BAYES 20 @LUND 19

A Mini-conference on Bayesian Methods at Lund University
7th of May, 2019
Palaestra, Lund University, Sweden
`bayesat.github.io/lund2019`

Program

09.00–09.55 Welcome and keynote presentation

▷ *Hierarchical models and their applications in astronomy*, Maggie Lieu, European Space Astronomy Centre, European Space Agency.

10.00–10.40 Session: Bayesian Learning

▷ *What cause successful learning in Bayesian methods?* George Moroz, National Research University Higher School of Economics.

▷ *Introducing Bayesian Stats through Signal Detection Theory,* Gerit Pfuhl, UiT The Arctic University of Norway, Department of Psychology.

10.40–11.00 Coffee break

11.00–12.00 Session: Bayesian Flexibility

▷ *Prior thoughts on mixed-membership models in linguistics,* Chundra Cathcart and Gerd Carling, Lund University, Centre for Languages and Literature.

▷ *A Bayesian method to localize lost gamma sources*, Antanas Bukartas, Christopher Rääf, Jonas Wallin, and Robert Finck, Lund University, Medical Radiation Physics.

▷ *Spatio-Temporal Reconstructions of Global CO2-Fluxes using Gaussian Markov Random Fields*, Unn Dahlén, Johan Lindström, and Marko Scholze, Lund University, Centre for Mathematical Sciences.

12.00–12.45 Lunch and mingle

12.45–13.30 Keynote presentation

▷ *Visualization for refining and communicating Bayesian analyses*, Robert Grant, BayesCamp.

13.30–13.50

▷ *Bayesian vs. Frequentism for experimentalists*, Jakob Lavröd, YPT Sweden.

13.50–14.00 Quick break

14.00–15.00 Session: Bayesian Decisions

▷ *Extending Bayes to Make Optimal Decisions*, Jonas Kristoffer Lindeløv, Aalborg University, Department of Communication and Psychology.

▷ *Rich-man's Monte Carlo: Uncertainty Analysis in Excel*, Dmytro Perepolkin, Lund University, Centre for Environmental and Climate Research.

▷ *Bayesian Deep Learning Applications in Biomedicine*, Nikolay Oskolkov, Lund University, Department of Biology.

15.00–15.30 Coffee and cake

15.30–16.50 Session: Bayesian Scattering

▷ *How to deal with a noisy zero – a simple Bayesian treatment for small angle neutron scattering* Alexander Holmes, European Spallation Source ERIC.

▷ *Bayesian inference of conformational ensembles from small-angle scattering data*, Wojciech Potrzebowski and Ingemar Andre, Lund University, European Spallation Source.

▷ *Bayesian determination of the effect of a deep eutectic solvent on the structure of lipid monolayers*, Andrew McCluskey, Andrew Jackson et al., Diamond Light Source / University of Bath.

▷ *Automatic Learning of Summary Statistics for Approximate Bayesian Computation Using Deep Learning*, Samuel Wiqvist, Pierre-Alexandre Mattei, Umberto Picchini, and Jes Frellsen, Lund University, Mathematical Statistics.

Keynote presentations

Hierarchical models and their applications in astronomy

Maggie Lieu, European Space Astronomy Centre, European Space Agency

Bayesian hierarchical models allow us to make inferences about populations based on many individuals. They can be particularly useful in astronomy where we have a lot of data, but the observations are often very noisy. It is not always easy to choose a principled prior for your problem, and event when you can it will dominate the posterior distribution if treated independently. Hierarchical bayesian methods model the joint posterior of the observations to self-consistently determine the prior from the data, improving constraints on data with large uncertainties. In this talk I will discuss the power of hierarchical models in astronomy, with case studies including noisy weak gravitational lensing and extragalactic proper motion.

Visualization for refining and communicating Bayesian analyses

Robert Grant, BayesCamp

Introductory statistics classes teach the value of basic plots to diagnose problems with regression models. Like most of us, I found these very boring and spent a long time avoiding them. Now that I focus on complex Bayesian models, I find visualization of variables, predictions and residuals more useful than ever. Analysts need to track down problems where the model is not fitting the data well, and also where the sampling algorithm is performing poorly, where the priors may be a bad choice, and where the likelihood may be a bad choice. In addition to this, Bayesian models fitted with simulation are fertile ground for visual communication of findings. In this talk, I will review relevant general principles of effective visualization, recent work on Bayesian workflow, and the role that interactive graphics have to play.

Contributed presentations

What causes successful learning in Bayesian methods?

George Moroz, National Research University Higher School of Economics

I would like to share the comparison of three groups of students that I taught Bayesian methods this year: (1) Mixed group (psychology, biology...) with good background in R and frequentist statistics; (2) linguistic group (3d year) with medium background in R and frequentist statistics; (3) further education group in Computer Linguistic with beginner R and no statistical background.

I expected first and second groups to be more successful than the third one, but the shocking result was that the third and the first groups were more successful than the second one. I will try to explain the obtained result.

Introducing Bayesian Stats through Signal Detection Theory

Gerit Pfuhl, UiT The Arctic University of Norway, Department of Psychology

To avoid starting with a formula a detour via utilized Signal detection theory (uSDT) familiarizes psychology undergraduates with some of the basic concepts in Bayesian statistics. uSDT includes payoffs (utility functions), base rates, and varies similarities, illustrated on perceptual decision processes. Payoffs and base rates influence bias, assisting students in understanding models and priors in Bayesian statistics.

Prior thoughts on mixed-membership models in linguistics

Chundra Cathcart, Lund University, Centre for Languages and Literature

Bayesian mixed-membership models are popular in linguistics, as they explicitly model contact between languages (Reesink et al 2009, Syrjänen et al 2016). Most linguistic applications use the biological Structure program (Pritchard et al 2000) with default presets, fixing the concentration parameter of the population-level Dirichlet prior over allele frequency (treated as an analog for the language-level prior over features) at 1. We show, using a cross-linguistic typological database, that there are linguistically meaningful consequences for the choice of this hyper-parameter (either fixed at different values or inferred from the data) using a series of posterior predictive checks designed for mixed-membership models (Mimno et al 2015).

A Bayesian method to localize lost gamma sources

Antanas Bukartas, Christopher Rääf, Jonas Wallin, and Robert Finck, Lund University, Medical Radiation Physics

Radioactive sources can sometimes be lost or misplaced despite the existing rigorous safety rules. Lost sources must be found as soon as possible to avoid inflicting harm to the public. Regardless of the type of equipment used it is desirable to use as much information as possible from the measurements to draw conclusions about the activity and location of a detected source. Using Bayesian inference it is possible to obtain a probability distribution for the position and activity of an unshielded gamma source in one pass with a mobile gamma spectrometry vehicle. The aim of this research was to investigate the feasibility of a Bayesian algorithm for mobile gamma spectrometry, test its accuracy in determining the location and activity.

Spatio-Temporal Reconstructions of Global CO2-Fluxes using Gaussian Markov Random Fields

Unn Dahlén, Johan Lindström, and Marko Scholze, Lund University, Centre for Mathematical Sciences

Atmospheric inverse modeling is a method for constraining Earth surface fluxes (sinks and sources) of greenhouse gases using measurements of atmospheric concentrations. The (linear) link between atmospheric concentration and fluxes are provided by an atmospheric transport model. Since the number of unknown surfaces fluxes is much larger than the number of observed atmospheric concentrations, the inverse problem is ill-conditioned. Requiring further assumption on the fluxes, leading to a Bayesian model. Historically, fluxes are discretized to a grid and modeled by a multivariate Gaussian distribution. Instead, we define the flux on a continuous spatial domain, with fluxes modeled as Gaussian Markov Random Fields, including both spatial and temporal dependence.

Bayesian vs. Frequentism for experimentalists

Jakob Lavröd, IYPT Sweden

A common rebuttal to Bayesian methods is that they are appropriate for large and complex problems (containing prior information and hidden variables), and most undergraduate teaching is often based upon frequentist methods. Starting from the context of the practical experimentalist, we explore the difference between the Bayesian and Frequentist methodology and highlights the advantages of teaching the Bayesian perspective already from the start. Many of the examples and the material come from the introductory course given for the students enrolled in the high school giftedness program "International Young Physicists' Tournament", an attempt at introducing Bayesian thinking into experimental practice.

Extending Bayes to Make Optimal Decisions

Jonas Kristoffer Lindeløv, Aalborg University, Department of Communication and Psychology

Utility Theory allow you to make optimal decisions in the face of uncertainty. For example, what bidding price would maximize your earnings, taking the chance of failure into account? Utility Theory latches nicely onto Bayesian Inference. Once you have a posterior distribution, you need only a few more lines of code to apply a utility function (aka loss function) and identify the decision that optimizes said utility. This approach scales well to more complex models and decisions. We will use R and rstanarm/brms for Bayesian inference and hand-code the utility. An R notebook with worked examples will accompany the tutorial.

Rich-man's Monte Carlo: Uncertainty Analysis in Excel

Dmytro Perepolkin, Lund University, Centre for Environmental and Climate Research

Adoption of uncertainty analysis in modern business environments is often challenging due to gaps in relevant skills and tooling (especially among decision makers). At the same time, Excel is ubiquitous and can be used to build decision maker's intuition about uncertainties. This talk will introduce typical business problem faced by businesses and organizations on daily basis and showcase a fully transparent formula-only Excel model with no dependency on external libraries or macros, that can enlighten and inform about the effect of uncertainties on business outcomes.

In this talk I will discuss framing, expert knowledge elicitation, modeling, visualization and communication of results to facilitate discussion about uncertainties and ultimately aid the decision making.

Bayesian Deep Learning Applications in Biomedicine

Nikolay Oskolkov, Lund University, Department of Biology

Next Generation Sequencing technologies gave rise to manifolds of Biomedical Big Data which is particularly manifested in the area of single cell transcriptomics where millions of cells are sequenced. Deep Learning (DL) is an ideal framework for analyzing large amounts of data and building predictive models for Clinical Diagnostics within the concept of Precision Medicine. Bayesian DL adds an important level of patient safety providing uncertainties to the biomedical predictions. Here, using single-cell transcriptomics data I demonstrate how Bayesian DL improves the accuracy of discovering novel cell sub-populations and dramatically outperforms classical methods when handling unknown cell sub-types.

How to deal with a noisy zero – a simple Bayesian treatment for small angle neutron scattering

Alexander Holmes, European Spallation Source ERIC

Neutron scattering measurements are an ideal case for Bayesian analysis – statistics are limited, measurement time is expensive and there is often relevant background information.

I will present an example of small angle neutron scattering from superconducting vortex lattices. Most of the signal detected is irrelevant and contributes nothing but noise to the final result. A Bayesian treatment of the results allows a huge enhancement of the signal to noise ratio and treats missing data sensibly. I will further argue that Bayesian analysis offers a huge amount to the scattering community and would benefit significantly from more systematic support from institutions such as the ESS.

Bayesian inference of conformational ensembles from small-angle scattering data

Wojciech Potrzebowski and Ingemar Andre, Lund University, European Spallation Source

Small-angle scattering (SAS) use x-ray or neutron scattering at small angles to investigate the structure of materials at the scale about 1-100nm. SAS is uniquely suited to study the conformational ensembles adopted by multi-domain proteins. However, analysis is complicated by the limited information content in SAS data and care must be taken to avoid constructing overly complex ensemble models and fitting to noise in the experimental data. To address these challenges, we developed a method based on Bayesian statistics that infers conformational ensembles from a structural library generated by all-atom Monte Carlo simulations. The method involves a fast model selection based on variational Bayesian inference that maximizes the model evidence, followed by a complete Bayesian inference of population weights.

Bayesian determination of the effect of a deep eutectic solvent on the structure of lipid monolayers

Andrew McCluskey et al., Diamond Light Source / University of Bath

We present a unique insight from Bayesian-driven modeling for a series of lipid monolayers at the air-deep eutectic solvent (DES) interface using reflectometry measurements. A chemically-consistent modeling approach shows that the lipid monolayers at the air-DES interface are similar to those on water while removing the need for water-specific constraints.

Furthermore, the use of Markov-chain Monte Carlo sampling enables the quantification of inverse uncertainties and parameter correlations in the modeling approach. Finally, we discuss limitations present in the use of Bayesian methods for reflectometry analysis and outline future work that will be conducted to overcome these.

Automatic Learning of Summary Statistics for Approximate Bayesian Computation Using Deep Learning

Samuel Wiqvist, Lund University, Mathematical Statistics

Learning summary statistics is a fundamental problem in Approximate Bayesian Computation (ABC). The problem of learning summary statistics is, in fact, the main challenge when applying ABC in practice, and affects the resulting inference considerably. Deep learning methods have previously been used to learn summary statistics for ABC. Here we introduce a novel deep learning architecture (Partially Exchangeable Networks, PENs), with the purpose of automatically learning summary statistics for ABC. Our case studies show that our methodology outperforms other popular methods, resulting in more accurate ABC inference for Markovian data.

Acknowledgement

The organizing committee of this year's Bayes@Lund consisted of Rasmus Bååth (King Digital Entertainment, Malmö), Ullrika Sahlin (Centre of Environmental and Climate Research, Lund University), and Alexander Holmes (European Spallation Source, Lund University), with help from Jakob Lavröd (YPT Sweden).

We are deeply grateful to the COMPUTE research school and the BECC research environment for financial support of this event.