Assignment

25 Januari 2020

Assignment description

In this assignment you will apply the Bayesian approaches taught in this course in a small research project.

- Come up with a dataset that is suitable for analysis with a linear regression model with (at least) two predictors
- Formulate (a) research question(s) that can be answered using one or multiple of the approaches presented in this course
- Program Bayesian linear regression model(s) of interest in R

1. Think about and specify prior distributions for your parameters
2. Use JAGS *(maximum 2 out of 10 points)* or program your own Gibbs sampler in R *(maximum 10 out of 10 points)*
3. Include a Metropolis-Hastings step included for at least one parameter *(JAGS or pre-existing R-script: maximum 2 out of 10 points; program yourself: maximum 10 out of 10 points)*
4. Assess convergence of the model *(maximum 5 out of 10 points for using available functions/code; maximum 10 out of 10 points for programming it yourself)*
5. Check a model assumption of linear regression by means of a posterior predictive p-value *(maximum 2 out of 10 points if you use pre-existing model assumption measures; maximum 10 out of 10 points if you program your own test statistic)* Do not forget to explain why your test-statistic has power against the model assumption you want to evaluate, that is, why will the test statistic have large values when the model assumption is violated.
6. Obtain parameter estimates, credible intervals and interpretation
7. Compare multiple models by means of DIC *(maximum 5 out of 10 points if you use the built in DIC function; maximum 10 out of 10 points if you program your DIC calculation)* and Bayes factor*(maximum 5 out of 10 points if you use Bain or other ready-to-use Bayes factor packages; maximum 10 out of 10 points if you program your own Bayes factor calculation)*

- Provide an interpretation of the results obtained using 1 through 7 and answer your research question(s).
- Consider the differences between Bayesian and frequentist analyses, do not do this "in general", make certain that your answer is relevant in the context of your research question(s) and your data.
- Other: 5% of your final grade is reserved for surprising content, exciting, extra or unique content. Examples are, but are not limited to, a small simulation study or a sensitivity analysis, or other output that shows additional individual effort for any of the required parts of the assignment.

Hand in a report of **at most** 6 pages, where you present and discuss the elements of the assignment presented above. When a scientist submits a paper to a journal with the hope of getting it published, it is never clear whether he will be successful. There is no rule book or list of criteria which, if adhered to, will ensure that the paper is published. The scientist will essentially have to come up with his own criteria: is the topic interesting, are the equations correct, are simulations correct and relevant, is there an interesting and correctly evaluated example, etc. The same holds for YOU when you hand in your assignment. YOU have to be the judge of what you will and won't include, and the lecturers will be the judge of whether you show sufficient understanding of Bayesian statistics and to which degree. This means that you should address all aspects of what you learn during this course, but do not have to be exhaustive in the examples you provide per aspect. On the other hand, journals usually do describe which kind of papers is suitable for them. Following the grading scheme you know how you will be evaluated.

What do you need to hand in?

Send an email to both h.hoijtink@uu.nl before **27 May 11:59 PM** with a zip file containing the four relevant files (**Do not send more than 4 files!!**):

- a report of **at most** 6 pages (it has to be a .pdf file). Use 1 page for your name, e-mail address, introduction of the data set, and elaboration of your research question(s) (all on the first page, no other material can be included on the first page). Use .5-1.5 page for each of "estimation, MH, convergence, interpretation of estimates and intervals"; "posterior predictive check", "model selection using the DIC", "model selection using the Bayes factor"; and "comparison of Bayesian and frequentist approaches".
- R code that can be used to replicate your presented results AND to check whether the steps are executed correctly. Very important: collect all the code in one file; use not more than 500 lines of code in total; only provide the code that renders results that you present in your report; highlight in the code and in the readme.txt (see below) which lines of code have to be executed to obtain which results that you present in your report.
- if it comes as a separate file: the dataset you used
- a readme file that explains what can be found in the zip file and which lines of code have to be executed to obtain which results presented in your report.

You have to hand in original work, that is, work executed by you, based on your ideas, and your choices. You cannot use code written by the other students, you cannot use a test-statistic that is used by the other students, you cannot apply model selection to models that are also considered by the other students, etc. etc. However, you can of course discuss Bayesian and frequentist statistics with the other students, as long as the work you hand in is yours and original.

Regarding the report

- Only the first 6 pages you hand in will be considered, so hand in only 6 pages!
- Use standard margins, single line spacing, font Calibri size 11, or equivalent
- Figures, references, tables, your name, etc. are all part of the 6 pages! Make choices in what to present, and keep readability in mind.

Regarding the R code

- Make sure your results are reproducible (think about setting a seed)
- The report is guiding in the grading, the R code is supplementary (to check whether the calculation is correct, and the output comes out the same).
- Only submit code that is used for output in the report. You probably work with more code while working on your assigment, but hand in only those pieces of code that create the output presented and discussed in the report! And: a maximum of 500 lines of code.
- Make sure the code is annotated, i.e., it is clear what a block of code does

Regarding the dataset

- Make sure it is clear how the data can be read in
- If a dataset is altered or subsetted, make sure this is clear

Regarding the documentation

- Organise your files and code. Make sure it is clear what can be found in the files, how to read in data, run the analysis, using a readme.txt file