Computational Biology Group

D BSSE

ETH zürich

Introduction to Bayesian Statistics with R

7: Exercises

Jack Kuipers

29 November 2022

A genetically encoded circuit was developed to detect the presence in cells of interleukin 4 (IL-4) and interleukin 13 (IL-13) and to produce secreted alkaline phosphatase (SEAP) in their presence: *Sensing and responding to allergic response cytokines through a genetically encoded circuit*, **Nature Communications** 8:1101 (2017).

The `genetic_circuit.csv` data contains a subset of the experimental results (e.g. without the 0 concentration) for the full genetic circuit (corresponding to Figure 1d in the article). Every experiment corresponds to a plate which consists of up to three measurements for every setting (i.e. up to three technical replicates per batch/experiment, cytokine and concentration).

For the analysis we have also

- log-transformed the concentration as the variable `log10conc`
- rescaled SEAP by dividing by 100, stored as the variable `seap_s`

NOTE: Since default priors are usually on the unit scale, we often want to scale our data to match typical prior widths, as above.

Exercise 7.1 - Bayesian multiple regression

Run a Bayesian multiple regression model akin to

$$\text{seap\_s} = \text{log10conc} + \text{experiment} + \text{cytokine}$$

with `brms` and the `brm()` function.

- Did you make your model robust?
- What did you select for your priors?
- Visualise the posterior distribution of the slope coefficient between SEAP and the log concentration.

Optional Exercise 7.2 - Multiple regression

Run a multiple regression of

$$\text{seap\_s} = \text{log10conc} + \text{experiment} + \text{cytokine}$$

- Visualise the residuals. Does a robust model for the Bayesian model in Exercise 7.1 make sense?
- Examine the regression coefficients.