85182 - Bayesian Inference
University of Bologna
November - December 2023

Pratical Exam Simulation

December 6, 2023

This example comes from Chapter 8.3 of Gelman and Hill (2007).

Suppose that we want to make inferences about the efficacy of a certain pest management system at reducing the number of roaches in urban apartments. Here is how Gelman and Hill describe the experiment (pg. 161):

> [...] the treatment and control were applied to 160 and 104 apartments, respectively, and the outcome measurement y_i in each apartment i was the number of roaches caught in a set of traps. [...]

In addition to an intercept, the regression predictors for the model are the pre-treatment number of roaches `roach1`, the treatment indicator `treatment`, and a variable `senior` indicating whether the apartment is in a building restricted to elderly residents.

Load the data and rescale the variable `roach1` using the following piece of code

```
library(rstanarm)
data(roaches)
# Rescale
roaches$roach1 <- roaches$roach1 / 100
```

Exercise questions:

1. Assume a simple Poisson regression model without random effects considering `roach1` and `senior` as independent variables (model a). Write the theoretical form of the model assuming weakly-informative prior distributions for the coefficients and program the model in `rstanarm`.

2. Monitor the convergence of the algorithm (referred to model a).

3. Assume now that we are interested in update `model a` in order to estimate also group-specific effects with respect to the type of treatment (model b). Write the theoretical form of the model assuming a $\mathcal{N}(0,10)$ prior distribution with automatic scale adjustments for the coefficients and program the model in `rstanarm`.

4. Monitor the convergence of the algorithm (referred to model b) and in particular discuss the interpretation of `n_eff`.

5. Compare the two models in terms of goodness of fit. Which one provides a better fit of the original data?

6. Suppose that we are interested in verifying that the chosen model is able to reproduce the mean number of post-treatment roaches. Check this assumption. How do the posterior predictive checks work? Provide a brief explanation.

7. Report the 90% credible interval of the predicted parameters.

8. Provide an overall comment of the performed analysis. How could you improve, if necessary, the fitted models in order to better fit the initial dataset?

Final remarks

- This simulation is longer than the usual exam.
- The "theoretical" questions that start with *discuss* or are highlighted with *provide a brief explanation* need to be answered.
- If the model described in the examination is similar to the examples carried out in the lab sessions you are supposed to know all the details and functions required to program them in **rstanarm**. Otherwise, if different functions are necessary to implement the model, a detailed description of these tools will be provided.