Advanced Data Analysis

Whiteboard Lecture

Bayesian Normal Mixed Models

The normal mixed model is

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\varepsilon} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{G} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{R} \end{bmatrix} \right)$$

The Bayesian version is:

$[\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{G}, \boldsymbol{R}] \sim N(\boldsymbol{X\beta} + \boldsymbol{Zu}, \boldsymbol{R})$

$[\boldsymbol{u}|\boldsymbol{G}] \sim N(\mathbf{0}, \boldsymbol{G})$

$[\boldsymbol{\beta}],\ [\boldsymbol{G}],\ [\boldsymbol{R}]$ requiring priors to be specified.

Let's focus first on the special case:

$$\boldsymbol{G} = \sigma_u^2 \boldsymbol{I}$$

$$\boldsymbol{R} = \sigma_\varepsilon^2 \boldsymbol{I}$$

(appropriate for the random intercept model).

Common priors are:

$$[\beta] \sim N(0, \sigma_\beta^2 I)$$
$$[\sigma_u^2] \sim \text{IG}(A_u, B_u)$$
$$[\sigma_\varepsilon^2] \sim \text{IG}(A_\varepsilon, B_\varepsilon)$$

where $[\sigma^2] \sim \text{IG}(A, B)$

$$\iff [\sigma^2] = \frac{B^A}{\Gamma(A)}(\sigma^2)^{-A-1}e^{-B/\sigma^2}, \quad \sigma^2 > 0$$

(Inverse Gamma distribution with parameters A and B).

Note: σ_β^2, A_u, B_u, A_ε and B_ε are called hyperparameters.

These have to be set by the analyst.In the absence of prior beliefs

AND ASSUMING DATA HAVE BEEN

APPROPRIATELY STANDARDISED

“reasonable” values are:

$\sigma_\beta^2 = 10^8$ and

$$A_u = B_u = A_\varepsilon = B_\varepsilon = \frac{1}{100}.$$

Wanted: Posterior distributions:

$$[\boldsymbol{\beta}|\boldsymbol{y}]$$

$$[\boldsymbol{u}|\boldsymbol{y}]$$

$$[\boldsymbol{\sigma}_u^2|\boldsymbol{y}]$$

$$[\boldsymbol{\sigma}_\varepsilon^2|\boldsymbol{y}]$$

BAD NEWS: None of these are available in closed form.

GOOD NEWS: The full conditionals are all available in closed form!

$\Longrightarrow$ Gibbs sampling (simplest type of MCMC) is available!

Treating $(\boldsymbol{\beta}, \boldsymbol{u})$ as an entity, the full conditionals are:

$$[\boldsymbol{\beta}, \boldsymbol{u}|\boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_\varepsilon^2, \boldsymbol{y}]$$
$$[\boldsymbol{\sigma}_u^2|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_\varepsilon^2, \boldsymbol{y}]$$
$$[\boldsymbol{\sigma}_\varepsilon^2|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_u^2, \boldsymbol{y}]$$

$$\begin{aligned}
[\boldsymbol{\sigma}_u^2|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_\varepsilon^2, \boldsymbol{y}] &= \frac{[\boldsymbol{\sigma}_u^2, \boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_\varepsilon^2, \boldsymbol{y}]}{[\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_\varepsilon^2, \boldsymbol{y}]} \\
&\propto [\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_\varepsilon^2] \\
&= [\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_\varepsilon^2] \\
&\quad \times [\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_\varepsilon^2] \\
&= [\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_\varepsilon^2] \\
&\quad \times [\boldsymbol{u}|\boldsymbol{\beta}, \boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_\varepsilon^2][\boldsymbol{\beta}, \boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_\varepsilon^2] \\
&= [\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_\varepsilon^2] \\
&\quad \times [\boldsymbol{u}|\boldsymbol{\sigma}_u^2][\boldsymbol{\beta}][\boldsymbol{\sigma}_u^2][\boldsymbol{\sigma}_\varepsilon^2] \\
&\propto [\boldsymbol{u}|\boldsymbol{\sigma}_u^2][\boldsymbol{\sigma}_u^2]
\end{aligned}$$

since $[y|\beta, u, \sigma_u^2, \sigma_\varepsilon^2]$, $[\beta]$ and $[\sigma_\varepsilon^2]$ do not involve σ_u^2.

Question

What do we mean by $[y|\beta, u, \sigma_u^2, \sigma_\varepsilon^2]$ does not involve σ_u^2?

Answer

$$[y|\beta, u, \sigma_u^2, \sigma_\varepsilon^2] \sim N(X\beta + Zu, \sigma_\varepsilon^2 I)$$

$\Longrightarrow$

$$[y|\beta, u, \sigma_u^2, \sigma_\varepsilon^2] = (2\pi\sigma_\varepsilon^2)^{-n/2} e^{-\frac{\|y - X\beta - Zu\|^2}{2\sigma_\varepsilon^2}}.$$

(where $\|v\|^2 = v^T v$ for any vector v).

The variable of current interest, σ_u^2 does not appear in this expression, so can be ignored in the above calculations (absorbed into the $\propto$ part).

So we have

$$
\begin{aligned}
[\boldsymbol{\sigma}_u^2|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_\varepsilon^2, \boldsymbol{y}] &\propto [\boldsymbol{u}|\boldsymbol{\sigma}_u^2][\boldsymbol{\sigma}_u^2] \\
&= (2\pi)^{-q/2}(\boldsymbol{\sigma}_u^2)^{-q/2} e^{-u^T u/(2\sigma_u^2)} \\
&\quad \times \frac{B_u^{A_u}}{\Gamma(A_u)}(\boldsymbol{\sigma}_u^2)^{-A_u-1} e^{-B_u/\sigma_u^2}
\end{aligned}
$$

where q is the length of the u vector (i.e. u is $q \times 1$).

This simplifies to

$$
[\boldsymbol{\sigma}_u^2|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_\varepsilon^2, \boldsymbol{y}] \propto (\boldsymbol{\sigma}_u^2)^{-A_u - \frac{q}{2} - 1} e^{-(\frac{1}{2}u^T u + B_u)/\sigma_u^2}.
$$

$$
\Longrightarrow
$$

$$
[\boldsymbol{\sigma}_u^2|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_\varepsilon^2, \boldsymbol{y}] \sim \mathsf{IG}(A_u + \tfrac{q}{2}, \tfrac{1}{2}\|\boldsymbol{u}\|^2 + B_u).
$$

Similarly,

$$
\begin{aligned}
&[\boldsymbol{\sigma}_\varepsilon^2|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_u^2, \boldsymbol{y}] \\
&\qquad \sim \mathsf{IG}(A_\varepsilon + \tfrac{n}{2}, \tfrac{1}{2}\|\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Zu}\|^2 + B_\varepsilon).
\end{aligned}
$$

Finally,

$$
\begin{aligned}
&[\boldsymbol{\beta}, \boldsymbol{u}|\boldsymbol{\sigma}_\varepsilon^2, \boldsymbol{\sigma}_u^2, \boldsymbol{y}] \propto [\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_\varepsilon^2] \\
= \;& [\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_\varepsilon^2][\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_\varepsilon^2] \\
= \;& [\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_\varepsilon^2][\boldsymbol{\beta}, \boldsymbol{u}|\boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_\varepsilon^2][\boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_\varepsilon^2] \\
= \;& [\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_\varepsilon^2][\boldsymbol{\beta}][\boldsymbol{u}|\boldsymbol{\sigma}_u^2][\boldsymbol{\sigma}_u^2][\boldsymbol{\sigma}_\varepsilon^2] \\
\propto \;& [\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_\varepsilon^2][\boldsymbol{\beta}][\boldsymbol{u}|\boldsymbol{\sigma}_u^2] \\
= \;& \phi_{\sigma_\varepsilon^2 I}(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Zu})\phi_{\sigma_\beta^2 I}(\boldsymbol{\beta})\phi_{\sigma_u^2 I}(\boldsymbol{u}) \\
= \;& \phi_{\sigma_\varepsilon^2 I}\left(\boldsymbol{y} - \boldsymbol{C}\begin{bmatrix}\boldsymbol{\beta} \\ \boldsymbol{u}\end{bmatrix}\right)\phi_{\left[\begin{smallmatrix}\sigma_\beta^2 I & 0 \\ 0 & \sigma_u^2 I\end{smallmatrix}\right]}\left(\begin{bmatrix}\boldsymbol{\beta} \\ \boldsymbol{u}\end{bmatrix}\right)
\end{aligned}
$$

where $\boldsymbol{C} = [\boldsymbol{X}\ \boldsymbol{Z}]$.

i.e.

$$[\boldsymbol{\beta}, \boldsymbol{u} | \boldsymbol{\sigma}_\varepsilon^2, \boldsymbol{\sigma}_u^2, \boldsymbol{y}] \propto$$
$$\phi_{\boldsymbol{\sigma}_\varepsilon^2 \boldsymbol{I}}\left(\boldsymbol{y} - \boldsymbol{C}\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}\end{bmatrix}\right)\phi_{\begin{bmatrix}\sigma_\beta^2 I & 0\\ 0 & \sigma_u^2 I\end{bmatrix}}\left(\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}\end{bmatrix}\right)$$

Now apply 'A Useful Result Involving Normal Distributions' (hand-out) with

$$\boldsymbol{\Sigma} = \boldsymbol{\sigma}_\varepsilon^2 \boldsymbol{I}, \quad \boldsymbol{\Lambda} = \begin{bmatrix}\boldsymbol{\sigma}_\beta^2 \boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{\sigma}_u^2 \boldsymbol{I}\end{bmatrix}$$
$$\boldsymbol{A} = \boldsymbol{C}, \quad \boldsymbol{a} = \begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{u}\end{bmatrix}, \quad \boldsymbol{x} = \boldsymbol{y}$$

to get

$$[\boldsymbol{\beta}, \boldsymbol{u} | \boldsymbol{\sigma}_\varepsilon^2, \boldsymbol{\sigma}_u^2, \boldsymbol{y}] \propto$$
$$\phi_{\left(C^T\left(\frac{1}{\sigma_\varepsilon^2}\right)C + \begin{bmatrix}\sigma_\beta^2 I & 0\\ 0 & \sigma_u^2 I\end{bmatrix}^{-1}\right)^{-1}}\left(\begin{bmatrix}\beta\\ u\end{bmatrix} - \left(C^T\left(\frac{1}{\sigma_\varepsilon^2}\right)C + \begin{bmatrix}\sigma_\beta^2 I & 0\\ 0 & \sigma_u^2 I\end{bmatrix}^{-1}\right)^{-1} C^T\left(\frac{1}{\sigma_\varepsilon^2}\right)y\right)$$

We have thus established that

$$[\boldsymbol{\beta}, \boldsymbol{u} | \boldsymbol{\sigma}_\varepsilon^2, \boldsymbol{\sigma}_u^2, \boldsymbol{y}]$$
$$\sim N\left(\left(\boldsymbol{C}^T\boldsymbol{C} + (\boldsymbol{\sigma}_\varepsilon^2/\boldsymbol{\sigma}_u^2)\begin{bmatrix}(\boldsymbol{\sigma}_u^2/\boldsymbol{\sigma}_\beta^2)\boldsymbol{I} & 0\\ 0 & \boldsymbol{I}\end{bmatrix}\right)^{-1}\boldsymbol{C}^T\boldsymbol{y},\right.$$
$$\left.\boldsymbol{\sigma}_\varepsilon^2\left(\boldsymbol{C}^T\boldsymbol{C} + (\boldsymbol{\sigma}_\varepsilon^2/\boldsymbol{\sigma}_u^2)\begin{bmatrix}(\boldsymbol{\sigma}_u^2/\boldsymbol{\sigma}_\beta^2)\boldsymbol{I} & 0\\ 0 & \boldsymbol{I}\end{bmatrix}\right)^{-1}\right)$$

Gibbs sampling involves:

Step 0: Get initial values $\boldsymbol{\beta}^{[0]}$, $\boldsymbol{u}^{[0]}$, $(\boldsymbol{\sigma}_u^2)^{[0]}$, $(\boldsymbol{\sigma}_\varepsilon^2)^{[0]}$.

Step 1: Generate $(\boldsymbol{\beta}^{[1]}, \boldsymbol{u}^{[1]})$ from

$$[\boldsymbol{\beta}, \boldsymbol{u} | (\boldsymbol{\sigma}_\varepsilon^2)^{[0]}, (\boldsymbol{\sigma}_u^2)^{[0]}, \boldsymbol{y}].$$

Step 2: Generate $(\boldsymbol{\sigma}_u^2)^{[1]}$ from

$$[\boldsymbol{\sigma}_u^2 | \boldsymbol{\beta}^{[1]}, \boldsymbol{u}^{[1]}, (\boldsymbol{\sigma}_\varepsilon^2)^{[0]}, \boldsymbol{y}].$$

Step 3: Generate $(\sigma_\varepsilon^2)^{[1]}$ from

$$[\sigma_\varepsilon^2 | \beta^{[1]}, u^{[1]}, (\sigma_u^2)^{[1]}, y].$$

Step 4: Generate $(\beta^{[2]}, u^{[2]})$ from

$$[\beta, u | (\sigma_\varepsilon^2)^{[1]}, (\sigma_u^2)^{[1]}, y].$$

Continue until we have

$$\beta^{[g]}, u^{[g]}, (\sigma_u^2)^{[g]}, (\sigma_\varepsilon^2)^{[g]}$$

for $1 \leq g \leq G$, where G is ‘large’ (e.g. G=5000).

After discarding the first 1000 (say) we should have samples from the posterior distributions:

$$[\beta | y],\ [u | y],\ [\sigma_u^2 | y],\ [\sigma_\varepsilon^2 | y]$$

with which to make inference.

♦ ♦ ♦