THE UNIVERSITY OF SYDNEY
ADVANCED BAYESIAN ANALYSIS

Semester 2	**Assignment 3**	2015

Consider the model

$$\mathbf{y}|\boldsymbol{\beta},\sigma^2 \sim N_n(\mathbf{X}\boldsymbol{\beta},\sigma^2\mathbf{I})$$

with priors

$$\begin{aligned}
\boldsymbol{\beta}|\sigma^2,g &\sim N_p(\mathbf{0}, g\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}) \\
p(\sigma^2) &= (\sigma^2)^{-1}\mathbb{I}(\sigma^2>0) \\
p(g) &= \frac{g^b(1+g)^{-a-b-2}}{\text{Beta}(a+1,b+1)}\mathbb{I}(g>0)
\end{aligned} \tag{1}$$

where $\mathbf{y}\in\mathbb{R}^n$, $\boldsymbol{\beta}\in\mathbb{R}^p$, $\sigma^2\in\mathbb{R}_+$, $\mathbf{X}\in\mathbb{R}^{n\times p}$, $\mathbf{I}$ is the $n\times n$ identity matrix, $g\in\mathbb{R}_+$, $a>-1$ and $b>-1$. Here Beta$(\cdot,\cdot)$ is the Beta function and the prior on g is called a Pearson Type VI or Beta prime distribution. Assume that the columns of $\mathbf{X}$, which we will denote $\mathbf{x}_j$, satisfy $\mathbf{x}_j^T\mathbf{1}=\mathbf{0}$ and $\mathbf{x}_j^T\mathbf{x}_j/n=1$ for $1\le j\le p$. Similarly assume that $\mathbf{y}^T\mathbf{1}=\mathbf{0}$ and $\mathbf{y}^T\mathbf{y}/n=1$. Finally, assume that $n>p$ and $(\mathbf{X}^T\mathbf{X})^{-1}$ exists.

(1) Show that
$$\int \exp\left\{-\tfrac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}+\mathbf{b}^T\mathbf{x}\right\} = |2\pi\boldsymbol{\Sigma}|^{1/2}\exp\left\{\tfrac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right\}$$
where $\boldsymbol{\mu}=\mathbf{A}^{-1}\mathbf{b}$ and $\boldsymbol{\Sigma}=\mathbf{A}^{-1}$.

(2) Hence or otherwise show that
$$p(\mathbf{y}|\sigma^2,g) = \exp\left\{-\tfrac{n}{2}\log(2\pi\sigma^2)-\tfrac{p}{2}\log(1+g)-\tfrac{1}{2\sigma^2}\|\mathbf{y}\|^2+\tfrac{g}{2\sigma^2(1+g)}\mathbf{y}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}\right\}$$
Note that the identities $|c\mathbf{A}|=c^d|\mathbf{A}|$ and $|\mathbf{A}^{-1}|=|\mathbf{A}|^{-1}$ when $\mathbf{A}\in\mathbb{R}^{d\times d}$ may be useful.

(3) Using the simplifying assumptions on $\mathbf{y}$ and $\mathbf{X}$ above show that
$$R^2 = \frac{\mathbf{y}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}}{\|\mathbf{y}\|^2}$$
where R^2 is the usual R-squared statistic associated with least squares regression.

(4) Hence, show that
$$p(\mathbf{y}|g) = \frac{\Gamma(n/2)}{\pi^{n/2}\|\mathbf{y}\|^n}(1+g)^{(n-p)/2}\left[1+g(1-R^2)\right]^{-n/2}.$$

(5) Assume that $b=(n-p)/2-2-a$. Show that
$$p(\mathbf{y}) = \frac{\Gamma(n/2)}{\pi^{n/2}\|\mathbf{y}\|^n}\frac{\text{Beta}(p/2+a+1,b+1)}{\text{Beta}(a+1,b+1)}(1-R^2)^{-(b+1)}.$$
The following result may be useful
$$\int_0^\infty \frac{x^{\mu-1}}{(1+\beta x)^\nu}dx = \beta^{-\mu}\text{Beta}(\mu,\nu-\mu) \qquad \text{(assuming } \mu,\nu>0 \text{ and } \nu>\mu).$$

(6) Suppose that we have observations $u_1,\ldots,u_{n_1}$ independently from $N(\mu_1,\sigma^2)$ and $v_1,\ldots,v_{n_2}$ independently from $N(\mu_2,\sigma^2)$. Let $\mathbf{w}=(\mathbf{1}_{n_1},\mathbf{0}_{n_2})\in\mathbb{R}^n$ where $n=n_1+n_2$. Consider the Bayesian hypothesis test
$$H_0\colon \boldsymbol{\mu}_1=\boldsymbol{\mu}_2 \qquad \text{vs} \qquad H_1\colon \boldsymbol{\mu}_1\neq\boldsymbol{\mu}_2.$$
Assume that both hypotheses are equally likely. If we use $\mathbf{y}=(\mathbf{u},\mathbf{v})$, for the null the design matrix $\mathbf{X}=\mathbf{1}_n\in\mathbb{R}^{n\times 1}$ and for the alternative $\mathbf{X}=[\mathbf{w},\mathbf{1}_n-\mathbf{w}]\in\mathbb{R}^{n\times 2}$ then we can use the model (1) to test H_0 vs H_1. Show that the Bayes factor for this test is given by
$$\text{BF} = \frac{p(\mathbf{y}|H_1)}{p(\mathbf{y}|H_0)} = \frac{\text{Beta}(a+2,n/2-2-a)}{\text{Beta}(a+1,n/2-1-a)}\frac{\text{Beta}(a+1,n/2-3/2-a)}{\text{Beta}(a+3/2,n/2-3/2-a)}(1-R_1^2)^{-n/2+a+2}.$$
where R_1^2 is the R-squared value corresponding to the alternative model.

(7) Suppose that

$$\mathbf{u} = (0.89, -0.59, -0.06, -0.62, 0.60, 1.95, -2.34, 1.07, 0.31, -1.24)$$

and

$$\mathbf{v} = (1.21, 2.21, 2.01, 2.09, -0.02, 1.15, 0.38, 0.27, -0.12, 1.81).$$

Set $a = -3/4$. Calculate BF for the above data. Which hypothesis is preferred?

(8) Consider m such hypotheses

$$H_{0j}\colon \boldsymbol{\mu}_{1j} = \boldsymbol{\mu}_{2j} \qquad \text{vs} \qquad H_{1j}\colon \boldsymbol{\mu}_{1j} \neq \boldsymbol{\mu}_{2j}, \quad 1 \leq j \leq m.$$

Let γ_j be a binary indicator which is 1 if the alternative holds and 0 if the null holds. Let $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_m)$. Consider the model

$$p(\mathbf{y}_1, \dots, \mathbf{y}_m|\boldsymbol{\gamma}) = \prod_{j=1}^{m} \left[p(\mathbf{y}_j|H_{1j})\right]^{\gamma_j} \left[p(\mathbf{y}_j|H_{0j})\right]^{1-\gamma_j}$$

where $\mathbf{y}_1, \dots, \mathbf{y}_m$ are the the response vectors for each of the individual hypothesis tests. Use the priors

$$\gamma_j|\rho \sim \text{Bernoulli}(\rho) \quad \text{and} \quad \rho \sim \text{Beta}(A, B).$$

Derive the variational Bayes approximation to the above model corresponding to the factorizaiton

$$q(\boldsymbol{\gamma}, \rho) = q(\rho) \prod_{j=1}^{m} q(\gamma_j).$$

Show that the VB q-densities are given by

$$q(\gamma_j) = \text{Bernoulli}(w_j) \quad \text{and} \quad q(\rho) = \text{Beta}(r, s)$$

where

$$\begin{aligned} w_j &= \text{logit}^{-1}(\eta_j) \\ \eta_j &= \log(BF) + \psi(r) - \psi(s) \\ r &= 1 + \sum_{j=1}^{m} w_j \\ s &= 1 + \sum_{j=1}^{m} (1 - w_j) \end{aligned}$$

and $\psi(x) = d\log\Gamma(x)/dx$ is the digamma function.

(9) Suppose that for $j = 1, \dots, 20$

$$\begin{aligned} u_{ij} &\sim N(0,1), \quad i = 1, \dots, n_1 = 10, \\ v_{ij} &\sim N(1,1), \quad i = 1, \dots, n_2 = 10, \end{aligned}$$

and for $j = 21, \dots, 1000$

$$\begin{aligned} u_{ij} &\sim N(0,1), \quad i = 1, \dots, n_1 = 10, \\ v_{ij} &\sim N(0,1), \quad i = 1, \dots, n_2 = 10. \end{aligned}$$

Initialize $r = 20$, $s = 980$. Repeat the above simulation 50 times and use VB to fit the model in Question 9. The null is preferred if $w_j < 0.5$ and teh alternative is preferred if $w_j > 0.5$. Draw a box plot of the number of true positives (the alternative is preferred when the alternative is true), true negatives (the null is preferred when the null is true), false positives (the alternative is preferred when the null is true) and false negatives (the null is preferred when the alternative is true) over the 50 simulations.