HOMEWORK 3

Handed out: Friday, Sept. 20, 2017 Due: Wednesday, Oct. 4 midnight

Notes:

1. We "**strongly encourage**" typed (Latex or Word) homework.
2. Submit a zipped directory **with all Latex or Word files** (**not just the PDF**) to ame70779ta@gmail.com. Include your computer programs & data files and a Readme file providing scripts on how to run them.
3. Form of your Email attachment/subject: **HW3_LastName_FirstName.zip**

In this homework, we consider the caterpillar regression problem discussed in detail in Chapter 3 (Regression and Variable Selection) of the Bayesian Core book by Jean-Michel Marin, and Christian P. Robert. Please read the chapter in detail as the homework set consists of reproducting all results (Tables) from this chapter. Additional notes and derivations can be found on Lecture 12.

Problem 1. The problem of interest consists of a scalar output y (the logarithmic transform of the number of processionary caterpillar colonies) and 10 explanatory variables $x_i, i = 1, \ldots, 10$ (their meaning not important here but is provided in the given reference). Thus your regression function has k+1 terms (including the bias). Using the data set caterpillar.mat, plot the pairs $(x_i, y), i = 1, \ldots, 10,$ i.e. the semi-log-y plots against each feature (except of course the intercept in the regression model).

Problem 2. Write a computer code that provides the MLE $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ of the regression coefficients and their standard significance analysis $\hat{\sigma}^2 = \frac{1}{n-k-1}\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) = \frac{s^2}{n-k-1}, where: s^2 \triangleq \left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)$.

Provide next the details of your t-values $t_{values(i)}$ calculation $T_i = \frac{\widehat{\beta}_i - \beta_i}{\sqrt{\widehat{\sigma}^2\omega_{ii}}} \sim \mathcal{T}(n-k-1, 0, 1), where: \omega_{(i,i)} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}|_{(i,i)}$ (t-statistic against null hypothesis) and p-values Pr $(|T_i| > |t_{values(i)}|)$ and reproduce the results shown in Figure 3.2 of the reference. The last two columns on this table are frequentist estimates.

Hint: To compute the p-values, Pr $(|T_i| > |t_{values(i)}|)$, add the left and right values using the CDF of the $(n-k-1, 0, 1)$.

Problem 3. Reproduce the results of Tables 3.1 and 3.2. This requires computing the posterior marginal statistics $\mathbb{E}_\pi\left(\sigma^2 \mid \boldsymbol{y}, \boldsymbol{X}\right), \mathbb{E}_\pi\left(\boldsymbol{\beta}_0 \mid \boldsymbol{y}, \boldsymbol{X}\right), V_\pi\left(\boldsymbol{\beta}_0 \mid \boldsymbol{y}, \boldsymbol{X}\right)$ and $\mathbb{E}_\pi\left(\boldsymbol{\beta}_i \mid \boldsymbol{y}, \boldsymbol{X}\right), V_\pi\left(\boldsymbol{\beta}_i \mid \boldsymbol{y}, \boldsymbol{X}\right), i = 1, .., 10.$ Use conjugate priors

$$\boldsymbol{\beta}|\sigma^2, \boldsymbol{X} \sim \mathcal{N}_{k+1}(\widetilde{\boldsymbol{\beta}}, \sigma^2\boldsymbol{M}^{-1}), \boldsymbol{M}\ a\ (k+1)\times(k+1)\ pos.def.symm.matrix$$
$$\sigma^2|\boldsymbol{X} \sim InvGamma(a, b), a, b > 0$$

with $\boldsymbol{M} = \frac{\boldsymbol{I}_{k+1}}{c}, c > 0$ and $\tilde{\beta} = \mathbf{0}_{k+1}, k = 10, a = 2.1, b = 2.0, c = 100$.

Problem 4. To reproduce the results of Tables 3.3 and 3.4, we now need to consider Zellner's informative G-prior.

$$\boldsymbol{\beta}|\sigma^2, \boldsymbol{X} \sim \mathcal{N}_{k+1}\left(\widetilde{\boldsymbol{\beta}}, c\sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}\right)$$
$$\sigma^2 \sim \pi(\sigma^2|\boldsymbol{X}) \propto \sigma^{-2} \text{ improper Jeffreys prior}$$

Consider $\widetilde{\boldsymbol{\beta}} = \mathbf{0}, and\ the\ cases\ c = 100\ and\ c = 1000.$ Compute the $\beta_i, i = 1, \dots 10$ tabulate the expectation $\mathbb{E}[\boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{X}] = \frac{1}{c+1}\left(\widetilde{\boldsymbol{\beta}} + c\widehat{\boldsymbol{\beta}}\right)$ and variance $V_\pi[\beta|\boldsymbol{y}, \boldsymbol{X}] = \frac{c}{c+1}\frac{\left(s^2 + (\widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}})^T\boldsymbol{X}^T\boldsymbol{X}(\widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}})/(c+1)\right)}{n-2}(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ of each coefficient as well as the log10 of the Bayes factors to account for the inclusion or not in the model of each of the input features.

$$B_{10}^{\pi} = \frac{f(\boldsymbol{y}|\boldsymbol{X})}{f(\boldsymbol{y}|\boldsymbol{X}_0, H_0)} = \frac{(c_0+1)^{(k+1-q)/2}}{(c+1)^{(k+1)/2}} \times$$
$$\left[\frac{\boldsymbol{y}^T\boldsymbol{y} - \frac{c_0}{c_0+1}\boldsymbol{y}^T\boldsymbol{X}_0(\boldsymbol{X}_0^T\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0^T\boldsymbol{y} + \frac{1}{c_0+1}\widetilde{\boldsymbol{\beta}}_0^T\boldsymbol{X}_0^T\boldsymbol{X}_0\,\widetilde{\boldsymbol{\beta}}_0 - \frac{2}{c_0+1}\boldsymbol{y}^T\boldsymbol{X}_0\,\widetilde{\boldsymbol{\beta}}_0}{\boldsymbol{y}^T\boldsymbol{y} - \frac{c}{c+1}\boldsymbol{y}^T\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y} + \frac{1}{c+1}\widetilde{\boldsymbol{\beta}}^T\boldsymbol{X}^T\boldsymbol{X}\,\widetilde{\boldsymbol{\beta}} - \frac{2}{c+1}\boldsymbol{y}^T\boldsymbol{X}\,\widetilde{\boldsymbol{\beta}}}\right]$$

This will require to compute the likelihood with all the features as well as with each feature removed out from your calculation.

Problem 5. Compute the 90% high posterior density (HPD) intervals for the $\beta_i, i = 1, \dots 10$ (note the reference wrongly indicates the given values to be the 95% HPD). Use again Zellner's informative G-prior with c=100.

Problem 6. Consider an uninformative Jeffrey's prior

$$\pi^J(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{X}) \propto \sigma^{-2}$$

In this setup we assume that c is now unknown, with prior

$$\pi(c) = c^{-1}\mathbb{I}_{N^*}(c)$$

Use as before $\widetilde{\boldsymbol{\beta}} = \mathbf{0}$. Verify that the marginal distribution is of the form:

$$f(\boldsymbol{y} \mid \boldsymbol{X}) = \sum_{c=1}^{\infty} f(\boldsymbol{y} \mid \boldsymbol{X}, c)c^{-1} \propto$$
$$\sum_{c=1}^{\infty} c^{-1}(c+1)^{-(k+1)/2}\left[\boldsymbol{y}^T\boldsymbol{y} - \frac{c}{c+1}\boldsymbol{y}^T\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}\right]^{-n/2}$$

Compute the expectation and variance of the regression coefficients as shown in Table 3.6 of the reference.

Problem 7. Duplicate Tables 3.7 – 3.11 of the reference computing the most likely models by decreasing probabilities using Zellner's G-prior (with c=100) and Zellner's uninformative G-prior. Implement the Gibbs sampler and compare the results for the top ten posterior probabilities.