Project Submission Instructions

Please complete and save your project in pdf file. Before submission, please use the following naming convention:

`firstname-lastname`-P2.pdf

Your pdf file should be submitted to the e-mail address below:

`BIOS276.31r8w8ndztx2a19n@u.box.com`

Please do not use my personal e-mail address. After submission you should receive confirmation of successful uploading. I will let you know if I am missing your project by the day following the due date.

Please describe all your work in clear terms, before implementing `R` code. Each question should include a description of your approach with clear indication of where I can find the associated source code. Your code should be attached to your assignment or uploaded online to a repository I can freely access.

1. Consider the data-set `diabetes` data available from `lars`, (use the matrix of predictors x). The outcome is indexed with y and predictors are indexed by x a matrix of 10 predictors.

Add a intercept column in the data design matrix X, and consider the following regression model

$$Y = X\beta + \epsilon$$

with $X : n \times p$, $\beta : p \times 1$ and $\epsilon \sim N(0, \sigma^2 I_n)$.

We are interested in model selection by including only a subset of the original p predictors. The number of possible models is 2^p.

Let $\gamma \in \{0,1\}^p$ define a p-dimensional vector of predictor indicators, such that $\gamma_j = 1$ if X_j is included in the regression, otherwise $\gamma_j = 0$, $(j = 1, \ldots, p)$. Also, let X_γ and β_γ denote the subset of predictors and regression coefficients associated with components of γ that are equal to 1.

Conditional on γ the sampling model is

$$Y \mid \gamma, \beta, \sigma^2 \sim N(X_\gamma \beta_\gamma, \sigma^2 I_n),$$

with improper prior:

$$p(\beta_\gamma, \sigma^2 \mid \gamma) \propto (\sigma^2)^{-\frac{q_\gamma}{2}-1} \exp\left[-\frac{1}{2c\sigma^2}\beta_\gamma'(X_\gamma' X_\gamma)\beta_\gamma\right]$$

where $q_\gamma = \sum_{j=1}^p \gamma_j$ is the model size, given a realization of γ and $c = n$, the sample size. Finally, assume

$$p(\gamma \mid \alpha) = \prod_{i=1}^p \alpha^{\gamma_j}(1-\alpha)^{(1-\gamma_j)}, \text{ with } \alpha \sim Beta(0.5, 0.5)$$

a. Develop and implement a Reversible Jumps MCMC algorithm exploring the model space $p(\gamma \mid Y)$. Report the estimated inclusion probabilities $p(\gamma_j = 1 \mid Y)$, for all $j = 1, 2, \ldots, p$.

b. Compare your results in (a) with results obtained sampling directly from $p(\gamma \mid Y)$. You should be able to implement this without the need for RJ Monte Carlo methods.