Supplementary Material for:

Sparse Linear Mixed Model Selection via Streamlined Variational Bayes

BY EMANUELE DEGANI[†], LUCA MAESTRINI[‡],
DOROTA TOCZYDŁOWSKA[♯] & MATT P. WAND[♯]

Università degli Studi di Padova[†],
The Australian National University[‡] *and University of Technology Sydney*[♯]

S.1 Distributions and Associated Useful Results

Many probability distributions are used throughout the paper. We provide details on their probability density functions together with additional useful results.

S.1.1 Inverse-Gaussian Distribution

A continuous random variable x has an Inverse Gaussian distribution with mean parameter $\mu > 0$ and rate parameter $\lambda > 0$, written $x \sim \text{Inverse-Gaussian}(\mu, \lambda)$, if the density function of x is

$$\mathfrak{p}(x) = \lambda^{1/2}(2\pi x^3)^{-1/2} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}, \quad x > 0.$$

S.1.2 Laplace Distribution

A continuous random variable x has a Laplace distribution (also known as *Double-Exponential* or *two-tailed Exponential* distribution) with mean parameter μ and scale parameter $\sigma > 0$, written $x \sim \text{Laplace}(\mu, \sigma)$, if the density function of x is

$$\mathfrak{p}(x) = \frac{1}{2\sigma} \exp\left(-\frac{|x-\mu|}{\sigma}\right), \quad x \in \mathbb{R}.$$

A useful result from Andrews & Mallows (1974) and West (1987) shows that

$$\text{if} \quad x|b \sim \text{N}(\mu, \sigma^2/b) \quad \text{and} \quad b \sim \text{Inverse-Gamma}(1, 1/2),$$
$$\text{then} \quad x \sim \text{Laplace}(\mu, \sigma).$$

S.1.3 Horseshoe Distribution

A continuous random variable x has a Horseshoe distribution with mean parameter μ and scale parameter $\sigma > 0$, written $x \sim \text{Horseshoe}(\mu, \sigma)$, if the density function of x is

$$\mathfrak{p}(x) = (2\pi^3)^{-1/2}\sigma^{-1} \exp\left\{\frac{(x-\mu)^2}{2\sigma^2}\right\} E_1\left\{\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}$$

where $E_1(x) \equiv \int_x^\infty t^{-1}e^{-t}\text{d}t$, with $x \neq 0$, is the exponential integral function of order 1. A useful result from Carvalho *et al.* (2010) shows that

$$\text{if} \quad x|b \sim \text{N}(\mu, \sigma^2/b), \quad b|c \sim \text{Gamma}(1/2, c) \quad \text{and} \quad c \sim \text{Gamma}(1/2, 1),$$
$$\text{then} \quad x \sim \text{Horseshoe}(\mu, \sigma).$$

S.1.4 Normal-Exponential-Gamma Distribution

A continuous random variable x has a Normal-Exponential-Gamma distribution with mean parameter μ, scale parameter $\sigma > 0$ and shape parameter $\lambda > 0$, written $x \sim \text{NEG}(\mu, \sigma, \lambda)$, if the density function of x is

$$\mathfrak{p}(x) = \pi^{-1/2}\sigma^{-1}\lambda 2^{\lambda}\Gamma(\lambda + 1/2) \exp\left\{\frac{(x-\mu)^2}{4\sigma^2}\right\} D_{-2\lambda-1}\left(\left|\frac{x-\mu}{\sigma}\right|\right), \quad x \in \mathbb{R},$$

where $D_\nu(x) \equiv 2^{\nu/2+1/4}W_{\nu/2+1/4,-1/4}(x^2/2)/\sqrt{x},\ x > 0$ is the parabolic cylinder function of order $\nu \in \mathbb{R}$ and $W_{k,m}$ is a confluent hypergeometric function or order k and m, as defined by Whittaker & Watson (1990). A useful result from Griffin & Brown (2011) shows that

$$\text{if} \quad x|b \sim \text{N}(\mu, \sigma^2/b), \quad b|c \sim \text{Inverse-Gamma}(1, c) \quad \text{and} \quad c \sim \text{Gamma}(\lambda/2, 1),$$
$$\text{then} \quad x \sim \text{NEG}(\mu, \sigma, \lambda).$$

S.1.5 Gamma Distribution

A continuous random variable x has a Gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$, written $x \sim \text{Gamma}(\alpha, \beta)$, if the density function of x is

$$\mathfrak{p}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\}, \quad x > 0.$$

S.1.6 Inverse-χ^2 and Inverse-Gamma Distributions

A continuous random variable x has an Inverse-χ^2 distribution with shape parameter $\xi > 0$ and scale parameter $\lambda > 0$, written $x \sim \text{Inverse-}\chi^2(\xi, \lambda)$, if the density function of x is

$$\mathfrak{p}(x) = \frac{(\lambda/2)^{\xi/2}}{\Gamma(\xi/2)} x^{-(\xi/2)-1} \exp\{-(\lambda/2)/x\}, \quad x > 0.$$

A continuous random variable x has an Inverse-Gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$, written $x \sim \text{Inverse-Gamma}(\alpha, \beta)$, if the density function of x is

$$\mathfrak{p}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\{-\beta/x\}, \quad x > 0.$$

Notice that

$$x \sim \text{Inverse-}\chi^2(\xi, \lambda) \quad \text{if and only if} \quad x \sim \text{Inverse-Gamma}(\xi/2, \lambda/2)$$

and, equivalently,

$$x \sim \text{Inverse-Gamma}(\alpha, \beta) \quad \text{if and only if} \quad x \sim \text{Inverse-}\chi^2(2\alpha, 2\beta).$$

S.1.7 Half-Cauchy Distribution

A continuous random variable x has a Half-Cauchy distribution with scale parameter $\sigma > 0$, written $x \sim \text{Half-Cauchy}(\sigma)$, if the density function of x is

$$p(x) = 2/[\sigma\pi\{1 + (x/\sigma)^2\}], \quad x > 0.$$

S.1.8 Half-t Distribution

A continuous random variable x has a Half-t distribution with $\nu > 0$ degrees of freedom and scale parameter $\sigma > 0$, written $x \sim \text{Half-}t(\sigma, \nu)$, if the density function of x is

$$\mathfrak{p}(x) = \frac{2\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\nu/2)\sigma\{1 + (x/\sigma)^2/\nu\}^{\frac{\nu+1}{2}}}, \quad x > 0.$$

If $\nu = 1$, then $x \sim \text{Half-Cauchy}(\sigma)$. Wand *et al.* (2011) show that

$$\text{if} \quad x|a \sim \text{Inverse-Gamma}(\nu/2, \nu/a) \quad \text{and} \quad a \sim \text{Inverse-Gamma}(1/2, 1/A^2),$$
$$\text{then} \quad \sqrt{x} \sim \text{Half-}t(A, \nu).$$

The same result can be equivalently formulated with:

$$x|a \sim \text{Inverse-Gamma}(\nu/2, \nu/(2a)) \quad \text{and} \quad a \sim \text{Inverse-Gamma}(1/2, 1/(2A^2)),$$

or with:

$$x|a \sim \text{Inverse-}\chi^2(\nu, 1/a) \quad \text{and} \quad a \sim \text{Inverse-}\chi^2(1, 1/(\nu A^2)).$$

S.1.9 Inverse-G-Wishart Distribution

The Inverse-G-Wishart distribution arises from the inverse of matrices with a *G-Wishart* distribution (e.g. Atay-Kayis & Massam, 2005; Maestrini & Wand, 2021). For any positive integer d, let G be an undirected graph with d nodes labeled $1, \ldots, d$ and set E consisting of sets of pairs of nodes that are connected by an edge. We say that the symmetric $d \times d$ matrix $\boldsymbol{M}$ *respects* G if

$$\boldsymbol{M}_{ij} = 0 \quad \text{for all} \quad \{i, j\} \notin E.$$

A $d \times d$ random matrix $\boldsymbol{X}$ has an Inverse G-Wishart distribution with graph G and parameters $\xi > 0$ and symmetric $d \times d$ matrix $\boldsymbol{\Lambda}$, written $\boldsymbol{X} \sim \text{Inverse-G-Wishart}(G, \xi, \boldsymbol{\Lambda})$, if and only if the density function of $\boldsymbol{X}$ satisfies

$$\mathfrak{p}(\boldsymbol{X}) \propto |\boldsymbol{X}|^{-(\xi+2)/2} \exp\left\{-\frac{1}{2}\text{tr}(\boldsymbol{\Lambda}\,\boldsymbol{X}^{-1})\right\}$$

over arguments $\boldsymbol{X}$ such that $\boldsymbol{X}$ is symmetric and positive definite and $\boldsymbol{X}^{-1}$ respects G. Two important special cases are

$$G = G_{\text{full}} \equiv \text{totally connected } d\text{-node graph},$$

for which the Inverse G-Wishart distribution coincides with the ordinary Inverse Wishart distribution $\boldsymbol{X} \sim \text{Inverse-Wishart}(\xi - d + 1, \boldsymbol{\Lambda})$, and

$$G = G_{\text{diag}} \equiv \text{totally disconnected } d\text{-node graph},$$

for which the Inverse G-Wishart distribution coincides with a product of independent Inverse Chi-Squared distributions. The subscripts of G_{full} and G_{diag} reflect the fact that $\boldsymbol{X}^{-1}$ is a full matrix and $\boldsymbol{X}^{-1}$ is a diagonal matrix in each special case. In the $d = 1$ special case, the Inverse G-Wishart distribution coincides with the Inverse Chi-Squared distribution.

S.1.10 Huang-Wand Distribution

A $d \times d$ symmetric positive definite matrix $\boldsymbol{\Sigma}$ has a Huang-Wand distribution with $\nu > 0$ degrees of freedom and scale parameters $s_1, \ldots, s_d$, written $\boldsymbol{\Sigma} \sim \text{Huang-Wand}(\nu; s_1, \ldots, s_d)$, if and only if the following augmented representation holds:

$$\boldsymbol{\Sigma}|\boldsymbol{A} \sim \text{Inverse-G-Wishart}(G_{\text{full}}, \nu + 2d - 2, \boldsymbol{A}^{-1})$$
$$\boldsymbol{A} \sim \text{Inverse-G-Wishart}(G_{\text{diag}}, 1, \{\nu\,\text{diag}(s_1^2, \ldots, s_d^2)\}^{-1}).$$

This distribution is defined in Huang & Wand, 2013). The formulation displayed here is provided in Maestrini & Wand (2021).

If $\nu = 2$, a marginally noninformative Huang-Wand prior specification for $\mathbf{\Sigma}$ can be obtained for arbitrarily large scale parameters. This corresponds to the standard deviation parameters $\sigma_j \equiv (\mathbf{\Sigma})^{1/2}_{jj}, 1 \leq j \leq d$, having Half-$t$ distributions with ν degrees of freedom and scale parameter given by s_j, and the correlation parameters $\rho_{jj'} \equiv (\mathbf{\Sigma})^{1/2}_{jj'}/(\sigma_j \sigma_{j'})$, $1 \leq j, j' \leq d$, having a Uniform distribution on the interval $(-1, 1)$.

S.2 Multilevel Sparse Matrix Problem Algorithms

Algorithms 2 and 3 described in Section 5 rely upon two matrix algebraic routines for efficiently solving the two-level and three-level versions of the *multilevel sparse matrix problems* defined in Nolan & Wand (2020). They correspond to the SOLVETWOLEVELSPARSEMATRIX and SOLVETHREELEVELSPARSEMATRIX algorithms that we list hereafter as Algorithms S.1 and S.2, respectively.

We briefly describe two-level and three-level sparse matrix structures, and give explicit definition of the aforementioned routines for efficiently solving the associated linear system problems, following Appendix A of Nolan *et al.* (2020).

S.2.1 Two-Level Sparse Matrix Problems

Two-level sparse matrix problems are summarized in Section 3. Such problems are efficiently solved by the SOLVETWOLEVELSPARSEMATRIX routine, which is here listed as Algorithm S.1 and is justified by Theorem 2.2 of Nolan & Wand (2020).

Algorithm S.1 *The* SOLVETWOLEVELSPARSEMATRIX *algorithm for solving the two-level sparse matrix problem* $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{a}$ *and sub-blocks of* $\boldsymbol{A}^{-1}$ *corresponding to the non-zero sub-blocks of* $\boldsymbol{A}$.

Inputs: $\boldsymbol{a}_1(p\times 1), \boldsymbol{A}_{11}(p\times p), \big\{\big(\boldsymbol{a}_{2,i}(q\times 1), \boldsymbol{A}_{22,i}(q\times q), \boldsymbol{A}_{12,i}(p\times q)\big):\ 1\leq i\leq m\big\}$

$\boldsymbol{\omega} \longleftarrow \boldsymbol{a}_1 \quad ; \quad \boldsymbol{\Omega} \longleftarrow \boldsymbol{A}_{11}$

For $i = 1, \ldots, m$:

$\qquad \boldsymbol{\omega} \longleftarrow \boldsymbol{\omega} - \boldsymbol{A}_{12,i}\boldsymbol{A}_{22,i}^{-1}\boldsymbol{a}_{2,i} \quad ; \quad \boldsymbol{\Omega} \longleftarrow \boldsymbol{\Omega} - \boldsymbol{A}_{12,i}\boldsymbol{A}_{22,i}^{-1}\boldsymbol{A}_{12,i}^T$

$\boldsymbol{A}^{11} \longleftarrow \boldsymbol{\Omega}^{-1} \quad ; \quad \boldsymbol{x}_1 \longleftarrow \boldsymbol{A}^{11}\boldsymbol{\omega}$

For $i = 1, \ldots, m$:

$\qquad \boldsymbol{x}_{2,i} \longleftarrow \boldsymbol{A}_{22,i}^{-1}(\boldsymbol{a}_{2,i} - \boldsymbol{A}_{12,i}^T\boldsymbol{x}_1) \quad ; \quad \boldsymbol{A}^{12,i} \longleftarrow -(\boldsymbol{A}_{22,i}^{-1}\boldsymbol{A}_{12,i}^T\boldsymbol{A}^{11})^T$

$\qquad \boldsymbol{A}^{22,i} \longleftarrow \boldsymbol{A}_{22,i}^{-1}\big(\boldsymbol{I} - \boldsymbol{A}_{12,i}^T\boldsymbol{A}^{12,i}\big)$

Outputs: $\boldsymbol{x}_1, \boldsymbol{A}^{11}, \big\{\big(\boldsymbol{x}_{2,i}, \boldsymbol{A}^{22,i}, \boldsymbol{A}^{12,i}\big):\ 1\leq i\leq m\big\}$.

S.2.2 Three-Level Sparse Matrix Problems

Three-level sparse matrix problems are described in Section 3 of Nolan & Wand (2020). An illustrative three-level sparse matrix example is

$$
\boldsymbol{A} = \begin{bmatrix}
\boldsymbol{A}_{11} & \boldsymbol{A}_{12,1} & \boldsymbol{A}_{12,11} & \boldsymbol{A}_{12,12} & \boldsymbol{A}_{12,2} & \boldsymbol{A}_{12,21} & \boldsymbol{A}_{12,22} & \boldsymbol{A}_{12,23} \\
\boldsymbol{A}_{12,1}^T & \boldsymbol{A}_{22,1} & \boldsymbol{A}_{12,1,1} & \boldsymbol{A}_{12,1,2} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\
\boldsymbol{A}_{12,11}^T & \boldsymbol{A}_{12,1,1}^T & \boldsymbol{A}_{22,11} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\
\boldsymbol{A}_{12,12}^T & \boldsymbol{A}_{12,1,2}^T & \boldsymbol{O} & \boldsymbol{A}_{22,12} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\
\boldsymbol{A}_{12,2}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_{22,2} & \boldsymbol{A}_{12,2,1} & \boldsymbol{A}_{12,2,2} & \boldsymbol{A}_{12,2,3} \\
\boldsymbol{A}_{12,21}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_{12,2,1}^T & \boldsymbol{A}_{22,21} & \boldsymbol{O} & \boldsymbol{O} \\
\boldsymbol{A}_{12,22}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_{12,2,2}^T & \boldsymbol{O} & \boldsymbol{A}_{22,22} & \boldsymbol{O} \\
\boldsymbol{A}_{12,23}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_{12,2,3}^T & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{A}_{22,23}
\end{bmatrix},
$$

which corresponds to level 2 group sizes $n_1 = 2$ and $n_2 = 3$, and a level 3 group size $m = 2$. A generic three-level sparse matrix $\boldsymbol{A}$ consists of the following components:

- A $p \times p$ matrix $\boldsymbol{A}_{11}$, which is assigned the $(1,1)$-block position.
- A set of partitioned matrices $\left\{\left[\begin{array}{c|c|c|c} \boldsymbol{A}_{12,i} & \boldsymbol{A}_{12,ij} & \dots & \boldsymbol{A}_{12,in_i} \end{array}\right] : 1 \le i \le m\right\}$, which is assigned the $(1,2)$-block position. For each $1 \le i \le m$, $\boldsymbol{A}_{12,i}$ is $p \times q_1$, and for each $1 \le j \le n_i$, $\boldsymbol{A}_{12,ij}$ is $p \times q_2$.
- A $(2,1)$-block, which is simply the transpose of the $(1,2)$-block.
- A block diagonal structure along the $(2,2)$-block position, where each sub-block is a two-level sparse matrix, as defined in Section 3.3. For each $1 \le i \le m$, $\boldsymbol{A}_{22,i}$ is $q_1 \times q_1$, and for each $1 \le j \le n_i$, $\boldsymbol{A}_{12,i,j}$ is $q_1 \times q_2$ and $\boldsymbol{A}_{22,ij}$ is $q_2 \times q_2$.

The three-level sparse matrix problem arising from the illustrative example above is defined as finding the vector $\boldsymbol{x}$ such that $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{a}$, where:

$$
\boldsymbol{a} \equiv \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_{2,1} \\ \boldsymbol{a}_{2,11} \\ \boldsymbol{a}_{2,12} \\ \boldsymbol{a}_{2,2} \\ \boldsymbol{a}_{2,21} \\ \boldsymbol{a}_{2,22} \\ \boldsymbol{a}_{2,23} \end{bmatrix} \quad \text{and} \quad \boldsymbol{x} \equiv \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_{2,1} \\ \boldsymbol{x}_{2,11} \\ \boldsymbol{x}_{2,12} \\ \boldsymbol{x}_{2,2} \\ \boldsymbol{x}_{2,21} \\ \boldsymbol{x}_{2,22} \\ \boldsymbol{x}_{2,23} \end{bmatrix},
$$

and determining the sub-blocks of $\boldsymbol{A}^{-1}$ corresponding to the non-zero sub-blocks of $\boldsymbol{A}$. The structure of $\boldsymbol{A}^{-1}$ is

$$\boldsymbol{A}^{-1}=\begin{bmatrix} \boldsymbol{A}^{11} & \boldsymbol{A}^{12,1} & \boldsymbol{A}^{12,11} & \boldsymbol{A}^{12,12} & \boldsymbol{A}^{12,2} & \boldsymbol{A}^{12,21} & \boldsymbol{A}^{12,22} & \boldsymbol{A}^{12,23} \\ \boldsymbol{A}^{12,1\,T} & \boldsymbol{A}^{22,1} & \boldsymbol{A}^{12,1,1} & \boldsymbol{A}^{12,1,2} & \times & \times & \times & \times \\ \boldsymbol{A}^{12,11\,T} & \boldsymbol{A}^{12,1,1\,T} & \boldsymbol{A}^{22,11} & \times & \times & \times & \times & \times \\ \boldsymbol{A}^{12,12\,T} & \boldsymbol{A}^{12,1,2\,T} & \times & \boldsymbol{A}^{22,12} & \times & \times & \times & \times \\ \boldsymbol{A}^{12,2\,T} & \times & \times & \times & \boldsymbol{A}^{22,2} & \boldsymbol{A}^{12,2,1} & \boldsymbol{A}^{12,2,2} & \boldsymbol{A}^{12,2,3} \\ \boldsymbol{A}^{12,21\,T} & \times & \times & \times & \boldsymbol{A}^{12,2,1\,T} & \boldsymbol{A}^{22,21} & \times & \times \\ \boldsymbol{A}^{12,22\,T} & \times & \times & \times & \boldsymbol{A}^{12,2,2\,T} & \times & \boldsymbol{A}^{22,22} & \times \\ \boldsymbol{A}^{12,23\,T} & \times & \times & \times & \boldsymbol{A}^{12,2,3\,T} & \times & \times & \boldsymbol{A}^{22,23} \end{bmatrix}.$$

For a general three-level sparse linear system problem, $\boldsymbol{a}_1$ and $\boldsymbol{x}_1$ are $p \times 1$ vectors. For each $1 \le i \le m$, $\boldsymbol{a}_{2,i}$ and $\boldsymbol{x}_{2,i}$ are $q_1 \times 1$ vectors. For each $1 \le i \le m$ and $1 \le j \le n_i$ the vectors $\boldsymbol{a}_{2,ij}$ and $\boldsymbol{x}_{2,ij}$ have dimension $q_2 \times 1$.

Such problems are efficiently solved by the SolveThreeLevelSparseMatrix routine, which is here listed as Algorithm S.2 and is justified by Theorem 3.2 of Nolan & Wand (2020).

Algorithm S.2 *The* SolveThreeLevelSparseMatrix *algorithm for solving the three-level sparse matrix problem* $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{a}$ *and sub-blocks of* $\boldsymbol{A}^{-1}$ *corresponding to the non-zero sub-blocks of* $\boldsymbol{A}$.

Input: $\boldsymbol{a}_1(p\times 1), \boldsymbol{A}_{11}(p\times p), \big\{\big(\boldsymbol{a}_{2,i}(q_1\times 1), \boldsymbol{A}_{22,i}(q_1\times q_1), \boldsymbol{A}_{12,i}(p\times q_1) : \ 1 \le i \le m\big\},$
$\big\{\boldsymbol{a}_{2,ij}(q_2\times 1), \boldsymbol{A}_{22,ij}(q_2\times q_2), \boldsymbol{A}_{12,ij}(p\times q_2), \boldsymbol{A}_{12,i,j}(q_1\times q_2)\big) : \ 1 \le i \le m, \ 1 \le j \le n_i\big\}.$

$\boldsymbol{\omega} \longleftarrow \boldsymbol{a}_1 \quad ; \quad \boldsymbol{\Omega} \longleftarrow \boldsymbol{A}_{11}$

For $i = 1, \ldots, m$:

$\quad \boldsymbol{h}_{2,i} \longleftarrow \boldsymbol{a}_{2,i} \quad ; \quad \boldsymbol{H}_{12,i} \longleftarrow \boldsymbol{A}_{12,i} \quad ; \quad \boldsymbol{H}_{22,i} \longleftarrow \boldsymbol{A}_{22,i}$

$\quad$ For $j = 1, \ldots, n_i$:

$\quad\quad \boldsymbol{h}_{2,i} \longleftarrow \boldsymbol{h}_{2,i} - \boldsymbol{A}_{12,i,j}\boldsymbol{A}_{22,ij}^{-1}\boldsymbol{a}_{2,ij} \quad ; \quad \boldsymbol{H}_{12,i} \longleftarrow \boldsymbol{H}_{12,i} - \boldsymbol{A}_{12,ij}\boldsymbol{A}_{22,ij}^{-1}\boldsymbol{A}_{12,i,j}^T$

$\quad\quad \boldsymbol{H}_{22,i} \longleftarrow \boldsymbol{H}_{22,i} - \boldsymbol{A}_{12,i,j}\boldsymbol{A}_{22,ij}^{-1}\boldsymbol{A}_{12,i,j}^T$

$\quad\quad \boldsymbol{\omega} \longleftarrow \boldsymbol{\omega} - \boldsymbol{A}_{12,ij}\boldsymbol{A}_{22,ij}^{-1}\boldsymbol{a}_{2,ij} \quad ; \quad \boldsymbol{\Omega} \longleftarrow \boldsymbol{\Omega} - \boldsymbol{A}_{12,ij}\boldsymbol{A}_{22,ij}^{-1}\boldsymbol{A}_{12,ij}^T$

$\quad \boldsymbol{\omega} \longleftarrow \boldsymbol{\omega} - \boldsymbol{H}_{12,i}\boldsymbol{H}_{22,i}^{-1}\boldsymbol{h}_{2,i} \quad ; \quad \boldsymbol{\Omega} \longleftarrow \boldsymbol{\Omega} - \boldsymbol{H}_{12,i}\boldsymbol{H}_{22,i}^{-1}\boldsymbol{H}_{12,i}^T$

$\boldsymbol{A}^{11} \longleftarrow \boldsymbol{\Omega}^{-1} \quad ; \quad \boldsymbol{x}_1 \longleftarrow \boldsymbol{A}^{11}\boldsymbol{\omega}$

For $i = 1, \ldots, m$:

$\quad \boldsymbol{x}_{2,i} \longleftarrow \boldsymbol{H}_{22,i}^{-1}(\boldsymbol{h}_{2,i} - \boldsymbol{H}_{12,i}^T\boldsymbol{x}_1) \quad ; \quad \boldsymbol{A}^{12,i} \longleftarrow -(\boldsymbol{H}_{22,i}^{-1}\boldsymbol{H}_{12,i}^T\boldsymbol{A}^{11})^T$

$\quad \boldsymbol{A}^{22,i} \longleftarrow \boldsymbol{H}_{22,i}^{-1}(\boldsymbol{I} - \boldsymbol{H}_{12,i}^T\boldsymbol{A}^{12,i})$

$\quad$ For $j = 1, \ldots, n_i$:

$\quad\quad \boldsymbol{x}_{2,ij} \longleftarrow \boldsymbol{A}_{22,ij}^{-1}\big(\boldsymbol{a}_{2,ij} - \boldsymbol{A}_{12,ij}^T\boldsymbol{x}_1 - \boldsymbol{A}_{12,i,j}^T\boldsymbol{x}_{2,i}\big)$

$\quad\quad \boldsymbol{A}^{12,ij} \longleftarrow -\big\{\boldsymbol{A}_{22,ij}^{-1}\big(\boldsymbol{A}_{12,ij}^T\boldsymbol{A}^{11} + \boldsymbol{A}_{12,i,j}^T\,\boldsymbol{A}^{12,i\,T}\big)\big\}^T$

$\quad\quad \boldsymbol{A}^{12,i,j} \longleftarrow -\big\{\boldsymbol{A}_{22,ij}^{-1}\big(\boldsymbol{A}_{12,ij}^T\,\boldsymbol{A}^{12,i} + \boldsymbol{A}_{12,i,j}^T\,\boldsymbol{A}^{22,i}\big)\big\}^T$

$\quad\quad \boldsymbol{A}^{22,ij} \longleftarrow \boldsymbol{A}_{22,ij}^{-1}\big(\boldsymbol{I} - \boldsymbol{A}_{12,ij}^T\boldsymbol{A}^{12,ij} - \boldsymbol{A}_{12,i,j}^T\,\boldsymbol{A}^{12,i,j}\big)$

Outputs: $\boldsymbol{x}_1, \boldsymbol{A}^{11}, \big\{\big(\boldsymbol{x}_{2,i}, \boldsymbol{A}^{22,i}, \boldsymbol{A}^{12,i}\big) : \ 1 \le i \le m\big\},$
$\big\{\big(\boldsymbol{x}_{2,ij}, \boldsymbol{A}^{22,ij}, \boldsymbol{A}^{12,ij}, \boldsymbol{A}^{12,i,j},\big) : \ 1 \le i \le m, \ 1 \le j \le n_i\big\}.$

S.3 Derivations

We derive explicit updates for the parameters of the optimal density functions in (16) using arguments similar to those provided in Neville *et al.* (2014) with some adjustments. Such updates are then combined with the derivations in Appendix B of Nolan *et al.* (2020) for deriving the streamlined MFVB algorithms for our two- and three- level linear mixed-effects models, i.e. Algorithms 2 and 3.

S.3.1 Derivation of $\mathfrak{q}^*(\beta_0, \boldsymbol{\beta})$ and Associated Parameter Updates

Given the model formulation (17), the full conditional density function for $(\beta_0, \boldsymbol{\beta})$ is:

$$
\begin{aligned}
\mathfrak{p}(\beta_0, \boldsymbol{\beta}|\text{rest}) &\propto \mathfrak{p}(\boldsymbol{y}|\beta_0, \boldsymbol{\beta}, \sigma^2)\, \mathfrak{p}(\beta_0)\, \mathfrak{p}(\boldsymbol{\beta}|\boldsymbol{\zeta}, \tau^2) \\
&\propto \exp\Bigg\{ -\frac{1}{2}\Bigg[\begin{bmatrix} \beta_0 \\ \boldsymbol{\beta} \end{bmatrix}^T \left(\frac{1}{\sigma^2} [\,\mathbf{1} \,|\, \boldsymbol{X}\,]^T [\,\mathbf{1} \,|\, \boldsymbol{X}\,] + \begin{bmatrix} \sigma_{\beta_0}^{-2} & \mathbf{0}^T \\ \mathbf{0} & \tau^{-2}\,\text{diag}(\boldsymbol{\zeta}) \end{bmatrix} \right) \begin{bmatrix} \beta_0 \\ \boldsymbol{\beta} \end{bmatrix} \\
&\qquad - 2 \begin{bmatrix} \beta_0 \\ \boldsymbol{\beta} \end{bmatrix}^T \left(\frac{1}{\sigma^2} [\,\mathbf{1} \,|\, \boldsymbol{X}\,]^T \boldsymbol{y} + \begin{bmatrix} \sigma_{\beta_0}^{-2} & \mathbf{0}^T \\ \mathbf{0} & \tau^{-2}\,\text{diag}(\boldsymbol{\zeta}) \end{bmatrix} \begin{bmatrix} \mu_{\beta_0} \\ \mathbf{0} \end{bmatrix} \right) \Bigg] \Bigg\}
\end{aligned}
$$

and application of (8) results in:

$$
\begin{aligned}
&\mathfrak{q}^*(\beta_0, \boldsymbol{\beta}) \propto \exp\{E_{\mathfrak{q}}\{\log \mathfrak{p}(\beta_0, \boldsymbol{\beta}|\text{rest})\}\} \\
&\propto \exp\Bigg\{ -\frac{1}{2}\Bigg[\begin{bmatrix} \beta_0 \\ \boldsymbol{\beta} \end{bmatrix}^T \left(E_{\mathfrak{q}}\{\sigma^{-2}\} [\,\mathbf{1} \,|\, \boldsymbol{X}\,]^T [\,\mathbf{1} \,|\, \boldsymbol{X}\,] + \begin{bmatrix} \sigma_{\beta_0}^{-2} & \mathbf{0}^T \\ \mathbf{0} & E_{\mathfrak{q}}\{\tau^{-2}\}\,\text{diag}(E_{\mathfrak{q}}\{\boldsymbol{\zeta}\}) \end{bmatrix} \right) \begin{bmatrix} \beta_0 \\ \boldsymbol{\beta} \end{bmatrix} \\
&\quad - 2 \begin{bmatrix} \beta_0 \\ \boldsymbol{\beta} \end{bmatrix}^T \left(E_{\mathfrak{q}}\{\sigma^{-2}\} [\,\mathbf{1} \,|\, \boldsymbol{X}\,]^T \boldsymbol{y} + \begin{bmatrix} \sigma_{\beta_0}^{-2} & \mathbf{0}^T \\ \mathbf{0} & E_{\mathfrak{q}}\{\tau^{-2}\}\,\text{diag}(E_{\mathfrak{q}}\{\boldsymbol{\zeta}\}) \end{bmatrix} \begin{bmatrix} \mu_{\beta_0} \\ \mathbf{0} \end{bmatrix} \right) \Bigg] \Bigg\}.
\end{aligned}
$$

After *completion of the square* manipulations for the multivariate Gaussian distribution and standard algebraic manipulations, it follows immediately that:

$$
\mathfrak{q}^*(\beta_0, \boldsymbol{\beta}) \text{ is a N}\left(\boldsymbol{\mu}_{\mathfrak{q}(\beta_0, \boldsymbol{\beta})}, \boldsymbol{\Sigma}_{\mathfrak{q}(\beta_0, \boldsymbol{\beta})}\right) \text{ density function}
$$

with

$$
\boldsymbol{\Sigma}_{\mathfrak{q}(\beta_0, \boldsymbol{\beta})} \longleftarrow \left(\mu_{\mathfrak{q}(1/\sigma^2)}\, [\,\mathbf{1} \,|\, \boldsymbol{X}\,]^T [\,\mathbf{1} \,|\, \boldsymbol{X}\,] + \begin{bmatrix} \sigma_{\beta_0}^{-2} & \mathbf{0}^T \\ \mathbf{0} & \mu_{\mathfrak{q}(1/\tau^2)}\,\text{diag}(\boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\zeta})}) \end{bmatrix} \right)^{-1}
$$

and

$$
\boldsymbol{\mu}_{\mathfrak{q}(\beta_0, \boldsymbol{\beta})} \longleftarrow \boldsymbol{\Sigma}_{\mathfrak{q}(\beta_0, \boldsymbol{\beta})} \left(\mu_{\mathfrak{q}(1/\sigma^2)}\, [\,\mathbf{1} \,|\, \boldsymbol{X}\,]^T \boldsymbol{y} + \begin{bmatrix} \mu_{\beta_0}/\sigma_{\beta_0}^2 \\ \mathbf{0} \end{bmatrix} \right).
$$

S.3.2 Derivation of $\mathfrak{q}^*(\sigma^2)$ and Associated Parameter Updates

Given the model formulation (17), the full conditional density function for σ^2 is:

$$
\begin{aligned}
\mathfrak{p}(\sigma^2|\text{rest}) &\propto \mathfrak{p}(\boldsymbol{y}|\beta_0, \boldsymbol{\beta}, \sigma^2)\, \mathfrak{p}(\sigma^2|a_{\sigma^2}) \\
&\propto (\sigma^2)^{-(\nu_{\sigma^2}+n)/2-1} \exp\left\{ -\frac{1}{\sigma^2} \frac{a_{\sigma^2}^{-1} + \|\boldsymbol{y} - \mathbf{1}\beta_0 - \boldsymbol{X}\boldsymbol{\beta}\|^2}{2} \right\}
\end{aligned}
$$

and application of (8) results in:

$$
\begin{aligned}
\mathfrak{q}^*(\sigma^2) &\propto \exp\left\{E_{\mathfrak{q}}\{\log \mathfrak{p}(\sigma^2|\text{rest})\}\right\} \\
&\propto (\sigma^2)^{-(\nu_{\sigma^2}+n)/2-1} \exp\left\{ -\frac{1}{\sigma^2} \frac{E_{\mathfrak{q}}\{a_{\sigma^2}^{-1}\} + E_{\mathfrak{q}}\{\|\boldsymbol{y} - \mathbf{1}\beta_0 - \boldsymbol{X}\boldsymbol{\beta}\|^2\}}{2} \right\}.
\end{aligned}
$$

After standard algebraic manipulations, it follows immediately that:

$$\mathfrak{q}^*(\sigma^2) \text{ is an Inverse-}\chi^2\left(\xi_{\mathfrak{q}(\sigma^2)}, \lambda_{\mathfrak{q}(\sigma^2)}\right) \text{ density function}$$

with

$$\xi_{\mathfrak{q}(\sigma^2)} \longleftarrow n + \nu_{\sigma^2}$$

and

$$\lambda_{\mathfrak{q}(\sigma^2)} \longleftarrow \mu_{\mathfrak{q}(1/a_{\sigma^2})} + \left\| \boldsymbol{y} - [\,\mathbf{1} \mid \boldsymbol{X}\,]\boldsymbol{\mu}_{\mathfrak{q}(\beta_0,\boldsymbol{\beta})} \right\|^2 + \text{tr}\Big\{\boldsymbol{\Sigma}_{\mathfrak{q}(\beta_0,\boldsymbol{\beta})}[\,\mathbf{1} \mid \boldsymbol{X}\,]^T[\,\mathbf{1} \mid \boldsymbol{X}\,]\Big\}.$$

S.3.3 Derivation of $\mathfrak{q}^*(a_{\sigma^2})$ and Associated Parameter Updates

Given the model formulation (17), the full conditional density function for a_{σ^2} is:

$$\begin{aligned}\mathfrak{p}(a_{\sigma^2}|\text{rest}) &\propto \mathfrak{p}(\sigma^2|a_{\sigma^2})\,\mathfrak{p}(a_{\sigma^2}) \\ &\propto (a_{\sigma^2})^{-(\nu_{\sigma^2}+1)/2-1}\exp\left\{-\frac{1}{a_{\sigma^2}}\frac{\sigma^{-2}+(\nu_{\sigma^2}s_{\sigma^2}^2)^{-1}}{2}\right\}\end{aligned}$$

and application of (8) results in:

$$\begin{aligned}\mathfrak{q}^*(a_{\sigma^2}) &\propto \exp\left\{E_{\mathfrak{q}}\{\log \mathfrak{p}(a_{\sigma^2}|\text{rest})\}\right\} \\ &\propto (a_{\sigma^2})^{-(\nu_{\sigma^2}+1)/2-1}\exp\left\{-\frac{1}{a_{\sigma^2}}\frac{E_{\mathfrak{q}}\{\sigma^{-2}\}+(\nu_{\sigma^2}s_{\sigma^2}^2)^{-1}}{2}\right\}.\end{aligned}$$

After standard algebraic manipulations, it follows immediately that:

$$\mathfrak{q}^*(a_{\sigma^2}) \text{ is an Inverse-}\chi^2\left(\xi_{\mathfrak{q}(a_{\sigma^2})}, \lambda_{\mathfrak{q}(a_{\sigma^2})}\right) \text{ density function}$$

with

$$\xi_{\mathfrak{q}(a_{\sigma^2})} \longleftarrow 1 + \nu_{\sigma^2} \quad \text{and} \quad \lambda_{\mathfrak{q}(a_{\sigma^2})} \longleftarrow \mu_{\mathfrak{q}(1/\sigma^2)} + (\nu_{\sigma^2}s_{\sigma^2}^2)^{-1}.$$

S.3.4 Derivation of $\mathfrak{q}^*(\tau^2)$ and Associated Parameter Updates

Given the model formulation (17), the full conditional density function for τ^2 is:

$$\begin{aligned}\mathfrak{p}(\tau^2|\text{rest}) &\propto \left\{\prod_{h=1}^H \mathfrak{p}(\beta_h|\tau^2)\right\}\mathfrak{p}(\tau^2|a_{\tau^2}) \\ &\propto (\tau^2)^{-(H+1)/2-1}\exp\left\{-\frac{1}{\tau^2}\frac{a_{\tau^2}^{-1}+\sum_{h=1}^H \zeta_h(\beta_h)^2}{2}\right\}\end{aligned}$$

and application of (8) results in:

$$\begin{aligned}\mathfrak{q}^*(\tau^2) &\propto \exp\left\{E_{\mathfrak{q}}\{\log \mathfrak{p}(\tau^2|\text{rest})\}\right\} \\ &\propto (\tau^2)^{-(H+1)/2-1}\exp\left\{-\frac{1}{\tau^2}\frac{E_{\mathfrak{q}}\{a_{\tau^2}^{-1}\}+\sum_{h=1}^H E_{\mathfrak{q}}\{\zeta_h\}E_{\mathfrak{q}}\{\beta_h^2\}}{2}\right\}.\end{aligned}$$

After standard algebraic manipulations, it follows immediately that:

$$\mathfrak{q}^*(\tau^2) \text{ is an Inverse-}\chi^2\left(\xi_{\mathfrak{q}(\tau^2)}, \lambda_{\mathfrak{q}(\tau^2)}\right) \text{ density function}$$

with

$$\xi_{\mathfrak{q}(\tau^2)} \longleftarrow H + 1 \quad \text{and} \quad \lambda_{\mathfrak{q}(\tau^2)} \longleftarrow \mu_{\mathfrak{q}(1/a_{\tau^2})} + \boldsymbol{\mu}_{\mathfrak{q}(\boldsymbol{\zeta})}^T\boldsymbol{\mu}_{\mathfrak{q}((\boldsymbol{\beta})^2)}.$$

S.3.5 Derivation of $\mathfrak{q}^*(a_{\tau^2})$ and Associated Parameter Updates

Given the model formulation (17), the full conditional density function for a_{τ^2} is:

$$\begin{aligned}\mathfrak{p}(a_{\tau^2}|\text{rest}) &\propto \mathfrak{p}(\tau^2|a_{\tau^2})\mathfrak{p}(a_{\tau^2})\\ &\propto (a_{\tau^2})^{-2}\exp\left\{-\frac{1}{a_{\tau^2}}\frac{\tau^{-2}+s_{\tau^2}^{-2}}{2}\right\}\end{aligned}$$

and application of (8) results in:

$$\mathfrak{q}^*(a_{\tau^2}) \propto \exp\left\{E_{\mathfrak{q}}\{\log\mathfrak{p}(a_{\tau^2}|\text{rest})\}\right\} \propto (a_{\tau^2})^{-2}\exp\left\{-\frac{1}{a_{\tau^2}}\frac{E_{\mathfrak{q}}\{\tau^{-2}\}+s_{\tau^2}^{-2}}{2}\right\}.$$

After standard algebraic manipulations, it follows immediately that:

$$\mathfrak{q}^*(a_{\tau^2}) \text{ is an Inverse-}\chi^2\left(\xi_{\mathfrak{q}(a_{\tau^2})},\lambda_{\mathfrak{q}(a_{\tau^2})}\right) \text{ density function}$$

with

$$\xi_{\mathfrak{q}(a_{\tau^2})} \longleftarrow 2 \quad \text{and} \quad \lambda_{\mathfrak{q}(a_{\tau^2})} \longleftarrow \mu_{\mathfrak{q}(1/\tau^2)} + s_{\tau^2}^{-2}.$$

S.3.6 Derivation of $\mathfrak{q}^*(\zeta_h)$ and Associated Parameter Updates

Given the model formulation (17), the full conditional density function for ζ_h, $1 \le h \le H$, is:

$$\mathfrak{p}(\zeta_h|\text{rest}) \propto \mathfrak{p}(\beta_h|\zeta_h,\tau)\,\mathfrak{p}(\zeta_h|a_{\zeta_h}),$$

where $\mathfrak{p}(\beta_h|\zeta_h)$ is the density function of a $\text{N}(0,\tau^2/\zeta_h)$ distribution, while $\mathfrak{p}(\zeta_h|a_{\zeta_h})$ depends on the hierarchical specification given in Table 1 for each of the considered global-local priors.

S.3.6.1 Laplace Prior Specification

If $\beta_h|\tau \sim \text{Laplace}(0,\tau)$ then:

$$\mathfrak{p}(\zeta_h|\text{rest}) \propto (\zeta_h)^{-3/2}\exp\left\{-\zeta_h\frac{\beta_h^2}{2\tau^2}-\frac{1}{2\zeta_h}\right\}$$

and application of (8) results in:

$$\mathfrak{q}^*(\zeta_h) \propto \exp\left\{E_{\mathfrak{q}}\{\log\mathfrak{p}(\zeta_h|\text{rest})\}\right\} \propto (\zeta_h)^{-3/2}\exp\left\{-\zeta_h\frac{E_{\mathfrak{q}}\{\beta_h^2\}E_{\mathfrak{q}}\{\tau^{-2}\}}{2}-\frac{1}{2\zeta_h}\right\}.$$

After standard algebraic manipulations, it follows immediately that:

$$\mathfrak{q}^*(\zeta_h) \text{ is an Inverse-Gaussian}(\mu_{\mathfrak{q}(\zeta_h)},1) \text{ density function}$$

with

$$\mu_{\mathfrak{q}(\zeta_h)} \longleftarrow \sqrt{1\Big/\left(\mu_{\mathfrak{q}(1/\tau^2)}\mu_{\mathfrak{q}(\beta_h^2)}\right)}.$$

S.3.6.2 Horseshoe Prior Specification

If $\beta_h^|\tau \sim \text{Horseshoe}(0,\tau)$ then:

$$\mathfrak{p}(\zeta_h|\text{rest}) \propto \exp\left\{-\zeta_h\left(\frac{\beta_h^2}{2\tau^2}+a_{\zeta_h}\right)\right\}$$

and application of (8) results in:

$$\mathfrak{q}^*(\zeta_h) \propto \exp\{E_\mathfrak{q}\{\log \mathfrak{p}(\zeta_h|\text{rest})\}\} \propto \exp\left\{-\zeta_h\left(\frac{E_\mathfrak{q}\{\beta_h^2\}E_\mathfrak{q}\{\tau^{-2}\}}{2} + E_\mathfrak{q}\{a_{\zeta_h}\}\right)\right\}.$$

After standard algebraic manipulations, it follows immediately that:

$$\mathfrak{q}^*(\zeta_h) \text{ is a Gamma}(1, \lambda_{\mathfrak{q}(\zeta_h)}) \text{ density function}$$

with

$$\lambda_{\mathfrak{q}(\zeta_h)} \longleftarrow \frac{\mu_{\mathfrak{q}(1/\tau^2)}\mu_{\mathfrak{q}(\beta_h^2)}}{2} + \mu_{\mathfrak{q}(a_{\zeta_h})}.$$

S.3.6.3 Normal-Exponential-Gamma Prior Specification

If $\beta_h|\tau \sim \text{NEG}(0, \tau, \lambda)$ then:

$$\mathfrak{p}(\zeta_h|\text{rest}) \propto (\zeta_h)^{-3/2} \exp\left\{-\zeta_h \frac{\beta_h^2}{2\tau^2} - \frac{a_{\zeta_h}}{\zeta_h}\right\}$$

and application of (8) results in:

$$\mathfrak{q}^*(\zeta_h) \propto \exp\{E_\mathfrak{q}\{\log \mathfrak{p}(\zeta_h|\text{rest})\}\} \propto (\zeta_h)^{-3/2} \exp\left\{-\zeta_h \frac{E_\mathfrak{q}\{\beta_h^2\}E_\mathfrak{q}\{\tau^{-2}\}}{2} - \frac{E_\mathfrak{q}\{a_{\zeta_h}\}}{\zeta_h}\right\}.$$

After standard algebraic manipulations, it follows immediately that:

$$\mathfrak{q}^*(\zeta_h) \text{ is an Inverse-Gaussian}(\mu_{\mathfrak{q}(\zeta_h)}, \lambda_{\mathfrak{q}(\zeta_h)}) \text{ density function}$$

with

$$\mu_{\mathfrak{q}(\zeta_h)} \longleftarrow \sqrt{2\mu_{\mathfrak{q}(a_{\zeta_h})} \Big/ \left(\mu_{\mathfrak{q}(1/\tau^2)}\mu_{\mathfrak{q}(\beta_h^2)}\right)} \quad \text{and} \quad \lambda_{\mathfrak{q}(\zeta_h)} \longleftarrow 2\mu_{\mathfrak{q}(a_{\zeta_h})}.$$

S.3.7 Derivation of $\mathfrak{q}^*(a_{\zeta_h})$ and Associated Parameter Updates

Given the model formulation (17), the full conditional density function for ζ_h, $1 \le h \le H$, is:

$$\mathfrak{p}(a_{\zeta_h}|\text{rest}) \propto \mathfrak{p}(\zeta_h|a_{\zeta_h})\, \mathfrak{p}(a_{\zeta_h}),$$

where both $\mathfrak{p}(\zeta_h|a_{\zeta_h})$ and $\mathfrak{p}(a_{\zeta_h})$ depend on the hierarchical specification given in Table 1 for each of the considered global-local priors.

S.3.7.1 Laplace Prior Specification

According to Table 1, for the Laplace prior it is not necessary to introduce the a_{ζ_h} auxiliary variables. Hence $\mathfrak{q}(a_{\zeta_h})$ is not present in (19) and does not need to be determined.

S.3.7.2 Horseshoe Prior Specification

If $\beta_h|\tau \sim \text{Horseshoe}(0, \tau)$ then:

$$\mathfrak{p}(a_{\zeta_h}|\text{rest}) \propto \exp\{-(\zeta_h + 1)a_{\zeta_h}\}$$

and application of (8) results in:

$$\mathfrak{q}^*(a_{\zeta_h}) \propto \exp\{E_\mathfrak{q}\{\log \mathfrak{p}(a_{\zeta_h}|\text{rest})\}\} \propto \exp\{-a_{\zeta_h}(E_\mathfrak{q}\{\zeta_h\} + 1)\}.$$

After standard algebraic manipulations, it follows immediately that:

$$\mathfrak{q}^*(a_{\zeta_h}) \text{ is a Gamma}(1, \lambda_{\mathfrak{q}(a_{\zeta_h})}) \text{ density function}$$

with

$$\lambda_{\mathfrak{q}(\zeta_h)} \longleftarrow \mu_{\mathfrak{q}(\zeta_h)} + 1.$$

S.3.7.3 Normal-Exponential-Gamma Prior Specification

If $\beta_h|\tau \sim \text{NEG}(0, \tau, \lambda)$ then:

$$\mathfrak{p}(a_{\zeta_h}|\text{rest}) \propto (a_{\zeta_h})^{(\lambda+1)-1} \exp\{a_{\zeta_h}(\zeta_h^{-1}+1)\}$$

and application of (8) results in:

$$\mathfrak{q}^*(a_{\zeta_h}) \propto \exp\left\{E_{\mathfrak{q}}\{\log \mathfrak{p}(a_{\zeta_h}|\text{rest})\}\right\} \propto (a_{\zeta_h})^{(\lambda+1)-1} \exp\{a_{\zeta_h}(E_{\mathfrak{q}}\{\zeta_h^{-1}\}+1)\}.$$

After standard algebraic manipulations, it follows immediately that:

$$\mathfrak{q}^*(\zeta_h) \text{ is a Gamma}(\lambda+1, \lambda_{\mathfrak{q}(a_{\zeta_h})}) \text{ density function}$$

with

$$\lambda_{\mathfrak{q}(\zeta_h)} \longleftarrow \mu_{\mathfrak{q}(1/\zeta_h)} + 1.$$

References

Andrews, D. F. & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society B*, **36**, 99–102.

Atay-Kayis, A. & Massam, H. (2005). A Monte Carlo method for computing marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika*, **92**, 317–335.

West, M. (1987). On scale mixtures of normal distributions. *Biometrika*, **74**, 646–648.

Whittaker, E. T. & Watson, G. N. (1990). *A Course in Modern Analysis*. Cambridge, U.K.: Cambridge University Press.