Bayesian Statistics and Probabilistic Programming – Spring 2023

Course project guidelines

Goals, extension, format

In addition to the merely academic purpose of grading, the course project should be an opportunity of practicing and orienting several skills to the common aim of gathering and sorting information on a specific subject and, then, composing a short, but self-contained and informative report on it.

An orientative figure for the expected length could be around fifteen pages, **plus code**. All submitted documents should be packed together as a .zip file, including:

(1) A written report in pdf format is required, including an abstract, a context introduction where the project scope is delimited and its goal stated, then a well-structured text body, plus full references, for articles or books, following the usual conventions in scholarly documents, for online documents specify an accessible, unbroken link. Difficult-to-find documents should be included in the submitted archive.

The length figure is just a *roughly orientative* estimate, meaning that a three page report, say, will certainly be deemed as much too short, whereas a forty five pages report would be an undue anticipation of your Master thesis.

(2) Code (a jupyter or Rmd notebook) is also required and should be fully operational, including data or pointer to an online source thereof. Any needed auxiliary material should be properly mentioned and described: I will run your submitted code expecting no errors or suspicious warnings. Try to be exhaustive, up to the point of fastidiousness in documenting code. It will be judged (and graded) better the more detailed, illustrative and clarifying you manage to be in this respect.

Both **individual or group projects** are possible. In collective projects each author is responsible of and should be able to answer questions about the

whole project. Meaning there should be no allocation of different project parts to different team members.

Send email with your intended project topic (and, if applicable, the list of group members) so I can give you the go ahead.

Possible subjects

In principle, any Data Analysis topic is possible, provided it is connected with, or seen from a Bayesian and/or Probabilistic Programming perspective. Thus if you know of, or have seen some possible subject, you can propose it by sending me an email. As possible suggestions of lines along which you can search for a theme, I propose the following classification:

A) Methodology-oriented projects:

1. Sampling schemes/approximation methods. Variants and extensions of sampling or approximation methods seen -or to be seen this month- in class. Some of these: Integrated Nested Laplace Approximation (INLA), Variational Inference, Langevin Markov Chain Monte Carlo, Adaptive MCMC, Reversible Jump MCMC.
2. Other models: Bayesian versions of Classification and regression trees (CART), or clustering, or neural networks, or Principal Components Analysis and other multivariate analysis techniques, or other aspects of inference: Bayesian model selection (in particular predictor selection, e.g. *spike-slab* method). Model averaging.
3. Centered on a probabilistic programming language, e.g., Greta, Edward(2), Pyro, Tensorflow-probability, Turing.

B) Data-oriented projects:

Centered on analyzing a given dataset or collection of datasets. Such project should include treatment applying several methods, with a detailed comparison and discussion of their respective strengths and weaknesses. Code should be thoroughly documented.

If in your job or research work you are using Bayesian Statistical methods, a possible course project is an informative description of some relevant problem and its treatment.

Caution

Beware of too ambitious projects which, as stated above, should better go to your Master thesis.

Projects in previous years

This list is just a glimpse of some Course Projects which have been submitted in previous editions of the course. By no means it is intended for you to pick these same subjects.

Some highly specialized titles are due to data or methods borrowed from a Master thesis or other research available to the authors.

- Prediction of football results by Bayesian Poisson regression.
- Bayesian Gausian mixture models. Comparison in JAGS, TensorFlow Probability and Edward.
- Bayesian hierarchical spatial models with INLA for the incidence of COVID-19 across Basic Health Areas 4in Catalonia.
- Prediction of number of days in a hospital ICU as a function of patient age and admission type.
- From the NBA Player Salary dataset for the 2017-2018 season, predict salary as a function of several performance metrics with different methods and computational tools.
- Bayesian Clustering: Finite and infinite Mixture Models.
- Bayesian estimation of the GARCH(1,1) model with Normal innovations.
- Introduction to Variational Inference.
- Bayesian acquisition functions for hyperparameter optimisation.
- Implementation of a Bayesian Neural Network for automated sleep scoring in rodents.
- Bayesian predictor selection.
- Impact of the Spanish COVID-19 lockdown in Twitter user activity.
- Bayesian AB testing
- Regularization from a Bayesian viewpoint.