Bayesian Statistics and Simulation Methods

Theodoros Georgiopoulos

Assignment

Bayesian Analysis of the Normal Regression Model

Exercise 2

(a) For the accompanying dataset, we have $p = 9$ and $n = 300$. So, $q = p + 1 = 10$, $\boldsymbol{X}$ is a 300×10 matrix, $\boldsymbol{y}$ has length 300, and the matrix $\boldsymbol{V}$ is 10×10. We assume that $\alpha = \beta = 0.01, \boldsymbol{\mu}_\beta = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^t$, and g hyper-parameter for the Zellner's g-prior is $g = n = 300$.

Our model is

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_9 x_9 + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma^2).$$

We fit the full model and we observe the following results

θ	$\mathrm{E}(\theta \mid \boldsymbol{y}, X)$	$\mathrm{Var}(\theta \mid \boldsymbol{y}, X)$	$\theta_{0.995}$	$\theta_{0.005}$
β_0	0.956	0.014	0.641	1.272
β_1	2.198	0.008	1.954	2.443
β_2	2.934	0.017	2.592	3.277
β_3	0.020	0.031	-0.435	0.475
β_4	-0.146	0.015	-0.464	0.171
β_5	0.271	0.053	-0.324	0.868
β_6	0.012	0.007	-0.204	0.229
β_7	-0.150	0.032	-0.612	0.312
β_8	-0.069	0.004	-0.240	0.101
β_9	-0.068	0.035	-0.552	0.415
σ^2	4.365	0.128	3.546	5.405

We see that $\beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8$, and β_9 credible intervals include the zero value. So, it is probable that the respective variables are not necessary for our model.

(b) The posterior probabilities are

$P(\beta_1 > 0 \mid \boldsymbol{y}, X) = 1$

$P(\beta_3 > 0 \mid \boldsymbol{y}, X) = 0.545$

$P(\sigma^2 > 3 \mid \boldsymbol{y}, X) = 0.999$

(c) We have the hypothesis

$H_0 : \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0,\ H_1 : \text{not } H_0.$

We test H_0 vs H_1 using Bayes factor. We calculate the following value

$$2 \log B_{10}(\boldsymbol{x}) = 2(\log m_0(\boldsymbol{y}) - \log m_1(\boldsymbol{y})) = 32.614$$

So, according to the Kass and Raftery table, we have very strong evidence against H_1. So, it is better to remove $x_3, x_4, x_5, x_6, x_7, x_8, x_9$ from our model.