Drawing samples from inverse Wishart distributions conditioning on the 1st block diagonal sub-matrix; with an application to variable selection on a GLMM model with nested random effects*

Georgios Karagiannis† and Christophe Andrieu

January 30, 2016

Abstract

We present an algorithm for sampling from inverse Wishart (and Wishart) distributions conditioning on the 1st block diagonal sub-matrix. The highlight of the proposed algorithm is that only the Cholesky decomposition derivatives of the involved symmetric positive matrices is used. This leads to more efficient algorithms in terms of computing. The numerical application under consideration involves a GLMM model with nested random effects and focuses on performing variable selection at the random effect design matrix via reversible jump (RJ) MCMC methods. The proposed algorithm is used as a part of the proposal generating mechanism of the RJ algorithm involved. Supplementary material for is available online at `https://github.com/georgios-stats`.

Keywords: Wishart distribution, inverse Wishard distribution, RJ-MCMC, Bayesian Statistics, GLMM, variable selection

---

*This report contains unpublished results produced during the PhD studies of the first author at the University of Bristol.

†Department of Mathematics, Purdue University, West Lafayette, IN 47907-2067, USA; gkaragia@purdue.edu ; georgios.stats@gmail.com

1 Introduction

Sampling from a Wishart or inverse Wishart distribution conditioned on the 1^{st} diagonal submatrix is not straightforward. Here, the need for sampling from an inverse Wishart distribution is motivated by the Bayesian variable selection problem on the design matrix corresponding to the random effects of a GLMM model, when reversible jump moves are used. For example, a possible birth move referring to the random effects leads to the expansion of the random effect design matrix.

A usual strategy in Bayesian problems, in order to simplify the calculations, is to assign conjugate priors to the random parameters when it is possible. In Bayesian GLMM models, a usual choice is to consider normal random effects and assign inverse Wishart distributions to the covariate matrices in order to take advantage of the conjugacy properties of the normal and inverse Wishart distributions. However, in reversible jump moves, this conjugacy can be easily violated as it is not easy for someone to find reversible jump dimension matching proposals leading to conjugate distributions, especially for the Wishart and inverse Wishart distributions.

Recent work has been done in this area by Korsgaard et al. (1999). Their work was motivated by the Bayesian multivariate analysis of models for analyzing jointly normally distributed and binary trails using Gibbs sampler and data augmentation methods. They have designed a method to sample from the conditional inverse Wishart distribution conditioned on the last block diagonal sub-matrix. Although that method can be more general, they set the last diagonal sub-matrix equal to the identity matrix because of the nature of the statistical application they apply it. The method they proposed is based on a matrix block-wise re-parametrisation using well known properties one can find in Bodnar and Okhrin (2008).

Korsgaard et al. (1999) consider re-parametrisation of the matrix Σ in terms of (T_1, T_2, T_3) sub-matrices in the following way:

$$\Sigma = \begin{bmatrix} T_1^{-1} + T_2 T_3^{-1} T_2^{\intercal} & -T_2 T_3^{\intercal} \\ -T_2^{\intercal} & T_3^{\intercal} \end{bmatrix}. \tag{1.1}$$

Thus, if one sets $t_3 = \mathbb{I}_{p_2}$, samples t_1 from $\mathrm{W}_{p_1}(S_{11}, m)$, next sample t_2 from $\mathrm{N}_{p_1 \times p_2}(S_{11}^{-1} S_{12}, t_1^{-1} \otimes S_{22\cdot 1})$, where $S_{22\cdot 1} = S_{22} - S_{21} S_{11}^{-1} S_{21}^{\intercal}$, and finally computes

$$r = \begin{bmatrix} t_1^{-1} + t_2 t_2^{\intercal} & -t_2 \\ -t_2^{\intercal} & I_{p_2} \end{bmatrix}, \tag{1.2}$$

then r is a draw from the conditional inverse Wishart distribution of $\Sigma|\Sigma_{22} = \mathbb{I}_{p_2}$.

However, this method does not produces draws in Cholesky form directly. In practice, because of the need to reduce the computing time when an algorithm is coded in a technical language, it is preferred the Cholesky decomposition derivatives of the the symmetric positive matrices to be used instead of them by themselves. This leads to more stable, fast and efficient algorithms. In MCMC methods, which require many iterations, generating samples from conditional Wishart or inverse Wishart distributions the need for computationally cheap random number generators is very big.

Here, we present a method to draw samples from Wishart and inverse Wishart distributions conditioned on the first block diagonal sub-matrix. The main advantage of those methods against the one developed by Korsgaard et al. (1999) is that here only the Cholesky decomposition derivatives of the involved symmetric positive matrices is used. This leads to more efficient algorithms in terms of computing. First we present the method for Wishart and secondly for inverse Wishart associated draws.

In Section 2, we discuss a general method for sampling from a Wishard distribution conditioning on the 1-st block diagonal sub-matrix. Sampling from an inverse Wishard distribution, conditioning on the 1-st block diagonal sub-matrix, is discussed in Section 3. A numerical application that uses the proposed algorithm is presented in Section 4.

2 Wishart Distribution

Definition 1. Suppose X is an $m \times p$, whose each row is independently drawn from a p-variate normal distribution with zero mean. $X_i^{\intercal} = \left(X_{i,1}^{\intercal}, ..., X_{i,p}^{\intercal}\right) \sim \mathrm{N}\left(0, V\right)$. Then, $W = X^{\intercal}X$ follows a Wishart distribution with parameters: m degrees of freedom, where m is a positive integer, and V scale matrix, where V is a positive definite matrix. We symbolise it as $W \sim \mathrm{W}_p\left(m, V\right)$.

The distribution of W follows a Wishart distribution with m degrees of freedom, V scale parameter and density

$$f_{\mathrm{W}_p(m,V)}(W) = \frac{1}{2^{mp/2}\det(V)^{m/2}\,\Gamma_p(m/2)}\det(W)^{\frac{m-p-1}{2}}\exp\left(-\frac{1}{2}\mathrm{tr}\left(V^{-1}W\right)\right), \tag{2.1}$$

where

$$\Gamma_p\left(\frac{m}{2}\right) = \pi^{\frac{p(p-1)}{4}}\prod_{j=1}^{p}\Gamma\left(\frac{m+1-j}{2}\right). \tag{2.2}$$

In order to sample from a Wishart distribution $\mathrm{W}_p(m,V)$, where $m > p$, Smith and Hocking (1972) proposed a method which makes use of the Bartlet decomposition (Algorithms 1 and 2).

Algorithm 1 Sampling G from $\mathrm{W}_p(m,\mathbb{I}_p)$

REQUIRE m the degrees of freedom and p the dimension of the matrix.

ENSURE G a draw from $\pi(G|m,\mathbb{I}_p,p) = \mathrm{W}_p(G|m,\mathbb{I}_p)$.

1. Sample a lower triangular matrix A s.t.

$$[A]_{i,j} = \begin{cases} \mathrm{N}(0,1) & \text{, if } j < i \\ 0 & \text{, if } j > i \\ \sqrt{\chi^2_{m+1-i}} & \text{, if } i = j \end{cases}, \tag{2.3}$$

   for $i,\ j = 1,...,p$.

2. Compute $G = AA^\intercal$

Algorithm 2 Sampling W from $\mathrm{W}_p(m, V)$

REQUIRE L the Cholesky decomposition of $V = LL^\intercal$, m the degrees of freedom and p the dimension of the matrix

ENSURE W, such that is a draw from $\pi(W|m, V, p) = \mathrm{W}_p(W|m, V)$.

1. Generate G from $\mathrm{W}_p(m, \mathbb{I}_p)$.
2. Compute the Cholesky decomposition A, s.t. $G = AA^\intercal$.
3. Compute the Cholesky decomposition L, s.t. $V = LL^\intercal$.
4. Compute $W = LA(LA)^\intercal$.

One can generate a matrix W from the same Wishart distribution $\mathrm{W}_p(m, \mathbb{I}_p)$, by using an upper triangular random matrix B, as it is presented in Algorithm 3.

Algorithm 3 Sampling W from $\mathrm{W}_p(m, \mathbb{I}_p)$

REQUIRE m the degrees of freedom and p the dimension of the matrix .

ENSURE W a draw from $\pi(W|m, \mathbb{I}_p, p) = \mathrm{W}_p(W|m, \mathbb{I}_p)$.

1. Generate B upper triangular matrix s.t.

$$[B]_{i,j} = \begin{cases} \mathrm{N}(0,1) & \text{, if } j \geqslant i \\ 0 & \text{, if } j \leqslant i \text{ ,} \\ \sqrt{\chi^2_{m-p+i}} & \text{, if } i = j \end{cases} \tag{2.4}$$

for $i,\ j = 1, ..., p$.

2. Compute $W = BB^\intercal$.

Proof. Consider matrices A from (2.3), W from (2.4) and P a $p \times p$ anti-diagonal matrix with ones on its anti-diagonal. Then,
$W = BB^\intercal = (PAP)(PAP)^\intercal = PAA^\intercal P^\intercal \sim \mathrm{W}_p(m, P\mathbb{I}_p P^\intercal) \equiv \mathrm{W}_p(m, \mathbb{I}_p)$. $\square$

Below we mention a basic property of the Wishart distribution.

Proposition 2. *If* $W \sim W_p(m, V)$ *and* C *a* $q \times p$ *matrix with rank* q *then* $CWC^\intercal \sim W_q(m, CVC^\intercal)$.

Proof. It can be proved directly by distribution transformation theorem. $\square$

2.1 Sampling from the conditional Wishart distribution

Let us assume, random matrix W partitioned into (W_{11}, W_{21}, W_{22}) submatrices as

$$W = \begin{bmatrix} W_{11} & W_{21}^\intercal \\ W_{21} & W_{22} \end{bmatrix},$$

where the dimensions of W, W_{11} and W_{22} are $p \times p$, $p_1 \times p_1$ and $p_2 \times p_2$ respectively. Considering that $W \sim \mathrm{W}_p(m, V)$ and $W_{22} \sim \mathrm{W}_{p_2}(m, V_{22})$, we wish to sample from the conditional distribution

$$\mathrm{W}_{p,p_1}(W_{21}, W_{22} | W_{11}, m, V) = \frac{\mathrm{W}_p(W|m, V)}{\mathrm{W}_{p_1}(W_{11}|m, V_{11})},$$

where

$$V = \begin{bmatrix} V_{11} & V_{21}^\intercal \\ V_{21} & V_{22} \end{bmatrix}.$$

Algorithm 4 shows how one can sample from the distribution of $W|W_{21}$. It is worth to mention that only the derivatives of the Cholesky decomposition of the involved symmetric positive definite matrices including the resulting one are used. This speeds up the calculations and leads to efficient algorithms in terms of computing.

Proof. Assume A_{11} and A random matrices similar to (2.3). Then:

$$C = LA = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} A_{11} & 0 \\ N_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} L_{11}A_{11} & 0 \\ L_{21}A_{11} + L_{22}N_{21} & L_{22}A_{22} \end{bmatrix},$$

and $W = CC^\intercal$. $\square$

Algorithm 4 Sampling from a conditional Wishart distribution

REQUIRE L, and C_{11} of $V = LL^{\intercal}$, and $W_{11} = C_{11}C_{11}^{\intercal}$ respectively from the Cholesky decomposition and that L is partitioned as $L = \begin{bmatrix} L_{11} & L_{21}^{\intercal} \\ L_{21} & L_{22} \end{bmatrix}$. Yet, A_{11} such that $C_{11} = L_{11}A_{11}$.

ENSURE C, such that $W = CC^{\intercal}$ is a draw from $\mathrm{W}_{p,p_1}(W_{21}, W_{22}|W_{11}, m, V) = \frac{\mathrm{W}_p(W|m,V)}{\mathrm{W}_{p_1}(W_{11}|m,V_{11})}$.

1. Generate a $p_2 \times p_1$ matrix N_{21} where $[N_{21}]_{i,j} \sim \mathrm{N}(0,1)$ for $i = 1, ..., p_2$ and $j = 1, ..., p_1$.
2. Generate a $p_2 \times p_2$ matrix A_{22} where:

$$[A_{22}]_{i,j} = \begin{cases} \mathrm{N}(0,1) & \text{, if } j < i \\ 0 & \text{, if } j > i \\ \sqrt{\chi^2_{m+1-(i+p_1)}} & \text{, if } i = j \end{cases},$$

   for $i,\ j = 1, ..., p_2$.
3. Compute

$$C = \begin{bmatrix} C_{11} & 0 \\ L_{21}A_{11} + L_{22}N_{21} & L_{22}A_{22} \end{bmatrix}.$$

   Where C is the Cholesky decomposition of $W = CC^{\intercal}$.
4. Compute $W = CC^{\intercal}$.

3 Inverse Wishart distribution

Definition 3. Σ follows an inverse Wishart distribution with m degrees of freedom and S parameter scale matrix, $\Sigma \sim \mathrm{IW}_p(m, S)$ if and only if $W = \Sigma^{-1}$ follows a Wishart distribution with m degrees of freedom and $V = S^{-1}$ parameter scale matrix, $W \sim \mathrm{W}_p(m, V)$.

The distribution of Σ follows an Inverse Wishart distribution with m degrees of freedom, S scale parameter and density function

$$f_{\mathrm{IW}_p(m,S)}(\Sigma) = \frac{\det(S)^{m/2}}{2^{mp/2}\Gamma_p(m/2)} \det(\Sigma)^{-\frac{m+p+1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(S\Sigma^{-1}\right)\right),$$

where

$$\Gamma_p\left(\frac{m}{2}\right) = \pi^{\frac{p(p-1)}{4}} \prod_{j=1}^{p} \Gamma\left(\frac{m+1-j}{2}\right).$$

Proposition 4. *If $\Sigma \sim IW_p(m, S)$ and C a $q \times p$ matrix with rank q then $C\Sigma C^\intercal \sim IW_q(m, CSC^\intercal)$.*

Proof. The proof comes directly from the Definition 3 and the Proposition 2. □

3.1 Sampling from the conditional inverse Wishart distribution

Let us assume, random matrix Σ partitioned into $(\Sigma_{11}, \Sigma_{21}, \Sigma_{22})$ as

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{21}^\intercal \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \tag{3.1}$$

where the dimensions of Σ, Σ_{11} and Σ_{22} are $p \times p$, $p_1 \times p_1$ and $p_2 \times p_2$ respectively. Considering that $\Sigma \sim \mathrm{W}_p(m, S)$ and $\Sigma_{22} \sim \mathrm{IW}_{p_2}(m, S_{22})$, we wish to sample from the conditional distribution

$$\mathrm{IW}_{p,p_1}(\Sigma_{21}, \Sigma_{22} | \Sigma_{11}, m, S) = \frac{\mathrm{IW}_p(\Sigma | m, S)}{\mathrm{IW}_{p_1}(\Sigma_{11} | m - (p - p_1), S_{11})}, \tag{3.2}$$

where

$$S = \begin{bmatrix} S_{11} & S_{21}^\intercal \\ S_{21} & S_{22} \end{bmatrix}. \tag{3.3}$$

Algorithm 5 Sampling from the conditional inverse Wishart distribution

REQUIRE L and C_{11} of $S = LL^\intercal$ and $W_{11} = C_{11}C_{11}^\intercal$ respectively from the Cholesky decomposition and that L is partitioned as $L = \begin{bmatrix} L_{11} & L_{21}^\intercal \\ L_{21} & L_{22} \end{bmatrix}$.

ENSURE C, such that $\Sigma = CC^\intercal$ is a draw from $\mathrm{IW}_{p,p_1}(\Sigma_{21}, \Sigma_{22}|\Sigma_{11}, m, S) = \frac{\mathrm{IW}_p(\Sigma|m,S)}{\mathrm{IW}_{p_1}(\Sigma_{11}|m-(p-p_1),S_{11})}$.

1. Generate the $p_2 \times p_1$ matrix N_{21} where $[N_{21}]_{i,j} \sim \mathrm{N}(0,1)$.
2. Generate the $p_2 \times p_2$ matrix B_{22} where:

$$[B_{22}]_{i,j} = \begin{cases} \mathrm{N}(0,1) & \text{, if } j > i \\ 0 & \text{, if } j < i \\ \sqrt{\chi^2_{m-p_2+i}} & \text{, if } i = j \end{cases},$$

   for $i,\ j = 1, ..., p_2$.
3. Compute

$$C = \begin{bmatrix} C_{11} & 0 \\ \left(L_{21} - L_{22}B_{22}^{-\intercal}N_{21}\right) L_{11}^{-1}C_{11} & L_{22}B_{22}^{-\intercal} \end{bmatrix}.$$

4. Compute $\Sigma = CC^\intercal$.

Algorithm 5 presents the general method we can follow. The main advantage of this method compared to the one developed by Korsgaard et al. (1999) is that here only the Cholesky decomposition derivatives of the involved symmetric positive matrices is used.

Proof. If $\Sigma \sim \mathrm{IW}_p(m, S)$, where $\Sigma = CC^\intercal$. From the Bartlett decomposition there exists B of the form (2.4) and $\Sigma^{-1} = (L^{-\intercal}B)(L^{-\intercal}B)^\intercal$ where $C_2^{-1} = \left(L_2^{-\intercal}B_2\right)^\intercal = B_2^\intercal L_2^{-1}$.

Now we observe that if $L = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix}$ and $B = \begin{bmatrix} B_{11} & N_{21}^{\intercal} \\ 0 & B_{22} \end{bmatrix}$ then

$$
\begin{aligned}
C^{-1} = B^{\intercal} L^{-1} &= \begin{bmatrix} B_{11}^{\intercal} & 0 \\ N_{21} & B_{22}^{\intercal} \end{bmatrix} \begin{bmatrix} L_{11}^{-1} & 0 \\ -L_{22}^{-1} L_{21} L_{1}^{-1} & L_{22}^{-1} \end{bmatrix}; \\
&= \begin{bmatrix} B_{11}^{\intercal} L_{11}^{-1} & 0 \\ N_{21} L_{11}^{-1} - B_{22}^{\intercal} L_{22}^{-1} L_{21} L_{11}^{-1} & B_{22}^{\intercal} Q_{22}^{-1} \end{bmatrix}; \\
&= \begin{bmatrix} C_{11}^{-1} & 0 \\ N_{21} L_{11}^{-1} - B_{22}^{\intercal} L_{22}^{-1} L_{21} L_{11}^{-1} & B_{22}^{\intercal} L_{22}^{-1} \end{bmatrix}.
\end{aligned}
$$

Thus

$$
C = \begin{bmatrix} C_{11} & 0 \\ \left(L_{21} - L_{22} B_{22}^{-\intercal} N_{21} \right) L_{11}^{-1} C_{11} & L_{22} B_{22}^{-\intercal} \end{bmatrix}.
$$

□

4 Numerical application: Variable selection on a GLMM model with nested effects

To illustrate the use of the proposed algorithm, we consider the problem of performing Bayesian inference in a simple Generalised linear mixed model (GLMM) model. We consider the classical data set of (Thall and Vail, 1990) that considers seizure counts for epileptics. Briefly the data set considers two-week seizure counts for 59 epileptics. The number of seizures was recorded for a baseline period of 8 weeks, and then patients were randomly assigned to a treatment group or a control group. Counts were then recorded for four successive two-week periods. The fixed covariates available are: $x^{(1)}$: 1 'constant'; $x^{(2)}$: 'placebo' or 'progabide' treatment; $x^{(3)}$: log-counts for the baseline period centred to have zero mean; $x^{(4)}$: log-ages centred to have zero mean; $x^{(5)}$: 0/1 indicator variable of period 4; and $x^{(6)}$: interaction of $x^{(2)}$ and $x^{(3)}$. The available random covariates are $\{z^{(k)} = x^{(k)}; k = 1 : 6\}$. The

GLMM model and the prior model considered are summarised as:

$$
\begin{aligned}
y_{i,j} &\sim \text{Poisson}(\lambda_{i,j}^{(\mathcal{I},\mathcal{J})});\\
\log(\lambda_{i,j}^{(\mathcal{I},\mathcal{J})}) &= \eta_{i,j}^{(\mathcal{I},\mathcal{J})};\\
\eta_{i,j}^{(\mathcal{I},\mathcal{J})} &= x_{i,j}^{(\mathcal{I}),\intercal}\beta^{(\mathcal{I})} + z_{i,j}^{(\mathcal{J}),\intercal}\gamma_i^{(\mathcal{J})};\\
\beta^{(\mathcal{I})} &\sim \text{N}_{|\mathcal{I}|}(0, \mathbb{I}_{|\mathcal{I}|}\sigma^2);\\
\sigma^2 &\sim \text{IG}(v, u);\\
\gamma_i^{(\mathcal{J})} &\sim \text{N}_{|\mathcal{J}|}(0, \Sigma^{(\mathcal{J})});\\
\Sigma^{(\mathcal{J})} &\sim \text{IW}_{|\mathcal{J}|}(m + |\mathcal{J}|, S^{(\mathcal{J})});\\
\mathcal{I} \sim \delta_{\{(1,...,6)\}}(\mathcal{I});\ \mathcal{J} &\sim \text{Uniform}(\{1, ..., 6\}).
\end{aligned}
$$

for $i = 1, ..., N$, $j = 1, ..., M_i$, where $y = (y_{i,j}; i = 1 : N, j = 1 : M_i)$, $n = (n_{i,j}; i = 1 : N, j = 1 : M_i)$, and $\lambda_{i,j} = (\lambda_{i,j}; i = 1 : N, j = 1 : M_i)$. Here, $v = 10^{-3}$, $u = 10^{-3}$, $m = 1$, $S^{(\mathcal{J})} = \mathbb{I}_{|\mathcal{J}|}$ $N = 59$, and $M_i = 4$. The upper indexes $\mathcal{I}$ and $\mathcal{J}$ indicate which covariates are included in the fixed and random effect design matrices, correspondingly. For the purpose of the article, we focus on the variable selection on the random effect design matrix only, and therefore we assign $\mathcal{I} \sim \delta_{\{(1,...,6)\}}(\mathcal{I})$ and $\mathcal{J} \sim \text{Uniform}(\{1, ..., 6\})$. However for the reader interested in variable selection on the fixed effect design matrix we cite (Dellaportas et al., 2008, 2002; Dellaportas and Forster, 1999). Regarding the variable selection on random effect design matrix in GLMM, we cite (Forster et al., 2012; Müller et al., 2013).

The joint posterior distribution admits intractable density, and cannot be sampled directly. For this reason, we can resort to MCMC methods to perform inference. We consider a systematic block-wise MCMC sweep that consists of blocks that simulates from the full conditionals of: $\beta^{(\mathcal{I})}|...$, and $\{\gamma_i^{(\mathcal{J})}|...; i = 1 : N\}$ via Metropolis random walk (Metropolis et al., 1953; Hastings, 1970); $\sigma^2|...$, and $\Sigma^{(\mathcal{J})}|...$ via direct sampling (Ripley, 2009); and $\mathcal{J}, \beta^{(\mathcal{I})}, \sigma^2, \gamma_{1:N}^{(\mathcal{J})}, \Sigma^{(\mathcal{J})}|...$ via reversible jump algorithm (Green, 1995). The first 4 updates are standard material and their implementation is straightforward, so they are not described further (Gamerman, 1997; Dellaportas et al., 2002, 2008). Regarding the reversible jump algorithm, a pair of birth and death moves is considered, where the death move removes a set of covariates randomly chosen, and the birth move adds a set of covariates randomly chosen. The birth move, $(\mathcal{I}_1, \mathcal{J}_1) \to (\mathcal{I}_2, \mathcal{J}_2)$, involves adding a set of covariates, denoted as $\mathcal{J}_0$, randomly chosen from those currently not included in the model of the current state, and setting $\mathcal{J}_2 = (\mathcal{J}_1, \mathcal{J}_0)$, $\mathcal{I}_2 = \mathcal{I}_1$. The dimension matching involves

1. Sample $\tilde{\Sigma}_{12}, \tilde{\Sigma}_{22}|\Sigma^{(\mathcal{J}_1)} \sim \mathrm{IW}_{|\mathcal{J}_2|,|\mathcal{J}_1|}\left(m + |\mathcal{J}_2|, S^{(\mathcal{J}_2)}\right)$, and $\tilde{\gamma}_{i,12}|\gamma_i^{(\mathcal{J}_1)} \sim \mathrm{N}_{|\mathcal{J}_0|}(\tilde{\Sigma}_{12}\Sigma^{(\mathcal{J}_1),-1}\gamma_i^{(\mathcal{J}_1)}, \tilde{\Sigma}_{22} - \tilde{\Sigma}_{12}\Sigma^{(\mathcal{J}_1),-1}\tilde{\Sigma}_{12}^{\intercal})$, for $i = 1, ..., N$.
2. Set $\beta^{(\mathcal{I}_2)} = \beta^{(\mathcal{I}_1)}$, $\sigma_2^2 = \sigma_1^2$, $\{\gamma_i^{(\mathcal{J}_2)} = (\gamma_i^{(\mathcal{J}_1)}, \tilde{\gamma}_{i,12}); i = 1 : N\}$, and $\Sigma^{(\mathcal{J}_2)} = \begin{bmatrix} \Sigma^{(\mathcal{J}_1)} & \tilde{\Sigma}_{12}^{\intercal} \\ \tilde{\Sigma}_{12} & \tilde{\Sigma}_{22} \end{bmatrix}$.

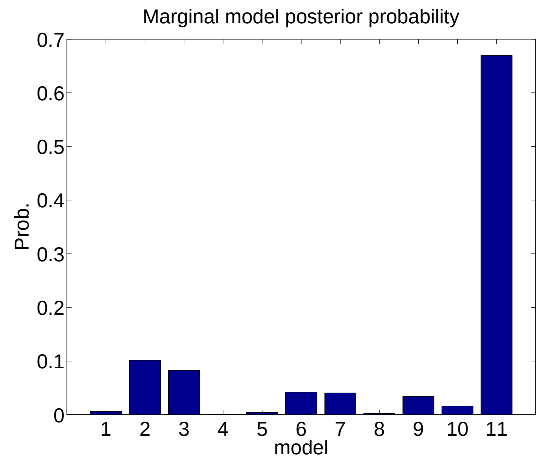
The birth move is accepted with probability $a_{(\mathcal{I}_1,\mathcal{J}_1)\to(\mathcal{I}_2,\mathcal{J}_2)} = \min(1, R_{(\mathcal{I}_1,\mathcal{J}_1)\to(\mathcal{I}_2,\mathcal{J}_2)})$, where

$$R_{(\mathcal{I}_1,\mathcal{J}_1)\to(\mathcal{I}_2,\mathcal{J}_2)} = \prod_{i=1}^{N}\prod_{j=1}^{M_i} \frac{\lambda_{i,j}^{(\mathcal{I}_2,\mathcal{J}_2),y_{i,j}}}{\lambda_{i,j}^{(\mathcal{I}_1,\mathcal{J}_1),y_{i,j}}} \exp(-\lambda_{i,j}^{(\mathcal{I}_2,\mathcal{J}_2)} + \lambda_{i,j}^{(\mathcal{I}_1,\mathcal{J}_1)}).$$

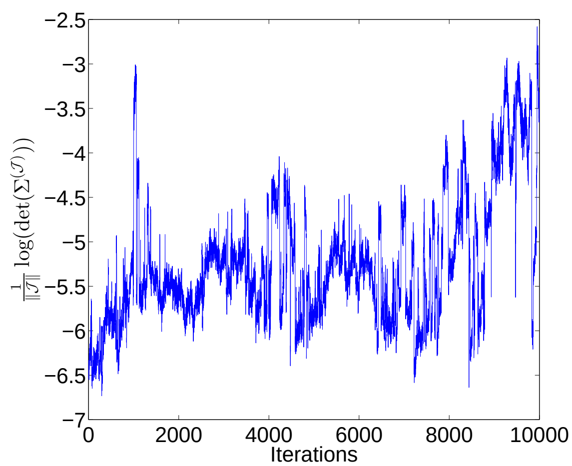
The death move is the reverse analogue $(\mathcal{I}_2, \mathcal{J}_2) \to (\mathcal{I}_1, \mathcal{J}_1)$. They involve removing covariates $\mathcal{J}_0$ randomly chosen from those currently included in the model of the current state, and setting $\mathcal{J}_1 = \mathcal{J}_2 \oplus \mathcal{J}_0$, $\mathcal{I}_2 = \mathcal{I}_1$. The move is accepted with probability $a_{(\mathcal{I}_2,\mathcal{J}_2)\to(\mathcal{I}_1,\mathcal{J}_1)} = \min(1, 1/R_{(\mathcal{I}_1,\mathcal{J}_1)\to(\mathcal{I}_2,\mathcal{J}_2)})$. Note that the RJ acceptance ratio simplifies to the likelihood ratio.

We run the aforementioned MCMC sampler for 11000 iterations where the first 1000 were discarded as a burn in. The algorithm started from the state with no random effects i.e. $\mathcal{J} = \emptyset$. The marginal model posterior probability distribution is presented in the Figure 4.1a. In Figure 4.1b, we present the trace plot of $\theta = \frac{1}{|\mathcal{J}|}\log(\det(\Sigma^{(\mathcal{J})}))$ generated. For θ, the estimate is -5.2494 with standard error $2.7246e-04$. We observe that the sampler 'mixes' adequately enough. The estimated expected acceptance probability was 0.035.

Remark 5. Performing variable selection on the design matrix of GLMM models with nested random effects through RJ methods can be computationally challenging because the competing models may have large differences in the dimensionality. Therefore, the 'jumps' between the models may be large in terms of dimension difference. For instance, in the RJ moves scheme considered here, the pairs of competing models considered differ $N \times |\mathcal{J}_0| + 0.5|\mathcal{J}_0|(|\mathcal{J}_0| + 1) + |\mathcal{J}_0||\mathcal{J}_1|$ dimensions. From the above numerical application, we observed that the proposed algorithm for sampling from inverse Wishart distributions conditional on the 1st diagonal allows an



(a)



(b)

Figure 4.1: [Section 4] Sub-Figure (a): Marginal model posterior probabilities, where the models have labels: 1 for $\mathcal{J} = \{1, 2, 3\}$; 2 for $\mathcal{J} = \{1, 3, 5\}$; 3 for $\mathcal{J} = \{1, 2, 3, 5\}$; 4 for $\mathcal{J} = \{1, 2, 3, 6\}$; 5 for $\mathcal{J} = \{1, 3, 4, 5\}$; 6 for $\mathcal{J} = \{1, 3, 5, 6\}$; 7 for $\mathcal{J} = \{1, 2, 3, 4, 5\}$; 8 for $\mathcal{J} = \{1, 2, 3, 4, 6\}$; 9 for $\mathcal{J} = \{1, 2, 3, 5, 6\}$; 10 for $\mathcal{J} = \{1, 3, 4, 5, 6\}$; 11 for $\mathcal{J} = \{1, 2, 3, 4, 5, 6\}$. The red line denotes the value 0.5. Sub-Figure (b): The trace plot of the log determinant of the covariance matrix $\Sigma^{(\mathcal{J})}$ scaled by the number of elements of its diagonal.

acceptable implementation of RJ moves in variable selection problems for GLMM models with nested random effects.

5 Conclusions

We presented an algorithmic procedure to generate random values from the inverse Wishart distribution conditioning on the 1st block diagonal sub-matrix. The highlight of the proposed algorithm is that only the Cholesky decomposition derivatives of the involved symmetric positive matrices are used. This leads to more efficient algorithms in terms of computing.

The numerical application considered involves a GLMM model with nested random effects and focusses on performing variable selection at the random effect design matrix via RJMCMC methods. The proposed algorithm was used as a part of the proposal generating mechanism of the RJ algorithm involved, and we observed that the resulted RJ performed adequately.

Supplementary material

Supplementary material for the article is available online. The MATLAB functions for the generation of random values and the computation of the densities of Wishart and inverse Wishard distributions conditioning on the 1st block diagonal sub-matrix are available on GitHub at
`https://github.com/georgios-stats`.

References

Bodnar, T. and Y. Okhrin (2008). Properties of the singular, inverse and generalized inverse partitioned wishart distributions. *Journal of Multivariate Analysis 99*(10), 2389–2405.

Dellaportas, P., J. Forster, and I. Ntzoufras (2008). Specification of prior distributions under model uncertainty.

Dellaportas, P. and J. J. Forster (1999). Markov chain Monte Carlo model determination for hierarchical and graphical log-linear models. *Biometrika 86*(3), 615–633.

Dellaportas, P., J. J. Forster, and I. Ntzoufras (2002). On Bayesian model and variable selection using MCMC. *Statistics and Computing 12*(1), 27–36.

Forster, J. J., R. C. Gill, and A. M. Overstall (2012). Reversible jump methods for generalised linear models and generalised linear mixed models. *Statistics and Computing 22*(1), 107–120.

Gamerman, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing 7*(1), 57–68.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*(4), 711–732.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika 57*(1), 97–109.

Korsgaard, I. R., A. H. Andersen, and D. Sorensen (1999). A useful reparameterisation to obtain samples from conditional inverse wishart distributions. *Genetics Selection Evolution 31*(2), 177–181.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics 21*(6), 1087–1092.

Müller, S., J. L. Scealy, and A. H. Welsh (2013). Model selection in linear mixed models. *Statistical Science 28*(2), 135–167.

Ripley, B. D. (2009). *Stochastic simulation*, Volume 316. John Wiley & Sons.

Smith, W. B. and R. R. Hocking (1972). Algorithm as 53: Wishart variate generator. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 21*(3), 341–345.

Thall, P. F. and S. C. Vail (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics*, 657–671.

A MATLAB code for generating random numbers from the Wishart distribution conditioning on the 1st block diagonal sub-matrix

```
function [Sig, cholSig] = wishart_cond_rng(m, p_1, p_2, Sig_11, S)

    % Inputs :
    %     p_1:       The 1st dimention of the 1st block diagonal sub-matrix.
    %     p_2:       p_2 = size(X,1)-p_1 .
    %     m:         Degrees of freedom of the big matrix
    %     Sig_11:    The p_1Xp_1 1st block diagonal sub-matrix we condition.
    %     S:         The complete (p_1+p_2)X(p_1+p_2) semi-positive scale matrix.
    % Outputs :
    %     Sig:       The complete random (p_1+p_2)X(p_1+p_2) semi-positive matrix.
    %     cholSig:   chol(Sig,'lower')

    C_11 = chol(Sig_11,'lower') ;
    Sig = chol(S,'lower') ;
    L_11 = Sig(1:p_1,1:p_1) ;
    L_21 = Sig((p_1+1):(p_1+p_2),1:p_1) ;
    L_22 = Sig((p_1+1):(p_1+p_2),(p_1+1):(p_1+p_2)) ;
    A_11 = L_11\C_11 ;
    N_21 = randn(p_2,p_1) ;
    A_22 = zeros(p_2,p_2) ;
    for i = 1:p_2
        for j = 1:p_2
            if (j<i)
                A_22(i,j) = randn() ;
            elseif (j==i)
                A_22(i,j) = sqrt( 0.5*randg(0.5*(m-p_2+i)) ) ;
            end
        end
    end
    cholSig = [C_11 , zeros(p_1,p_2) ;
    L_21*A_11+L_22*N_21 , L_22*A_22] ;
    Sig = cholSig*(cholSig') ;

end
```

B MATLAB code for generating random numbers from the inverse Wishart distribution conditioning on the 1st block diagonal sub-matrix

```
function [Sig, cholSig] = invwishart_cond_rng(m, p_1, p_2, Sig_11, S)

    % Inputs :
    %     p_1:      The 1st dimention of the 1st block diagonal sub-matrix.
    %     p_2:      p_2 = size(X,1)-p_1.
    %     m:        Degrees of freedom of the big matrix.
    %     Sig_11:   The p_1Xp_1 1st block diagonal sub-matrix we condition.
    %     S:        The complete (p_1+p_2)X(p_1+p_2) semi-positive scale matrix.
    %
    % Outputs :
    %     Sig:      The complete random (p_1+p_2)X(p_1+p_2)
    %                  semi-positive matrix.
    %     cholSig:  chol(Sig,'lower')

    C_11 = chol(Sig_11,'lower') ;
    Sig = chol(S,'lower') ;
    L_11 = Sig(1:p_1,1:p_1) ;
    L_21 = Sig((p_1+1):(p_1+p_2),1:p_1) ;
    L_22 = Sig((p_1+1):(p_1+p_2),(p_1+1):(p_1+p_2)) ;
    N_21 = randn(p_2,p_1) ;
    B_22 = zeros(p_2,p_2) ;
    for i = 1:p_2
        for j = 1:p_2
            if (j>i)
                B_22(i,j) = randn() ;
            elseif (j==i)
                B_22(i,j) = sqrt( 0.5*randg(0.5*(m-p_2+i)) ) ;
            end
        end
    end
    B_22 = L_22/(B_22') ;
    cholSig = [C_11 , zeros(p_1,p_2) ;
    (L_21-B_22*N_21)*(L_11\C_11) , B_22] ;
    Sig = cholSig*(cholSig') ;

end
```