First PlanetS workshop on Bayesian Statistics

Abstract Booklet

Bern University.
November 26th 2015.

Statistical Inference

Adam Amara (ETHZ)

- Probabilities: Joint, conditional and marginal; on the road to Bayes Theorem.
- Dealing with full probability density functions in the case where a likelihood can be written: grids, adaptive grids and MCMC.
- Linear models with Gaussian errors: generalised least square or regressions.
- Understanding the χ^2 distribution and what it can tell you about your data.

Basic MCMC: implementing the Metropolis Algorithm

Daniel Mortlock (Imperial College London)

Abstract: Bayesian parameter inference consists of calculating $\Pr(\theta \mid \text{data, prior})$, the posterior distribution of the parameters, theta, conditional on whatever data is available (and some prior knowledge). In many situations it is hard to guess from the data what values of parameters are favoured, i.e., where the function $\Pr(\theta \mid \text{data, prior})$ peaks and how wide the peak(s) is/are. One generic way around this is to use a set of samples from this posterior distribution, θ_i, in place of the function itself, but some algorithm is needed to generate these samples. A simple, but very general, option is the Metropolis algorithm, one of the most basic forms of Markov chain Monte Carlo (MCMC). The Metropolis algorithm is a guided random walk through parameter space that focuses on the theta values for which the likelihood is highest and can produce samples from the posterior given only the ability to evaluate a function that is proportional to $\Pr(\theta \mid \text{data, prior})$, the standard option being the product of the likelihood, $\Pr(\text{data} \mid \theta)$ and the prior, $\Pr(\theta \mid \text{prior})$.

This lecture will describe the basics of the Metropolis algorithm, giving all the ingredients necessary to implement it and do full Bayesian parameter estimation for ”moderate” problems (particularly where there is just a single peak in the posterior and there are less than ten parameters). For more challenging problems it will generally be necessary to use some more bespoke software packages, or other algorithms like Hamiltonian Monte Carlo and nested sampling (some of which are covered in other lectures at this school).

Performing Bayesian Model Selection with Nested Sampling

Farhan Feroz (Cambridge)

Abstract: Astrophysics and cosmology have increasingly become data driven with the availability of large amount of high quality data from missions like WMAP, Planck and LHC. This has resulted in Bayesian inference methods being widely used to analyse observations, but they can be extremely computationally demanding. I will discuss the challenges involved in performing Bayesian model selection and discuss how it is done in practice. In particular, I will describe the MultiNest algorithm, which is based on a Monte Carlo technique called Nested Sampling. MultiNest has been applied successfully to numerous challenging problems in cosmology and astroparticle physics due to its capability of efficiently exploring multi-modal parameter spaces. MultiNest can also calculate the Bayesian evidence and therefore provides means to carry out Bayesian model selection. I will also review some of the recent applications of MultiNest in astrophysics.

A journey through Gaussian random fields with a view towards Bayesian optimization

David Ginsbourger (IDIAP / Uni Bern)

Abstract: Gaussian random fields have been used as a flexible and practical family of priors in the context of Bayesian statistics when the parameter of interest is functional. In particular, these priors are now massively used in machine learning as a building block of adaptive strategies for approximating and optimizing functions based on limited evaluation budgets. We will review the basics of GRFs and give an introduction to Bayesian optimization.

Outline:

- Basics of Gaussian random fields (GRFs).
- On the use of GRFs in the context of Bayesian statistics, e.g., in curve fitting.
- Introduction to Bayesian optimization of deterministic functions relying on GRF priors.
- Discussion on some applications of Bayesian optimization relevant to physical modelling.

Conservative estimate for the excursion set of a deterministic function under a Gaussian random field model

Dario Azzimonti (Uni Bern)

Abstract: The problem of estimating the excursion set of a deterministic function under a limited evaluation budget can be approached with Gaussian random field (GRF) modelling. Here we review two recent techniques based on Gaussian random field priors and we propose a fast algorithm to compute conservative estimates from the joint distribution of the posterior field.

Outline:

- Introduction to GRF modelling in the context of estimating excursion sets of deterministic functions;
- Definition of conservative estimates of excursion sets;
- Conservative excursion set estimate using a novel MCQMC algorithm.

The role of model selection, summary statistics and multiple-point statistics in Bayesian inverse problems: Examples from geophysics

Niklas Linde (UNIL)

Abstract: Model selection and multiple-point statistics allow for comparisons of alternative conceptual descriptions of the system under study, while summary statistics (Approximate Bayesian Computation) allows us to relax some of the assumptions made in terms of observational, modelling and prior uncertainties. These concepts will be introduced and motivated with examples from geophysics.

Gaussian processes for modelling stellar activity and detecting planets

Vinesh Rajpaul (Oxford)

Abstract: To date, the radial-velocity (RV) method has been one of the most productive techniques for detecting extrasolar planetary candidates. Unfortunately, stellar activity can induce RV variations which can drown out or even mimic planetary signals, and it is extremely difficult to model and thus mitigate these stellar effects. This is expected to be a major obstacle to using next-generation instruments to detect lower mass planets, planets with longer periods, and planets around more active stars. Enter Gaussian processes (GPs), which have a number of attractive features that make them very well suited to the joint modelling of stochastic activity processes and dynamical (e.g. planetary) signals. In this talk I will present briefly a GP framework developed to model RV time series jointly with ancillary activity indicators, allowing the activity component of RV time series to be constrained and disentangled from planetary components. I will talk briefly about how it is used in practice, and demonstrate its performance using both synthetic and real data sets.