STAT 578 — Advanced Bayesian Modeling — Fall 2017

Project Information

General

This project will give you an opportunity to apply the Bayesian methods you have learned to a data set of your choice. You will develop a non-trivial Bayesian model, which you will fit and assess using Bayesian computational tools. You will have the opportunity to work and exchange ideas with your classmates.

Key parts of your project will be formation of your project group, submission of a written proposal, presentation of your work to your classmates, and submission of a written report.

You should individually expect to spend at least 20 hours on the project, not including time required to prepare documents or videos.

Groups

You will work on your project in a group of 2 to 4 people, all of whom must be officially enrolled for credit in this course at the time you form your group.

Please try to find partners. A suggested method is to post your project interests on Piazza using the `project_ideas` folder. If you cannot find partners, please ask the instructor to assign you a group. To stay on schedule, your group should be formed by October 11.

Once you have formed a group, have someone post a private message to the course staff on Piazza, using the `project` folder, stating the names and e-mail addresses of all group members.

All submitted work must be jointly produced by everyone in your group. Please attempt to structure your work such that everyone makes an equal contribution. Each group member must make sufficient effort. A breakdown of individual contributions is to be included in the written report.

Topic

Your group's first task will be to decide on a data set and modeling approach. Your group will perform a complete and non-trivial Bayesian hierarchical model analysis of an interesting real data set. For computation, you should use Markov chain Monte Carlo.

Your group may choose any real data set, with the following restriction: It may *not* be from a textbook or software manual, nor from any web sites associated with them. You are encouraged to find data of particular relevance to your occupation or area of study. However, please make sure that you can obtain permission to use and present any data that may be restricted (confidential or proprietary).

You will not have studied all of the course topics by the time you submit your proposal, and you may even need techniques that go beyond what is covered. For this reason, you may want to consult the example analyses in the OpenBUGS manual: `http://www.openbugs.net/w/Examples`. There you may be able to find an analysis similar to one that will suit your purposes. (Please note that OpenBUGS code usually needs to be altered to run in JAGS.)

Some questions to consider when choosing a topic:

- Is this an actual (not artificial) data set? Has it been pre-processed?
- Are there any interesting or important questions that can be answered with this data?
- What is known about the circumstances surrounding collection of this data?
- Is there existing literature documenting the data or offering an analysis of it?
- Is any data missing or incorrect? What should be done about that?
- What hierarchical statistical models might reasonably apply to the data? (Remember, a hierarchical model has more than one level of priors.)

Choice of data and modeling approach is subject to the approval of the course staff.

Proposal

Your group will submit a one- or two-page written summary of your research questions, data, and preliminary plan for analysis, complete with expected software and any relevant references. The description should be as specific as possible, subject to the page limit.

You are invited to post an early version of your proposal on Piazza using the `project` folder, for a preliminary review by your classmates. Please do so by October 18 to allow ample time for comments. After that date, you should spend some time reviewing other proposals and offering comments. Those proposals and comments from your classmates may give you some ideas to help you shape your own final proposal.

Final proposals are due to Coursera, by October 25. Late submissions require approval of the instructor, or a penalty may be imposed.

Proposals should include names and e-mail addresses of all group members and designation of one member as a contact person for correspondence.

You will receive a review with any comments, suggestions, or required modifications. If modifications are extensive, you may be asked to re-submit. Aspects of your review may be relevant to the grading of your written report.

Presentation

You will submit a voiced slide presentation (e.g. PPT) or something similar, uploaded as a video.

Your group's presentation must include the names of all group members. It must be between 5 and 10 minutes in length, and should therefore focus on the main ideas that might be of interest to your classmates and leave less important details for the written report. Effective presentation techniques include using pictures, graphs, illustrations, and examples.

Your group should submit your presentation via Illinois Media Space at `https://mediaspace.illinois.edu/`. You will need to log in (Active Directory) and upload your presentation: Select "Add New" and then "Media Upload" (at the top right corner of the page). You will be asked to fill in a form describing the uploaded video. Make sure to:

1. Use the tag "STAT578-Fall-2017".
2. Add all group member names to the "team members" field (using "Add" to expand the form for additional members).
3. Check "unlisted" as the type of upload at the bottom (which would allow others to see your video).

After uploading, your presentation should appear under "My Media". Make sure to test it. If you are satisfied, click "Share" beneath the video-playing frame. This gives a web link to your presentation, which you should submit to Coursera.

Your presentation *must* be uploaded by December 8, with the web link submitted to Coursera.

You should expect your presentation to be peer graded. You will rate several presentations (to be assigned), and several other students will rate yours.

Grading will be based on following instructions (uploading on time, including group member names, making sure the length is right), degree of clarity of presentation, and how affirmatively the following questions can be answered:

- Are questions of interest clearly stated (at least two)?
- Is the nature of the data set described?
- Is the Bayesian hierarchical model clearly specified?
- Are relevant results from the analysis presented?
- Are the questions of interest addressed in the data analysis?

In general, you should leave more detailed information for the written report.

Your score on the presentation portion is subject to your submission of ratings for all the presentations you are assigned to review. This is an individual responsibility. Your ratings must be submitted by December 15.

Written Report

Your report should be typed and in at least 11 point font, with at least one inch margins on all sides. One-and-a-half line spacing (50% leading) is recommended. The main text portion (excluding references and appendices) may not exceed 10 pages.

It is recommended that you structure your report as follows:

1. Purpose: brief motivation for data collection, questions of interest or importance
2. Data: source, how collected, structure, important variables and their meanings, any implications for data analysis, useful summaries (perhaps graphs), any pre-processing or reduction performed (with justification), references and previous analyses (if any)
3. Model: full mathematical specification (data model and hierarchical priors)
4. Computation: summary of important computational details (initialization, number of chains, convergence diagnostics, sample size and Monte Carlo error)
5. Model Assessment: posterior predictive checks and (possibly) sensitivity analysis
6. Results: explanation of your findings, answering the questions of interest using Bayesian inference techniques
7. Contributions: a sentence for each member of your group detailing their contribution (including to the proposal, presentation, and written report)
8. References: full citations
9. Appendices: data-related details, important computer code and output (JAGS or R, carefully documented), other computing details, any non-essential items you choose to submit

Your group will upload the report to Coursera by the deadline of December 17. Its score will be subject to a penalty of 10% (of the possible total) for each day it is late.

Important Dates

October 11	Group should be formed
October 18	Preliminary written proposal to be posted on Piazza
October 25	Written proposal due (possible penalty if late)
December 8	Presentation due (hard deadline)
December 15	Deadline to review other presentations
December 17	Written report due (penalty if late)

Grading

The project determines 20% of your final composite score in the course, allocated as follows:

Proposal: 10% for a complete and timely written submission.

Presentation: 30% Peer graded (rubrics to be provided).

Report: 60% Graded by course staff, allocated as follows:

Specifications: 20% adherence to the report requirements and expectations, including addressing the proposal review

Quality of report: 20% clarity, conciseness, structure, focus, efficiency

Quality of analysis: 20% sophistication, correctness, completeness, appropriateness, value of results

All group members receive the same project score, provided everyone contributed substantially, and provided everyone completed all individual project responsibilities.