Fully Conditional Distributions for Implementing the BRL (Bayesian Regression Coupled with LASSO) Model

The fully conditional distributions of the unknowns in the Bayesian model of [8] and [9] are given. The derivation uses standard results for Bayesian linear models (e.g., SORENSEN and GIANOLA, 2002), and results for the Bayesian LASSO (BL, PARK and CASELLA, 2008).

The joint posterior distribution is

\[
p(\mu, \beta, \tau, \lambda | y) \propto \left\{ \prod_{i=1}^{n} N(y_i | \mu + x_i' \beta_i + x_i' \beta_i + u_i, \sigma_\varepsilon^2) \right\} \times N(\mu | 0, \sigma_\mu^2) \times \prod_{j=1}^{p} N(\beta_j | 0, \sigma_\beta^2 \tau_j^2) \times N(u | 0, A \sigma_u^2) \times \chi^{-2}(\sigma_\varepsilon^2 | S_\varepsilon, df_\varepsilon) \chi^{-2}(\sigma_\beta^2 | S_\beta, df_\beta) \chi^{-2}(\sigma_u^2 | S_u, df_u)
\]

The fully conditional distribution of any unknown is obtained by removing from the right-hand-side of [10] the components that do not involve that unknown. When conjugate priors were chosen, the remaining components are kernels of known distributions, as described next.

1. Intercept

\[
p(\mu | \text{else}) \propto \left\{ \prod_{i=1}^{n} N(y_i | \mu + x_i' \beta + x_i' \beta + u_i, \sigma_\varepsilon^2) \right\} \times N(\mu | 0, \sigma_\mu^2)
\]

where, \( y_i' = y_i - x_i' \beta - x_i' \beta - u_i \). The right-hand side of the above expression is recognized as the kernel of a normal distribution with mean \( \left[ n \sigma_\varepsilon^{-2} + \sigma_\mu^{-2} \right]^{\frac{1}{2}} \sum \sigma_\varepsilon^{-2} \) and variance \( \left[ n \sigma_\varepsilon^{-2} + \sigma_\mu^{-2} \right]^{\frac{1}{2}} \). In practice one can set \( \sigma_\mu^2 \) large enough so that an effectively flat prior is assigned to the intercept.
2. Regression coefficient with homogeneous-variance Normal prior ($\beta_r$).

\[
p(\beta_r | y) \propto \prod_{i=1}^{n} N(y_i | \mu + x_i^r \beta_r + x_i^h \beta_h + u_i, \sigma_r^2) \prod_{i=1}^{n} N(\beta_r | 0, I) \sigma_r^2
\]

\[
= \prod_{i=1}^{n} N(y_i | x_i^r \beta_r, \sigma_r^2) \prod_{i=1}^{n} N(\beta_r | 0, I) \sigma_r^2
\]

where, $y_i = y_i - \mu - x_i^r \beta_r - u_i$. This is recognized as a multivariate-normal distribution (MVN) with mean vector (co-variance matrix) equal to the solution (inverse of the coefficient matrix) of the system of equations,

\[
[X_r X_r \sigma_r^{-2} + I] \hat{\beta}_r = X_r y_i^\ast \sigma_r^{-2}
\]

[11a]

Alternatively, if one wants to draw samples from the fully conditional distributions of each element of $\beta_r$, one has,

\[
p(\beta_r | y) \propto \prod_{i=1}^{n} N(y_i | \mu + x_i^r \beta_r + x_i^h \beta_h + u_i, \sigma_r^2) \prod_{i=1}^{n} N(\beta_r | 0, I) \sigma_r^2
\]

\[
= \prod_{i=1}^{n} N(y_i | x_i^r \beta_r, \sigma_r^2) \prod_{i=1}^{n} N(\beta_r | 0, I) \sigma_r^2
\]

where, $y_i^\ast = y_i - \sum_{k \neq r} x_{ik} \beta_k$. The right-hand-side of the above expression is recognized as the kernel of a normal distribution with mean and variance equal to the solution (inverse of the coefficient in the left-hand-side) of,

\[
[\sigma_r^{-2} \sum_{i=1}^{n} x_{ij}^2 + \sigma_r^{-2}] \hat{\beta}_{ij} = \sigma_r^{-2} \sum_{i=1}^{n} x_{ij} y_i^\ast
\]

[11b]
3. Regression coefficient with heterogeneous-variance Normal prior \((\beta_j)\).

\[
p(\beta_j | \text{else}) \propto \prod_{i=1}^{n} N(y_i | \mu + \mathbf{x}_i \beta_r + \mathbf{x}_i \beta_i + u_i, \sigma^2) \prod_{j=1}^{p} N(\beta_{ij} | 0, \sigma^2 \tau_j^2)
\]

\[
\propto \prod_{i=1}^{n} N(y_i^{**} | \mathbf{x}_i \beta_i, \sigma^2) \prod_{j=1}^{p} N(\beta_{ij} | 0, \sigma^2 \tau_j^2),
\]

where, \(y_i^{**} = y_i - \mu - \mathbf{x}_i \beta_r - u_i\). This is recognized as a MVN distribution with mean vector (co-variance matrix) equal to the solution (inverse of the coefficient matrix) of the system of equations,

\[
\left[\mathbf{X}' \mathbf{X} + \text{Diag} \left( \sigma^2 \tau_j^2 \right) \right]\hat{\beta}_i = \mathbf{X}' y^{**} \sigma^2.
\]

[12a]

Again, if one wishes to draw samples form the fully conditional distribution of each of the elements of \(\beta_i\), one has,

\[
p(\beta_{ij} | \text{else}) \propto \prod_{i=1}^{n} N(y_i | \mu + \mathbf{x}_i \beta_r + \mathbf{x}_i \beta_i + u_i, \sigma^2) \prod_{j=1}^{p} N(\beta_{ij} | 0, \sigma^2 \tau_j^2)
\]

\[
\propto \prod_{i=1}^{n} N(y_i^{**} | \mathbf{x}_i \beta_{ij}, \sigma^2) N(\beta_{ij} | 0, \sigma^2 \tau_j^2),
\]

where \(y_i^{**} = y_i^{**} - \sum_{k \neq j} x_{ik} \beta_{jk}\). The right-hand-side of the above expression is recognized as the kernel of a normal distribution with mean and variance equal to the solution (inverse of the coefficient of the left-hand-side) of the following equation,

\[
\left[\sigma^2 \sum_{i=1}^{n} x_{ij}^2 + \sigma^2 \tau_j^2 \right] \hat{\beta}_{ij} = \sigma^2 \sum_{i=1}^{n} x_{ij} y_i^{**}.
\]

[12b]

4. Infinitesimal additive effects \((u)\).

\[
p(u | \text{else}) \propto \prod_{i=1}^{n} N(y_i | \mu + \mathbf{x}_i \beta_r + \mathbf{x}_i \beta_i + u_i, \sigma^2) N(u | 0, \mathbf{A} \sigma_u^2)
\]

\[
\propto \prod_{i=1}^{n} N(y_i^{****} | u_i, \sigma^2) N(u | 0, \mathbf{A} \sigma_u^2),
\]
where, \( y_i^{\*\*\*} = y_i - \mu - x'_i \beta_r - x'_i \beta_l \). This is recognized as a MVN distribution with mean vector (co-variance matrix) equal to the solution (inverse of the coefficient matrix) of the system,

\[
\begin{bmatrix}
I \sigma^2_\epsilon + A^{-1} \sigma^2_u
\end{bmatrix} \hat{\mathbf{u}} = \mathbf{y}^{\*\*\*} \sigma^2_\epsilon.
\]

[13a]

It follows that each of the entries of \( \mathbf{u} \) also has a fully conditional distribution that is normal (e.g., SORENSEN and GIANOLA, 2002), with the following mean and variance,

\[
E(u_i | \text{else}) = c_{ii}^{-1} \left( r_i s_i - \sum_{k \neq i} c_{ik} y_k \right) ; \quad \text{Var}(u_i | \text{else}) = c_{ii}^{-1}
\]

[13b]

where, \( c_{ii} \) and \( r_i s_i \) are the \( i \)th diagonal element and the \( i \)th entry of \( \mathbf{C} = [I \sigma^2_\epsilon + A^{-1} \sigma^2_u] \) and \( \mathbf{rhs} = \mathbf{y}^{\*\*\*} \sigma^2_\epsilon \), respectively.

5. Residual variance \((\sigma^2_\epsilon)\).

\[
p(\sigma^2_\epsilon | \text{else}) \propto \prod_{i=1}^{n} N(y_i \mid \mu + x'_i \beta_r + x'_i \beta_l + u_i, \sigma^2_\epsilon) \prod_{i=1}^{p} N(\beta_j \mid 0, \sigma^2_\tau_j) \chi^2(\sigma^2_\epsilon | S_\epsilon, df_\epsilon)
\]

\[
= \chi^2(\sigma^2_\epsilon | S_\epsilon + \mathbf{e} \mathbf{e}' + \mathbf{\beta} \mathbf{\tilde{\beta}}, df = df_\epsilon + n + p)
\]

[14]

where \( \mathbf{e} = (e_1, ..., e_n)' = \mathbf{y} - \mathbf{1} \mu - \mathbf{X} \beta_r - \mathbf{X} \beta_l - \mathbf{u} \), \( \mathbf{\tilde{\beta}}_j = \left( \frac{\beta_j}{\tau_j} \right) \) and \( \mathbf{\tilde{\beta}} = (\mathbf{\tilde{\beta}}_1, ..., \mathbf{\tilde{\beta}}_p)' \).

6. Additive-genetic variance \((\sigma^2_u)\).

\[
p(\sigma^2_u | \text{else}) \propto N(\mathbf{u} \mid \mathbf{0}, \mathbf{A} \sigma^2_u) \chi^2(\sigma^2_u | S_u, df_u)
\]

\[
= \chi^2(\sigma^2_u | S_u + \mathbf{u}' \mathbf{A}^{-1} \mathbf{u}, df = df_u + q)
\]

[15]

where \( q \) is the order of the square matrix \( \mathbf{A} \).

7. Scaling variables associated to marker effects \((\tau^2)\).

\[
p(\tau^2 | \text{else}) \propto \prod_{j=1}^{p} N(\beta_j \mid 0, \sigma^2_\tau \tau_j^2) \prod_{j=1}^{p} \text{Exp}(\tau_j^2 | \lambda),
\]

so that, each of the \( \tau_j^2 \)'s are conditionally independent,
\[ p(\tau_j^2 | \text{else}) \propto N(\beta_j, \sigma_j^2, \tau_j) \exp(\tau_j^2 | \lambda) \]

From results presented in PARK and CASELLA, the fully conditional distributions of the reciprocal of the \( \tau_j^2 \)s are Inverse-Gaussian.

\[ p(\tau_j^{-2} | \text{else}) \propto IG(\tau_j^{-2} | \mu_j = \frac{\alpha_j \lambda}{\beta_j}, S = \lambda^2) \]  

[16]

8. LASSO smoothing parameter (\( \lambda \)).

PARK and CASELLA choose a Gamma prior for \( \lambda^2 \), in this case,

\[ p(\lambda^2 | \text{else}) \propto \left[ \prod_{j=1}^{p} \exp(\tau_j^2 | \lambda) \right] G(\lambda^2 | \alpha, \beta). \]

The above is the kernel of the density of a Gamma distribution, with parameters,

\[ p(\lambda^2 | \text{else}) = G(\lambda^2 | b + \alpha_1, \frac{1}{2} \sum_{j=1}^{p} \tau_j^2 + \alpha_2). \]  

[17]

An alternative is to use a uniform prior: \( p(\lambda^2) = U[a, b] \) with \( 0 \leq a < b \). In this case, the fully conditional distribution is,

\[ p(\lambda^2 | \text{else}) \propto G(\lambda^2 | b, \frac{1}{2} \sum_{j=1}^{p} \tau_j^2) 1(a \leq \lambda^2 \leq b). \]

Samples from the above distribution can be obtained by using rejection sampling, which may be implemented as follows:

(i) sample a candidate from \( G(\lambda^2 | b, \frac{1}{2} \sum_{j=1}^{p} \tau_j^2) \);

(ii) accept the candidate if \( a \leq \lambda^2 \leq b \), otherwise go back to (i).

A more general formulation may be obtained by using the Beta distribution. Let \( \tilde{\lambda} = \frac{\lambda}{\lambda_{\text{max}}} \), where \( \lambda_{\text{max}} > 0 \) is an upper bound on \( \lambda \), and \( p(\tilde{\lambda}) = \text{Beta}(\tilde{\lambda} | \alpha_3, \alpha_4) \). It follows that,
\[ p(\lambda) = \text{Beta}\left[ \lambda | \alpha_3, \alpha_4, \frac{\partial \ln \lambda}{\partial \lambda} \right] \propto \text{Beta} \left( \frac{\lambda}{\max \{\alpha_3, \alpha_4\}} \right). \]

If \( \alpha_3 = \alpha_4 = 1 \) this gives a uniform prior for \( \lambda \) (note that this is different than a uniform prior on \( \lambda^2 \)). When a Beta distribution is used the fully conditional distribution \( p(\lambda | \theta) \) does not have closed form, however one can use the Metropolis-Hastings algorithm to obtain samples. In our implementation, at iteration \( t \), we use the following algorithm:

(i) Sample a candidate, \( \lambda^c \), from \( G\left( \lambda^2 | p, \frac{1}{2} \sum_j \tau_j^2 \right) \).

(ii) Let \( \lambda_t \) and \( \tau_{jt}^2 \) the current sample of the corresponding parameters, evaluate:

\[
\begin{align*}
    r &= \frac{\prod_{j=1}^p \text{Exp}(\tau_{jt}^2 | \lambda_t) \text{Beta} \left( \frac{\lambda_t}{\max \{\alpha_3, \alpha_4\}} \right) G\left( \lambda^2 | p, \frac{1}{2} \sum_j \tau_j^2 \right)}{\prod_{j=1}^p \text{Exp}(\tau_{jt}^2 | \lambda_t) \text{Beta} \left( \frac{\lambda_t}{\max \{\alpha_3, \alpha_4\}} \right) G\left( \lambda^2 | p, \frac{1}{2} \sum_j \tau_j^2 \right)}
\end{align*}
\]

(iii) Sample \( u \sim U(0,1) \)

(iv) Set \( \lambda^{t+1} = \lambda^c \) if \( r \geq u \), otherwise set \( \lambda^{t+1} = \lambda' \).