Applied Bayesian Statistics
Practical 3
Linear regression. Variable selection. Gibbs' sampling.

Robin Ryder

Aim: Selection of explanatory variables in a linear regression setting, through exact computation and Gibbs' sampling.

Reference: *Bayesian Essentials with R* (Marin & Robert), chapter 3.

In this practical, we would like to perform Bayesian linear regression to explain the death rate in various American urban areas in the 1960s[1].

Download the data and load it with

```
> read.csv(deathrate2.csv)
```

In this data set, there are 21 columns: the variable to explain and 20 explanatory variables.

1. the death rate (variable to explain)
2. the average annual precipitation;
3. the average January temperature;
4. the average July temperature;
5. the size of the population older than 65;
6. the number of members per household;
7. the number of years of schooling for persons over 22;
8. the number of households with fully equipped kitchens;
9. the population per square mile;
10. the size of the nonwhite population;
11. the number of office workers;
12. the number of families with an income less than $3000;
13. the hydrocarbon pollution index;
14. the nitric oxide pollution index;
15. the sulfur dioxide pollution index;
16. the degree of atmospheric moisture;
17. 5 columns of gaussian noise.

Our linear model is (with $p = 20$)

$$y_i = \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \ldots + \beta_p x_i^p + \epsilon_i$$
$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

1. **Frequentist analysis** Give the frequentist estimates of β and σ^2, which we denote $\hat{\beta}$ and s^2.

---

[1]Data from Gunst & Mason 1980, McDonals & Schwing 1973, Spaeth 1991.

```
> summary(lm(y~X))
```

2. **Bayesian inference using Zellner's G-prior** Consider the following prior:

$$\begin{aligned}
\beta | \sigma^2, X &\sim \mathcal{N}_{p+1}\left(0, g\sigma^2 ({}^t X X)^{-1}\right) \\
\pi(\sigma^2) &\propto \sigma^{-2}
\end{aligned}$$

This prior is conjugate; the associated posterior is

$$\begin{aligned}
\beta | \sigma^2, y, X &\sim \mathcal{N}_{p+1}\left(\frac{g}{g+1}\hat{\beta}, \frac{\sigma^2 g}{g+1}(X^T X)^{-1}\right) \\
\sigma^2 | y, X &\sim \mathcal{IG}\left(\frac{n}{2}, \frac{s^2}{2} + \frac{1}{2(g+1)}\hat{\beta}^T X^T X \hat{\beta}\right)
\end{aligned}$$

hence

$$\beta | y, X \sim \mathcal{T}_{p+1}\left(n, \frac{g}{g+1}\hat{\beta}, \frac{g(s^2 + \hat{\beta}^T X^T X \hat{\beta}/(g+1))}{n(g+1)}(X^T X)^{-1}\right).$$

a) For $g = 0.1, 1, 10, 100, 1000$, give $E[\sigma^2 | y, X]$ and $E[\beta_0 | y, X]$. What can you conclude about the impact of the prior on the posterior?

b) We would like to test the hypothesis $H_0 : \beta_7 = \beta_8 = 0$. For a model with $p - q$ non-zero coefficients, the marginal likelihood is

$$(g+1)^{-(p+1-q)/2}\pi^{-n/2}\Gamma(n/2) \times \left[y^T y - \frac{g_0}{g_0 + 1}y^T X_0 (X_0^T X_0)^{-1} X_0^T y\right]^{-n/2}$$

Compute Bayes' factor given our data and conclude, using Jeffreys' scale of evidence.

3. **Model choice: exact computation** In this question, we restrict ourselves to the first 3 explanatory variables

```
> X1 = X[ , 1:4]
```

We would like to know which variables to include in our model, and assume that the intercept is necessarily included. We have $2^3 = 8$ possible models. To each model, we associate the variable $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ where $\gamma_i = 1$ if x^i is included in the model, and $\gamma_i = 0$ otherwise. Let X_γ be the submatrix of X where we only keep the columns i such that $\gamma_i = 1$.

Using the marginal probability formula of question 2b, compute for each model its marginal likelihood $m(y|X_\gamma)$. Deduce the most likely model a posteriori.

4. **Model choice using Gibbs' sampling** We now consider all p explanatory variables. We thus need to choose between 2^p models. Use the `bms` function from package `BMS` to sample from the posterior distribution of γ, and conclude on the most likely model.

5. **Prediction** For a new value of your choice of the covariates, perform prediction using (a) the most probable model a posteriori; (b) the best model as selected by AIC; (c) a mixture of models weighted by their posterior probabilities.