Stat 238, Fall 2025
Project Minis

Alexander Strang
Due: by **5:00 pm** Friday, April 4th, 2025

Mini Project 6: Structured Models

Policies

- This is a group project. You may work with up to 2 partners.
- All submissions must be properly type-set and sourced. This means providing formal citations and a complete bibliography when sources are used.
- Project minis should be uploaded to Gradescope under the appropriate mini-assignment.
- Please export any notebooks as a single pdf and merge all components into one file. Only one group member should submit to gradescope. They must tag the other members when they submit.

Prompts

This is a model and inference project. You will be asked to select an inference problem of interest to your group, then propose either a generalized linear model as a likelihood, or a hierarchical model as a prior. You are welcome to look at chapters 5 and 16 in BDA for examples, but should select your own model specification.

(a) (8 points) Select an inference problem to study. The inference problem must admit a reasonable Bayesian model with either a structured likelihood (e.g. a generalized linear model), or a structured prior (e.g. a hierarchical model). This model must be relevant to the example, and of interest. Interest can be motivated by, (1) necessity (e.g. the model has the minimal components needed to perform inference), (2) standard practice (if so, proceed skeptically), or (3) curiosity. You should be able to clearly articulate why the model is interesting to you, course relevant, and appropriate for the problem you selected.

You may look for examples in BDA, in Bayesian Methods for Data Analysis (Carlin and Louis), or the reference source list on Ed. You are also welcome, and encouraged, to use an inference problem from your own research work. If you get stuck, please contact us on Ed. We can suggest sources, or point you to the course librarian.

In your write-up please:

1. Clearly explain your inference problem, its importance/your interest in it, and its course relevance. Clearly motivate Bayesian methods.
2. Clearly cite your source.
3. Explain the statistical model used in the source, then motivate the assumptions made. Place each with respect to the three modeling aims discussed in class (veracity, simplicity, tractability). This is a chance to think critically.
   - If the model uses a structured likelihood, clearly identify the model components. For example, given a generalized linear model, identify the features, link/inverse link, and conditional distribution of the data given the parameter predicted by passing the linear combination of the features through the link. How were the features, link, and distribution motivated?
   - If the model uses a structured prior, clearly identify the components. Discuss any exchangeability assumptions made, or mixture models used. Explain how the model structure shares information between observations. How were the distributions at each stage of the model motivated? How did the authors choose when to stop adding priors? (Standard answers are: stop when you run out of exchangeable variables to group, stop when you reach a prior that you know empirically, stop when you reach an uninformative prior, or stop when you think that further modeling is counterproductive.)
4. Check that the data used for inference is available. If it is not, but you can find an equivalent data set, cite the alternate data source, and explain why any differences in format may be ignored, or accounted for with model adjustments.

(b) (6 points) Implement the model. You do not need to implement everything from scratch. You will likely find it easier to adopt a probabilistic programming language (see, for example, pymc3, bambi, arviz). You may not borrow code provided by the authors. You are welcome to use it as reference if attached as supplement to the source, but should attempt to build your own implementation.

(c) (6 points) Try to replicate an inferential analysis performed in the source you selected. Note: replication is often frustratingly difficult, even with extensive documentation. Try to make this work. If it doesn't work, document your efforts as clearly as possible. This part will be graded based on demonstrated effort.

Your write-up should discuss:

1. What you did to replicate the study. Discuss any information missing here, and what you did to fill in gaps in the documentation.
2. Whether or not you could replicate your source's results. If you couldn't, explain why you think your results differed from the published results.

(d) (4 points) Describe a follow-up study you would be interested in performing based on your experience in this project. Alternate studies could include changes to the model architecture, application to related data sets, or changes to the inferential aim. Clearly provide one research question in the style of a proposal abstract.