Bayesian models 2 - Stan

October 14, 2019

Example: Linear regression with Normal responses

Example: average household income
Households in 284 Swedish municipalities were sampled. For each municipality we have the average household income (INCOME) as a response variable, and the average age of the head of household (AVAGE) and a measure of how rural the municipality is (RURAL: 1 urban, 2 mixed, 3 rural) as predictors.

`income-lm-data.txt`

What is the model? Priors? Posterior means and credible intervals?
Looking at credible intervals, should we drop the RURAL variable?

Linear regression with Normal responses

Model:

$$\begin{aligned} Y_i &\sim \text{Normal}(\mu_i, \sigma^2) \\ & \qquad i = 1, \ldots, n \\ \mu_i &= \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} \end{aligned}$$

Prior:

$$\begin{aligned} \beta_j &\sim \text{Normal}(0, (100)^2) \\ & \qquad j = 0, 1, \ldots, k \\ \tau (= 1/\sigma^2) &\sim \text{Gamma}(shape = 0.5, rate = 0.0005) \end{aligned}$$

Note that β_j has variance 10000 and τ has variance 20000.

Remarks: For these models, sampling from the posterior using MCMC works better if the x_{ij} are centered. That is, for $\overline{x}_{\cdot j} = \sum_i x_{ij}/n$ we put

$$\mu_i = \beta_0 + \sum_{j=1}^{k} \beta_j (x_{ij} - \overline{x}_{\cdot j})$$

Stan example: See Stan_linear_model.pdf and income-lm.stan.