in missing data through multiple imputation strategies. We find that VAEs provide poor em- pirical coverage of missing data, with underestimation and overconfident impu- tations, particularly for more extreme missing data values. To overcome this, we employ -VAEs, which viewed from a generalized Bayes framework, provide ro- bustness to model misspecification. Assigning a good value of is critical for uncertainty calibration and we demonstrate how this can be achieved using cross- validation. In downstream tasks, we show how multiple imputation with -VAEs can avoid false discoveries that arise as artefacts of imputation.

Presenter: **Brieuc Lehmann** (University College London)

Title: A Predictive Approach to Bayesian Nonparametric Survival Analysis

Abstract: Bayesian nonparametric methods are a popular choice for analysing survival data due to their ability to flexibly model the distribution of survival times. These methods typically employ a nonparametric prior on the survival function that is conjugate with respect to right-censored data. Eliciting these priors, particularly in the presence of covariates, can be challenging and inference typically relies on computationally intensive Markov chain Monte Carlo schemes. Here, we build on recent work that recasts Bayesian inference as assigning a predictive distribution on the unseen values of a population conditional on the observed samples, thus avoiding the need to specify a complex prior. We describe a copula-based predictive update which admits a scalable sequential importance sampling algorithm to perform inference that properly accounts for right-censoring. We provide theoretical justification through an extension of Doob's consistency theorem and illustrate the method on a number of simulated and real data sets, including an example with covariates. Our approach enables analysts to perform Bayesian nonparametric inference through only the specification of a predictive distribution. Joint work with Dr Edwin Fong.

Presenter: **Charles Margossian** (Flatiron Institute)

Title: Assessing the convergence of Markov chains Monte Carlo when running many chains

Abstract: The growing availability of hardware accelerators such as GPUs has generated interest in Markov chains Monte Carlo (MCMC) workflows which run a large number of chains in parallel. Each chain still needs to forget its initial state but the subsequent sampling phase can be almost arbitrarily short. To determine if the resulting short chains are reliable, we need to assess how close the Markov chains are to convergence to their stationary distribution. A battle-tested diagnostic is the $\widehat{R}$ statistic, where one appraises convergence by checking that $\widehat{R} < 1 + \epsilon$ for some small ϵ. Unfortunately, convergence of $\widehat{R}$ to 1 requires a large ESS per chain, a condition which is incompatible with the many short chains regime. We present a generalization, nested $\widehat{R}$, which offers a useful diagnostic for many short chains, provided we meet certain initialization conditions. Our approach also opens the door to more efficient MCMC workflows in which the warmup length can be chosen adaptively rather than preset, as is current standard practice.

Presenter: **Charly Andral** (Université Paris Dauphine)

Title: The Importance Markov Chain

Abstract: The Importance Markov chain is a new algorithm bridging the gap between rejection sampling and importance sampling, moving from one to the other through a tuning parameter. Based on a modified sample of an instrumental Markov chain targeting an instrumental

distribution (typically with a MCMC kernel), the Importance Markov chain aims to construct an extended Markov chain where the marginal distribution of the first component converges to the target distribution. We obtain geometric ergodicity for this extended kernel, under mild assumptions on the instrumental kernel. For example, when targeting a multimodal distribution, the instrumental distribution can be chosen as a tempered version of the target which allows the algorithm to explore its modes more efficiently. A law of large numbers and a central limit theorem are also obtained. Computationally, the algorithm is easy to implement and preexisting librairies can be used to sample from the instrumental distribution.

Presenter: **Conor Osborne** (University of Edinburgh)

Title: Convergence rates of non-stationary and deep Gaussian process regression

Abstract: Gaussian processes have proved to be a powerful and flexible tool in the reconstruction of functions given a set of known data points, with applications in machine learning, optimisation and data assimilation. However, they can be limited when the functions being reconstructed are of a non-stationary or anisotropic nature. Deep Gaussian processes, constructed using a hierarchical process where the inputs to a Gaussian process are themselves Gaussian processes, aim to give a more flexible approach to function reconstruction. We look at convergence rates of these deep Gaussian processes in terms of the number of known data points. We also show that deep Gaussian process regression achieves considerably better results than standard Gaussian process regression when reconstructing non-stationary and anisotropic functions.

Presenter: **Constandina Koki** (Warwick University)

Title: Bayesian computational analysis of cell division dynamics

Abstract: Millions of cell divisions take place in our body on a daily basis, divisions that have to accurately replicate the cell's chromosomes. In particular newly duplicated chromosomes – called sister chromatids – must be segregated into daughter cells with high fidelity. Errors in this process, or mis-segregation is a leading cause of cancer. Segregation is a complex mechanical process requiring sisters to correctly attach to the bipolar mechanical spindle machine, thereby moving the chromosomes. Recent innovations in imaging allow 3D high resolution tracking of sister chromatids in living human cells throughout cell division (40 mins). We deploy Bayesian computational techniques to analyse these time-series by fitting biophysically motivated models to infer attachment states. We utilise and compare models with hidden discrete mechanical states and a continuum. Our analysis reveals that sister dynamics are significantly asymmetric in 20% of pairs, likely resulting from asymmetric attachments. To

deal with the missing data (detection loss during tracking) we utilise a "reduced'' likelihood that is based only on the actual data. Model inference is carried out in STAN.

Presenter: **Cristian Castiglione** (University of Padova)

Title: Bayesian Non-Conjugate Regression via Variational Belief Updating

Abstract: A new variational algorithm is presented as a flexible tool for approximating the general posterior distribution of Bayesian models that combine subjective prior beliefs with an empirical risk function. Particular attention is delivered to regression and classification models linking data and parameters through a continuous convex loss function and a linear predictor. Many remarkable examples belonging to this class are of particular interest for statistical applications, such as generalized linear models, support vector machines, quantile and expectile regression. The proposed iterative procedure lies in the family of semiparametric variational Bayes and enjoys closed–form updating formulas along with an efficient integration of the evidence lower bound. Neither conjugacy nor elaborate data augmentation strategies are required. Structured prior distributions, e.g., nested random effects, spatial or temporal processes, or inducing shrinkage and sparsity priors, can be easily accommodated into such a framework without additional effort since the modularity of mean field variational Bayes is preserved. The properties of the algorithm are then assessed through a simulation study, where the proposed method is compared with Markov chain Monte Carlo and conjugate mean field variational Bayes in terms of posterior approximation accuracy, prediction error and computational runtime.

Presenter: **Daniel Winkler** (Vienna University of Economics and Business)

Title: Arianna: A Domain-Specific Language for MCMC Algorithms

Abstract: The development of MCMC algorithms involves an implementation in a mathematical language, in addition to one in a programming language. Often, further versions are written in faster, lower-level languages (e.g., C++). This development cycle comes with obvious drawbacks (e.g., multiple manual implementations).

We contribute "Arianna" (honoring Rosenbluth), a system for MCMC algorithms based on domain-specific languages (DSLs; special-purpose languages like SQL). Our DSL allows the implementation of algorithms using mathematical notation, which can be translated to different programming languages.

Our DSL constitutes a concise, easily comprehensible, and extensible yet powerful system to streamline MCMC development. We highlight different non-trivial MCMC algorithms, including global-local shrinkage priors for factor models.

While researchers retain complete control over the algorithm (no black box), highly optimized backends (e.g., GPU) can be provided by domain specialists. In our presentation, we will highlight several backends (Rust for speed, Javascript for in-browser execution, among others).

Presenter: **David Edwards** (Virginia Tech)

Title: Multi-Set Model Averager and Predictor

Abstract: Model-based prediction and forecasting is an inherently challenging task, especially in a high-dimensional setting. Predictions based on the average of several models routinely outperform predictions based on a single model. With this in mind, we look to extend the Multi-Set Sampler to include a model averaging feature based on indicator variables. The Multi-Set Sampler, developed by S. Leman, is an MCMC technique that efficiently explores model space with multi-modal posterior distributions or non-continuous nuisance parameters. Previously the Multi-Set Sampler was shown to perform well on high-dimensional model selection problems by creating indicator variables for each covariate in the full model. In this way, model selection can be accomplished by computing the posterior probability of each model and selecting the model with the highest probability. A natural extension of this Multi-Set Model Selector (MSMS) is to average predictions over models based on their posterior probabilities, thus allowing models with a higher posterior probability to influence the prediction more than simply equally weighting the prediction of all models. This presentation will introduce and demonstrate the Multi-Set Model Averager and Predictor and compare its performance to similar prediction models.

Presenter: **David Helekal** (University of Warwick)

Title: Bayesian Inference of Fitness Cost of Antimicrobial Resistance from Genomic Data

Abstract: Antimicrobial resistance of bacterial pathogens is a pressing public health problem. Resistance has a clear benefit for pathogen transmission in the presence of antimicrobial use. Increased antimicrobial use therefore leads to an increased net value of adaptions that confer resistance to it. However, often adaptations that confer resistance to antimicrobials come at a fitness cost in the absence of sufficient antimicrobial use. Quantifying the fitness cost of antimicrobial resistance as a function antimicrobial usage through time, along with the uncertainty associated is essential to understanding the emergence and dispersal of antimicrobial resistance as well for public health decision making. Using genomic data for this purpose has several advantages over traditional data. One such advantage is much greater resolution, avoiding averaging over lineages with distinct fitness parameters. Another advantage is possibility of using the coalescent process as an observational process, as the coalescent process is less

sensitive to sampling considerations. It however also comes with additional challenges, such as the complex interplay between different lineages being only partially observed. We investigate how fitness cost and benefit of antimicrobial resistance can be inferred from genomic data in a Bayesian framework.

Presenter: **El Mahdi Khribch** (ESSEC Business School)

Title: On mixing Times of Metropolized Algorithm with Optimization Step (MAO)

Abstract: In this paper, we consider the problem of sampling from a class of d dimensional distributions with thin tails supported and make two primary contributions. First, we propose a new algorithm capable of working in regimes where the Metropolis-adjusted Langevin algorithm (MALA) is not geometrically ergodic. In addition, we derive upper bounds on the mixing time of the Markov chain generated by the said algorithm. We then compare the empirical performance of our proposed algorithm to other sampling methods (MALA, HMC, and RWMH).

Presenter: **Estevão Batista do Prado** (Lancaster University)

Title: Bayesian additive regression trees with model trees

Abstract: Bayesian additive regression trees (BART) is a tree-based machine learning method that has been successfully applied to regression and classification problems. BART assumes regularisation priors on a set of trees that work as weak learners and is very flexible for predicting in the presence of non-linearity and high-order interactions. In this paper, we introduce an extension of BART, called model trees BART (MOTR-BART), that considers piecewise linear functions at node levels instead of piecewise constants. In MOTR-BART, rather than having a unique value at node level for the prediction, a linear predictor is estimated considering the covariates that have been used as the split variables in the corresponding tree. In our approach, local linearities are captured more efficiently and fewer trees are required to achieve equal or better performance than BART. Via simulation studies and real data applications, we compare MOTR-BART to its main competitors. R code for MOTR-BART implementation is available at https://github.com/ebprado/MOTR-BART.

Presenter: **Fernando Zepeda** (University of Warwick)

Title: Non-revesibility and Tempering

Abstract: Okabe et al's (2001) tempering methodology incorporating some degree of non-reversible dynamics has been recovered and expanded by Syed et al (2022), showing good properties. More explicit non-reversibility has also been proposed by Sutton et al (2022). We incorporate these types of techniques into recently proposed methodologies within the tempering literature and exemplify their performance.

Presenter: **Fiona Turner** (King's College London)

Title: Sensitivity of land ice projections to experimental design: the effects of input parameterisation on emulator performance

Abstract: The performance of a Gaussian process emulator is often determined by the quality of the experimental design used for the simulator in question. Building on Edwards et al. (2021), we build a series of emulators of ice sheet and glacier models to project land ice contributions to future global sea level rise this century. Our validation techniques show the performance of the emulators are strongly affected by the number of simulations available, as well as the sampling of input parameters. We wish to better understand how much of the uncertainty is represented by each model ensemble and the minimum number of models needed to encompass the majority of this uncertainty. We present probability density functions of each land ice region's contribution to sea level at 2100 and consider the most appropriate way to weight the projections from different models.

References [1] Edwards, Tamsin L., et al. ""Projected land ice contributions to twenty-first-century sea level rise."" Nature 593.7857 (2021): 74-82.

Presenter: **Hanlin Yu** (University of Helsinki)

Title: Geometric SG-MCMC Samplers for Bayesian Deep Learning

Abstract: Bayesian inference is increasingly being used also in deep learning, for quantifying uncertainty of large neural networks. Due to the large parameter space this needs efficient methods, such as Stochastic Gradient MCMC (SG-MCMC) algorithms which utilize stochastic gradients while still sample from the posterior distribution. These algorithms can be further improved by accounting for the curvature of the posterior, but the current best methods only do this separately for each element, corresponding to a diagonal approximation.

Building on the recently proposed Monge metric, we present a way to account for the local curvature jointly over all dimensions, using a position-dependent preconditioner matrix without

resorting to the diagonal approximation. Even though we use a full matrix preconditioner that depends on the current parameters and hence need to be re-computed during the algorithm, the approach remains computationally tractable. The preconditioner is estimated from first-order gradient information alone, and all of the computations can be carried out using element-wise operations. We explain how different SG-MCMC methods, including Stochastic Gradient Langevin Dynamics (SGLD) and Stochastic Gradient Hamiltonian Monte Carlo (SGHMC), can be implemented using the proposed preconditioner, opening up new opportunities for efficient Bayesian deep learning.

Presenter: **Harrison Wilde** (University of Warwick)

Title: MCMCTempering.jl: An Adaptable and Modular Approach to Overcoming Multi-Modality in Bayesian Modelling Problems

Abstract: Bayesian Neural Ordinary Differential Equations (BNODEs) are an emerging class of models that are adept at modelling physical and dynamical systems through machine learning assisted methods, without the need for an explicit definition of a system's underlying ODE. The adoption of a Bayesian learning paradigm provides uncertainty estimates for model parameters: a desirable property for a myriad of applications, such as climate or epidemiological models. However, BNODEs are intrinsically computationally complex and their posteriors can often exhibit severe multi-modalities, causing standard approaches for high-dimensional parameter inference such as No-U-Turn or HMC samplers to perform poorly or fail altogether. We apply several algorithms leveraging parallel tempering schemes and novel extensions to achieve state of the art sampling performance for this family of models. Our tempering methods can be applied widely via a generic "internal" MCMC kernel, and they are parallelisable with minimal additional overhead cost relative to sampling efficiency.

Presenter: **Houda Yaqine** (Bielefeld university)

Title: Estimation of time/age-dependent ROC curves using Pólya trees based models

Abstract: To determine whether a biomarker or a test will accurately identify diseased from healthy individuals, a receiver operating characteristic (ROC) curve is used. But the construction of such curves is challenging when the true status of the disease is unknown (non-gold standard data). The reason is the significant uncertainty regarding the distributions of measurements of healthy and diseased individuals. In this setting, the best way to estimate unknown distributions is to use non-parametric methods. Still, only a few have been developed, including Bayesian tools based on finite Pólya trees and Dirichlet processes. Moreover, these methods assume the test measurements to be fixed over time or across age groups, which

is rarely the case in practice. In fact, during the course of a disease, some individuals may respond or even evolve differently from the rest of the population. Consequently, distributions of measurements of healthy and diseased individuals display skewness, heterogeneity, and multimodality. As part of our research, we develop a nonparametric Bayesian model called a mixture of dependent finite Pólya trees (MDFPT) to estimate ROC curves based on time or age. Our model introduces more flexibility for different types of data and can also be extended to estimate ROC curves dependently on any categorical variables with levels.

Presenter: **Ilaria Pia** (University of Helsinki)

Title: Robust count regression models

Abstract: In Bayesian statistics, outlier observations can heavily affect posterior distributions, leading to biased predictive results. Robust models are based on the assumption that outliers are atypical manifestations of the underlying data-generating system and hence, should be kept in the data. Robustness is then reached by using flexible response ditributions. In continuous cases, adding heteroschedasticity to a Gaussian response process leads to a robust Student-t distribution. In count regression, the response generating process is commonly assumed to follow a Poisson distribution, which is heavily affected by outliers. Adding some Gamma distributed noise to the Poisson rate parameter leads to a Negative Binomial response which is still not robust against outliers. In our work, we present a novel heavy tailed distribution, the Student-t Poisson mixture derived by adding heteroschedasticity in the response generating process distribution, and assuming the logarithm of the varying component of the rate parameter follows a Student-t distribution. Preliminary results show how such distribution greatly overcomes Poisson, Negative Binomial, and other mixture distributions in handling the presence of outliers, both when the rate is independent of covariates and in a GLM framework, when the rate is expressed in terms of covariates.

Presenter: **Ilina Iliyanova Yozova** (University College London)

Title: Bayesian Causal Forests for Heterogeneous Treatment Effects Estimation from randomised and observational data

Abstract: The Bayesian Causal Forest (BCF) of Hahn et al. (2020) was developed to perform a nonparametric causal regression model using Bayesian Additive Regression Trees (Chipman et al. (2010)). BCF is specifically designed to estimate heterogeneous treatment effects using observational data, imposing regularisation directly on the treatment effect rather than the response of interest. BCF carefully teases apart the model into three distinct pieces: the prognostic effect, which is the influence of the covariates directly on the response, the treatment

effect, which represents the inhomogeneous effect of treatment on the outcome, and the traditional propensity score, which captures the treatment assignment mechanism. However, when using data from different sources, the treatment assignment mechanism might differ greatly between each dataset. Additionally, the prognostic and/or treatment effects, as well as the covariate profile of the data, may also be different. A notable example arises when combining randomised control trial (RCT) data with observational data: RCT data have (ideally) balanced propensity, often correspond to a narrower profile of covariates, and tend to include less measurement error than observational studies. We extend the BCF model formulation to account for different prognostic effects across datasets, using the Lalonde (1986) data as an illustrative example.

Presenter: **Imke Botha** (Queensland University of Technology)

Title: Adaptively selecting the mutation kernel in SMC^2

Abstract: Sequential Monte Carlo squared (SMC^2) methods give exact parameter inference of state-space models where the likelihood of the model parameters is unknown in closed form. SMC methods propagate a set of particles through a sequence of distributions using a combination of reweighting, resampling and mutation steps. In the Bayesian setting, this sequence often starts at the prior and ends at the posterior. SMC^2 is similar to particle Markov chain Monte Carlo (MCMC) methods in the sense that it replaces the intractable likelihood in the sequence of distributions being traversed with a particle filter estimator. As a result, particle MCMC methods are a natural choice to mutate the particles within SMC^2. We introduce a method that adaptively chooses between two particle MCMC algorithms for the mutation step: particle marginal Metropolis-Hastings and particle Gibbs. As the most efficient mutation kernel greatly depends on both the model and the target distribution, i.e. the current distribution in the sequence, adaptively switching between mutation kernels can greatly improve the performance of SMC^2. This is joint work with Chris Drovandi, Leah South and Robert Kohn.

Presenter: **Javier Enrique Aguilar Romero** (University of Stuttgart)

Title: Intuitive Joint Priors for Bayesian Linear Multilevel Models: The R2-D2-M2 prior

Abstract: The training of high-dimensional regression models on comparably sparse data is an important yet complicated topic, especially when there are many more model parameters than observations in the data. From a Bayesian perspective, inference in such cases can be achieved with the help of shrinkage prior distributions, at least for generalized linear models. However, real-world data usually possess multilevel structures, such as repeated measurements

or natural groupings of individuals, which existing shrinkage priors are not built to deal with. We generalize and extend one of these priors, the R2-D2 prior by Zhang et al. (2020), to linear multilevel models leading to what we call the R2-D2-M2 prior. The proposed prior enables both local and global shrinkage of the model parameters. It comes with interpretable hyperparameters, which we show to be intrinsically related to vital properties of the prior, such as rates of concentration around the origin, tail behavior, and amount of shrinkage the prior exerts. We offer guidelines on how to select the prior's hyperparameters by deriving shrinkage factors and measuring the effective number of non-zero model coefficients. Hence, the user can readily evaluate and interpret the amount of shrinkage implied by a specific choice of hyperparameters. Finally, we perform extensive experiments on simulated and real data, showing that our prior is well calibrated, has desirable global and local regularization properties and enables the reliable and interpretable estimation of much more complex Bayesian multilevel models than was previously possible.

Presenter: **Jiale Tao** (University of Nottingham)

Title: Bayesian detection and prediction of disruptive events from Twitter data

Abstract: The volume of tweets on Twitter is increasing exponentially, thus providing us with numerous opportunities for detecting the occurrence of major events in real-time. We develop a state-space model for detecting disruption on the National Railway in Great Britain in a timely fashion, by using the content and volume of tweets referring to delays and disturbance in the railway. A time-inhomogeneous Poisson process, $\lambda(t)$, is proposed to model the number of tweets whose time-dependent intensity function is parameterized such that it captures the observed periodic pattern in the data. A hidden Markov process that represents the state of the railway through time ('normal'/ 'abnormal') then modulates the Poisson process. We develop a computationally efficient MCMC algorithm to learn the parameters governing $\lambda(t)$ and infer the state of the railway through time by utilizing a Forward-Backward algorithm to efficiently update the unobserved process. We demonstrate through extensive simulation studies that (i) we can successfully recover the model's unknown parameters, (ii) predict the unobserved state with high accuracy framework our results are robust model misspecification. Finally, we illustrate via Bayesian filtering how to predict the future state of railway in real-time given the observed number of tweets.

Presenter: **Jonas Bauer** (Universität Bielefeld)

Title: Efficient sampling from high-dimensional Bayesian mixture models for scRNA data

Abstract: The roots of many diseases may be traced back to differences on the level of biological cells. Technological advances in the past decades allowed us to sequence ribonucleic acids for

single cells (scRNA-Seq) and enabled rich statistical analyses to distinguish healthy from diseased cells. However, such data sets consist of thousands of microbiological features which lead to high-dimensional parameter spaces. Common clustering approaches like Dirichlet-process mixture models suffer especially from the curse of dimensionality. Hence, the assessment of posterior distributions remains a computational bottleneck even with efficient samplers based on Hamiltonian dynamics. In our work, we discuss various concepts to reduce the computational effort of one popular gradient-based approach: the No-U-turn sampler. We focus on the trade-off between computational costs and chain convergence speed while using, for instance, random and systematic blocked gradient derivation, partial objective evaluation, and penalized gene-specific step lengths. Our findings aim to shed light on how well these popular performance-oriented approaches yield reliable sampling in this challenging biological context.

Presenter: **Jose Pablo Folch** (Imperial College London)

Title: SnAKe: Bayesian Optimization via Pathwise Exploration

Abstract: Bayesian Optimization is a very effective tool for optimizing expensive black-box functions. Inspired by applications developing and characterizing reaction chemistry using droplet microfluidic reactors, we consider a novel setting where the expense of evaluating the function can increase significantly when making large input changes between iterations. We further assume we are working asynchronously, meaning we have to select new queries before evaluating previous experiments. This paper and talk formalizes the problem and introduces 'Sequential Bayesian Optimization via Adaptive Connecting Samples' (SnAKe), which provides a solution by considering large batches of queries and preemptively building optimization paths that minimize input costs. We investigate some convergence properties and empirically show that the algorithm is able to achieve regret similar to classical Bayesian Optimization algorithms in both the synchronous and asynchronous settings, while reducing the input costs significantly. We show the method is robust to the choice of its single hyper-parameter and provide a parameter-free alternative.

Presenter: **Josefina Correa** (Massachusetts Institute of Technology)

Title: A State-Space Random Effects Point Process Model for Estimating Spike Rate Functions

Abstract: A common approach for analyzing spike train data in stimulus-response experiments is to estimate spike rates relative to the onset of the stimulus. These experiments typically

involve collecting repeated measurements from the same subject across several trials and sessions, and performing the same experiment across multiple subjects. Accounting for these different sources of variability is important in order to accurately estimate the population response. However, current spike rate estimation methods only allow for estimating trial-level, session-level or group-level spike rates, neglecting the variability between levels. We develop a state-space random effects point-process (SSREPP) model to estimate population and subject-level spike rate functions from multi-subject, multi-session and multi-trial spike train data. The statistical model allows for including stimulus information, and for quantifying the effect of the stimulus on the spiking propensity. Our approach provides goodness-of-fit assessment and a Monte Carlo algorithm for computing confidence intervals for population and subject-level spike rate functions. We assess the accuracy of our SSREPP model in simulation. We apply our SSREPP model to data from published stimulus-response experiments. We compare results obtained using an Empirical Bayes framework to those obtained using a fully Bayesian framework.

Presenter: **Joseph Marsh** (University of Nottingham)

Title: A Bayesian approach to integrate epidemiological and whole genome sequence data for analysing infectious disease outbreaks

Abstract: Advances in sequencing technology and the reduction in associated costs have enabled scientists to obtain highly detailed genomic data on disease-causing pathogens on a scale never seen before. Combining genomic data with traditional epidemiological data (e.g. incidence data) provides a unique opportunity to determine the actual transmission pathway of the pathogen through a population. Despite recent advances, existing approaches have their own limitations, such as simplifications to the underlying biological processes, arbitrary phenomenological models or approximations to the likelihood function, to name a few.

We present a modelling framework for integrating epidemiological and whole genome sequence data to overcome the above limitations where (i) we use the matrix of pairwise horizontal distances between sequences as a summary statistic for the genetic data and (ii) explicitly derive the joint probability distribution of pairwise genetic distances under the assumption of a mutation model. We develop computationally efficient data-augmentation MCMC algorithms to infer the transmission network and the unobserved pathogen distances at the time of transmission as well as the times of transmission. Finally, we demonstrate the performance of our framework on simulated data and also analyse an outbreak of S. aureus in an intensive care unit in Brighton during 2011-2012.

Presenter: **Karel Kaurila** (University of Helsinki)

Title: Model Uncertainty Quantification for Environmental Policy Planning

Abstract: Deterministic physical, biological and chemical simulators are widely used to study real world phenomena and to aid decision making. Hence, uncertainty quantification of such simulators poses an important computational challenge for modern statistics. We will present a method for calibrating a Finnish Coastal Nutrient Load (FICOS) model, which is used for studying bio-geo-chemical processes in the Finnish Archipelago sea and to predict water quality under alternative management scenarios. We estimate the posterior distributions of a set of FICOS model parameters, and use them to quantify the uncertainties in model outputs. FICOS is computationally costly making traditional Markov chain Monte Carlo methods unviable. Hence, we emulate the log posterior density function with a Gaussian Process (GP) and train the GP emulator with fixed number of FICOS model runs. In order to reduce the number of FICOS model runs we locate first the posterior mode with Bayesian Optimization. We then construct a space filling design around the mode and run FICOS model with the parameter values at the design points to train the GP emulator. We use the emulator for constructing posterior density estimates for the FICOS model parameters and predictions with a set of alternative management scenarios.

Presenter: **Kostas Tsampourakis** (University of Edinburgh)

Title: Augmented Gaussian Sum Filters

Abstract: Bayesian filtering is the problem of recursive approximation of the posteriors of the latent states of a nonlinear dynamical system, known as filtering distributions.

An important family of Bayesian filtering methods are Gaussian filters which rely on Gaussian approximations of the state distributions. Gaussian filters however are unable to approximate multimodal distributions that commonly arise in nonlinear systems. They can nonetheless be used to construct Gaussian mixture approximations of the filtering distribution, known as Gaussian sum filters (GSF), which are able to approximate any distribution. Previous works have established that GSF's performance heavily relies on the covariance matrices of the mixture components.

In the work we are going to present, we devise a way for controlling the covariances of the components. Our approach relies on a well-known Gaussian identity, which is used to break down the "parent" components of a GSF into "children" components of smaller width. Since the children components have smaller width, when they are propagated through the nonlinearities they can reduce linearisation error significantly, compared to standard GSF. The novel approach, has the standard GSF and as a special case and allows different approximation methods to be combined into a unified algorithm.

We provide a theoretical justification of the proposed methodology. Moreover, we provide criteria for choosing the covariances of the 'children' components. Finally we use numerical simulations to compare the performance our approach with standard GSFs and particle filters.

Presenter: **Lars Skaaret-Lund** (Norwegian university of life sciences (NMBU))

Title: Boosting Performance of latent binary Bayesian neural networks with the local reparametrizaion trick and normalizing flows

Abstract: Artificial neural networks (ANN) are powerful machine learning methods that are used in many modern applications such as facial recognition, machine translation and cancer diagnostics. A common issue with ANNs is that they usually have millions or billions of trainable parameters, and therefore tend to overfit to the training data. This is especially problematic in applications where it is important to have reliable uncertainty estimates. Bayesian neural networks (BNN) can improve on this, since they incorporate parameter uncertainty. In addition, latent binary Bayesian neural networks (LBBNN) also take into account model uncertainty, enabling inference in both model and parameter space. In this paper, we will consider two extensions to the LBBNN method: Firstly, by using the local reparametrization trick (LRT) to sample the hidden units directly, we get a more computationally efficient algorithm. Secondly, by using normalizing flows on the variational posterior distribution of the LBBNN parameters, the network learns a more flexible variational posterior distribution than the mean field Gaussian. Experimental results show that this improves significantly on predictive power compared to the LBBNN method, while also obtaining a more sparse network. Additionally, we perform a simulation study where the normalizing flow method performs best at variable selection.

Presenter: **Laura Battaglia** (University of Oxford)

Title: The Variational Meta-Posterior for the Learning Rate and Prior Hyper-Parameters: fitting misspecified spatial models with missing measurement locations

Abstract: We devise a probabilistic model targeted at a spatial inference problem where we simultaneously learn English dialect fields and estimate missing locations of parts of the texts on a map of England. We depart from traditional Bayesian inference and rather select a Semi-Modular inference (SMI) approach to module the feedback between estimated fields and imputed locations, thus mitigating model misspecification concerns. For estimation, we use a (flow-based) Variational approximation to the SMI posterior that allows for good approximation, controllable feedback and end-to-end optimisation. In this context, we propose a

novel method to jointly learn model hyperparameters by minimising the distance between the SMI-predictive distribution and the true generative model on held-out data, in a LOOCV framework tailored to the specific inference objectives.

Presenter: **Lorenzo Rimella** (Lancaster University)

Title: Inference on Extended-Spectrum Beta-Lactamase Escherichia coli and Klebsiella pneumoniae data through SMC

Abstract: Many epidemic models are naturally defined as individual-based models: where we track the state of each individual within a susceptible population. Inference for individual-based models is challenging due to the high-dimensional state-space of such models, which increases exponentially with population size. We consider sequential Monte Carlo algorithms for inference for individual-based epidemic models where we make direct observations of the state of a sample of individuals. Standard implementations, such as the bootstrap filter or the auxiliary particle filter are inefficient due to mismatch between the proposal distribution of the state and future observations. We develop new efficient proposal distributions that take account of future observations, leveraging the properties that (i) we can analytically calculate the optimal proposal distribution for a single individual given future observations and the future infection rate of that individual; and (ii) the dynamics of individuals are independent if we condition on their infection rates. Thus we construct estimates of the future infection rate for each individual, and then use an independent proposal for the state of each individual given this estimate. Empirical results show order of magnitude improvement in efficiency of the sequential Monte Carlo sampler for both SIS and SEIR models.

Presenter: **Louis Grenioux** (École Polytechnique (CMAP))

Title: Assisted Learning of Energy-Based Models with Normalizing Flows

Abstract: Energy-based models (EBM) are versatile unnormalized density estimation models. EBMs can be estimated using maximum likelihood but this procedure inherently requires sampling the learned model. Meanwhile, the energy function is often parametrized by a deep neural network so as to model complicated target distributions, which ultimately defines a high-dimensional sampling problem inheriting the multimodality of the target. As a result maximum likelihood training of EBMs remains a challenge for common MCMC samplers. In this work, we propose to use recent algorithms from the adaptive MCMC literature by jointly training a normalizing flow to enhance the sampling of the intermediate energy function. Unlike commonly used Langevin samplers, adaptive samplers combine local transitions and global

jumps in the Markov chain thus enabling much better mixing in multi-modal contexts. Moreover, this auxiliary generative model provides an easy way to get diverse samples realistically reflecting the multi-modality of the energy-based model. We show that the trained models are capable of sampling and modeling complex distributions in high-dimensions.

Presenter: **Luca Alessandro Silva** (Bocconi university)

Title: Robust Leave-one-out cross validation for high dimensional Bayesian models

Abstract: Leave-one-out cross-validation (LOO-CV) is a popular method for estimating out-of-sample predictive accuracy. However, computing LOO-CV criteria can be computationally expensive due to the need to fit the model multiple times. In the Bayesian context, importance sampling provides a possible solution but classical approaches can easily produce estimators whose variance is infinite, making them potentially unreliable. Here we propose and analyze a novel mixture estimator to compute Bayesian LOO-CV criteria. Our method retains the simplicity and computational convenience of classical approaches, while guaranteeing finite variance of the resulting estimators. Both theoretical and numerical results are provided to illustrate the improved robustness and efficiency. The computational benefits are particularly significant in high-dimensional problems, allowing to perform Bayesian LOO-CV for a broader range of models. The proposed methodology is easily implementable in standard probabilistic programming software and has a computational cost roughly equivalent to fitting the original model once.

Presenter: **Luca Danese** (University of Milano-Bicocca)

Title: Model based clustering of time series with common historical change times

Abstract: The developments of the last decades in theory and methods of Bayesian model-based clustering paved the way to perform analyses with complex and structured data. Within this framework, we propose a novel approach to perform clustering among different time series on the base of their structural changes. Two distinct time series belong to the same cluster if they share the same change points, regardless of whether their local behaviours are homogeneous. Our modelling strategy relies on a hierarchical specification in a Bayesian framework, where the latent partitions describing the change points are assumed to be exchangeable realizations from a discrete distribution. Posterior inference with the proposed model can be done by resorting to a collapsed Gibbs sampler based on a Pòlya urn scheme. Our studies are mainly motivated by an epidemiological application where we want to cluster together structural changes of the COVID-19 diffusion for different states.

Presenter: **Luca Presicce** (University of Milano-Bicocca)

Title: Measurement Error Models for Spatial Network Lattice Data: Analysis of Car Crashes in Leeds

Abstract: Road casualties represent an alarming concern for modern societies demanding evidence-based interventions. In the last years, several authors developed sophisticated statistical approaches to help local authorities implement new policies and mitigate the problem. These models typically include socio-economic variables, such as population density, and traffic volumes. However, the latter variables usually suffer from measurement errors (ME), which can severely bias the statistical inference. This paper presents a Bayesian hierarchical model to analyse car crashes occurrences on a linear network level taking into account ME in the spatial covariates and the lattice structure of the road segments. Using a CAR prior, this work introduces a spatial dependence structure at the network lattice level within the classical ME framework. The suggested methodology is exemplified considering all collisions in the road network of Leeds (UK) from 2011 to 2019. Traffic volumes are approximated at the street segment level using an extensive dataset of road counts obtained from mobile devices, and the estimates are corrected using a ME model. Estimation was carried out with the INLA methodology, which allows for computational advantages. Our results show that omitting ME adjustment considerably worsens the model's fit and attenuates the effects of imprecise covariates.

Presenter: **Luke Hardcastle** (University College London)

Title: MCMC for automatic polyhazard model selection

Abstract: Polyhazard models, where individual event times are viewed as the minimum of multiple latent times, are an attractive method for modelling hazards over extended time horizons. Their additive hazard structure allows for flexible, non-proportional hazards, whilst also providing a natural framework for external data or strong prior information to inform long-term survival. Significant user input is required, however, in selecting the number of latent hazards to model, their distributions and the choice of which variables to associate with each hazard. The resulting set of models is too large to explore manually, limiting their practical usefulness. In this work we extend the standard polyhazard model by introducing spike and slab priors for automatic variable selection, and a prior over the number of hazards and their corresponding distributions allowing for the joint inference of both the model parameters and structure. To sample from the resulting posterior, we have developed a sampling scheme that utilises state of the art sticky Piecewise Deterministic Markov Processes for variable selection, embedded within a bespoke reversible jump MCMC scheme to automatically select the number of components and their corresponding distributions.

Presenter: **Mahasen B. Dehideniya** (University of Peradeniya, Sri Lanka)

Title: A gradient boosting regression approach to learn summary statistics for likelihood-free Bayesian inference

Abstract: Constructing informative yet low-dimensional summary statistics is a critical task when using likelihood-free Bayesian inference for real-world problems, particularly with a large number of observations. Due to the inherent ability to approximate multi-target non-linear functions, neural networks are extensively used as an automatic tool to convert observed and simulated data into useful summary statistics despite the high computational cost associated with the model training. While tree-based regression methods have outperformed neural networks in real-world problems, tree-based methods have only been used in constructing summary statistics for single-parameter estimation. This study proposes a novel approach to construct informative low-dimensional summary statistics via multi-target gradient boosting regression. The performance of the proposed approach was compared with the existing neural network-based methods using two sequential inferential algorithms, ABC-SMC and SNL. We considered three standard benchmarking examples with multiple parameters found in the literature, namely Ornstein-Uhlenbeck process, Moving average model with observational noise, and Lotka–Volterra model. According to the results, the proposed method outperforms the existing methods in terms of time and amount of training data required to train the model with minimal tuning. Thus, the proposed approach enables non-expert practitioners to estimate parameters of complex models in a computationally efficient manner.

Presenter: **Marcelo Hartmann** (University of Helsinki)

Title: Only Riemann can learn an invariant-preserving Bayes MAP estimate

Abstract: In Bayesian statistical inference, the maximum a posterior (MAP) estimate is known to be dependent on the choice of parameterization of the data probability distribution. Thus, it has long been criticized as not being a proper Bayesian measure for the posterior summary. In this work we propose a solution to this problem by looking at the posterior density as a density on a manifold. Most commonly, the posterior density is defined as the Radon-Nikodym derivative of the underlying Lebesgue measure. By modifying this underlying measure using a specific metric-tensor G from an embedded Riemannian manifold, we show that this new function is a proper density if the original one is. With this modification the new posterior is invariant over reparameterizations and preserves the map estimate of the original posterior density. Compared to the widely known Fisher-Rao information matrix as a metric-tensor, the proposed metric offers easy implementation and fast computation. Furthermore, the Fisher-Rao information matrix does not guarantee preservation of the MAP estimate.

Presenter: **Martin Jørgensen** (University of Oxford)

Title: Bayesian Quadrature with Log-Gaussian Bézier Processes

Abstract: We present a new class of stochastic processes, namely Bézier Processes, inspired by Bézier surfaces. This class can perform variational inference on both high-dimensional datasets with millions of observations. On top of this, they allow for analytical integration which makes them suitable for task the task of Bayesian Quadrature, which has historically been limited in both dimensions and number of observations. We show how we can both perform adaptive Bayesian Quadrature and do posterior inference on models with many parameters. This scalability can potentially make Bayesian Quadrature a competitor to MCMC methods.

Presenter: **Marvin Schmitt** (University of Stuttgart)

Title: Detecting Model Misspecification in Amortized Simulation-Based Inference

Abstract: Recent advances in probabilistic deep learning enable amortized Bayesian inference in settings where the likelihood function is implicitly defined by a simulation program. But how faithful is such inference when simulations represent reality somewhat inaccurately? In this paper, we conceptualize the types of model misspecification arising in simulation-based inference and systematically investigate the performance of SNPE-C (APT) and the BayesFlow framework under these misspecifications. We propose an augmented optimization objective which imposes a probabilistic structure on the learned latent data summary space and utilize maximum mean discrepancy (MMD) to detect potentially catastrophic misspecifications during inference undermining the validity of the obtained results. We verify our detection criterion on a number of artificial and realistic misspecifications, ranging from toy conjugate models to complex models of decision making and disease outbreak dynamics applied to real data. Further, we show that posterior inference errors increase when the distance between the latent summary distributions of the true data-generating process and the training simulations grows. Thus, we demonstrate the dual utility of MMD as a method for detecting model misspecification and as a proxy for verifying the faithfulness of amortized simulation-based Bayesian inference.

Presenter: **Mateus Maia Marques** (Maynooth University)

Title: GP-BART: an extension to Bayesian additive regression trees approach using Gaussian processes

Abstract: The Bayesian additive regression trees (BART) model is an ensemble method extensively and successfully used in regression tasks due to its consistently strong predictive performance and its ability to quantify uncertainty. BART combines "weak" tree models through a set of shrinkage priors, whereby each tree explains a small portion of the variability in the data. However, the lack of smoothness and the absence of a covariance structure over the observations in standard BART can yield poor performance in cases where such assumptions would be necessary. We propose Gaussian processes Bayesian additive regression trees (GP-BART) as an extension of BART which assumes Gaussian process (GP) priors for the predictions of each terminal node among all trees. We illustrate our model on simulated and real data and compare its performance to traditional modelling approaches. An implementation of our method is available in the R package rGPBART available at: https://github.com/MateusMaiaDS/gpbart

---

Presenter: **Matthew Heiner** (Brigham Young University)

Title: Quantile Slice Sampling with Transformations to Approximate Targets

Abstract: While general-purpose slice samplers can be more efficient than Metropolis-type alternatives, they are often overlooked as candidates for MCMC algorithms in complex modeling scenarios, in part because they also require tuning parameters. Generalized elliptical slice samplers address this issue by substituting the problem of tuning with that of approximating the target distribution. We apply this trade-off to conventional slice samplers by extending Neal's shrinkage procedure to general continuous distributions via transformations that automatically bound the slice region and can eliminate the need for a length-scale tuning parameter. This, together with a suitable approximated target that expands the slice region, yields an automatic and efficient rejection algorithm. We extend the transformation method to multivariate slice samplers that achieve high efficiency when natural approximations to the target are available. We demonstrate the method with a constrained state-space model for which a readily available chain of unconstrained forward-filter, backward-sampling densities provides the approximate target.

Presenter: **Maxim Fedotov** (Universitat Pompeu Fabra)

Title: Sequential Monte Carlo for Approximate Variable Selection in Generalized Linear Models

Abstract: In the context of high-dimensional generalized linear regression, model selection methods must use heuristics or approximations to explore an exponentially growing search space in a time-efficient manner. The classical Laplace Approximation (LA) facilitates calculation of integrated likelihoods which are used in Bayes factors to compare models. It has

been shown to be model-selection consistent with a fast convergence rate with respect to the number of observations (Rossell and Rubio, 2019; Kass, 1990). Nevertheless, it quickly becomes computationally infeasible, since an optimization problem has to be solved for every possible model that is being explored. The Approximate Laplace Approximation (ALA) is a computationally cheap alternative, which is obtained via a quadratic log-likelihood expansion at an appropriate initial guess on the model parameters. However, it does not recover the same model as the LA asymptotically. We outline and implement a Sequential Monte Carlo (SMC) algorithm that aims to quickly sample models from the LA posterior, using the ALA to get a starting distribution and subsequently applying it to sampled models. The algorithm exploits the fact that the ALA makes it cheaper to compute Bayes factors, and that the samples start to concentrate on a set of most likely models relatively quickly.

Presenter: **Maxim Panov** (Technology Innovation Institute, UAE)

Title: Stochastic normalizing flows with applications to variational autoencoders and beyond

Abstract: Variational inference recently became a popular tool to train Bayesian machine learning models. The training is usually performed by maximizing an Evidence Lower Bound (ELBO). To obtain tighter ELBO and hence better variational approximations, it was proposed to enhance variational distributions by normalizing flows. However, normalizing flows have several limitations including high computational complexity and limited flexibility due to their parametric nature. Recently, it was suggested to add stochasticity to normalizing flows that greatly improves their approximation properties. In this talk I will review the existing approaches to construction stochastic normalizing flows and will introduce some new approaches. In particular, I will consider the application of Annealed Importance Sampling (AIS) and its Sequential Importance Sampling (SIS) extensions to this problem. The potential benefits brought by these advanced techniques have never been realized for Bayesian ML models and Variational Autoencoders (VAEs): the AIS estimate cannot be easily differentiated, while SIS requires the specification of carefully chosen backward Markov kernels. In this talk, we address both issues and demonstrate the performance of the resulting Monte Carlo VAEs on a variety of applications.

Presenter: **Michael Komodromos** (Imperial College London)

Title: Mean-field group spike-and-slab variational Bayes

Abstract: In recent years variational Bayes (VB) has presented itself as a viable alternative to MCMC, particularly in situations where scalability is key. We follow such developments

and present a VB approximation to group sparse high-dimensional Bayesian linear and logistic regression.

Under our approximation, we consider two variational families, the first a mean-field group spike-and-slab variational family where the slab is multivariate Gaussian with diagonal covariance, and the second where the slab is Gaussian with an unrestricted covariance.

Under the later variational family in the linear regression setting, we demonstrate that the optimum can be parametrized with O(p) parameters rather than O(p^2), greatly reducing the computational complexity. Additionally, we study the Frequentist properties of the VB posterior and introduce theoretical results for our methodology.

Finally, in the logistic regression setting we introduce a new bound on the expectation of the log-likelihood for Gaussian variational families and assess the performance against others in the VI literature. Notably, this bound can be made arbitrarily tight, albeit at high computational cost, without the need of variational parameters.

Presenter: **Michael Whitehouse** (University of Bristol)

Title: Consistent and fast inference in compartmental models of epidemics using Poisson Approximate Likelihoods

Abstract: Addressing the challenge of scaling-up epidemiological inference to complex and heterogeneous models, we introduce Poisson Approximate Likelihood (PAL) methods. In contrast to the popular ODE approach to compartmental modelling, in which a large population limit is used to motivate a deterministic model, PALs are derived from approximate filtering equations for finite-population, stochastic compartmental models, and the large population limit drives the consistency of maximum PAL estimators. Our theoretical results appear to be the first likelihood-based parameter estimation consistency results applicable across a broad class of partially observed stochastic compartmental models concerning the large population limit. Compared to simulation-based methods such as Approximate Bayesian Computation and Sequential Monte Carlo, PALs are simple to implement, involving only elementary arithmetic operations and no tuning parameters; and fast to evaluate, requiring no simulation from the model and having computational cost independent of population size. Through examples, we demonstrate how PALs can be: embedded within Delayed Acceptance Particle Markov Chain Monte Carlo to facilitate Bayesian inference; used to fit an age-structured model of influenza, taking advantage of automatic differentiation in Stan; and applied to calibrate a spatial meta-population model of measles.

Presenter: **Mikko Koivisto** (University of Helsinki)

Title: Scalable Bayesian inference of causal DAGs

Abstract: We advance computational methods for Bayesian inference of Bayesian networks given data over all variables in the network. Our methods build on a recent Markov chain Monte Carlo scheme, which enables efficient approximate sampling from the posterior distribution of directed acyclic graphs, DAGs, provided that each node is assigned a small number of candidate parents. We present algorithmic techniques to significantly lower the space and time requirements both in the preprocessing phase and during the Markov chain simulation. Our scheme admits substantially larger sets of candidate parents as well as efficient incorporation of an edge-reversal move that is vital for fast mixing of the Markov chain. We apply the methods to the discovery of cause–effect relations and to the inference of causal effects in linear DAG models. Numerical experiments show that our method outperforms previous Bayesian methods and is superior to non-Bayesian methods in causal effect estimation. (Joint work with Jussi Viinikka, Antti Hyttinen, and Johan Pensar.)

Presenter: **Mikolaj Kasprzak** (University of Luxembourg & MIT)

Title: How good is your Laplace approximation? Finite-sample error bounds for a variety of useful divergences

Abstract: The Laplace approximation is a popular method for providing posterior mean and variance estimates. But can we trust these estimates for practical use? One might consider using rate-of-convergence bounds for the Bayesian Central Limit Theorem (BCLT) to provide quality guarantees for the Laplace approximation. But the bounds in existing versions of the BCLT either: require knowing the true data-generating parameter, are asymptotic in the number of samples, do not control the Bayesian posterior mean, or apply only to narrow classes of models. Our work provides the first closed-form, finite-sample quality bounds for the Laplace approximation that simultaneously (1) do not require knowing the true parameter, (2) control posterior means and variances, and (3) apply generally to models that satisfy the conditions of the asymptotic BCLT. In fact, our bounds work even in the presence of misspecification. We compute exact constants in our bounds for a variety of standard models, including logistic regression, and numerically demonstrate their utility. We provide a framework for analysis of more complex models. This is joint work with Ryan Giordano (MIT) and Tamara Broderick (MIT). A preprint is available at arXiv:2209.14992.

Presenter: **Nicola Branchini** (The University of Edinburgh)

Title: Generalized Self-Normalized Importance Sampling

Abstract: Importance sampling (IS) estimators can maximize the efficiency of Monte Carlo estimation of expectations, i.e. integrals involving a probability density function (pdf). The integrand is often known only up to a normalizing constant. In these cases, the *self-normalized* IS (SNIS) estimator is used. SNIS has fundamental limitations: its variance is lower bounded by a nonzero constant, and its recycling of samples implicitly induce correlations whose effect is not well studied. We first consider a new perspective on SNIS that views is as an estimator for a ratio of integrals. Then, we introduce an IS proposal in an extended space that allows us to control the dependence in the numerator and denominator estimates. Our formulation generalizes SNIS, and provides guidance on how to design appropriate proposals to minimize the resulting mean squared error.

Presenter: **Nikolas Siccha** (Aalto University)

Title: A novel warm-up for HMC based samplers

Abstract: We propose a novel type of warm-up for HMC based samplers which alleviates many of the obstacles to an automatic and efficient exploration of diverse posteriors.

The warm-up is generally

- computationally more efficient than the current state of the art,
- avoids insignificant modes,
- facilitates automatic reparameterizations,
- allows automatic adaptation of discretizations for approximate posteriors as encountered e.g. for ordinary differential equation models or approximate Gaussian process models

and supports iterative model building by

- quickly revealing coding errors, prior misspecifications and obstacles to efficient posterior exploration,
- firmly incorporating prior, intermediate and posterior predictive checks and
- allowing user intervention at every step of the process.

Reductions in warm-up wall time and in required computational work can be tenfold or more, while generally not negatively impacting sampling efficiency and sometimes even increasing sampling efficiency as measured in the number of leapfrog steps needed per minimal estimated effective sample size for a range of posteriors.

Examples posteriors include hierarchical pharmacological ODE models, Gaussian process approximations via basis functions and epidemiological models. We present comparisons against Stan's current warm-up, combined and extended with options for using Pathfinder and diagonal + low-rank mass matrices.

Presenter: **Noa Kallioinen** (Aalto University)

Title: priorsense: Intuitive and efficient Bayesian prior sensitivity analysis

Abstract: Determining the sensitivity of the posterior to perturbations of the prior and likelihood is an important part of the Bayesian workflow. We introduce a practical and computationally efficient sensitivity analysis approach using importance sampling to estimate properties of posteriors resulting from power-scaling the prior or likelihood. On this basis, we suggest a diagnostic that can indicate the presence of prior-data conflict or likelihood noninformativity and discuss strengths and limitations of the power-scaling approach. The approach can be easily included in modern MCMC-based Bayesian workflows with minimal effort by the model builder. We present an implementation in our new R package priorsense, along with examples and guidelines for use.

Presenter: **Özge Özenoglu** (Ludwig-Maximilians-University)

Title: faintr: An R package for interpreting factors in Bayesian regression models

Abstract: Categorical predictor variables in regression models are traditionally encoded in a way that allows easy comparison of different constellations of factor cells. However, there is not always a single contrast coding scheme that allows all relevant research questions to be directly addressed, such that further processing or model fitting is necessary. Moreover, in Bayesian regression models, a contrast coding that facilitates setting priors over model coefficients might not be practical for answering the research questions. To support the (exploratory) probing of research questions in factorial designs, we have developed the R package faintr (FActorINTerpreteR), which provides convenience functions for fitted Bayesian regression models with categorical predictor variables. By disassociating the priors, model parameters and the research questions from the contrast coding, the package allows flexible extraction and comparison of posterior draws of arbitrary subsets of factor cells.

Presenter: **Paolo Ceriani** (Bocconi University)

Title: Unbiased estimator for crossed random effect models via couplings

Abstract: We exploit couplings to obtain unbiased posterior estimates for Gaussian crossed random effects models with a linear computational cost in the number of datapoints and parameters. The proposed technique matches the cost of state of the art methods such as the

backfitting algorithm by Ghosh et al. (2022) and the collapsed Gibbs sampler by Papaspiliopoulos et al. (2019) but achieves unbiasedness for a finite number of iterations. We derive bounds on the expected number of iterations needed before coalescence for the "two steps" Gaussian coupling and study its asymptotic behaviour. We then estimate the mean coalescence time in real and synthetic datasets for different asymptotic regimes, showing strong agreement with our theoretical findings. Although our theory applies only in the Gaussian case, the proposed methodology proves effective also in the general case of non-Gaussian likelihood as shown in simulated studies.

Presenter: **Patric Dolmeta** (Bocconi University)

Title: Efficient Thompson sampling for fractional-factorial multi-armed bandits with binary rewards

Abstract: Multi-armed bandits are sequential experiments aimed at maximizing the reward from some payoff distribution with unknown parameters. At each stage, the experimenter decides which experimental configuration to observe next. In doing so, she needs to control the exploding number of configurations as factors are added to the experiment. A possible solution is to assume a probit regression model with fractional factorial design as reward likelihood. In this work, we analyze a Bayesian approach to this sequential learning problem. We first outline a closed form posterior result under prior normality for regression coefficients. Further, we motivate why i.i.d. sampling from the Unified Skew Normal (SUN) posteriors is appealing in Thompson sampling, the policy of choice for selecting which strategy to play. Finally, we suggest a twofold approach to face the computational bottlenecks given by the time horizon expansion. On one hand, we derive a Sequential Monte Carlo scheme whose performance challenges the common MCMC procedures in literature. On the other, we exploit closed form expressions of quantities needed by our policy, to speed up computations. In both cases, we exploit the knowledge on the structure of the posterior to enhance precision and computational tractability.

Presenter: **Paul Rosa** (University of Oxford)

Title: Estimating a density near an unknown submanifold : a Bayesian nonparametric approach

Abstract: We study the Bayesian density estimation of data living in the offset of an unknown submanifold of the Euclidean space. In this perspective, we introduce a new notion of anisotropic Hölder for the underlying density, new Dirichlet process location-scale mixtures of

Gaussians priors and obtain posterior rates that are minimax optimal, adaptive to the regularity of the density, to the intrinsic dimension of the manifold, and to the size of the offset. We implement our Bayesian procedure using either a Markov chain Monte-Carlo algorithm (with Gibbs sampling) or maximum a posteriori. It yields good practical results even for quite singular data, especially compared to the standard Normal inverse Wishart conjugate prior.

Presenter: **Petar Jovanovski** (Chalmers University of Technology)

Title: Approximate Bayesian Computation with Forward-Backward Simulation for Stochastic Differential Equations

Abstract: Stochastic differential equations (SDE) are employed in many areas of science as a powerful tool for modeling processes that are subject to random fluctuations. Bayesian inference for SDEs is problematic because in the majority of cases the likelihood function is analytically intractable. Approximate Bayesian Computation (ABC) methods can be employed because forward simulation is made possible with numerical methods. We propose a simulation scheme for SDE models that is based on processing the observation in the forward and then in the backward direction, effectively utilizing the information provided by the observed data. This leads to the simulation of sample paths that are consistent with the observation, and thereby increase the acceptance rate. We additionally leverage partial exchangeability of Markov processes and employ invariant neural networks to learn the summary statistics that are needed in ABC. The summary statistics are sequentially learned by exploiting an ABC-SMC sampler, which povides new training data with each iteration. Our ambition is therefore to provide a learning tool for the SDEs model parameters while simultaneously learning the summary statistics.

Presenter: **Philipp Reiser** (University of Stuttgart)

Title: Quantifying Uncertainty in Surrogate-based Bayesian Inference

Abstract: Surrogate models are statistical approximations for highly complex simulations. In this context, it is crucial to propagate the uncertainty due to limited simulations and due to the approximation error of the surrogate model to predictions and subsequent decision-relevant quantities. However, quantifying the uncertainty of surrogates is usually limited to the use of special analytic cases or is otherwise very computationally expensive. In this poster, we will explore a scalable, fully Bayesian approach to surrogate modeling and uncertainty propagation using probabilistic programming languages. We will also present a method for Bayesian inverse modeling with surrogate models, where we propagate the uncertainty to the unknown input given measurement output. Finally, we validate our inverse method and introduce an adapted

version of Simulation-Based Calibration (SBC) that handles non-identified models and multiple posterior modes.

Presenter: **Prakhar Verma** (Aalto University)

Title: Gaussian Variational Inference for Diffusion Processes Revisited

Abstract: Non-linear diffusion processes are a class of stochastic differential equations (SDEs) models providing a rich family of expressive latent prior models that arise naturally in dynamic modelling tasks. Variational inference in such models was introduced by Archambeau et al. (2007), whereby the posterior process is approximated by a Markovian Gaussian process via maximizing the classic variational lower bound to the marginal likelihood. An algorithm was proposed to learn the variational distribution, parameterized via a time-dependent linear drift function, and used in multiple subsequent extensions. In this work, we explain why this algorithm is slow to converge and propose an alternative parameterization and optimization algorithm that significantly speeds up approximate state inference and learning of the prior hyperparameters. Our work combines 1) recent advances in the efficient use of natural gradient descent for variational inference in Gaussian Process models (Chang et al., 2020) with 2) iterative posterior linearization of the prior process (Garcia -Fernandez et al., 2016). Essentially, we adapt a fast, exact algorithm for linear diffusions to the non-linear setting by locally linearizing the drift of non-linear diffusions along the optimization trajectory.

Presenter: **Rayleigh Lei** (University of Washington)

Title: Geometrically Fitting Tree-directed Topic Modeling

Abstract: Topic models have been extensively used to analyze corpora. However, when corpora with many subjects are examined, many topics might be needed despite only a subset being relevant for a particular subject. A more parsimonious extension to address this while giving further meaning to the topics is to organize them into a tree structure. Here, the topic corresponding to the root node is shared by all subjects, the topics corresponding to nodes at the next level are shared by subsets of subjects, and so on until the topics corresponding to the leaf nodes are unique to each subject. Not only can this reveal how subjects are related through their topics, but also show how much a topic helps to define a subject. Such a model is challenging to fit using traditional approaches even if the tree topology is known. As an alternative, we propose geometric algorithms to estimate the topic simplices and merge them to learn the topics and their relationships. Simulation studies demonstrate that this approach is more efficient and can outperform traditional methods according to various metrics. We apply this algorithm to analyze a subset of articles in the New York Times from 2000, uncovering meaningful semantic relationships.

Presenter: **Riccardo Cogo** (University of Milano-Bicocca)

Title: Hierarchical neutral to the right priors

Abstract: The Beta-Stacy process introduced by Walker and Muliere (1997) is a well-known Bayesian nonparametric prior for survival functions, which is typically used in presence of censored survival times. This process belongs to the more general class of neutral to the right priors, which are suitable for a single group of homogeneous, i.e. exchangeable, survival times. We introduce a hierarchical version of neutral to the right priors, which is tailored for heterogeneous, though related, groups of survival times. We investigate the theoretical properties of the new class of processes and we provide a posterior characterization of the processes. As a noteworthy example, we focus on a hierarchical version of the Beta-Stacy process, and we specify all the general results in this particular case. We also provide Bayesian estimators of the random survival functions. The proposed methodology allows us to develop an efficient conditional algorithm to estimate the survival functions, which is tested on simulated and real data.

Presenter: **Richard Everitt** (University of Warwick)

Title: ABC-SMC^2

Abstract: We propose a new algorithm based on a combination of an SMC sampler for estimating the ABC likelihood in the case of high-dimensional data (Prangle et al, 2018) and ABC-SMC for exploring the parameter space. The new method has a similar structure to SMC2 (Chopin et al, 2012). To automate the approach, we make use of an adaptive scheme for both the sequence of ABC tolerances in the SMC, and also for the MCMC rejuvenation steps of the external parameter space SMC. This is joint work with Ivis Kerama and Tom Thorne.

Presenter: **Rocco Caprio** (University of Warwick)

Title: Markov Chain Monte Carlo Calculus

Abstract: Markov Chain Monte Carlo (MCMC) algorithms are based on the construction of a Markov Chain with transition probabilities $P_\mu(x, \cdot)$ where μ is an invariant distribution of interest. In this work, we look at these transition probabilities, within a suitable family of such transitions, as functions of their invariant distributions and we develop a notion of *derivative with respect to the invariant distribution*. We are building around this concept a set of tools

we refer to as *Markov Chain Monte Carlo Calculus*. We present some connections between these new concepts and ergodicity and we explain how MCMC Calculus is naturally suited to the study of approximation-based MCMC, also illustrating how it suggests practical guidelines for MCMC algorithms efficiency. Finally, we present implications for the convergence of interacting and sequential MCMC (joint work with Adam Johansen).

Presenter: **Rui Li** (Aalto University)

Title: Towards Improved Learning in Gaussian Processes: The Best of Two Worlds

Abstract: Gaussian process training decomposes into inference of the (approximate) posterior and learning of the hyperparameters. For non-Gaussian (non-conjugate) likelihoods, two common choices for approximate inference are Expectation Propagation (EP) and Variational Inference (VI), which have complementary strengths and weaknesses. While VI's lower bound to the marginal likelihood is a suitable objective for inferring the approximate posterior, it does not automatically imply it is a good learning objective for hyperparameter optimization. We design a hybrid training procedure where the inference leverages conjugate-computation VI and the learning uses an EP-like marginal likelihood approximation. We empirically demonstrate on binary classification that this provides a good learning objective and generalizes better.

Presenter: **Sam Duffield** (Quantinuum)

Title: Ensemble Kalman Inversion for General Likelihoods

Abstract: In this work, we generalise Ensemble Kalman inversion techniques to general Bayesian models where previously they were restricted to additive Gaussian likelihoods — all in the difficult setting where the likelihood can be sampled from, but its density not necessarily evaluated.

Presenter: **Sanket Agrawal** (University of Warwick)

Title: Optimal scaling of MCMC beyond Metropolis

Abstract: The problem of optimally scaling the proposal distribution in a Markov chain Monte Carlo algorithm is critical to the quality of the generated samples. Much work has gone into obtaining such results for various Metropolis-Hastings (MH) algorithms. Recently, acceptance

probabilities other than MH are being employed in problems with intractable target distributions. There is little resource available on tuning the Gaussian proposal distributions for this situation. We obtain optimal scaling results for a general class of acceptance functions, which includes Barker's and Lazy-MH. In particular, optimal values for the Barker's algorithm are derived and found to be significantly different from that obtained for the MH algorithm. Our theoretical conclusions are supported by numerical simulations indicating that when the optimal proposal variance is unknown, tuning to the optimal acceptance probability remains an effective strategy.

Presenter: **Seth D. Axen** (University of Tübingen)

Title: Pathfinder.jl: approximate inference for the statistical workflow

Abstract: Building a good model in probabilistic programming languages (PPLs) often requires iterative model adjustments motivated by early inference results, where Markov Chain Monte Carlo (MCMC) is usually employed as a versatile inference method. MCMC can be especially computationally expensive for problematic or poorly parameterized models, where most of the time is spent in the lengthy MCMC warm-up phase. Pathfinder is a recent variational inference method designed to replace all or part of MCMC's warm-up phase, facilitating quicker evaluation and improvement of such challenging models. Furthermore, while some parameters computed by Pathfinder have been proposed to to replace the metric often tuned during the warm-up phase of Hamiltonian Monte Carlo (HMC) as implemented in PPLs like Stan, Turing, and PyMC, this potential remains unexplored. Here, we introduce Pathfinder.jl, an extensible, generic implementation of Pathfinder in the Julia language. We use the implementation to explore the performance gains of using Pathfinder to initialize HMC relative to tuning the metric on a variety of models. Finally, we explore several modifications of the algorithm and compare their performance with the default settings.

Presenter: **Simon Urbainczyk** (Heriot-Watt University)

Title: Computational Methods for Bayesian Imaging with Deep Gaussian Process Priors

Abstract: In image reconstruction, an accurate quantification of uncertainty is of great importance for informed decision making. Here, the the Bayesian approach to inverse problems can be used: the image is represented through a random function that incorporates prior information which is then updated through Bayes' formula. Finding a prior is difficult. Images often exhibit non-stationary effects and multiscale behaviour. Thus, usual Gaussian process priors are not suitable. Deep Gaussian processes, on the other hand, encode non-stationary behaviour in a natural way through their hierarchical structure. To apply Bayes' formula, one

commonly employs a Markov chain Monte Carlo method that requires sampling from the prior. In the case of deep Gaussian processes, sampling is especially challenging in high dimensions: the associated covariance matrices are large, dense, and changing from sample to sample. A popular strategy towards decreasing computational complexity is to view Gaussian processes as the solutions to a fractional stochastic partial differential equation (SPDE). In this work, we investigate efficient computational strategies to solve the fractional SPDEs occurring in deep Gaussian process sampling. Indeed, we employ rational approximations to represent the fractional operators through sparse matrices and reduce computational cost from cubic to near-linear. We test our techniques in standard Bayesian image reconstruction problems.

Presenter: **Stanislas du Ché** (Ecole Polytechnique - Columbia University)

Title: Parallelization for Markov chains Monte Carlo with heterogeneous runtimes

Abstract: A recent trend in Markov chains Monte Carlo sampling is the deployment of hardware accelerators to run many chains in parallel. This approach presents several benefits, including short sampling phases, ensemble-chains warmup, and better computation of diagnostics. But the parallelization of stochastic operations presents an important challenge: the runtime for fitting a model is determined, not by the average runtime of the chains, but by the worst runtime which follows the distribution of an ordered statistic. This is particularly problematic for Bayesian models with highly varying fitting times. Such examples can be found in Pharmacokinetics, where likelihoods are based on ordinary differential equations. We examine examples and demonstrate that, because of unstable runtimes, running more chains in parallel can lead to a decrease in performance. We study various strategies to stabilize the runtime of MCMC and propose a multiple chains sampling scheme, which stops early once a convergence criterion is met and a target effective sample size achieved, without waiting for all the chains to finish. The runtime is then determined not by the slowest running chains but by the fastest running ones.

Presenter: **Stefan T. Radev** (Heidelberg University)

Title: Towards Compressed Bayesian Inference for Exchangeable Sequences

Abstract: We investigate a method to compress exchangeable sequences with varying sizes and use the compressed representations for downstream Bayesian parameter estimation and model comparison tasks. We employ information maximizing variational autoencoders (VAEs) which we augment with normalizing flows for more expressive likelihood learning. We showcase the ability of our method to learn informative representations on toy examples and two real world modeling scenarios.

Presenter: **Stephen Berg** (Penn State University)

Title: Control Variates and Rao-Blackwellization for Deterministic Sweep Markov Chains

Abstract: A variety of techniques, including control variates and conditioning, exist for reducing the asymptotic variance of estimators from Markov chain Monte Carlo (MCMC) simulations. In this paper, we develop new methods and theory for control variates and conditioning-based approaches in the setting of deterministic sweep Markov chain samplers where multiple transition kernels are cycled through in a fixed order. Specifically, we devise control variates that are time inhomogeneous and applicable to commonly used deterministic sweep MCMC samplers, with simplifications arising for deterministic sweep Gibbs samplers. Compared with existing approaches in the literature such as those for data augmentation and random sweep, our estimators based on control variates and a related Rao-Blackwellization type of estimator provide variance reduction guarantees in new settings. We conduct a simulation study which demonstrates the theoretical benefits of our proposed methods in practice for popular models such as the bivariate Gaussian as an example of a multivariate Gaussian distribution and the Ising model in the family of Markov random fields.

Presenter: **Tamas Papp** (Lancaster University)

Title: Asymptotically optimal coupling for the random walk Metropolis

Abstract: The reflection-maximal coupling of the random walk Metropolis (RWM) algorithm was recently proposed for use within unbiased MCMC. Numerically, when the target is spherical this coupling has been shown to perform well even in high dimensions. We derive high-dimensional ODE limits for Gaussian targets, which confirm this behaviour in the spherical case. However, we then extend our theory to the elliptical case and find that as the dimension increases the reflection coupling performs increasingly poorly relative to the mixing of the underlying RWM chains. To overcome this obstacle, we introduce gradient common random number (GCRN) couplings, which leverage gradient information. We show that: (1) the behaviour of GCRN couplings does not break down with the ellipticity or the dimension, (2) GCRN couplings are asymptotically optimal for contraction, and (3) GCRN couplings scale in proportion to the mixing of the underling RWM chains. Numerically, we apply GCRN couplings for convergence and bias quantification, and demonstrate that our theoretical findings extend beyond the Gaussian case.

Presenter: **Theodore Kypraios** (University of Nottingham)

Title: Latent Branching Trees: Modelling and Bayesian Computation

Abstract: We present a novel class of semi-parametric time series models, in which one can specify the marginal distribution of the observations in advance and then build their dependence structure around them by introducing an underlying stochastic process termed as a Latent Branching Tree (LBT).

We first discuss the properties of the model and then demonstrate how to fit it to time series data within a Bayesian framework and infer the parameters governing the model's structure. We develop a suite of efficient Markov Chain Monte Carlo algorithms which exploit the properties of the tree offering computational speed and good mixing. The proposed model and methods are illustrated using synthetic data and are also used to analyse genome scheme data. Finally, we discuss how general and flexible this class of models is and how it can be extended to a very general class of models.

Presenter: **Ti John** (Aalto University)

Title: Link functions in Cox processes

Abstract: Modelling the occurrence of events - points in space and/or time - is important to fields such as epidemiology, ecology, and neuroscience. To build Bayesian models appropriate for these data modalities, a popular choice is a Cox process, an inhomogeneous Poisson process whose positive intensity function is connected to a latent Gaussian process via a link function. Commonly, the link function is chosen for computational convenience or adopted without discussion. However, the link function is a modelling choice and should also get due consideration. We highlight the differences in effective priors on the events induced by different link functions and investigate to what extent the link function itself can be inferred from observations.

Presenter: **Tsuyoshi Ishizone** (Meiji University)

Title: Nonlinear Embeddings for Biomolecular Structure Data

Abstract: The conformational space of biomolecules is very high-dimensional. The intrinsic conformational space is a low dimension when considering slow time-scale motions such as folding. A linear method, tICA, has been widely used as an embedding method to capture such slow time-scale motions. With breakthroughs in deep learning, nonlinear embedding methods have also attracted attention in recent years. In line with this trend, we propose a new nonlinear embedding method. The proposed method is characterized by statistical modeling of the time transition structure. In the presentation, we will introduce the method and show the experiments of its application to several biomolecules.

Presenter: **Vesa Kaarnioja** (Freie Universität Berlin)

Title: Domain uncertainty quantification using periodic random variables with application to elliptic PDEs

Abstract: Computational measurement models may involve several uncertain simulation parameters: not only can the material properties of a heterogeneous medium be unknown, but the shape of the structure itself can be uncertain as well. We discuss a parameterization for an uncertain domain using a random perturbation field in which a countable number of independent random variables enter the random field as periodic functions. The random field can be constructed to have a prescribed mean and covariance function. As an application, we study how uncertainty in the domain shape impacts the stochastic response of an elliptic PDE. The periodic structure in the random field enables us to develop computationally simple quasi-Monte Carlo cubature rules which achieve higher-order convergence rates for high-dimensional numerical integration problems associated with the output of the PDE problem subject to uncertainty in the domain shape. The theoretical rates are assessed by numerical experiments.

This is joint work with Harri Hakula (Aalto University), Helmut Harbrecht (University of Basel), Frances Kuo (University of New South Wales), and Ian Sloan (University of New South Wales).

Presenter: **Xiaoyue Xi** (Univerisity of Cambridge)

Title: A hierarchical framework for inferring and leveraging node-level information in Bayesian networks

Abstract: Bayesian graphical models are powerful tools to infer complex associations in high dimensions, yet are often fraught with computational and statistical challenges. If exploited in a principled way, the increasing information collected alongside the data of primary interest constitutes an opportunity to mitigate these difficulties by guiding the detection of dependence structures. For instance, estimation of conditional independence in gene co-expression networks may be informed by the encoding of (publicly available) summary statistics on the regulation of genes by genetic variants. Here we present a hierarchical modelling framework to identify and leverage the variable-level information that affects the propensity of nodes to have high degree. Specifically, we consider a two-level spike-and-slab formulation for the simultaneous inference of (1) precision matrix elements and (2) effects of node- level information encoded as external covariates through a probit submodel. As efficient exploration of multimodal posterior spaces is a prerequisite to ensure applicability in real-world settings, we propose a variational EM inference strategy which utilises simulated annealing and dynamic

posterior explorations. We illustrate and exploit the advantages of our approach in simulations and in a gene co-expression study.

Presenter: **Xitong Liang** (University College London)

Title: Adaptive Random Neighbourhood Informed MCMC on Bayesian Variable Selection

Abstract: Bayesian variable selection methods (BVS) are growing in popularity. BVS both enables quantification of model uncertainty and also helps investigate underlying low-dimensional structure in predicting the response of interest. However, the increasing adoption of BVS is hindered by a lack of efficient computational algorithms for discrete posterior distributions and large-scale datasets. Our main contribution is developing a new MCMC scheme named PARNI. It first randomly constructs a neighbourhood using ideas from [1], then proposes a new model within this neighbourhood according to a locally balanced proposal [2]. Our results related to linear models from both synthetic and real data-sets show that the PARNI scheme outperforms many state-of-the-art algorithms and it achieves extraordinary mixing rate especially for high-dimensional problems. In addition, we use our scheme to explore the posterior distributions for other problems such as Bayesian variable selection on generalised linear models and Bayesian causal discovery.

[1] Griffin, J., Łatuszyński, K. and Steel, M., 2020. In search of lost mixing time: adaptive Markov chain Monte Carlo schemes for Bayesian variable selection with very large p. Biometrika, 108(1), pp.53-69.

[2] Zanella, G., 2019. Informed Proposals for Local MCMC in Discrete Spaces. Journal of the American Statistical Association, 115(530), pp.852-865.

Presenter: **Yann McLatchie** (Aalto University)

Title: Kulprit: Kullback-Leibler projections for Bayesian model selection in Python

Abstract: Often when modelling we wish to identify those covariates most important for prediction, or to eliminate those that are unimportant. From a predictive standpoint, this amounts to identifying submodels whose posterior predictive distributions are as similar to the true data-generating distribution as possible. In projection predictive inference, instead of reasoning directly on this true data-generating distribution, we approximate it with a rich reference model containing all available data, fit with appropriate priors, and satisfying all diagnostic checks. We then look to find a restricted posterior such that its induced posterior predictive distribution is as close to that of the reference model as possible in Kullback-Leibler divergence. By implementing a search heuristic, we can project our reference model onto a collection of

submodels in our model space. From this set of inferred parsimonious posteriors we can then determine which submodel is the most appropriate alternative to our reference model.

We present kulprit, a Python implementation of projection predictive inference for Bambi-fitted models. Our proposed architecture allows for an extensible API and efficient computations through just-in-time compilation.

Presenter: **Yuexi Wang** (University of Chicago Booth School of Business)

Title: Semiparametric Bayesian Bootstrap

Abstract: Modeling individual heterogeneity has always been one of the central topics of applied research in economics and social sciences. Advances in deep learning make it possible to recast the parameters as fully flexible nonparametric functions. While previous work by Farrell et al. (2020) has illustrated the success of deep learning in structured modeling of heterogeneity, performing statistical inference on the estimated parameter functions remains challenging. We utilize the Bayesian bootstrap (BB) framework, which passes random bootstrap weights to loss functions. To avoid repeatedly re-fitting neural networks, we adopt a semi-parametric linear approximation to the dependence of the parameter functions on the weights. Once the vanilla network is trained, the approximated bootstrap samples can be obtained with negligible costs. Under mild regularity conditions, we show our approximation consistently estimates the Bayesian bootstrap posterior. We illustrate the performance of our method on both simulated and real datasets.

Presenter: **Yunyi Shen** (Massachusetts Institute of Technology)

Title: Fitting sparse Gaussian chain graph models with spike-and-slab LASSO priors

Abstract: The Gaussian chain graph model simultaneously parametrizes (i) the direct effect of each of p predictors on each of q possibly related outcomes and (ii) the residual partial covariance for each pair of outcomes. We introduce a new method for fitting sparse Gaussian chain graph models with spike-and-slab LASSO (SSL) priors. We develop an Expectation-Conditional Maximization algorithm to obtain sparse estimates of the pxq matrix of direct effects and the qxq residual precision matrix. Our algorithm iteratively solves a sequence of penalized maximum likelihood problems with self-adaptive penalties that gradually filter out negligible regression coefficients and partial covariances. Because it adaptively penalizes model parameters, our method is seen to outperform fixed-penalty competitors on simulated data. We establish the posterior concentration rate for our model, buttressing our method's excellent empirical performance with strong theoretical guarantees. We additionally investigate several bootstrap-like methods for quantifying uncertainty about our final estimates. We use our

method to reanalyze a dataset from a study of the effects of diet and residence type on the composition of the gut microbiome of elderly adults.
