Bayesian Modeling and Computing

General Information

This workshop is organized by the working group on 'Bayesian Modeling and Computing' of the CRC 1456:

The whole meeting takes place completely virtually via zoom. The zoom-details are:

Invitation link:
`https://uni-goettingen.zoom.us/j/95331975127?pwd=bjk0dWs0S2tuQmhZWkZOQlNFdEtxQT09`
Meeting ID: 953 3197 5127
Passcode: 683951

For small discussions and coffee chats we also have a gather town environment.

Gather town link: `https://gather.town/invite?token=rWHNr99LAxf5NJdbcN9o79jQs62YJpdl`

The capacity of people which are able to join this gathering is restricted to 25. Therefore only the first 25 which use this opportunity get in. If there is more space required we can quickly set up a second gathering.

Below you can find the schedule and the abstracts of the two days workshop. Let us mention here that the schedule is w.r.t. Berlin time (CEST).

If there are any questions or problems contact Michael Habeck (michael.habeck@uni-jena.de) or Daniel Rudolf (daniel.rudolf@uni-goettingen.de).

Wednesday, 8th of September

10:25–10:30	**Welcome remarks**		
10:30–11:30	CT	**Krzysztof Łatuszyński** Warwick, UK	Adaptive MCMC - how to teach your MCMC to self-tune?
11:30–12:00	CT	**Jan Münch** Jena, Germany	Bayesian selection of Hidden Markov models for multi-dimensional ion channel data
12:00–13:00	**Break**		
13:00–13:30	**Virtual coffee chat and gathering**		
13:30–14:00	CT	**Han Cheng Lie** Potsdam, Germany	Random forward models and log-likelihoods in Bayesian inverse problems
14:00–15:00	CT	**Juergen Koefinger** Frankfurt, Germany	Integration of experiments and molecular simulations by Bayesian ensemble refinement

Thursday, 9th of September

10:30–11:30	CT	**Johannes Soeding** Göttingen, Germany	Adaptive MCMC and stochastic variational inference for sparse linear regression in statistical genetics
11:30–12:00	CT	**Philipp Wacker** Erlangen, Germany	Laplace-based importance sampling
12:00–13:00	**Break**		
13:00–13:30	**Virtual coffee chat and gathering**		
13:30–14:00	CT	**Benjamin Eltzner** Göttingen, Germany	Modeling protein ensembles with doubly intractable distributions
14:00–15:00	CT	**Matti Vihola** Jyväskylä, Finland	The conditional particle filter - scalable MCMC for dynamic models

List of Abstracts

Wednesday 8th

Adaptive MCMC - how to teach your MCMC to self-tune?

K. Łatuszyński

University of Warwick, Coventry, UK

The big data and big models era in statistical inference posed new challenges to Markov chain Monte Carlo (MCMC) in that off the shelf algorithms increasingly require user intensive tuning and tweaking to be applicable in real data problems. Such tuning often requires expert knowledge of the algorithm and of the computational problem at hand and may not be practical. Adaptive MCMC is aimed at automating this task in real time based on the ongoing simulation output. Yet, how to design self-tuning MCMC algorithms, and are they even valid?

Bayesian selection of Hidden Markov models for multi-dimensional ion channel data

J. Münch

Friedrich-Schiller-University, Jena, Germany

TBA

Random forward models and log-likelihoods in Bayesian inverse problems

H. C. Lie

University Potsdam, Potsdam, Germany

In many Bayesian inverse problems, the forward model is expensive to solve. This motivates the development of computationally cheaper approximations, with approximations based on Gaussian processes being popular in many applications. However, until recently, a rigorous error analysis of the associated random approximate posterior measures was lacking. Stuart and Teckentrup (Math. Comput., 2017) proved error bounds for random approximate posteriors arising from Gaussian process approximations. In this talk, we present error bounds that apply to a more general class of random approximations, and illustrate how the theory may be applied to random projection methods for processing high-dimensional data. Joint work with T. J. Sullivan (U. Warwick) and A. Teckentrup (U. Edinburgh).

Integration of experiments and molecular simulations by Bayesian ensemble refinement

J. Koefinger

Max Planck Institute of Biophysics, Frankfurt, Germany

Molecular simulations provide ensembles of structural models of flexible and dynamic (bio)molecules. These ensembles are determined by the molecular interactions encoded in empirical force fields. The opposing goals of force field accuracy and sampling efficiency limit the predictive power of these simulations. Ensemble refinement alleviates these limitations by integrating experimentally measured ensemble averages. I will introduce the task at hand and present the Bayesian inference of ensembles (BioEn,) method [10.1063/1.4937786], which epitomizes various superficially different but strongly related approaches to ensemble refinement. To directly encode the information provided by the experimental data in the force field, we adapted BioEn for force field refinement. In the Bayesian inference of force fields (BioFF) method [10.33774/chemrxiv-2021-tsbj3], the BioEn entropic prior serves to compensate for essentially unavoidable overs-simplifications in the prior on the force field parameters. I will introduce the key challenges to force field refinement and illustrate a way forward applying BioFF in a simple example.

Thursday 9th

Adaptive MCMC and stochastic variational inference for sparse linear regression in statistical genetics

J. Soeding

Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

In the last $10 - 15$ years, datasets for hundreds of thousands of patients have become available to study the origin of complex diseases (coronary artery disease, Alzheimer's, depression etc.). For example genome-wide association studies (GWAS) measure the genotype of thousands of diseased and healthy patients to identify genetic risk variants that predispose to higher disease risk. Analysis tasks include finding the variants that are causal for an increased risk to develop a certain disease, or finding genes whose genetically predisposed higher expression in a certain tissue predisposes to higher disease risk. Since the number of training samples N is typically much smaller than the number of explanatory variables P, $N \ll P$, and we can assume that most of them have no effect, these tasks are addressed with sparse regression. Here I will report on our group's first Bayesian method for sparse multiple regression using a spike-and-slab prior on the effect sizes. We use Metropolis-Hastings MCMC with an adaptive proposal function, an undirected graphical network, to sample the active-set indicator vector. We will also give an introduction to stochastic variational inference, a modern alternative to MCMC, by explaining the similarities to adaptive MCMC at the example of sparse linear regression.

Laplace-based importance sampling

P. Wacker

FAU Erlangen-Nürnberg , Erlangen, Germany

A frequent task in inference is the one of sampling from a measure, for example if we have constructed a posterior measure from data and we want to interpret it. There is a large number of sampling methods and in this talk we will take a closer look at importance sampling. Importance sampling's performance is strongly dependent on a suitable choice of reference measure. Broadly speaking, the reference measure should be as similar to the measure of interest as possible while still being elementary enough such that we can explicitly sample from it. One possible choice for the reference measure is the prior measure but it is readily observed that this does not work for high-dimensional problems. For this reason we propose choosing the Laplace approximation to the measure of interest: It is both relatively close to the posterior measure and can be explicitly sampled as it is a Gaussian measure. In this talk, we will talk about the Laplace approximation, about importance sampling including scenarios where prior-based importance sampling does not work well and how Laplace-based importance sampling can improve computation.

Modeling protein ensembles with doubly intractable distributions

B. Eltzner

Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Molecular dynamics has become an important tool to model and understand protein dynamics. However, in some cases properties measured from a protein ensemble, like atom distances, are not correctly recovered in simulations. To remedy this problem, ensemble refinement methods have been developed and Bayesian Monte Carlo methods have been applied. We approach the problem from the maximum entropy point of view. The problem then presents as doubly intractable and thus requires sophisticated Monte Carlo methods for approximate sampling. In addition to an ensemble refinement, this approach also provides an energy refinement and variance estimates for the energy parameters.

The conditional particle filter - scalable MCMC for dynamic models

M. Vihola

University of Jyväskylä, Jyväskylä, Finland

Particle filters were developed in 1990s by the engineering community for the so-called 'filtering problem': on-line estimation of time-varying latent state based on a stream of noisy observations. More recently, particle filters have been combined with Markov chain Monte Carlo (MCMC) in order to perform Bayesian estimation of state space model (SSM) parameters and 'smoothing', that is, estimation of the latent state progress through time, given a fixed (but potentially long) observation record.

We introduce the particle filter, and then focus on one 'particle MCMC' method: the conditional particle filter (CPF). The CPF is a slight algorithmic variant of the original particle filter, but serves a different purpose: it defines an MCMC transition targeting the SSM smoothing distribution. The empirical evidence suggests that certain versions of the CPF mix well even in high dimension (with long observation records). We review some theoretical insights that consolidate such empirical findings, and justify why the CPF is often efficient for SSM inference.