
MA 578: BAYESIAN STATISTICS

Final Project

Proposal Due Date: November 7th, Beginning of the class

Final Report Due Date: December 12th, Beginning of the class

Instructor: AmirEmad Ghassami

Department of Mathematics and Statistics

Boston University

Fall 2023

- An individual project constitutes 25% of the overall grade for the course.
- You will analyze a real-world dataset using Bayesian techniques.
- By November 7th, you are expected to submit a concise 1-page project proposal.
- Upon approval of the project proposal, a comprehensive 5-page project final report is due on December 12th. This 5-page limit encompasses all appendices and figures, excluding references and code.
- You should turn in a .pdf file for both the project proposal and final report.

Good places to find interesting datasets include

- <https://github.com/awesomedata/awesome-public-datasets>
- <https://www.icpsr.umich.edu/web/pages/ICPSR/index.html>
- <https://data.europa.eu/data/datasets?locale=en&minScoring=0>
- <https://www.kaggle.com>

The project evaluation will consider the ambition, significance, originality, technical depth, results, relevance, writing quality, overall complexity of the goals outlined in the project and the degree to which those objectives are achieved.

Your final report should have the following components:

1. **Introduction** (10 points)

- Clearly describe the problem. Describe why the problem you are trying to solve is important, useful, and interesting.
- Describe/visualize the dataset that you are using: How is the data collected, what is the potential defect in the data, etc.

2. **Method and analysis** (40 points)

- At least two different models with different sampling models and priors should be fitted. Each model should be appropriate and well-motivated. For instance, a Poisson model should not be used for a continuous variable.
- Clearly describe your statistical analysis. Describe the method you are using for inference. For example, your loss function, your estimator, your predictor, approximation methods that you use, etc.
- Sensitivity analysis with respect to prior choices and model checking should be performed.
- MCMC should be used in your analysis.

3. **Presentation and interpretation of the results** (20 points)

- Visualization should be used to report findings. Results and outputs (e.g., for posterior predictive checks, sensitivity analysis, point estimates, etc.) should be properly interpreted not merely stated.
- Your report should have a conclusion section where you state what you learned in doing this project.

4. **Quality of the report's composition and presentation** (30 points)

- Report should be clear, well-written, and use academic language. It should be well-organized and properly divided into different sections and subsections.

Examples

- Here are some examples:
 - <https://mc-stan.org/users/documentation/case-studies>
 - https://github.com/stan-dev/stancon_talks
- The purpose of these examples is to inspire project ideas and offer insights into report presentation and techniques that could be useful. It's important to note that these examples may not encompass all the necessary workflow steps for your report, and certain sections may contain much more detail than what you need to provide in your project.

Project Proposal (due November 7th)

- You should essentially provide a picture of how the completed project will look like
- You should state:
 - Clear description of the problem
 - Describe why the problem is important, useful, and interesting
 - Describe the dataset that you will use
 - A high level description of how you will analyze the data and models you may use
 - A high level description of results you intend to report