Course Outline

Markov chain Monte Carlo (MCMC) algorithms are one of the most dynamic areas of research in modern statistics and are key to Bayesian inference. The recent trends in statistical modeling that focus on big models and big data render Markov chain Monte Carlo more challenging and motivate new exciting developments. This course will introduce and present the theory, methodology and practice of Markov chain Monte Carlo comprising the following indicative list of topics:

- Main concepts of Bayesian modelling and inference; uncertainty quantification; exploring the posterior distribution: the need for Markov chain Monte Carlo and related tools.
- The classical Monte Carlo computational method and its validity via probabilistic limit theorems.
- Markov chains, stationary distributions, reversibility and the Metropolis algorithm.
- Markov chain Monte Carlo, the toolbox of algorithms: Metropolis-Hastings, the Gibbs sampler, Metropolis adjusted Langevin Algorithm (MALA), Hamiltonian Monte Carlo (HMC), the slice sampler, hybrid algorithms.
- Markov chain Monte Carlo and related algorithms for multimodal distributions.
- Validity of Markov chain Monte Carlo: limit theorems, theoretical properties and practical implications.
- Optimal scaling and the adaptive Metropolis algorithm.
- Adaptive MCMC - theoretical background.
- Adapting the Gibbs Sampler.
- Adaptive MCMC for variable selection problems.
- Adapting increasingly rarely Markov chain Monte Carlo (AirMCMC).
- Markov chain Monte Carlo and Intractable Likelihoods.
- Pseudo-marginal MCMC.
- Approximate Bayesian Computation (ABC) and ABC-MCMC.
- Continuous time MCMC: the Zig-Zag.

The course will be illustrated with Bayesian inference examples and will be accessible to students with varied backgrounds. The students will be encouraged to familiarize themselves with R (or an equivalent programming language), implement some of the discussed algorithms and perform simulations.

Objectives

After the course the students will have acquired the following skills:

- Failiarity with mainstream MCMC theory and methodology and their applicability in Bayesian statistics
- Ability to collaborate in developing an understanding and investigating properties of a new methodology
- Ability to implement mainstream MCMC computational algorithms in R or another programming language
- Experience of reading and discussing code; understanding the benefits of producing clear well-written code
- Critically assessing someone else's code and work, giving constructive feedback
- Receiving critical feedback and using it constructively to improve work
- Ability to present analysis and research

Assessment

Students will be divided into teams of 2 or 3 to work on a miniproject. Each team will produce a report of maximum 8 pages, and return the project including code and graphs in the format of an R vignette or an overleaf LaTeX document. Each group will present their analysis at the end of the course. The report will be marked and there will be feedback on the analysis and presentation.

0.1 Timeline of the miniproject (tentative)

- week 1: students will be split into teams; a list of projects will be given; projects will be assigned to teams
- week 2: work on the the project
- week 3: first half - produce code and draft reports; submit for peer review by Thursday noon;
- week 3: provide constructive feedback to peer reviewed projects by Friday noon
- week 4: first half - youse feedbck constructively to revise code and reports; submit the final version for assessment by Wednesday noon
- week 4 second half - work on your presentation

Q&A about your project

- Q1: What do you need to submit?
  A: Code written as R package and report written as an R vignette (main text limited to 8 pages). Alternatively you can use a different programming language and write the report in LaTeX using a collaborative platform (e.g. overleaf). Please split the report in two parts: theory (where you explain the statistical and methodological context), and simulations

- Q2: What is the purpose of the project?
  A: The purpose of the project is to give you a taste what it is like to implement a fairly complicated statistical methodology of computational nature from a research paper, and to give you confidence that you will be able to do this in future, working over a short time span

- Q3. What do you need to demonstrate in your project?
  A: In your submitted project you need to demonstrate that:
  - you have understood the methodology of the paper, and what is its main contribution
  - you have understood what are the methodological difficulties or weak points of the approach
  - you have understood what are implementational difficulties of the methodology
  - you are able to carefully investigate a methodological and/or implementational aspect of the approach and discuss the results

- Q4: What are the "aspects" that can be investigated?
  A: For some projects I will tell you what is the aspect to investigate, but in general this depends on the project you choose, examples include:
  - how does the methodology scale with the size of data?
  - how does the methodology scale with the dimension of the parameter space?
  - is the methodology robust to violating some of the assumptions (especially unrealistic assumptions)?