Baysian Random Covariance Model

Definition & Posterior Sampling Algorithm

Quinton Neville

10/25/2022

Table of contents

1. Model Specification

1.1 Likelihood

For the k^{th} subject and the i^{th} replicate or fMRI volume, let

$$y_{ki} \sim \mathcal{N}(X_{ki}\beta + Z_{ki}\boldsymbol{u},\, \Omega_k^{-1})$$

1.2 RCM Prior

$$
\begin{aligned}
\pi(\beta) &\propto 1 \\
\boldsymbol{u} &\sim \mathcal{N}(0, P) \\
\Omega_k &\sim GWish(G_k, \tau_k, \Omega_0/\tau_k),
\end{aligned}
$$

where Ω_0 is the group-level or overall precision matrix for the entire sample.

1.3 Hyperpriors

$$
\begin{aligned}
\pi(G_k) &\propto \exp\{-\lambda_1 |G_k|\} \\
\tau_k &\sim Exp(\lambda_2) \\
\pi(\Omega_0) &\propto \exp\{-\lambda_3 ||\Omega_0||_1\}.
\end{aligned}
$$

1.4 Hyperpriors for Regularization

$$
\{\lambda_j\}_{1:3} \sim \Gamma(a_j, b_j).
$$

1.5 GWishart Distribution

Assuming that $\mathbb{E}[y_{ki}] = 0$, it follows then for $Y_{ki} \sim \mathcal{N}(0, \Omega_k^{-1})$ that the posterior depends only on the distribution of Ω_k through it's associated hyperpriors and parameters. Let Θ denote the entire parameter space for the model above, without fixed or random mean effects, and $(\cdot)$ denote all other parameters in conditional forms below. Further, for the k^{th} subject and the i^{th} replicate or fMRI volume, let $|Y_{ki}| = p$ such that Ω_k is a positive-definite $p \times p$ precision matrix. Now, each Ω_k is constrained by the undirected graph $G_k = (V_k, E_k)$, imposing strict conditional dependence such that network connections (j, l) are conditionally independent if and only if the corresponding entry $\omega_{jl} = 0$; equivalent to $(j, l) \notin E_k$ in G_k.

Next, recalling the definition of our likelihood and hyperpriors above, it follows directly from Wang et. al., 2012; equation §1.1, that the density $\Omega_k \,|G_k, \cdot \sim GWish(G_k, \tau_k, \Omega/\tau_k)$ is given by

$$
\pi(\Omega_k \,|\, G_k, \cdot) = I_G(\tau_k, \Omega_o/\tau_k)^{-1} \;\; \cdot |\Omega_k|^{\frac{\tau_k - 2}{2}} \; \exp\left\{\frac{1}{2} tr(\Omega_k \cdot \Omega_0/\tau_k)\right\} \; I\{\Omega \in P_G^+\}
$$

with normalizing constant

$$
I_G(\tau_k, \Omega_o/\tau_k) = \int_{\Omega \in P_{G_k^+}} |\Omega_k|^{(\tau_k - 2)/2} \; \exp\left\{ -\frac{1}{2}\, \mathrm{tr}(\Omega_k \cdot \Omega_0/\tau_k)\right\} \, \mathrm{d}\Omega,
$$

where $P_{G_k^+}$ denotes the set of all $p \times p$ symmetric, positive definite matrices with off diagonal elements $\omega_{jl} = 0 \iff (j,l) \notin E_k$.

2. Full Conditional Distribution & Posterior Sampling Outline

2.1 Shrinkage/Penalty Parameters λ_j (Direct Sampling)

2.1.1 L1 Penalty on $|G_k|$ (λ_1)

$$\begin{aligned}
\pi(\lambda_1 \,|\, Y_{ki}, \cdot) &\propto \pi(G_k \,|\, \lambda_1)\pi(\lambda_1) \\
&\propto \Gamma(a_1, |G_k| + b_1) \\
&\implies \\
\pi(\lambda_1 \,|\, \boldsymbol{Y}_i, \cdot) &\propto \Gamma\Big(a_1 + K, \sum_k^K |G_k| + b_1\Big).
\end{aligned}$$

2.1.2 Rate Parameter of Regularization $\tau_k \sim Exp(\lambda_2)$

$$\begin{aligned}
\pi(\lambda_2 \,|\, Y_{ki}, \cdot) &\propto \pi(\tau_k \,|\, \lambda_2)\pi(\lambda_2) \\
&\propto \Gamma(a_2 + 1, \tau_k + b_2) \\
&\implies \\
\pi(\lambda_2 \,|\, \boldsymbol{Y}_i, \cdot) &\propto \Gamma\Big(a_2 + K + 1, \sum_k^K \tau_k + b_2\Big);
\end{aligned}$$

2.1.3 L1 Matrix Norm Penalty on $||\Omega_0||_1$ (λ_1)

$$\begin{aligned}
\pi(\lambda_3 \,|\, \cdot) &\propto \pi(\Omega_0 \,|\, \lambda_3)\pi(\lambda_3) \\
&\propto \Gamma\Big(a_3 + |\Omega_0|, ||\Omega_0||_1 + b_3\Big);
\end{aligned}$$

2.2 Regularization Penalties τ_k (MH with log-Normal proposal)

The full conditional distribution for the regularization parameter(s) τ_k is dependent on Ω_k, Ω_0 proportional to the prior dependent on λ_2 –

$$\begin{aligned}
\pi(\tau_k \,|\, Y_{ki}, \cdot) &\propto \pi(\Omega_k \,| G_k, \tau_k, \Omega_0)\pi(\tau_k) \\
&\propto -\lambda_2 W_{\Omega_k}(\tau_k, \Omega_0/\tau_k)\, \exp\{-\lambda_2\tau_k\};
\end{aligned}$$

However, this distribution does not have a closed form so instead we propose an MH algorithm, with a lognormal proposition, to sample from the full conditional posterior –

MH Algorithm for τ_k Update

1. Let $q(\tau_k^{(i+1)} \,|\, \tau_k^{(i)}) =$ Lognormal$(0, \tau_{\text{step}}^2)$ and draw $\tau_k^{(i+1)} \sim \exp\{N(0, \tau_{\text{step}}^2)\}$.
2. Accept $\tau_k^{(i+1)}$ with probability α, given by

$$\alpha = \min\left\{1, \frac{\pi(\tau_k^{(i+1)} \,|\, Y_{ki}, \cdot)\, q(\tau_k^{(i+1)} \,|\, \tau_k^{(i)})}{\pi(\tau_k^{(i)} \,|\, Y_{ki}, \cdot)\, q(\tau_k^{(i)} \,|\, \tau_k^{(i+1)})}\right\}$$

In addition, during burn-in adaptively adjust τ_{step} based on the acceptance rate of the last pre-specified amount of samples (i.e. increase or decrease the variance of the proposal distribution q). *Note* - This requires some tuning on my part, seems to get stuck at the boundary (low, near zero) –> which blows up Ω_k/τ_k.

2.3 Ω_k & G_k (MH within Graph Sampler + Direct Posterior)

Recalling the normality of Y_{ki}, this yields the full conditional posterior distribution outlined in Wang et. al., 2012; equation §1.2, given by

$$\pi(\Omega_k \,|\, Y_{ki}, G_k, \cdot) = I_G(\tau_k + p,\, \Omega_0/\tau_k + Y_{ki}Y_{ki}^T) \;\cdot |\Omega_k|^{(\tau_k+p-2)/2} \exp\left\{-\frac{1}{2}\operatorname{tr}\Big[(\Omega_0/\tau_k + Y_{ki}Y_{ki}^T)\Omega_k\Big]\right\} I\{\Omega_k \in P_{G_k^+}\}$$

such that

$$\Omega_k \,\Big|\, Y_{ki}, G_k, (\cdot) \sim Gwish\Big(G_k, \tau_k + p, (\Omega_0/\tau_k + Y_{ki}Y_{ki}^T)\Big).$$

2.4 Ω_0 (MH with Step-function Proposal)

Let $D = \sum_k^K \Omega_k^{(i+1)}/\tau_k$ and $\nu = \sum_k^K \tau_k$. While the full conditional may be given by

$$\begin{aligned}\pi(\Omega_0 \,|\, Y_{ki}, \cdot) &\propto \pi(\Omega_k \,|G_k, \tau_k, \Omega_0)\pi(\Omega_0)\\ &\propto W_{\Omega_k}(\tau_k, \Omega_0/\tau_k)\, \exp\{-\lambda_3 ||\Omega_0||_1\},\end{aligned}$$

this does not in general have closed form and we will not be updating Ω_0 as a whole anyways. Instead, we propose a sampling algorithm which updates each diagonal / off-diagonal element individually. To do so, given $|\Omega| = p \times p$ –

2.4.1 Diagonal Elements

For j in 1 to p:
1. Shuffle Ω such that Ω^* has the rows/cols of the first $j-1$ rows/cols of Ω, not including j, with j as the p^{th} row/col (and same with D^* for D).

$\Omega' = \Omega_{[\{(1:j)/j, \quad j\}, \ \{(1:j)/j, \quad j\}]}$

2. Compute necessary components for proposal.

$$\begin{aligned}
\omega &:= \Omega^*_{[\{1:p\}/p\}, \ \cdot \quad]} \\
\Omega_{\cdot/p} &:= \left\{\Omega^*_{V/p}\right\}^{-1}, \\
c &:= \omega\,\Omega_{\cdot/p}\,\omega^T, \quad \text{(Step)} \\
v &:= \left(\omega\,\Omega_{\cdot/p}\,, -1\right)_{1\times p} \\
d &:= vD^*v^T.
\end{aligned}$$

3. Sample update

$\Omega^{(i+1)}_{0_{j,j}} \sim \mathrm{GIG}(1-\nu/2, d, \lambda_3) \ + c,$

where GIG is the generalized inverse gaussian distribution.

2.4.2 Off-Diagonal Elements

For j in 1 to $(p-1)$:
For l in $(j+1)$ to p: 1. Shuffle Ω such that Ω^* has the rows/cols of the first $j-1$ rows/cols of Ω, not including j, with j as the p^{th} row/col (and same with D^* for D).

$\Omega' = \Omega_{[\{(1:j)/j, \quad j\}, \ \{(1:j)/j, \quad j\}]}$

2. Compute necessary components for proposal.

$$\begin{aligned}
\omega &:= \Omega^*_{[\{(p-1):p\}/p\}, \ 1:(p-2)]} \\
\Omega_{\cdot/p} &:= \left\{\Omega^*_{[\{1:(p-2)\,,\,1:(p-2)\}]}\right\}^{-1}, \\
c &:= \omega\,\Omega_{\cdot/p}\,\omega^T, \quad \text{(Step)} \\
v &:= \left(\omega\,\Omega_{\cdot/p}\,, -I_2\right)_{2\times p} \\
d &:= vD^*v^T.
\end{aligned}$$

3. Sample update

$\Omega_{0_{j,j}}^{(i+1)} \sim \text{GIG}(1-\nu/2, d, \lambda_3) \ + c,$

where GIG is the generalized inverse gaussian distribution.

3. Posterior MCMC-within-Gibbs Sampling Algorithm (old)

Next, to start the MCMC within Gibbs procedure set $i = 0$. While $i <$ max iterations + burn, do –

1. Sample regularization hyperparameters $\{\lambda_j^{(i+1)}\}_{1:3}$ from the full conditional $\Gamma(\cdot)$ posterior(s) given in §2.1.3. (Need to address update of a_j, b_j hyper-hyperparameters).

2. Propose and accept/update $G_k^{(i+1)}$ and $\Omega_k^{(i+1)}$ 'simulataneously' via the PAS algorithm with exchangeability to avoid $GWish$ normalizing constant estimation outlined in §5 of Wang & Li (2012) –

i. For computational purposes, proposals will be made column/rowwise rather than elementwise or via the BIPS algorithm, so select the i^th column - As upper triangular form is necessary for computation & the $Gwish$ distribution, we must first permute the rows and columns such that our i^th selected column comes first - Next compute $b_{\text{post}} = \tau_k^{(i)} + K + 2$ and $D_{\text{post}} = \{\Omega_0^{(i)}\}^{-1}\tau_k + Y_k Y_k^T$

ii. For j on off diagonal:

   a) Compute $w = \log H_{i,j}(b_{\text{post}}, D_{\text{post}}) + \lambda_1^{(i+1)}$ and $p = \ \{\text{expit}(1 + \exp(w))\}^{-1}$
   b) If `runif(1)` $\leq p$ accept the proposed edge
   c) If accepted, sample proposed $\Omega_{k;\ i,j}^{(i+1)}$ from $GWish$ prior given in §1.2
   d) Update $G^{(i+1)}$ via MH step as outlined in §5.2 of Wang & Li (2012)
   e) If accepted, update $\omega_{k;\ i,j}^{(i+1)}$ via posterior outlined in §2.1.1
   f) If not symmetric, force symmetry in upper triangle

3. Update $\tau_k^{(i+1)}$ via usual MH step via posterior given in 2.1.2

   - Propose from lognormal distribution
   - Add in adaptive update given acceptange rate

4. Update $\Omega_0^{(i+1)}$ via the scheme outlined in §3 of Wang & Li (2012)

   - Elementwise
   - Diagonal reduces to posterior sample from inverse gaussian
   - Off-diagonal requires proposal from step function
     - Optimize posterior (GWish) with bounds obtained from §3 Wang & Li scheme
     - Compute MH step similar to Step (3.) via elementwise posterior given in 2.1.1 above