Homework #4

Bayesian Psychometric Modeling (PSQF 7375, Fall 2024)

Due Friday, November 15th, 2024 at 11:59pm in ICON

## Bayesian Psychometric Models

The objectives of this homework are:

1. To learn how to estimate a polytomous item response model with MCMC
2. To see how each type of observed data assumption impacts the latent variable estimates
3. To see the impact of different prior distributions on all model parameters

### General Instructions:

The data for homework 3 were collected at a casino from a sample of persons who were gambling. The content and response options for 24 items measuring pathological gambling are given in the excel file GRI\_Items.xlsx, along with the item responses in the excel file GRI\_Data.xlsx. Use Stan for all models in this assignment.

#### For all questions below, when estimates are requested, report:

* The posterior mean (EAP)
* The posterior standard deviation
* The 95% highest density posterior interval (lower and upper bound)

#### To estimate each model, use Stan with the following options:

* Four chains
* Warmup of 5000 iterations
* Sampling of 5000 iterations
* Starting values drawn from for all factor loading/discrimination parameters

You may work with other students in building the homework, but your answers and syntax must not be copied from anyone else.

### Submission Instructions:

Please submit your homework as an R Markdown file where all R syntax is embedded as chunks and each answer is provided using your words (text answers are required, not just output) following each question in the section marked **My Answer**.

**Name your file with your first and last name and submit your file to ICON.**

### Data File

The data file is available for download at <https://jonathantemplin.com/wp-content/uploads/2022/11/GRI_Data.xlsx>. The list of items are available at <https://jonathantemplin.com/wp-content/uploads/2022/11/GRI_Items.xlsx>

The variables in the data are:

* ID: The ID number of the person
* gri1-gri24: The response to a 6-point Likert item for each person

The purpose of this homework is for you to practice polytomous item response models (the graded response model), including writing the syntax for the model, as well as finding and and interpreting the relevant pieces of output. The graded response model assumes each observed item follows a conditional multinomial (categorical) distribution with a logistic link function and a proportional odds assumption and each factor follows a normal distribution.

Using Stan, estimate a single-factor graded response model for all items. Use the z-score method of identification, in which the factor mean is fixed to 0, the factor variance is fixed to 1, and all item parameters are estimated.

Use Stan’s default prior distributions for each model parameter that is not a factor score.

**1. What is the model for each observed item response (i.e., the linear predictor, the link function, the submodels, the distribution, and any other parameters)**

**My Answer**

**2. Did your Markov chains converge? What evidence do you have to support your answer?**

**My Answer**

**3. What are your estimates for each item intercept parameter?** (see estimate details above for what to report)

**My Answer**

**4. What are your estimates for each factor loading/discrimination parameter** (see estimate details above for what to report)

**My Answer**

**5. What are your estimates for the factor scores?** (see estimate details above for what to report)

**My Answer**

**6. Provide a scatter plot of the EAP estimates of latent variables () from the analysis from section 1 (x-axis) and the analysis from section 3 (y-axis)$**

**My Answer**

**7. In which ways do the results for the latent variables when using the CFA assumptions versus the GRM assumptions?**

**My Answer**

**8. What would you predict would happen to the latent variables if you were to use an empirical prior (as in section 2)?**

**My Answer**

**9. Provide a scatter plot of the EAP estimates of the latent variables () from the analysis of section 3 (x-axis) and the posterior standard deviation estimates of the latent variables from the analysis of section 3 (y-axis)**

**My Answer**

**10. Where (i.e., for what range of EAP estimates) is the estimation of the latent variables most precise?**

**My Answer**