

Homework 3

Data

The data in ‘`data(GasolineYield)`’ (from the `betareg` R package) represent quantitative characteristics of $n = 32$ samples of crude oil. For this homework assignment, we seek to construct a Bayesian beta regression model to predict the proportion of crude oil converted to gasoline (`yield`) based on crude oil gravity (`gravity`), vapor pressure of crude oil (`pressure`), and temperature of crude oil at which gasoline has vaporized (`temp`). A critical aspect of this study is that the response variable $y_i = \text{yield}_i$ is a proportion (i.e., $0 < y_i < 1$).

Questions

Prepare a written response to the following, using Overleaf. The assignment shouldn’t be longer than 10 (double-spaced, excluding title page, references, and appendices). Due Tues., March 7, at the beginning of the class period. Please submit the assignment as a PDF through CANVAS.

1. Develop a MCMC algorithm to fit the Bayesian regression model:

$$y_i \sim \text{Beta}(a_i, b_i) \text{ , for } i = 1, \dots, n \text{ ,} \quad (1)$$

$$a_i = \mu_i \tau \text{ ,} \quad (2)$$

$$b_i = (1 - \mu_i) \tau \text{ ,} \quad (3)$$

where $\tau > 0$ controls the precision, and the mean μ_i is linked to a set of covariates \mathbf{x}_i using $\text{logit}(\mu_i) = \mathbf{x}_i' \boldsymbol{\beta}$. For this model, assume the priors

$$\boldsymbol{\beta} \sim \text{N}(\boldsymbol{\mu}_\beta, \sigma_\beta^2 \mathbf{I}) \text{ ,} \quad (4)$$

$$\tau \sim \text{Gamma}(\gamma_1, \gamma_2) \text{ .} \quad (5)$$

2. Use the MCMC algorithm to fit the beta regression model to the gasoline data set using `yield` as the response variable and the three covariates below:

(a) `gravity`

(b) `pressure`

(c) `temp`

You may want to standardize the covariates before fitting the model. Tune the algorithm as needed and check the trace plots for convergence.

3. Calculate the posterior predictive p-value using MSE as a statistic for this model and data set. Is this model appropriate for these data based on the p-value?
4. If the model is appropriate for these data, what inference can you make about the predictors of yield proportion?

References

- Cribari-Neto F. and A. Zeileis. (2010). Beta Regression in R. *Journal of Statistical Software*, 34(2): 1–24.