Linear Mixed Effects Models

STA702

Merlise Clyde
Duke University

Random Effects Regression

- Easy to extend from random means by groups to random group level coefficients:

$$Y_{ij} = \boldsymbol{\theta}_j^T \mathbf{x}_{ij} + \epsilon_{ij}$$
$$\epsilon_{ij} \overset{\text{iid}}{\sim} \textsf{N}(0, \sigma^2)$$

- $\boldsymbol{\theta}_j$ is a $d \times 1$ vector regression coefficients for group j
- $\mathbf{x}_{ij}$ is a $d \times 1$ vector of predictors for group j
- If we view the groups as exchangeable, describe across group heterogeneity by

$$\boldsymbol{\theta}_j \overset{\text{iid}}{\sim} \textsf{N}(\boldsymbol{\beta}, \boldsymbol{\Sigma})$$

- $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$ and σ^2 are population parameters to be estimated.
- Designed to accommodate correlated data due to nested/hierarchical structure/repeated measurements: students w/in schools; patients w/in hospitals; additional covariates

Linear Mixed Effects Models

- We can write $\boldsymbol{\theta} = \boldsymbol{\beta} + \boldsymbol{\gamma}_j$ with $\boldsymbol{\gamma}_j \overset{\text{iid}}{\sim} \textsf{N}(\mathbf{0}, \boldsymbol{\Sigma})$
- Substituting, we can rewrite model

$$
\begin{aligned}
& Y_{ij} = \boldsymbol{\beta}^T \mathbf{x}_{ij} + \boldsymbol{\gamma}_j^T \mathbf{x}_{ij} + \epsilon_{ij}, \qquad \epsilon_{ij} \overset{iid}{\sim} \textsf{N}(0, \sigma^2) \\
& \boldsymbol{\gamma}_j \overset{iid}{\sim} \textsf{N}_d(\mathbf{0}_d, \boldsymbol{\Sigma})
\end{aligned}
$$

- Fixed effects contribution $\boldsymbol{\beta}$ is constant across groups
- Random effects are $\boldsymbol{\gamma}_j$ as they vary across groups
- called **mixed effects** as we have both fixed and random effects in the regression model

More General Model

- No reason for the fixed effects and random effect covariates to be the same

$$
\begin{aligned}
Y_{ij} = \boldsymbol{\beta}^T \mathbf{x}_{ij} + \boldsymbol{\gamma}_j^T \mathbf{z}_{ij} + \epsilon_{ij}, \qquad \epsilon_{ij} \overset{\text{iid}}{\sim} \mathsf{N}(0, \sigma^2) \\
\boldsymbol{\gamma}_j \sim \mathsf{N}_p(\mathbf{0}_p, \boldsymbol{\Sigma})
\end{aligned}
$$

- dimension of $\mathbf{x}_{ij}$ $d \times 1$
- dimension of $\mathbf{z}_{ij}$ $p \times 1$
- may or may not be overlapping
- $\mathbf{x}_{ij}$ could include predictors that are constant across all i in group j. (can't estimate if they are in $\mathbf{z}_{ij}$)
- features of school j that

Likelihoods

- Complete Data Likelihood $(\boldsymbol{\beta}, \{\boldsymbol{\gamma}_j\}, \sigma^2, \boldsymbol{\Sigma})$

$$\mathcal{L}(\boldsymbol{\beta}, \{\boldsymbol{\gamma}_j\}, \sigma^2, \boldsymbol{\Sigma}) \propto \prod_j \mathsf{N}(\boldsymbol{\gamma}_j; \mathbf{0}_p, \boldsymbol{\Sigma}) \prod_i \mathsf{N}(y_{ij}; \boldsymbol{\beta}^T \mathbf{x}_{ij} + \boldsymbol{\gamma}_j^T \mathbf{z}_{ij}, \sigma^2)$$

- Marginal likelihood $(\boldsymbol{\beta}, \{\boldsymbol{\gamma}_j\}, \sigma^2, \boldsymbol{\Sigma})$

$$\mathcal{L}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Sigma}) \propto \prod_j \int_{\mathbb{R}^p} \mathsf{N}(\boldsymbol{\gamma}_j; \mathbf{0}_p, \boldsymbol{\Sigma}) \prod_i \mathsf{N}(y_{ij}; \boldsymbol{\beta}^T \mathbf{x}_{ij} + \boldsymbol{\gamma}_j^T \mathbf{z}_{ij}, \sigma^2) \, d\boldsymbol{\gamma}_j$$

- Option A: we can calculate this integral by brute force algebraically
- Option B: (lazy option) We can calculate marginal exploiting properties of Gaussians as sums will be normal - just read off the first two moments!

Marginal Distribution

- Express observed data as vectors for each group j: $(\mathbf{Y}_j, \mathbf{X}_j, \mathbf{Z}_j)$ where $\mathbf{Y}_j$ is $n_j \times 1, \mathbf{X}_j$ is $n_j \times d$ and $\mathbf{Z}_j$ is $n_j \times p$;
- Group Specific Model (1):

$$\begin{aligned} \mathbf{Y}_j &= \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\boldsymbol{\gamma}_j + \boldsymbol{\epsilon}_j, \qquad \boldsymbol{\epsilon}_j \sim \mathsf{N}(\mathbf{0}_{n_j}, \sigma^2\mathbf{I}_{n_j}) \\ \boldsymbol{\gamma}_j &\overset{\text{iid}}{\sim} \mathsf{N}(\mathbf{0}_p, \boldsymbol{\Sigma}) \end{aligned}$$

- Population Mean $\mathsf{E}[\mathbf{Y}_j] = \mathsf{E}[\mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\boldsymbol{\gamma}_j + \boldsymbol{\epsilon}_j] = \mathbf{X}_j\boldsymbol{\beta}$
- Covariance $\mathsf{Var}[\mathbf{Y}_j] = \mathsf{Var}[\mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\boldsymbol{\gamma}_j + \boldsymbol{\epsilon}_j] = \mathbf{Z}_j\boldsymbol{\Sigma}\mathbf{Z}_j^T + \sigma^2\mathbf{I}_{n_j}$
- Group Specific Model (2)

$$\mathbf{Y}_j \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \sigma^2 \overset{\text{ind}}{\sim} \mathsf{N}(\mathbf{X}_j\boldsymbol{\beta}, \mathbf{Z}_j\boldsymbol{\Sigma}\mathbf{Z}_j^T + \sigma^2\mathbf{I}_{n_j})$$

Priors

- Model (1) leads to a simple Gibbs sampler if we use conditional (semi-) conjugate priors on $(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \phi = 1/\sigma^2)$

$$\begin{aligned}
\boldsymbol{\beta} &\sim \textsf{N}(\mu_0, \Psi_0^{-1}) \\
\phi &\sim \textsf{Gamma}(v_0/2, v_o \sigma^2_0/2) \\
\boldsymbol{\Sigma} &\sim \textrm{IW}_p(\eta_0, \boldsymbol{S}_0^{-1})
\end{aligned}$$

MCMC Sampling

- Model (1) leads to a simple Gibbs sampler if we use conditional (semi-) conjugate priors on $(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \phi = 1/\sigma^2)$

$$\begin{aligned}
\boldsymbol{\beta} &\sim \mathsf{N}(\mu_0, \Psi_0^{-1}) \\
\phi &\sim \mathsf{Gamma}(v_0/2, v_o\sigma_0^2/2) \\
\boldsymbol{\Sigma} &\sim \mathrm{IW}_p(\eta_0, \boldsymbol{S}_0^{-1})
\end{aligned}$$

- Model (2) can be challenging to update the variance components! no conjugacy and need to ensure that MH updates maintain the positive-definiteness of $\boldsymbol{\Sigma}$ (can reparameterize)
- Is Gibbs always more efficient?
- No - because the Gibbs sampler can have high autocorrelation in updating the $\{\boldsymbol{\gamma}_j\}$ from their full conditional and then updating $\boldsymbol{\beta}, \sigma^2$ and $\boldsymbol{\Sigma}$ from their full full conditionals given the $\{\boldsymbol{\gamma}_j\}$
- slow mixing

Blocked Gibbs Sampler

- sample $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$'s as a block! (marginal and conditionals) given the others
- update $\boldsymbol{\beta}$ using (2) instead of (1) (marginalization so is independent of $\boldsymbol{\gamma}_j$'s

3 Block Sampler at each iteration

1. Draw $\boldsymbol{\beta}, \boldsymbol{\gamma}_1, \ldots \boldsymbol{\gamma}_J$ as a block given $\phi, \boldsymbol{\Sigma}$ by
   a. Draw $\boldsymbol{\beta} \mid \phi, \boldsymbol{\Sigma}, \mathbf{Y}_1, \ldots \mathbf{Y}_j$ then
   b. Draw $\boldsymbol{\gamma}_j \mid \boldsymbol{\beta}, \phi, \boldsymbol{\Sigma}, \mathbf{Y}_j$ for $j = 1, \ldots J$
2. Draw $\boldsymbol{\Sigma} \mid \boldsymbol{\gamma}_1, \ldots \boldsymbol{\gamma}_J, \boldsymbol{\beta}, \phi, \mathbf{Y}_1, \ldots \mathbf{Y}_j$
3. Draw $\phi \mid \boldsymbol{\beta}, \boldsymbol{\gamma}_1, \ldots \boldsymbol{\gamma}_J, \boldsymbol{\Sigma}, \mathbf{Y}_1, \ldots \mathbf{Y}_j$

- Reduces correlation and improves mixing!

Marginal update for β

$$\mathbf{Y}_j \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}, \sigma^2 \overset{ind}{\sim} \mathsf{N}(\mathbf{X}_j\boldsymbol{\beta}, \mathbf{Z}_j\boldsymbol{\Sigma}\mathbf{Z}_j^T + \sigma^2\mathbf{I}_{n_j})$$
$$\boldsymbol{\beta} \sim \mathsf{N}(\mu_0, \Psi_0^{-1})$$

- Let $\Phi_j = (\mathbf{Z}_j\boldsymbol{\Sigma}\mathbf{Z}_j^T + \sigma^2\mathbf{I}_{n_j})^{-1}$ (precision in model 2)

$$\pi(\boldsymbol{\beta} \mid \boldsymbol{\Sigma}, \sigma^2, \boldsymbol{Y}) \propto |\Psi_0|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\mu_0)^T\Psi_0(\boldsymbol{\beta}-\mu_0)\right\} \cdot \prod_{j=1}^{J} |\Phi_j|^{1/2} \exp\left\{-\frac{1}{2}(\mathbf{Y}_j - \mathbf{X}_j\boldsymbol{\beta})^T\Phi_j(\mathbf{Y}_j - \mathbf{X}_j\boldsymbol{\beta})\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left((\boldsymbol{\beta}-\mu_0)^T\Psi_0(\boldsymbol{\beta}-\mu_0) + \sum_j(\mathbf{Y}_j - \mathbf{X}_j\boldsymbol{\beta})^T\Phi_j(\mathbf{Y}_j - \mathbf{X}_j\boldsymbol{\beta})\right)\right\}$$

Marginal Posterior for $\boldsymbol{\beta}$

$$\pi(\boldsymbol{\beta} \mid \boldsymbol{\Sigma}, \sigma^2, \boldsymbol{Y})$$
$$\propto \exp\left\{-\frac{1}{2}\left((\boldsymbol{\beta} - \mu_0)^T \Psi_0 (\boldsymbol{\beta} - \mu_0) + \sum_j (\mathbf{Y}_j - \mathbf{X}_j \boldsymbol{\beta})^T \Phi_j (\mathbf{Y}_j - \mathbf{X}_j \boldsymbol{\beta})\right)\right\}$$

- precision $\Psi_n = \Psi_0 + \sum_{j=1}^J \mathbf{X}_j^T \Phi_j \mathbf{X}_j$
- mean

$$\mu_n = \left(\Psi_0 + \sum_{j=1}^J \mathbf{X}_j^T \Phi_j \mathbf{X}_j\right)^{-1} \left(\Psi_0 \mu_0 + \sum_{j=1}^J \mathbf{X}_j^T \Phi_j \mathbf{X}_j \hat{\boldsymbol{\beta}}_j\right)$$

- where $\hat{\boldsymbol{\beta}}_j = (\mathbf{X}_j^T \Phi \mathbf{X}_j)^{-1} \mathbf{X}_j^T \Phi_j \mathbf{Y}_j$ is the generalized least squares estimate of $\boldsymbol{\beta}$ for group j

Full conditional for σ^2 or ϕ

$$
\begin{aligned}
\mathbf{Y}_j \mid \boldsymbol{\beta}, \boldsymbol{\gamma}_j, \sigma^2 &\overset{\text{ind}}{\sim} \mathsf{N}(\mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\boldsymbol{\gamma}_j, \sigma^2\mathbf{I}_{n_j}) \\
\boldsymbol{\gamma}_j \mid \boldsymbol{\Sigma} &\overset{\text{iid}}{\sim} \mathsf{N}(\mathbf{0}_d, \boldsymbol{\Sigma}) \\
\boldsymbol{\Sigma} &\sim \mathrm{IW}_p(\eta_0, \boldsymbol{S}_0^{-1}) \\
\boldsymbol{\beta} &\sim \mathsf{N}(\mu_0, \Psi_0^{-1}) \\
\phi &\sim \mathsf{Gamma}(v_0/2, v_o\sigma_0^2/2)
\end{aligned}
$$

$$
\pi(\phi \mid \boldsymbol{\beta}, \{\boldsymbol{\gamma}_j\}\{Y_j\}) \propto \mathsf{Gamma}(\phi; v_0/2, v_o\sigma_0^2/2) \prod_j \mathsf{N}(\mathbf{Y}_j; \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\boldsymbol{\gamma}_j, \phi^{-1}\mathbf{I}_{n_j}))
$$

$$
\phi \mid \{Y_j\}, \boldsymbol{\beta}, \{\boldsymbol{\gamma}_j\} \sim \mathsf{Gamma}\left(\frac{v_0 + \sum_j n_j}{2}, \frac{v_o\sigma_0^2 + \sum_j \|\mathbf{Y}_j - \mathbf{X}_j\boldsymbol{\beta} - \mathbf{Z}_j\boldsymbol{\gamma}_j\|^2}{2}\right)
$$

Conditional posterior for $\boldsymbol{\Sigma}$

$$
\begin{aligned}
\mathbf{Y}_j \mid \boldsymbol{\beta}, \boldsymbol{\gamma}_j, \sigma^2 &\overset{\text{ind}}{\sim} \mathsf{N}(\mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\boldsymbol{\gamma}_j, \sigma^2\mathbf{I}_{n_j}) \\
\boldsymbol{\gamma}_j \mid \boldsymbol{\Sigma} &\overset{\text{iid}}{\sim} \mathsf{N}(\mathbf{0}_d, \boldsymbol{\Sigma}) \\
\boldsymbol{\Sigma} &\sim \mathrm{IW}_p(\eta_0, \boldsymbol{S}_0^{-1}) \\
\boldsymbol{\beta} &\sim \mathsf{N}(\mu_0, \Psi_0^{-1}) \\
\phi &\sim \mathsf{Gamma}(v_0/2, v_o\sigma_0^2/2)
\end{aligned}
$$

- The conditional posterior (full conditional) $\boldsymbol{\Sigma} \mid \boldsymbol{\gamma}, \mathbf{Y}$, is then

$$
\begin{aligned}
\pi(\boldsymbol{\Sigma} \mid \boldsymbol{\gamma}, \boldsymbol{Y}) &\propto \pi(\boldsymbol{\Sigma}) \cdot \pi(\boldsymbol{\gamma} \mid \boldsymbol{\Sigma}) \\
&\propto \underbrace{|\boldsymbol{\Sigma}|^{\frac{-(\eta_0+p+1)}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}_0\boldsymbol{\Sigma}^{-1})\right\}}_{\pi(\boldsymbol{\Sigma})} \cdot \underbrace{\prod_{j=1}^{J} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\boldsymbol{\gamma}_j^T\boldsymbol{\Sigma}^{-1}\gamma_j\right]\right\}}_{\pi(\boldsymbol{\gamma}|\boldsymbol{\Sigma})}
\end{aligned}
$$

Posterior Continued

- Full conditional $\boldsymbol{\Sigma} \mid \{\gamma_j\}, \boldsymbol{Y} \sim \mathrm{IW}_p\left(\eta_0 + J, (\boldsymbol{S}_0 + \sum_{j=1}^J \gamma_j \gamma_j^T)^{-1}\right)$
- Work

$$\begin{aligned}\pi(\boldsymbol{\Sigma} \mid \boldsymbol{\gamma}, \boldsymbol{Y}) &\propto \pi(\boldsymbol{\Sigma}) \cdot \pi(\boldsymbol{\gamma} \mid \boldsymbol{\Sigma}) \\ &\propto |\boldsymbol{\Sigma}|^{\frac{-(\eta_0+p+1)}{2}} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}_0 \boldsymbol{\Sigma}^{-1})\right\} \cdot \prod_{j=1}^{J} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\boldsymbol{\gamma}_j^T \boldsymbol{\Sigma}^{-1} \gamma_j\right]\right\}\end{aligned}$$

Full conditional for $\{\gamma_j\}$

$$\begin{aligned}
\mathbf{Y}_j \mid \boldsymbol{\beta}, \boldsymbol{\gamma}_j, \sigma^2 &\overset{\text{ind}}{\sim} \mathsf{N}(\mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\boldsymbol{\gamma}_j, \sigma^2\mathbf{I}_{n_j}) \\
\boldsymbol{\gamma}_j \mid \boldsymbol{\Sigma} &\overset{iid}{\sim} \mathsf{N}(\mathbf{0}_d, \boldsymbol{\Sigma}) \\
\boldsymbol{\Sigma} &\sim \text{IW}_p(\eta_0, \boldsymbol{S}_0^{-1}) \\
\boldsymbol{\beta} &\sim \mathsf{N}(\mu_0, \Psi_0^{-1}) \\
\phi &\sim \mathsf{Gamma}(v_0/2, v_o\sigma_0^2/2)
\end{aligned}$$

$$\pi(\boldsymbol{\gamma}_j \mid \boldsymbol{\beta}, \phi, \boldsymbol{\Sigma}) \propto \mathsf{N}(\boldsymbol{\gamma}_j; 0, \boldsymbol{\Sigma}) \prod_j \mathsf{N}(\mathbf{Y}_j; \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\boldsymbol{\gamma}_j, \phi^{-1}\mathbf{I}_{n_j}))$$

- work out as HW

Other Questions

- How do you decide what is a random effect or fixed effect?
- Design structure is often important
- Other priors ?
- How would you implement MH in Model 2? (other sampling methods?)
- What if the means are not normal? Extensions to Generalized linear models
- more examples in