Efficient Implementation of Bayesian Hierarchical Linear Models

September 18, 2019

Bayesian hierarchical linear mixed model

$$p(\boldsymbol{\theta}) \times N(\beta \mid \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \times N(\boldsymbol{\alpha} \mid \mathbf{0}, \mathbf{K}(\boldsymbol{\theta})) \times N(\mathbf{y} \mid \mathbf{X}\beta + \mathbf{Z}(\boldsymbol{\theta})\boldsymbol{\alpha}, \mathbf{D}(\boldsymbol{\theta}))$$

- $\mathbf{y}$ is an $n \times 1$ vector of possibly irregularly located observations,
- $\mathbf{X}$ is a known $n \times p$ matrix of regressors ($p < n$),
- $\mathbf{K}(\boldsymbol{\theta})$ and $\mathbf{D}(\boldsymbol{\theta})$ are families of $r \times r$ and $n \times n$ covariance matrices, respectively,
- $\mathbf{Z}(\boldsymbol{\theta})$ is $n \times r$ with $r \leq n$, all indexed by a set of unknown process parameters $\boldsymbol{\theta}$.
- $\boldsymbol{\alpha}$ is the $r \times 1$ random vector and β is the $p \times 1$ slope vector.

Bayesian hierarchical linear mixed model

$p(\boldsymbol{\theta}) \times N(\beta \,|\, \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \times N(\boldsymbol{\alpha} \,|\, \mathbf{0}, \mathbf{K}(\boldsymbol{\theta})) \times N(\mathbf{y} \,|\, \mathbf{X}\beta + \mathbf{Z}(\boldsymbol{\theta})\boldsymbol{\alpha}, \mathbf{D}(\boldsymbol{\theta}))$

- $\mathbf{y}$ is an $n \times 1$ vector of possibly irregularly located observations,
- $\mathbf{X}$ is a known $n \times p$ matrix of regressors ($p < n$),
- $\mathbf{K}(\boldsymbol{\theta})$ and $\mathbf{D}(\boldsymbol{\theta})$ are families of $r \times r$ and $n \times n$ covariance matrices, respectively,
- $\mathbf{Z}(\boldsymbol{\theta})$ is $n \times r$ with $r \leq n$, all indexed by a set of unknown process parameters $\boldsymbol{\theta}$.
- $\boldsymbol{\alpha}$ is the $r \times 1$ random vector and β is the $p \times 1$ slope vector.

Space-varying intercept model is a special case where $\mathbf{D}(\boldsymbol{\theta}) = \tau^2 I_n$, $\boldsymbol{\alpha} = (w(\mathbf{s}_1), w(\mathbf{s}_2), \ldots, w(\mathbf{s}_n))^\top$, $\mathbf{Z}(\boldsymbol{\theta}) = I_n$, and the $n \times n$ $\mathbf{K}(\boldsymbol{\theta}) = \sigma^2 \mathbf{R}(\phi)$.

For faster convergence, we integrate out β and α from the model and first sample from

$$p(\boldsymbol{\theta} \,|\, \mathbf{y}) \propto p(\boldsymbol{\theta}) \times N(\mathbf{y} \,|\, \mathbf{X}\boldsymbol{\mu}_{\beta}, \boldsymbol{\Sigma}_{y \,|\, \theta}),$$

where $\boldsymbol{\Sigma}_{y \,|\, \theta} = \mathbf{X}\boldsymbol{\Sigma}_{\beta}\mathbf{X}^{\top} + \mathbf{Z}(\theta)\mathbf{K}(\theta)\mathbf{Z}(\theta)^{\top} + \mathbf{D}(\theta)$.

This involves evaluating

$$\log p(\boldsymbol{\theta} \,|\, \mathbf{y}) = \mathrm{const} + \log p(\boldsymbol{\theta}) - \frac{1}{2}\log|\boldsymbol{\Sigma}_{y \,|\, \theta}| - \frac{1}{2}Q(\boldsymbol{\theta}) \ ,$$

where $Q(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\mu}_{\beta})^{\top}\boldsymbol{\Sigma}_{y \,|\, \theta}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\mu}_{\beta})$.

For faster convergence, we integrate out β and α from the model and first sample from

$$p(\boldsymbol{\theta} \,|\, \mathbf{y}) \propto p(\boldsymbol{\theta}) \times N(\mathbf{y} \,|\, \mathbf{X}\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_{y\,|\,\theta}),$$

where $\boldsymbol{\Sigma}_{y\,|\,\theta} = \mathbf{X}\boldsymbol{\Sigma}_\beta\mathbf{X}^\top + \mathbf{Z}(\theta)\mathbf{K}(\theta)\mathbf{Z}(\theta)^\top + \mathbf{D}(\theta)$.

This involves evaluating

$$\log p(\boldsymbol{\theta} \,|\, \mathbf{y}) = \text{const} + \log p(\boldsymbol{\theta}) - \frac{1}{2}\log|\boldsymbol{\Sigma}_{y\,|\,\theta}| - \frac{1}{2}Q(\boldsymbol{\theta}) \ ,$$

where $Q(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\mu}_\beta)^\top \boldsymbol{\Sigma}_{y\,|\,\theta}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\mu}_\beta)$.

1. $\mathbf{L} = \texttt{chol}(\boldsymbol{\Sigma}_{y\,|\,\theta})$, lower-triangular Cholesky factor $\mathbf{L}$ of $\boldsymbol{\Sigma}_{y\,|\,\boldsymbol{\theta}}$ ($O(n^3/3)$ flops)

For faster convergence, we integrate out β and α from the model and first sample from

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto p(\boldsymbol{\theta}) \times N(\mathbf{y} \mid \mathbf{X}\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_{y \mid \theta}),$$

where $\boldsymbol{\Sigma}_{y \mid \theta} = \mathbf{X}\boldsymbol{\Sigma}_\beta \mathbf{X}^\top + \mathbf{Z}(\theta)\mathbf{K}(\theta)\mathbf{Z}(\theta)^\top + \mathbf{D}(\theta)$.

This involves evaluating

$$\log p(\boldsymbol{\theta} \mid \mathbf{y}) = \text{const} + \log p(\boldsymbol{\theta}) - \frac{1}{2}\log|\boldsymbol{\Sigma}_{y \mid \theta}| - \frac{1}{2}Q(\boldsymbol{\theta})\ ,$$

where $Q(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\mu}_\beta)^\top \boldsymbol{\Sigma}_{y \mid \theta}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\mu}_\beta)$.

1. $\mathbf{L} = \texttt{chol}(\boldsymbol{\Sigma}_{y \mid \theta})$, lower-triangular Cholesky factor $\mathbf{L}$ of $\boldsymbol{\Sigma}_{y \mid \boldsymbol{\theta}}$ ($O(n^3/3)$ flops)
2. $\mathbf{u} = \texttt{trsolve}(\mathbf{L}, \mathbf{y} - \mathbf{X}\boldsymbol{\mu}_\beta)$, solves $\mathbf{L}\mathbf{u} = \mathbf{y} - \mathbf{X}\boldsymbol{\mu}_\beta$ ($O(n^2)$ flops)

For faster convergence, we integrate out $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ from the model and first sample from

$$p(\boldsymbol{\theta} \,|\, \mathbf{y}) \propto p(\boldsymbol{\theta}) \times N(\mathbf{y} \,|\, \mathbf{X}\boldsymbol{\mu}_{\beta}, \boldsymbol{\Sigma}_{y\,|\,\theta}),$$

where $\boldsymbol{\Sigma}_{y\,|\,\theta} = \mathbf{X}\boldsymbol{\Sigma}_{\beta}\mathbf{X}^{\top} + \mathbf{Z}(\boldsymbol{\theta})\mathbf{K}(\boldsymbol{\theta})\mathbf{Z}(\boldsymbol{\theta})^{\top} + \mathbf{D}(\boldsymbol{\theta})$.

This involves evaluating

$$\log p(\boldsymbol{\theta} \,|\, \mathbf{y}) = \text{const} + \log p(\boldsymbol{\theta}) - \frac{1}{2}\log|\boldsymbol{\Sigma}_{y\,|\,\theta}| - \frac{1}{2}Q(\boldsymbol{\theta}) \;,$$

where $Q(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\mu}_{\beta})^{\top}\boldsymbol{\Sigma}_{y\,|\,\theta}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\mu}_{\beta})$.

1. $\mathbf{L} = \texttt{chol}(\boldsymbol{\Sigma}_{y\,|\,\theta})$, lower-triangular Cholesky factor $\mathbf{L}$ of $\boldsymbol{\Sigma}_{y\,|\,\boldsymbol{\theta}}$ ($O(n^3/3)$ flops)
2. $\mathbf{u} = \texttt{trsolve}(\mathbf{L}, \mathbf{y} - \mathbf{X}\boldsymbol{\mu}_{\beta})$, solves $\mathbf{L}\mathbf{u} = \mathbf{y} - \mathbf{X}\boldsymbol{\mu}_{\beta}$ ($O(n^2)$ flops)
3. $Q(\boldsymbol{\theta}) = \mathbf{u}^{\top}\mathbf{u}$ ($2n$ flops)

For faster convergence, we integrate out $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ from the model and first sample from

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto p(\boldsymbol{\theta}) \times N(\mathbf{y} \mid \mathbf{X}\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_{y \mid \theta}),$$

where $\boldsymbol{\Sigma}_{y \mid \theta} = \mathbf{X}\boldsymbol{\Sigma}_\beta\mathbf{X}^\top + \mathbf{Z}(\boldsymbol{\theta})\mathbf{K}(\boldsymbol{\theta})\mathbf{Z}(\boldsymbol{\theta})^\top + \mathbf{D}(\boldsymbol{\theta})$.

This involves evaluating

$$\log p(\boldsymbol{\theta} \mid \mathbf{y}) = \text{const} + \log p(\boldsymbol{\theta}) - \frac{1}{2}\log|\boldsymbol{\Sigma}_{y \mid \theta}| - \frac{1}{2}Q(\boldsymbol{\theta}) \ ,$$

where $Q(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\mu}_\beta)^\top\boldsymbol{\Sigma}_{y \mid \theta}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\mu}_\beta)$.

1. $\mathbf{L} = \texttt{chol}(\boldsymbol{\Sigma}_{y \mid \theta})$, lower-triangular Cholesky factor $\mathbf{L}$ of $\boldsymbol{\Sigma}_{y \mid \boldsymbol{\theta}}$ ($O(n^3/3)$ flops)
2. $\mathbf{u} = \texttt{trsolve}(\mathbf{L}, \mathbf{y} - \mathbf{X}\boldsymbol{\mu}_\beta)$, solves $\mathbf{Lu} = \mathbf{y} - \mathbf{X}\boldsymbol{\mu}_\beta$ ($O(n^2)$ flops)
3. $Q(\boldsymbol{\theta}) = \mathbf{u}^\top\mathbf{u}$ ($2n$ flops)
4. log-determinant is $2\sum_{i=1}^{n}\log l_{ii}$, where l_{ii} are the diagonal entries in $\mathbf{L}$ (n flops)

Given marginal posterior samples $\boldsymbol{\theta}$ from $p(\boldsymbol{\theta} \mid \mathbf{y})$, we can draw posterior samples of β and $\boldsymbol{\alpha}$ using *composition sampling*.

We'll consider a portion of this algorithm in a subsequent exercise.

For more details see Finley, A.O., S. Banerjee, A.E. Gelfand. (2015) spBayes for large univariate and multivariate point-referenced spatio-temporal data models. *Journal of Statistical Software*, **63**:1–28.