BAYESIAN STATISTICAL MODELS - University of Milano-Bicocca

Case study - the frogs data

Lecturer: Riccardo Corradin

We consider a set of data contained in the file *frogs.csv*, available at the module page. The data consist in a set of measures for different amphibious. Specifically, for each observed animal we measure

- The scientific *family*.
- The *body weight*.
- The *nose-to-tail length*.
- The *brain weight*.

We are interested study the body weight, here playing the role of the response variable, as function of the other observed quantities. Specifically, we first transform on suitable scales the positive real-valued observed quantities, then we consider a linear model of the form

$$y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,4},$$

where $x_{i,1} = 1$ for all $i = 1, \ldots, n$, $x_{i,2}$ is a binary variable denoting if an observation is in the *Ranidae* family or not, $x_{i,3}$ is the logarithm of the nose-to-tail length, $x_{i,4}$ is the logarithm brain weight. We assume an hierarchical model of the form

$$\begin{aligned} Y_i \mid \boldsymbol{x}_i, \boldsymbol{\beta}, \sigma^2 &\sim N(\boldsymbol{x}_i^\intercal \boldsymbol{\beta}, \sigma^2), \qquad i = 1, \ldots, n, \\ \boldsymbol{\beta} \mid \sigma^2 &\sim N(\boldsymbol{b}_0, \sigma^2 \Sigma_0), \\ \sigma^2 &\sim IG(a_0, b_0). \end{aligned}$$

a) Implement a function in R to sample from the posterior distribution of interest.

b) Produce a sample of size $1\,000$ from the posterior distribution of interest.

c) For each regression coefficient, obtain point estimate and credible interval of your choice, and compute

$$\min\{P(\beta_j > 0), P(\beta_j < 0)\}.$$

Use the previous to identify which coefficients have no significative effect in the model. You can obtain the credible intervals, for example, with the *ci* function in *bayestestR* library.

d) Produce a second model estimate without the covariates previously identified, check the posterior distributions of interest. Then, perform a test to compare the two models.