MATH 8820

Homework Assignment 5

Problem 1: (Bayes factors) In multiple linear regression, consider the problem of attempting to choose between one of the following two models

$$
\begin{aligned}
\text{Model 1:} \quad & Y_i = \mathbf{x}_{1i}'\boldsymbol{\beta}_1 + \epsilon_{1i}, \quad \text{for } i = 1, ..., n, \\
\text{Model 2:} \quad & Y_i = \mathbf{x}_{2i}'\boldsymbol{\beta}_2 + \epsilon_{2i}, \quad \text{for } i = 1, ..., n,
\end{aligned}
$$

where Y_i is the response variable measured on the ith unit, $\mathbf{x}_{ji} = (1, x_{j1i}, ..., x_{jp_ji})'$ is a $(p_j + 1) \times 1$ dimensional vector of covariates, $\boldsymbol{\beta}_j$ is the corresponding vector of regression coefficients, and $\epsilon_{ji} \overset{iid}{\sim} N(0, \sigma_j^2)$. Note, this formulation encompasses many different modeling scenarios; e.g., Model 1 could be a “full” model and Model 2 could be a “reduced: model, or Models 1 and 2 could consider different covariate sets, etc. The goal of this problem will be to develop techniques to compare these 2 models through the use of Bayes factors. To complete the Bayesian model we assume the following priors

$$
\begin{aligned}
\boldsymbol{\beta}_j|\sigma_j^2 &\sim N(\boldsymbol{\beta}_{j0}, \boldsymbol{\Sigma}_{j0}) \\
\sigma_j^2 &\sim IG(a_{j0}, b_{j0}),
\end{aligned}
$$

for $j = 1, 2$, where $IG(\cdot, \cdot)$ denotes the inverse-gamma distribution.

1. Derive an analytical expression for $m(\mathbf{y}|M = j)$; i.e., the marginal distribution of the data under the jth model.

2. Derive a Bayes factor for comparing Model 1 to Model 2.

3. From lecture, we discussed different Monte Carol based techniques that could be used to approximate Bayes factors. Choose three of these techniques and discuss how they could be used for this problem.

4. Design and implement a small simulation study where you evaluate the Monte Carlo based techniques you discuss in part 3. Note, the analytical expression you specify in 2. is your target. Which technique works better? Justify your assertion based on the simulation you have run.