GR5065 Homework 6

Ben Goodrich

Due April 25, 2022 at 4PM

```
# call the set.seed function once here to make the knitting conditionally deterministic
# but remember to also pass a seed to any MCMC function
```

1 Social Surveys

In the past week, data were released that allow researchers to join data from the post-2020 American National Election Survey (ANES) with the General Social Survey (GSS) panel survey where the same people were first asked questions in either 2016 or 2018 and then asked a largely identical set of questions a few months before the 2020 election. However, these surveys were conducted over the internet in 2020 due to the covid pandemic, whereas they were historically conducted with face-to-face interview methods in previous years.

To load these data, it is best to start with the Stata version of the GSS panel

```
if (!file.exists("GSS_2020_panel_stata_1a.zip")) {
  download.file("https://gss.norc.org/Documents/stata/GSS_2020_panel_stata_1a.zip",
                destfile = "GSS_2020_panel_stata_1a.zip")
  unzip("GSS_2020_panel_stata_1a.zip")
}
library(haven)
GSS <- as_factor(read_dta(file.path("GSS_2020_panel_stata_1a", "gss2020panel_r1a.dta")))
```

The `GSS` data.frame is organized in "wide" form with the fields that have a `_1a` suffix being collected in 2016 the fields that have a `_1b` suffix were collected in 2018, and the fields with a `_2` suffix were collected before the 2020 election.

To obtain the 2020 ANES data, you have to register with your Columbia email at

https://electionstudies.org/data-center/anes-gss-2020-joint-study/

and then download the Stata version of the dataset from the top left to your working directory and unzip it. At which point, you can execute

```
ANES <- as_factor(read_dta(file.path("anes_timeseries_2020_gss_stata_20220408",
                                     "anes_timeseries_2020_gss_stata_20220408.dta")))
```

These variables were collected in late 2020, after the election.

To merge the two datasets, use

```
library(dplyr)
ANES_GSS <- inner_join(ANES, GSS, by = c(YEARID = "yearid"))
```

Refer to the documentation of the ANES and GSS for more information on the variables.

For this problem, you should choose one ordinal outcome variable in `ANES_GSS` to model and predictor variables from `ANES_GSS` that include at least one ordinal variable whose relationship with the outcome

variable can be assumed monotonic. However, you will need to use the `as.ordered` function to coerce these factor variables in `ANES_GSS` to ordered factors before calling `brms::brm`.

1.1 Prior Predictive Distribution

Call `brm` with the non-default `sample_prior = "only"` argument to draw from the prior distribution of all the parameters and then pass that to `posterior_predict` to draw from the prior predictive distribution of the ordinal outcome. Specify the `prior` argument to `brm` so that the prior predictive distribution is roughly discrete uniform across survey respondents, which is to say that each ordinal outcome category has about the same chance of occurring under the model.

1.2 Posterior Distribution

Condition on the data to draw from the posterior distribution of the parameters. Adjust the `control` and / or `iter` arguments to `brm` as necessary to ensure that there are no warning messages emitted by the Stan algorithm during sampling.

1.3 Interpretation

Plot the `conditional_effects` for the ordinal predictor whose effect was presumed to be monotonic. How would you describe your posterior beliefs about this relationship?

1.4 PSISLOOCV

Call the `loo` function. Are there any observations that have an estimated Pareto k value of greater than 0.7? Compare the estimated ELPD with a similar model that relaxes the monotonicity assumption for the predictor in the previous subproblem. Which is expected to best predict future data?

1.5 Posterior Predictive Distribution

Call `posterior_predict` on the object produced by `brm` that conditioned on the data and the monotonicity assumption. How does this posterior predictive distribution differ from the roughly discrete uniform prior predictive distribution that you obtained previously?

2 The Replication Crisis in Psychology

Read this profile of Amy Cuddy whose research figured prominently — although many other articles were fundamentally similar — in the methodological debate over the past decade among social psychology researchers and others. You do not need to read any of the papers, blog posts, etc. referred to in the article in order to answer the following questions, although you could.

How would social psychology research, such as Cuddy's, differ if it were conducted with Bayesian methods? In particular, what sort of prior distributions would you use for the treatment effects? How would you evaluate whether the results of a Bayesian study in this area were worth publishing and how would you evaluate whether the theory of the paper was supported by the data being conditioned on? What would constitute a successful or unsuccessful "replication" of a Bayesian study and would that replication study itself need to be Bayesian?

A Google Scholar search for "brms" will reveal about $20,000$ results many of which are in psychology or related fields, and most of which were written after the literature review on Bayesian articles in the social sciences that we referred to at the start of the first class. And yet, very little of the debate about the replication crises in psychology has considered the role of Bayesian methods. What do you think are the obstacles to using Bayesian methods in social psychology and what would it take to overcome them?