Assignment 4

STAT40850 Bayesian Analysis (online)

2024/2025

AUTHOR

Alberto Caimo

Instructions

- Hand-in date: **Wednesday 30th April 2025 at 11:59pm.**

- You should submit it to the Assignment 4 object in Brightspace.

- You should submit a pdf file and corresponding source file (e.g., rmd or qmd file) containing your answers to the questions.

- You may submit it multiple times before the deadline, but only the last version will be marked.

Dataset

A referendum is taking place in a certain country. To analyse the distribution of the 'Yes' votes and the probability that the Yes vote will win, we have information on an exit poll involving 50 constituencies localised in 5 different regions in the country (10 constituencies were sampled in every region).

The dataset includes:

- `n_yes` number of Yes votes in each constituency;
- `n_votes` number of votes in each constituency;
- `constituency` constituency indicator;
- `region` region indicator.

To load the dataset:

```
load(url("https://acaimo.github.io/teaching/data/referendum.RData
```

Questions

1. Specify and estimate (using Stan) a pooled Binomial model with **logit transformation** of the parameter associated to the probability of voting yes in the referendum. Interpret the results obtained.

   [20]

2. Relax the complete pooling assumption of the previous model and specify and estimate (using Stan) a Bayesian hierarchical model capturing the

average and the variation of the probability of observing a Yes vote within and between constituencies. Interpret the results obtained.

[20]

3. Now consider a **three-level** hierarchical model capturing the average and variation of the probability of observing a Yes vote within and between regions and constituencies of each region. Interpret the meaning of the parameters included in this model.

[25]

4. Estimate the model defined in the previous question (using Stan) and interpret the results obtained.

[20]

5. Perform a goodness-of-fit check for each model estimated above and briefly comment on the results obtained.

[10]

6. Compare the three models estimated above using LOO-CV and draw your conclusions. (**NB:** you can improve Pareto k diagnostic values by setting `moment_match = TRUE` but it will increase the computation time).

[5]