Inference and Estimation in GLMMs

Recall GLMMs

Conditional on cluster specific random effects $\boldsymbol{b}_i$,

$$\begin{aligned} E(Y_{ij}|b_i) &= \mu_{ij}^{\mathbf{b}} \\ var(Y_{ij}|b_i) &= \phi a_{ij}^{-1} v(\mu_{ij}^{\mathbf{b}}) \end{aligned}$$

Random Effects Model :

$$g(\mu_{ij}^{\mathbf{b}}) = \mathbf{X}_{ij}^T \beta + \mathbf{Z}_{ij}^T \mathbf{b}_i$$

where

$$\mathbf{b}_i \sim N\{0, \mathbf{D}_0(\boldsymbol{\theta})\}$$

Likelihood

$$e^{\ell(\beta,\boldsymbol{\theta})} \propto |\mathbf{D}|^{-\frac{1}{2}} \int exp^{\{\sum_{i=1}^{n} \ell_i(Y_i|\mathbf{b};\beta) - \frac{1}{2}\mathbf{b}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{b}\}} d\mathbf{b}$$

How do we do estimation?

$$e^{\ell(\beta,\boldsymbol{\theta})} \propto |\mathbf{D}|^{-\frac{1}{2}} \int exp^{\{\sum_{i=1}^{n} \ell_i(Y_i|\mathbf{b};\beta) - \frac{1}{2}\mathbf{b}^\mathbf{T}\mathbf{D}^{-1}\mathbf{b}\}} d\mathbf{b}$$

Main challenge: Difficult to take that integral!

Conditional Inference

Recall: this is potential approach for dealing with random effects

- Idea: Calculate the sufficient statistic for $\mathbf{b}_i$s and make inference using the conditional likelihood on the sufficient statistics.
- Pro: robustness (no distributional assumptions on $\mathbf{b}_i$) and likelihood has closed form
- Con: only works in special cases (e.g. logistic GLMM and poisson log linear GLMM)
- Con: you've lost the random effects

Instead: can we just take the MLE?

Full MLE Approach

$$e^{\ell(\beta,\boldsymbol{\theta})} \propto |\mathbf{D}|^{-\frac{1}{2}} \int exp^{\{\sum_{i=1}^{n} \ell_i(Y_i|\mathbf{b};\beta) - \frac{1}{2}\mathbf{b}^{\mathbf{T}}\mathbf{D}^{-\mathbf{1}}\mathbf{b}\}} d\mathbf{b}$$

Idea: To directly evaluate the integrated loglikelihood $\ell(\boldsymbol{\beta}, \boldsymbol{\theta})$ using numerical integration methods, such as adaptive Gaussian quadrature or Monte-Carlo.

Numerical Integration: Adaptive Guassian Quadrature

Idea: To numerically evaluate the integral in $\ell(\boldsymbol{\beta}, \boldsymbol{\theta})$ using a weighted sum over predefined set of quadrature points, i.e., by replacing the integral by a sum.

Advantages:

- Full MLE if the integral can be evaluated precisely using sufficient number of quadrature points.
- For a one-dimensional integral, e.g., random intercept models, 10-20 quadrature points give a very precise evaluation of the integral and calculations are very quick.

Numerical Integration: Adaptive Gaussian Quadrature (2)

Disadvantages:

- It is only feasible if the dimension of integral is small, e.g., in simple random effect models for longitudinal/clustered data.
- For a multi-dimensional integral, the number of quadrature points grows exponentially and such a direct quadrature approximation can be prohibited, e.g., for a 5-dimensional integral, the number of quadratures is 20^5.
- For spatial data, the dimension of integration is the sample size. Direct numerical integration is infeasible.

References:

Golub and Welsch (1969); Pinheiro and Bates (1995)

Adaptive Gaussian Quadrature: Operationally

Quadrature replaces

$$I = \int w(u)g(u)du \text{ by } \widetilde{I} = \sum_{i=1}^{n} w_i g(u_i)$$

for nodes u_i and weights w_i (note n is not sample size)

- For GLMM: we have integrals w.r.t. a Gaussian density
- Gauss-Hermite quadrature is designed for problems of this type

Adaptive Gaussian Quadrature: Operationally (2)

Gauss-Hermite Quadrature:

$$I = \int_{-\infty}^{\infty} \exp(-u^2) g(u) du \approx \sum_{i=1}^{n} w_i g(u_i) = \widetilde{I}$$

where $g(\cdot)$ is a polynomial of degree $2n - 1$ or less.

u_i is the i^{th} of n roots of the Hermite polynomial $H_n(u)$ with weight

$$w_i = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 [H_{n-1}(u_i)]^2}$$

where $H_n(u)$ are orthogonal polynomial sequence given by

$$H_n(u) = (-1)^n \exp(-u^2) \frac{\partial^n}{\partial u^n} \exp(-u^2)$$

Adaptive Gaussian Quadrature: Getting Weights (3)

R functions to get the weights

```
> library(statmod)
> quad <- gauss.quad(4, kind="hermite")
> round(quad$nodes, 3)
[1] -1.651 -0.525 0.525 1.651
> round(quad$weights, 3)
[1] 0.081 0.805 0.805 0.081
> quad <- gauss.quad(5, kind="hermite")
> round(quad$nodes, 3)
[1] -2.020 -0.959 0.000 0.959 2.020
> round(quad$weights, 3)
[1] 0.020 0.394 0.945 0.394 0.020
```

Adaptive Gaussian Quadrature: Example (3)

$$\int_{-\infty}^{\infty} \exp(-u^2)\exp(cos(u)^2 + sin(u)^3)du$$

```
> u <- seq(-100,100,0.1)
> sum(exp(-u^2)*exp(cos(u)^2+sin(u)^3)*0.1)
[1] 3.829421
> quad <- gauss.quad(5, kind="hermite")
> sum(quad$weights*exp(cos(quad$nodes)^2+sin(quad$nodes)^3)
[1] 3.895256
> quad <- gauss.quad(10, kind="hermite")
> sum(quad$weights*exp(cos(quad$nodes)^2+sin(quad$nodes)^3)
[1] 3.830933
```

Adaptive Gaussian Quadrature: Multiple Dimensions

Suppose $\boldsymbol{\theta}$ is two dimensional:

$$I = \int f(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int \int f(\theta_1, \theta_2) d\theta_2 d\theta_1 = \int f^*(\theta_1) d\theta_1$$

where $f^*(\theta_1) = \int f(\theta_1, \theta_2) d\theta_2$. Then

$$\widetilde{I} = \sum_{i=1}^{m} w_i f^*(\theta_{1_i}) = \sum_{i=1}^{n_1} w_i \sum_{j=1}^{n_2} w_j f(\theta_{1_i}, \theta_{2_j}) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} w_i w_j f(\theta_{1_i}, \theta_{2_j})$$

which is a Cartesian product.

Question: Where is the "adaptive" part?

Adaptive Gaussian Quadrature

- Note that the nodes are selected without regard to the function $g(\cdot)$
  - Good if $w(u)g(u)$ behaves like a normal
  - But in general, quadrature points may not lie in interesting or important region

- Adaptive: rescale and shift quadrature points s.t. more quadrature points in region of interest

- Adaptive version requires fewer quadrature points, but requires calculating mode of $w(u)g(u)$ which can be slow!

Recall: gets complicated when dimensionality of the integral gets higher $\rightarrow$ Approximations

More Numerical Integration

- Tools in R:
  - `integrate()` and `cubature` package for integration
  - `nlm()` and `optim()` for optimization
  - `numDeriv()` for getting Hessians
- lme4 bundles a lot of this stuff for you
- SAS is (again) probably a little bit better
- None of the R packages are perfect and need to look at the output carefully
- GEE is a bit more stable (numerically)

Estimating β and $\boldsymbol{\theta}$

$$e^{\ell(\beta,\boldsymbol{\theta})} \propto |\mathbf{D}|^{-\frac{1}{2}} \int e^{\left\{\sum_{i=1}^{n} \ell_i(Y_i|\mathbf{b};\beta) - \frac{1}{2}\mathbf{b}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{b}\right\}} d\mathbf{b}$$

So far, to estimate β and $\boldsymbol{\theta}$:

1. Conditional inference (condition on sufficient statistic)
2. Full MLE using numerical integration (Gaussian Quadrature)

Other strategies:

1. Approximate Inference
2. Expectation Maximization algorithm
3. Gibbs Sampling

Approximate Inference

Idea: To approximate the integrated log-likelihood $\ell(\boldsymbol{\beta}, \boldsymbol{\theta})$ using various lower-order approximations and maximize the approximate log-likelihood wrt $(\boldsymbol{\beta}, \boldsymbol{\theta})$.

- Laplace approximation
- Solomon-Cox approximation
- Penalized Quasilikelihood (PQL)
- Corrected PQL

Note: These approximation procedures do not always give consistent estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ except for normal data.

Laplace Approximation

Idea: To expand the integrand about the mode $\mathbf{b} = \widehat{\mathbf{b}}$ in a lower-order Taylor series before integration.

Then we have

$$\ell(\beta, \mathbf{b}) \approx \ell(\beta, \widehat{\mathbf{b}}) - \frac{1}{2}(\mathbf{b} - \widehat{\mathbf{b}})^T \left[-\ell''_{\mathbf{bb}}(\beta, \boldsymbol{\theta}, \mathbf{b})|_{\mathbf{b}=\widehat{\mathbf{b}}} \right] (\mathbf{b} - \widehat{\mathbf{b}})$$

and using this approximation we can calculate

$$\begin{aligned}
e^{\ell(\beta,\boldsymbol{\theta})} &\propto |\mathbf{D}|^{-\frac{1}{2}} \int e^{\{\sum_{i=1}^{n} \ell_i(Y_i|\mathbf{b};\beta) - \frac{1}{2}\mathbf{b}^{\mathsf{T}}\mathbf{D}^{-1}\mathbf{b}\}} d\mathbf{b} \\
&\approx \int e^{\ell(\beta,\widehat{\mathbf{b}}) - \frac{1}{2}(\mathbf{b}-\widehat{\mathbf{b}})^T \left[-\ell''_{\mathbf{bb}}(\beta,\boldsymbol{\theta},\mathbf{b})|_{\mathbf{b}=\widehat{\mathbf{b}}} \right] (\mathbf{b}-\widehat{\mathbf{b}})} d\mathbf{b} \\
&= L(\beta, \widehat{\mathbf{b}}) \int e^{-\frac{1}{2}(\mathbf{b}-\widehat{\mathbf{b}})^T \left[-\ell''_{\mathbf{bb}}(\beta,\boldsymbol{\theta},\mathbf{b})|_{\mathbf{b}=\widehat{\mathbf{b}}} \right] (\mathbf{b}-\widehat{\mathbf{b}})} d\mathbf{b} \\
&= L(\beta, \widehat{\mathbf{b}}) \sqrt{\frac{(2\pi)^q}{\left| -\ell''_{\mathbf{bb}}(\beta, \boldsymbol{\theta}, \mathbf{b}) \Big|_{\mathbf{b}=\widehat{\mathbf{b}}} \right|}}
\end{aligned}$$

Laplace Approximation (2)

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) \approx \ell(\boldsymbol{\beta}, \widehat{\mathbf{b}}) - \frac{1}{2} \log \left\{ \left| -\ell''_{\mathbf{bb}}(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{b}) \Big|_{\mathbf{b}=\widehat{\mathbf{b}}} \right| \right\} + \frac{q}{2} \log(2\pi)$$

- Laplace likelihood only approximates: some amount of error in the resulting estimates
- Usually "accurate enough"

References

Tierney and Kadane (1986, JASA, 82-86)
Breslow and Clayton (1993, JASA);
Breslow and Lin (1995, Biometrika, 81-91)

Soloman-Cox Approximation

Idea: To expand the integrand about the mode $\mathbf{b} = 0$ before integration.

Similar idea to Laplace approximation: simpler than Laplace, but also less accurate

References
Barndorff-Niolsen and Cox, 1989, 3.3
Solomon and Cox, 1992, Biometrika
Breslow and Lin, 1995, Biometrika, 81-91

Penalized Quasilikelihood (PQL)

Idea: Modified Laplace in which we replace the GLM aspect with a nonlinear least-squares model

Main feature is that it iteratively fits linear mixed models using GLM-type working weights and working vectors

Usual PQL does not work well for sparse (e.g binary) data since Laplace doesn't work so well $\rightarrow$ corrected PQL

References

Schall, 1991, Biometrika
Breslow and Clayton, 1993, JASA
Breslow and Lin, 1995, Biometrika
Lin and Breslow, 1996, JASA

Expectation-Maximization (EM) Algorithm

Complete data: $\mathbf{Y}, \mathbf{Z}$ Observed data: $\mathbf{Y}$

We want to estimate β which has likelihood $L(\beta; \mathbf{Y}, \mathbf{Z})$ by maximizing the marginal likelihood

$$L(\beta; \mathbf{Y}) = \int L(\beta; \mathbf{Y}, \mathbf{Z}) d\mathbf{Z}.$$

Expectation (E) Step: Define $Q(\beta|\beta^{(k)})$ as expected log likelihood wrt current distribution of $\mathbf{Z}|\mathbf{Y}$ and current parameters ($\beta^{(k)}$)

$$Q(\beta|\beta^{(k)}) = E_{\mathbf{Z}|\mathbf{Y},\beta^{(k)}}[\ell(\beta; \mathbf{YZ})]$$

Maximization (M) Step: Find parameters that maximize

$$\beta^{(k+1)} = \underset{\beta}{\operatorname{argmax}}\, Q(\beta|\beta^{(k)})$$

EM for G/LMM

Complete data: **Y**, **b**

Observed data: **Y**

E-step:

$$Q(\beta, \boldsymbol{\theta}|\beta^{[k]}, \boldsymbol{\theta}^{[k]}) = E\{\ell(\mathbf{Y}|\mathbf{b}; \beta) + \ell(\mathbf{b}; \boldsymbol{\theta})|\mathbf{Y}; \beta^{[k]}, \boldsymbol{\theta}^{[k]}\}$$

Involves the same dimension of integration as the likelihood but the terms are relatively easier to calculate.

- Gaussian approximation (Stiratelli,et al, 1982, Biometrika)
- 2nd order Laplace approximation (Steele, 1996, Biometrics)
- Monte-Carlo simulation (Metropolis) (McCulloch, 1994, 1997, JASA; Waller, et al, 1997, JASA)

M-step: Maximize $Q(\beta, \boldsymbol{\theta}|\beta^{[k]}, \boldsymbol{\theta}^{[k]})$ wrt β and $\boldsymbol{\theta}$.

Remarks

- Implementation of EM in practice is done backward, as it is harder to deal with maximization but easier to deal with equation solving.
- One starts from the M-step by calculating using the complete data loglikelihood the score equations for the model parameters, i.e., $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, and identify terms that involve the missing data, i.e., the terms involving $\mathbf{b}_i$.
- At the E-step, calculate the expectations of the identified terms that involve the missing data $\mathbf{b}_i$ and evaluate the expectations at $\hat{\boldsymbol{\beta}}^{[k]}$ and $\hat{\boldsymbol{\theta}}^{[k]}$.
- Iterate between the M-step and the E-step until convergence.

Example: EM for LMM

Model:

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i$$

where $e_i \sim N(0, \sigma^2\mathbf{I})$, $\quad \mathbf{b}_i \sim N_q(0, \mathbf{D})$.

Observed data: Y **Complete data: Y, b**
Parameters: $\beta, \sigma^2, \mathbf{D}$

Complete Data Loglikelihood

$$\sum_{i=1}^{m} \ell(\mathbf{Y}_i|\mathbf{b}_i; \beta, \sigma^2) + \ell(\mathbf{b}_i; \mathbf{D})$$

$$= \sum_{i=1}^{m} \{-\frac{n_i}{2} ln\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{Y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i)^T(\mathbf{Y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i)$$
$$-\frac{1}{2} ln|\mathbf{D}| - \frac{1}{2}\mathbf{b}_i^T\mathbf{D}^{-1}\mathbf{b}_i\}$$

Example: EM for LMM (2)

Score equations for complete data:

$$\sum_{i=1}^{m} \mathbf{X}_i^T (\mathbf{Y}_i - \mathbf{X}_i \beta - \mathbf{Z}_i \mathbf{b}_i) = 0$$

$\Rightarrow$

$$\widehat{\beta} = (\sum_{i=1}^{m} \mathbf{X}_i^T \mathbf{X}_i)^{-1} \sum_{i=1}^{m} (\mathbf{Y}_i - \mathbf{Z}_i \mathbf{b}_i)$$

$$\hat{\mathbf{D}} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{b}_i \mathbf{b}_i^T$$

$$\hat{\sigma}^2 = \frac{1}{\sum n_i} \sum_{i=1}^{m} (\mathbf{Y}_i - \mathbf{X}_i \beta - \mathbf{Z}_i \mathbf{b}_i)^T (\mathbf{Y}_i - \mathbf{X}_i \beta - \mathbf{Z}_i \mathbf{b}_i)$$

Example: EM for LMM - E-step

Need to calculate

$$E[\mathbf{b}_i|\mathbf{Y}_i; \hat{\beta}^{[k]}, \hat{\mathbf{D}}^{[k]}, \hat{\sigma^2}^{[k]}]$$
$$E[\mathbf{b}_i\mathbf{b}_i^T|\mathbf{Y}_i, \hat{\beta}^{[k]}, \hat{\mathbf{D}}^{[k]}, \hat{\sigma^2}^{[k]}]$$

$$E[\mathbf{e}_i^T\mathbf{e}_i|\mathbf{Y}_i, \hat{\beta}^{[k]}, \hat{\mathbf{D}}^{[k]}, \hat{\sigma^2}^{[k]}],$$

where $\mathbf{e_i} = \mathbf{Y_i} - \mathbf{X_i}\beta - \mathbf{Z_i}\mathbf{b_i}$.

Example: EM for LMM - E-step (2)

Recall LMMs:

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \mathbf{e}_i, \;\; \mathbf{b}_i \sim N(0, \mathbf{D})$$

Fact 1:

$$\begin{pmatrix} \mathbf{Y}_i \\ \mathbf{b}_i \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{X}_i\beta \\ 0 \end{pmatrix}, \begin{pmatrix} \mathbf{V}_i & \mathbf{Z}_i\mathbf{D} \\ \mathbf{D}\mathbf{Z}_i^T & \mathbf{D} \end{pmatrix}\right]$$

where $\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \sigma^2\mathbf{I}$. Then

$$\begin{aligned} \hat{\mathbf{b}}_i = E(\mathbf{b}_i|\mathbf{Y}_i) &= \mathbf{D}\mathbf{Z}_i^T\mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\beta) \\ \hat{\mathbf{V}}_{b_i} = cov(\mathbf{b}_i|\mathbf{Y}_i) &= \mathbf{D} - \mathbf{D}\mathbf{Z}_i^T\mathbf{V}_i^{-1}\mathbf{Z}_i\mathbf{D} \end{aligned}$$

Fact 2:

If a random variable $\mathbf{c} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$E(\mathbf{c}^T\mathbf{A}\mathbf{c}) = tr(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T\boldsymbol{A}\boldsymbol{\mu}$$

Example: EM for LMM - E-step (3)

$$
\begin{aligned}
\hat{\boldsymbol{b}}_{\boldsymbol{i}}^{[\boldsymbol{k}]} = \quad & E[\mathbf{b}_i|\mathbf{Y}_i;\hat{\beta}^{[k]},\hat{\mathbf{D}}^{[k]},\hat{\sigma^2}^{[k]}] && = \mathbf{D}^{[k]}\mathbf{Z}_i^T\mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\beta^{[k]}) \\
\hat{\mathbf{V}}_{b_i}^{[k]} = \quad & cov[\mathbf{b}_i|\mathbf{Y}_i;\hat{\beta}^{[k]},\hat{\mathbf{D}}^{[k]},\hat{\sigma^2}^{[k]}] && = \mathbf{D}^{[k]} - \mathbf{D}^{[k]}\mathbf{Z^T}_i\mathbf{V}_i^{-1[k]}\mathbf{Z}_i\mathbf{D}^{[k]} \\
& E[\mathbf{b}_i\mathbf{b}_i^T|\mathbf{Y}_i;\hat{\beta}^{[k]},\hat{\mathbf{D}}^{[k]},\hat{\sigma^2}^{[k]}] && = \hat{\mathbf{V}}_{b_i}^{[k]} + \hat{\mathbf{b}}_i^{[k]}\hat{\mathbf{b}}_i^{[k]T} \\
& E[\mathbf{e}_i^T\mathbf{e}_i|\mathbf{Y}_i,\hat{\beta}^{[k]},\hat{\mathbf{D}}^{[k]},\hat{\sigma^2}^{[k]}] && = tr(\mathbf{Z^T}_i\hat{\mathbf{V}}_{b_i}^{[k]}\mathbf{Z}_i) + \hat{\mathbf{e}}_i^{[k]T}\hat{\mathbf{e}}_i^{[k]},
\end{aligned}
$$

where $\hat{\mathbf{e}}_{\mathbf{i}}^{[\mathbf{k}]} = \mathbf{Y_i} - \mathbf{X_i}\beta^{[\mathbf{k}]} - \mathbf{Z_i}\mathbf{b}_{\mathbf{i}}^{[\mathbf{k}]}$.

Example: EM for LMM - M-step

$$
\begin{aligned}
\widehat{\beta}^{[k+1]} &= (\sum_{i=1}^{m} \mathbf{X}_i^T \mathbf{X}_i)^{-1} \sum_{i=1}^{m} (\mathbf{Y}_i - \mathbf{Z}_i \hat{\mathbf{b}}_i^{[k]}) \\
\widehat{\mathbf{D}}^{[k+1]} &= \frac{1}{m} \sum_{i=1}^{m} E(\mathbf{b_i} \mathbf{b_i^T} | \mathbf{Y}_i; \beta, \hat{\mathbf{D}}^{[k]}, \hat{\sigma^2}^{[k]}) \\
&= \frac{1}{m} \sum_{i=1}^{m} (\hat{\mathbf{V}}_{b_i}^{[k]} + \hat{\mathbf{b}}_i^{[k]} \hat{\mathbf{b}}_i^{[k]T})
\end{aligned}
$$

$$
\begin{aligned}
\hat{\sigma^2}^{[k+1]} &= \frac{1}{\sum n_i} \sum_i E(\mathbf{e_i^T} \mathbf{e_i} | \mathbf{Y}_i; \widehat{\beta}^{[k]}, \hat{\mathbf{D}}^{[k]}, \hat{\sigma^2}^{[k]}) \\
&= \frac{1}{\sum n_i} \sum_{i=1}^{m} \{ tr(\mathbf{Z}_i^T \hat{\mathbf{V}}_{b_i}^{[k]} \mathbf{Z}_i) + \hat{\mathbf{e}}_i^{[k]T} \hat{\mathbf{e}}_i^{[k]} \}.
\end{aligned}
$$

Gibbs Sampling

A popular Bayesian inference procedure in hierarchical models.

Prior for β: nearly non-informative prior, i.e., $\beta \sim (0, 1000\mathbf{I})$.

Prior for $\mathbf{D}(\boldsymbol{\theta})$: Gamma/Wishart (Jeffery prior does not work, since the posterior is not proper).

Objective: Generate the joint distribution of $[\beta, \boldsymbol{\theta}, \mathbf{b} \mid \mathbf{Y}]$

How: Generate a series of conditional distributions $[\beta \mid \boldsymbol{\theta}, \mathbf{b}, \mathbf{Y}]$, $[\mathbf{b} \mid \beta, \boldsymbol{\theta}, \mathbf{Y}]$, $[\boldsymbol{\theta} \mid \beta, \mathbf{b}, \mathbf{Y}]$

References: Zeger and Karim (1991, JASA); McCulloch (1994, JASA)