Bayesian Data Analysis Guide

Fletcher G.W. Christensen

1 Understand the Data

For your response variable, ask:

- What was measured?
- How was it measured?

Then, for each predictor variable, ask:

- What was measured?
- How was it measured?
- How does this variable relate to the other variables in the study (response and predictor)?

Are there any missing observations? If so, ask:

- Is there partial information on those observations?
- How is missingness coded in the dataset?
- Is *missingness* related to any of the other variables in the study (response or predictor)?

Are there any outliers / unusual observations? If so, ask:

- Are these observations correctly recorded?
- Are these observations from my population of interest?

2 Understand the Science

What is the scientific goal being pursued?

For what scientific goal were the data collected? Does this impact investigation of the current scientific goal?

How does the study build on other work in the same area? What is new in this study? What is already well understood?

3 Pre-Modeling

What type of analysis is desired? (Exploratory, model construction, model validation, prediction)

What are the critical summaries that best address the scientific goal? (NB: Preliminary decisions should be made here *before* models are composed.

What do the data look like?

- Plot the data and view summary statistics
- Create a scatterplot matrix and/or a correlation matrix for all variables
- Identify factor variables
- Identify which variables are candidates for transformation

4 Compose the (Full) Model

What sampling model best fits these data?

- Distributional model
- Simple linear model
- Generalized linear model
- Other

How exchangeable are the data?

- Fully exchangeable $\longrightarrow$ Basic models
- Exchangeable within subsets $\longrightarrow$ Mixed effects models
- Not exchangeable $\longrightarrow$ Correlated data models w/ informative correlation structure

What parameters are necessary to specify the desired model?

What prior information can we assume about these parameters?

- Use informative priors wherever possible.
- Use BCJ method / partial prior information when direct elicitation is untenable.
- When all else fails, use reference priors—*but use them cautiously!*

5 Post-Modeling

Assess quality of fit:

- Try using your model to predict a new dataset, and compare your original dataset to the new dataset using summaries like the empirical CDF and density estimation plots.
- Does your predictive dataset look like your original data, or are there important feature differences your model hasn't accounted for?
- Check for over/underdispersion.
- Check for heteroscedasticity.
- Re-compose full model until fit seems adequate.

If variable selection is called for, create appropriate reduced models, fit them, and compare to full model.

Perform inferences as necessary, focusing on pre-specified critical summaries.

Perform sensitivity analyses relative to priors and outlier inclusion to investigate sensitivity of results to these choices.

6 Write-Up

Based on understanding of scientific goals, decide what results are most important.

Choose figures, tables, and summary statistics that best highlight important results. Remember to focus on critical summaries, not just naive parameters.

Explain *important* parts of the analysis process, but skip past *unimportant* parts.

- E.g. report details about sensitivity analyses if they show that results *are* highly sensitive to prior and outlier-inclusion choices.
- If results *are not* highly sensitive to these choices, report that sensitivity analyses were performed without going into great detail.
- E.g. If variable selection was performed, report (1) full model, (2) final model, and (3) variable selection procedure—but *do not* report on all intermediate models considered.

Explain scientific goals and how your results pertain to those goals.

Use pictures and clear, concise English.