Maastricht University Leading in Learning!

ASSIGNMENT

Introduction to Software in Econometrics and Operations Research (EBS2072)

Academic Year

Coordinator: Dr. Nalan Baştürk
Department of Quantitative Economics, Maastricht University
Contact: n.basturk@maastrichtuniversity.nl

Office: TS53 A3.15, by appointment

Contents

Introduction

The assignment deadline is Monday, January 30th, 2023, 16:00 hours. Late submissions will be penalized.

Study material: Tutorial exercises and study material mentioned in the tutorials are the first batch of study material for this assignment. An additional study material is Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. Journal of the Royal Statistical Society: Series A (Statistics in Society), 182(2).
`https://research.monash.edu/en/publications/visualization-in-bayesian-workflow`

Reminder on deliverables: You should submit your assignments individually via the assignment on the student portal.

- All files you submit (R functions and your pdf file) should include your name and student number on top of the file. As an example:

(.pdf file)	*(.R file)*
Assignment, EBS2072 Jane Roe Student number: 1234567	# R code for assignment EBS2072 # Author: Jane Roe # Student number: 1234567

- If you use R libraries within your code, you do not need to submit these library functions. However, you should specify in the code which libraries or R packages are used with a command `library('XX')`.
- Make sure that your R script runs from an empty workspace.
- Each R script you submit should have an explanation of what the script aims, and brief comments to make sure that your code is easily readable. See provided R scripts for the tutorial meetings as examples.
- Not adhering to assignment submission rules will be penalized.

See the course book on Canvas for further details.

Structuring your report: The following should be included in your report.

- Data definition. The data can be simulated or real data which are appropriate for the model and the analysis. Your data definition should include a plot of the data and/or summary of data properties (e.g. number of observations, data source, interesting properties of data that affect your econometric model choice etc)
- The econometric model.
- Definition of the selected prior / priors and explanation of why you choose a specific prior.
- The simulation algorithm you use for the analysis.
- Important computational aspects of your simulation algorithm. E.g. did you use burn-in draws? Why/why not? Which simulation algoritm did you use (e.g. Nuts sampler).
- Results of the simulation algorithm. Traceplots of parameter draws, posterior mean, standard deviation, 95% intervals for the parameters of interest (this depends on your research question).
- Answers to the specified tasks below.
- General conclusion about your findings. I.e. what your conclusion is in relation the research question you started with and/or answers to the sub-questions in the topic you choose.

Data: You may prefer to simulate data and analyze simulated data or analyze a real dataset. Some useful resources for freely economic data are included below:

- https://www.economicsnetwork.ac.uk/links/data_free
- https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/00Index.html
- https://www.rpubs.com/mperlow/552127

Tasks

Complete the sub-tasks below. For each sub-task, also include some explanation in your pdf report. You can insert your answers in *Rmarkdown* if you prefer to generate your report using *Rmarkdown*.

1. Choose your datase, dependent variable and independent variables. You can choose to generate data from a given model or analyze a dataset, for example from the list given above. Summarize the dataset, its general properties such as the type of dependent variable, number of observations and other relevant statistics. Write down a research question for these data.

    - For real data, a research question can be 'What is the effect of x on y?'
    - For simulated data, a research question can be 'Can we retreive the true parameters using Bayesian inference?'

2. Choose the logit regression or a linear regression as your model. Explain why this model is chosen for the data in the previous section. If none of these models is appropriate for your data, it is possible that you choose a rather difficult dataset in terms of model choice. Re-consider the choice of your data in sub-task 1.

3. Define the priors of your model. Clearly motivate your choice of priors.

4. Estimate the model using RStan. Clearly state the choices you made for estimation.

5. Report the posterior results from your estimations.

6. Replicate steps [3-5] for half of your dataset. Compare the obtained results to those obtained in sub-task 5. How does the sample size affect the results you obtained? Are the differences or similarities in results intuitive?

7. Report the convergence properties of the estimation in sub-task 5. Do you think the algorithm converged? Are there any indicators of a badly mixing algorithm?

8. Perform a preliminary data analysis to see if the model you choose in sub-task 2 can be extended. If yes, propose an extended model. If not, explain why the model you had in sub-task 2 is the best possible model for the dataset you have.

9. (Optional) Estimate the model in sub-task 8. Perform posterior predictive checks to see if the models in sub-task 2 and sub-task 8 are suitable for the data. Based on the posterior predictive checks, does one of these models perform better than the other?