Supplementary material for “Selection of tuning parameters, solution paths and standard errors for Bayesian lassos”

Vivekananda Roy∗ and Sounak Chakraborty†

Appendix A: Proofs of Lemmas

Form of the conditional densities:

Below we provide the expressions of the full conditional densities of the elastic net Gibbs sampler presented in Section 2.1 of the paper. These densities are used in the proof of Lemma 1 and Lemma 2.

The conditional density \( \pi(\beta | \tau^2, \sigma^2, y) \) corresponding to (8) is given by

\[
\pi(\beta | \tau^2, \sigma^2, y) = \frac{X^TX + D^*_\tau^{-1}}{(2\pi\sigma^2)^{p/2}} e^{-\frac{(\beta - (X^TX + D^*_\tau^{-1})^{-1}X^Ty)^T(X^TX + D^*_\tau^{-1})^{-1}X^Ty}{2\sigma^2}}.
\]

The conditional density of \( \tau^2 \) given \( \beta, \sigma^2, y \) is

\[
\pi(\tau^2 | \beta, \sigma^2, y) = \prod_{j=1}^{p} \left( \frac{\lambda_j^2}{2\pi \tau^2_j} \right) e^{-\frac{(y_j - X_j\beta)^2}{2\tau^2_j}}.
\]

Finally, the conditional density \( \pi(\sigma^2 | \beta, \tau^2, y) \) corresponding to (10) is given by

\[
\pi(\sigma^2 | \beta, \tau^2, y) = \frac{((y - X\beta)^T(y - X\beta) + \beta^TD^*_\tau^{-1}\beta + 2\xi)(n-1+p+2\alpha)/2}{(\sigma^2)^{(n-1+p+2\alpha)/2+1}(n-1+p+4\alpha)/2} e^{-\frac{(y - X\beta)^T(y - X\beta) + \beta^TD^*_\tau^{-1}\beta + 2\xi}{2\sigma^2}}.
\]

Proof of Lemma 1. We have

\[
(KV)(\beta_0, \tau^2_0, \sigma^2_0) = \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \int_{\mathbb{R}^p} V(\beta, \tau^2, \sigma^2) k((\beta, \tau^2, \sigma^2) | (\beta_0, \tau^2_0, \sigma^2_0)) d\beta d\tau^2 d\sigma^2
= E[E[V(\beta, \tau^2, \sigma^2)|\tau^2, \sigma^2][\beta_0, \sigma^2]|\beta_0, \tau^2_0],
\]

where \( E[|\tau^2, \sigma^2] \) denotes the expectation with respect to the conditional density \( \pi(\beta | \tau^2, \sigma^2) \) given in (8), and the other two conditional expectations are with respect to (9) and (10) respectively. As in Khare and Hobert (2013), we evaluate the three conditional expectations one at a

∗Department of Statistics, Iowa State University, Ames, Iowa e-mail: vroy@iastate.edu
†Department of Statistics, University of Missouri, Columbia, Missouri e-mail: chakrabortys@missouri.edu
time. Starting with the innermost expectation, doing similar calculations as in Khare and Hobert (2013), we have

\[ E[V(\beta, \tau^2, \sigma^2)|\tau^2, \sigma^2] = E[(\tilde{y} - X\beta)^T(\tilde{y} - X\beta) + \beta^T D_{\tau^0}^{-1}\beta|\tau^2, \sigma^2] + \sum_{j=1}^{p} \tau_{j}^2 \]

\[ \leq \tilde{y}^T \tilde{y} + \sum_{j=1}^{p} \tau_{j}^2 + p \sigma^2. \quad (A.2) \]

Next, using the fact that if \( W \sim \text{Inverse-Gaussian} \ (\delta, \xi) \) then \( E(W^{-1}) = \delta^{-1} + \xi^{-1} \), we have

\[ E[\tau_{j}^2|\beta_0, \sigma^2] = \sqrt{\frac{\beta_{0j}^2}{\lambda_i^2 \sigma^2} + 1} \]

\[ = \sqrt{\frac{n - 1 + p + 2\alpha}{\lambda_i^2} \frac{\beta_{0j}^2}{(n - 1 + p + 2\alpha)\sigma^2} + 1} \]

\[ \leq \frac{n - 1 + p + 2\alpha}{2\lambda_i^2} + \frac{\beta_{0j}^2}{2(n - 1 + p + 2\alpha)\sigma^2} + 1. \quad (A.3) \]

Finally, for the outermost expectation in (A.1), note that

\[ E[\sigma^2|\beta_0, \tau_{0j}^2] = \frac{(\tilde{y} - X\beta_0)^T(\tilde{y} - X\beta_0) + \beta_0^T D_{\tau_0}^{-1}\beta_0 + 2\xi}{n + p + 2\alpha - 3}, \quad (A.4) \]

and

\[ E\left[ \frac{1}{\sigma^2}|\beta_0, \tau_{0j}^2 \right] = \frac{n - 1 + p + 2\alpha}{(\tilde{y} - X\beta_0)^T(\tilde{y} - X\beta_0) + \beta_0^T D_{\tau_0}^{-1}\beta_0 + 2\xi}. \quad (A.5) \]

Using (A.2) to (A.5), from (A.1) we have

\[ (KV)(\beta_0, \tau_{0j}^2, \sigma_0^2) \leq \tilde{y}^T \tilde{y} + \frac{p(n - 1 + p + 2\alpha)}{2\lambda_i^2} + \frac{p}{\lambda_i^2} + \frac{p}{2(n - 1 + p + 2\alpha)\sum_{j=1}^{p} \beta_{0j}^2} \]

\[ + \frac{2(n - 1 + p + 2\alpha)}{(n - 1 + p + 2\alpha + 1)\beta_{0j}^2} \left\{ (\tilde{y} - X\beta_0)^T(\tilde{y} - X\beta_0) + \beta_0^T D_{\tau_0}^{-1}\beta_0 + 2\xi \right\} \]

\[ + \frac{p}{n + p + 2\alpha - 3} \left\{ (\tilde{y} - X\beta_0)^T(\tilde{y} - X\beta_0) + \beta_0^T D_{\tau_0}^{-1}\beta_0 \right\} \]

\[ \leq \tilde{y}^T \tilde{y} + \frac{p(n - 1 + p + 2\alpha)}{2\lambda_i^2} + \frac{p}{\lambda_i^2} + \frac{p}{2\beta_0^T D_{\tau_0}^{-1}\beta_0} + \frac{2p\xi}{n + p + 2\alpha - 3} + \]

\[ + \frac{p}{n + p + 2\alpha - 3} \left\{ (\tilde{y} - X\beta_0)^T(\tilde{y} - X\beta_0) + \beta_0^T D_{\tau_0}^{-1}\beta_0 \right\} \]

\[ \leq \tilde{y}^T \tilde{y} + \frac{p(n - 1 + p + 2\alpha)}{2\lambda_i^2} + \frac{p}{\lambda_i^2} + \frac{2p\xi}{n + p + 2\alpha - 3} + \]

\[ + \frac{p}{n + p + 2\alpha - 3} \left\{ (\tilde{y} - X\beta_0)^T(\tilde{y} - X\beta_0) + \beta_0^T D_{\tau_0}^{-1}\beta_0 \right\} + \frac{1}{2} \sum_{j=1}^{p} \tau_{0j}^2. \quad (A.6) \]
The second inequality is due to the fact that \((y - X\beta_0)^T(y - X\beta_0) + 2\xi \geq 0\). The last inequality follows since \(\beta_0^TD_{\tau_0}^{-1}\beta_0 = \sum_{j=1}^p \beta_{0j}^2(\tau_{0j}^2 + \lambda_2) \geq \sum_{j=1}^p \beta_{0j}^2\tau_{0j}^2\) and from Khare and Hobert (2013) we have \(\sum_{j=1}^p \beta_{0j}^2\tau_{0j}^2 \geq \sum_{j=1}^p \beta_{0j}^2/\sum_{j=1}^p \tau_{0j}^2\).

From (A.6) we see that (16) holds with \(\gamma = \max(1/2, p/(n + p + 2\alpha - 3))\) and

\[
d = \tilde{y}^T\tilde{y} + \frac{p(n - 1 + p + 2\alpha)}{2\lambda_1^2} + \frac{p}{\lambda_1^2} + \frac{2p\xi}{n + p + 2\alpha - 3}.
\]

\[\Box\]

**Proof of Lemma 2.** Recall from (2.9) that

\[
k((\beta, \tau^2, \sigma^2)|(\beta_0, \tau_0^2, \sigma_0^2)) = \pi(\beta|\tau^2, \sigma^2, y)\pi(\tau^2|\beta_0, \sigma^2, y)\pi(\sigma^2|\beta_0, \tau_0^2, y).
\]

As in Khare and Hobert (2013), we begin with analyzing the density \(\pi(\tau^2|\beta_0, \sigma^2, y)\). In fact, doing similar calculations as in Khare and Hobert (2013), and the fact that \(\beta_0^TD_{\tau_0}^{-1}\beta_0 \geq \sum_{j=1}^p \beta_{0j}^2/\sum_{j=1}^p \tau_{0j}^2\), we have for every \((\beta_0, \tau_0^2, \sigma_0^2) \in B_{V,L}\)

\[
\pi(\tau^2|\beta_0, \sigma^2, y) \geq \prod_{j=1}^p g_1(\tau_j^2|\sigma^2)e^{-\sqrt{\frac{\lambda_1^2\tau_j^2}{\sigma^2}}}, \tag{A.7}
\]

where \(g_1(\tau_j^2|\sigma^2)\) is the pdf of the reciprocal of an Inverse-Gaussian random variable with parameters \(\lambda_1\sigma/L\) and \(\lambda_1^2\).

Next we consider the conditional density \(\pi(\sigma^2|\beta_0, \tau_0^2, y)\). For every \((\beta_0, \tau_0^2, \sigma_0^2) \in B_{V,L}\) we have

\[
(y - X\beta_0)^T(y - X\beta_0) + \beta_0^TD_{\tau_0}^{-1}\beta_0 + 2\xi \leq L + 2\xi, \tag{A.8}
\]

and from Khare and Hobert (2013) it follows that

\[
(y - X\beta_0)^T(y - X\beta_0) + \beta_0^D_{\tau_0}^{-1}\beta_0 + 2\xi \geq \tilde{y}^T\tilde{y} - y^TX(X^TX + D_{\tau_0}^{-1})^{-1}X^T\tilde{y} + 2\xi. \tag{A.9}
\]

Now for all \((\beta_0, \tau_0^2, \sigma_0^2) \in B_{V,L}, 1/\tau_{0j}^2 \geq 1/L\) for all \(j = 1, \ldots, p\). So \(D_{\tau_0}^{-1} - (\lambda_2 + 1/L)I_p\) is a positive definite matrix and hence from (A.9) we have

\[
(y - X\beta_0)^T(y - X\beta_0) + \beta_0^D_{\tau_0}^{-1}\beta_0 + 2\xi \geq \tilde{y}^T\tilde{y} - y^TX(X^TX + (\lambda_2 + 1/L)I_p)^{-1}X^T\tilde{y} + 2\xi. \tag{A.10}
\]

From (A.8) and (A.10) we have for every \((\beta_0, \tau_0^2, \sigma_0^2) \in B_{V,L},

\[
e^{-\frac{\sqrt{\lambda_1^2\tau_j^2}}{\sigma^2}}\pi(\sigma^2|\beta_0, \tau_0^2, y)
\]

\[
\geq \frac{((y - X\beta_0)^TX^TX + (\lambda_2 + 1/L)I_p)^{-1}X^T\tilde{y} + 2\xi/2)^{n-1+p+2\alpha/2}}{(\sigma^2)^{(n-1+p+2\alpha)/2+1}\Gamma((n - 1 + p + 2\alpha)/2)}e^{-\frac{L+2\xi}{2\sigma^2}}e^{-\frac{p\sqrt{\lambda_1^2\tau_j^2}}{2\sigma^2}}
\]

\[
\geq \frac{((y - X\beta_0)^TX^TX + (\lambda_2 + 1/L)I_p)^{-1}X^T\tilde{y} + 2\xi/2)^{n-1+p+2\alpha/2}}{(\sigma^2)^{(n-1+p+2\alpha)/2+1}\Gamma((n - 1 + p + 2\alpha)/2)}e^{-\frac{L+2\xi}{2\sigma^2}}e^{-\frac{1}{2}p\sqrt{\lambda_1^2\tau_j^2}}
\]
Selection of tuning parameters and solution paths for Bayesian lassos

\[ g_2(\sigma^2) \]

where \( g_2(\sigma^2) \) is the density of the Inverse-Gamma distribution with parameters \((n - 1 + p + 2\alpha)/2) \) and \((L + 2\xi + p^2\lambda_1^2 L^2)/2\). Combining (A.7) and (A.11), from (11) we have

\[ k((\beta, \tau^2, \sigma^2)|((\beta_0, \tau_0^2, \sigma_0^2)) \geq \epsilon u(\beta, \tau^2, \sigma^2), \]

where

\[ \epsilon = e^{-\frac{1}{2}} \left( \frac{\beta^T \beta - \beta^T X (X^T X + (\lambda_2 + 1/L)I_p)^{-1} X^T \beta + 2\xi (\lambda_2 + 1/L)^{-1}}{L + 2\xi + p^2\lambda_1^2 L^2} \right)^{(n-1+p+2\alpha)/2} g_2(\sigma^2), \quad (A.11) \]

and \( u(\beta, \tau^2, \sigma^2) = (\prod_{j=1}^{p} g_1(\tau_j^2|\sigma^2))g_2(\sigma^2). \) Thus (17) is established.

Proof of Lemma 3. We first consider the ratio

\[ \frac{\pi(\beta, \tau^2, \sigma^2|\lambda)}{\pi(\beta, \tau^2, \sigma^2|\lambda')} = \left( \frac{C(\lambda_1, \lambda_2)}{C(\lambda_1, \lambda_2')} \right)^p \prod_{j=1}^{p} \exp \left\{ -\frac{(\lambda_2 - \lambda_2')}{2\sigma^2} \sum_{j=1}^{p} \beta_j \right\} \exp \left\{ -\frac{(\lambda_1 - \lambda_1')}{2} \sum_{j=1}^{p} \tau_j^2 \right\} \leq \left( \frac{C(\lambda_1, \lambda_2)}{C(\lambda_1, \lambda_2')} \right)^p. \]

where the last inequality follows from the facts that \( \lambda_1' < \lambda_1, \lambda_2' < \lambda_2. \)

Next the proof of the lemma follows since

\[ Z(\theta) = \frac{\pi(\beta, \tau^2, \sigma^2|\lambda)}{\sum_{i=0}^{k-1} \alpha_i \pi(\beta, \tau^2, \sigma^2|\lambda_i)/\tau_i} \leq \frac{m_{\lambda^*}(\lambda)}{m_{\lambda^*}(\lambda')} g(\beta, \tau^2, \sigma^2|\lambda^*). \]

Appendix B: Summary of the steps involved in inference

Below we summarize the steps involved in the estimation of the tuning parameters and the solution paths in the context of EN.

- Finding EB estimate of \( \lambda = (\lambda_1, \lambda_2) \) and plotting the solution paths.

Stage 1 Generate MCMC samples \( \{\beta(j), \tau(j), \sigma(j, l)\}_{l=1}^{M_j} \) with stationary density \( \pi(\beta, \tau^2, \sigma^2|y, \lambda^*) \) for \( j = 0, \ldots, k - 1 \), and use these samples to estimate \( \mathbf{r} \) by the reverse logistic regression method.
Stage 2 Independently of Stage 1, again generate MCMC samples \( \{ \beta^{(j:t)}, \tau^{2(j:t)}, \sigma^{2(j:t)} \}_{t=1}^{M} \) with stationary density \( \pi(\beta, \tau^2, \sigma^2 | y, \lambda^j) \) for \( j = 0, \ldots, k-1 \). Estimate the BFs \( B_{\lambda, \lambda^0} \) by (21) based on these new samples and \( \hat{r} \) computed in Stage 1. Find \( \hat{\lambda} \) by maximizing \( \hat{B}_{\lambda, \lambda^0}(\hat{r}) \).

Estimate the whole solution path by \( \hat{\eta}^{[\beta]} \) given in (24) using the stage 2 samples.

- Estimating \( \beta, \sigma^2, \tau^2 \) and predicting \( y \).

Once the EB estimate \( \hat{\lambda} \) of \( \lambda \) is formed as described above, generate new MCMC samples \( \{ \beta^i, \tau^{2i}, \sigma^{2i} \}_{i=1}^{M} \) with stationary density \( \pi(\beta, \tau^2, \sigma^2 | y, \hat{\lambda}) \). Estimate \( \beta, \sigma^2, \tau^2 \) by the time average estimators, \( \bar{\beta} \equiv \frac{1}{M} \sum_{i=1}^{M} \beta^i / M \), etc. The minimum mean squared error predictor of \( y \) at a new \( x \) is given by \( \bar{\mu} + x^T \bar{\beta} \).

References