Bayesian Analysis (0000) **00**, Number 0, pp. 1

Supplement to "On variance estimation for Bayesian variable selection"

Gemma E. Moran*, Veronika Ročková† and Edward I. George*

1 Gibbs sampler for Bayesian ridge regression

We derive the Gibbs sampler used to obtain posterior estimates for the independent Bayesian ridge regression model in Section 3.1 of the main paper. The model is:

$$\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \tag{1.1}$$

$$\boldsymbol{\beta} \sim N_p(0, \tau^2\mathbf{I}) \tag{1.2}$$

$$\pi(\sigma) \propto 1/\sigma. \tag{1.3}$$

The full conditional distributions of the parameters $\boldsymbol{\beta}$ and σ^2 are:

$$\boldsymbol{\beta}|\mathbf{Y}, \sigma^2 \sim N_p(\sigma^{-2}\mathbf{V}\mathbf{X}^T\mathbf{Y}, \mathbf{V}) \tag{1.4}$$

$$\sigma^2|\mathbf{Y}, \boldsymbol{\beta} \sim IG(n/2, \|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2/2) \tag{1.5}$$

where $\mathbf{V} = \left[\sigma^{-2}\mathbf{X}^T\mathbf{X} + \tau^{-2}\mathbf{I}_p\right]^{-1}$. The Gibbs sampling algorithm alternates sampling from (1.4) and (1.5). After burn-in, the posterior mean estimate is the mean of the samples.

2 EM Algorithm for the Spike-and-Slab lasso with a scale-invariant prior

We provide the details of the EM algorithm for the Spike-and-Slab Lasso with a scale-invariant prior in Section 6.1 of the main paper. The model is given by:

$$\pi(\boldsymbol{\beta}|\boldsymbol{\gamma}, \sigma^2) \sim \prod_{j=1}^{p} \left(\gamma_j \frac{\lambda_1}{2\sigma} e^{-|\beta_j|\lambda_1/\sigma} + (1-\gamma_j)\frac{\lambda_0}{2\sigma} e^{-|\beta_j|\lambda_0/\sigma} \right) \tag{2.1}$$

$$\boldsymbol{\gamma}|\theta \sim \prod_{j=1}^{p} \theta^{\gamma_j}(1-\theta)^{1-\gamma_j}, \quad \theta \sim \text{Beta}(a, b) \tag{2.2}$$

$$p(\sigma^2) \propto \sigma^{-2}. \tag{2.3}$$

Then, the "complete" data log posterior is given by

$$\log \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma, \theta|\mathbf{Y}) = -\frac{1}{2\sigma^2}\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2 - (n+2)\log\sigma$$

*Department of Statistics, University of Pennsylvania, gmoran@wharton.upenn.edu, edgeorge@wharton.upenn.edu

†Booth School of Business, University of Chicago, veronika.rockova@chicagobooth.edu

© 0000 International Society for Bayesian Analysis DOI: 0000

$$
\begin{aligned}
&+ \sum_{j=1}^{p} \log\left(\gamma_j \frac{\lambda_1}{2\sigma} e^{-|\beta_j|\lambda_1/\sigma} + (1-\gamma_j)\frac{\lambda_0}{2\sigma} e^{-|\beta_j|\lambda_0/\sigma}\right) \\
&+ \sum_{j=1}^{p} \log\left(\frac{\theta}{1-\theta}\right)\gamma_j + (a-1)\log(\theta) \\
&+ (p+b-1)\log(1-\theta) + C \qquad (2.4)
\end{aligned}
$$

The EM algorithm then proceeds as follows: treat $\boldsymbol{\gamma}$ as unknown and iteratively maximize

$$E[\log \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma, \theta | \mathbf{Y}) | \boldsymbol{\beta}^{(k)}, \sigma^{(k)}, \theta^{(k)}] \qquad (2.5)$$

where $\boldsymbol{\beta}^{(k)}, \sigma^{(k)}, \theta^{(k)}$ are the parameter values after the kth iteration.