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:

$$egin{aligned} Y_{ij} &= oldsymbol{ heta}_j^T \mathbf{x}_{ij} + \epsilon_{ij} \ \epsilon_{ij} &\stackrel{ ext{iid}}{\sim} \mathsf{N}(0,\sigma^2) \end{aligned}$$

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

$$oldsymbol{ heta}_j \stackrel{ ext{iid}}{\sim} \mathsf{N}(oldsymbol{eta}, oldsymbol{\Sigma})$$

- β , Σ 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

- ullet We can write $oldsymbol{ heta}=oldsymbol{eta}+oldsymbol{\gamma}_j$ with $oldsymbol{\gamma}_j\stackrel{ ext{iid}}{\sim} \mathsf{N}(oldsymbol{0},oldsymbol{\Sigma})$
- Substituting, we can rewrite model

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

- Fixed effects contribution $oldsymbol{eta}$ is constant across groups
- Random effects are $oldsymbol{\gamma}_j$ 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

$$egin{aligned} Y_{ij} &= oldsymbol{eta}^T \mathbf{x}_{ij} + oldsymbol{\gamma}_j^T \mathbf{z}_{ij} + \epsilon_{ij}, \qquad \epsilon_{ij} \overset{ ext{iid}}{\sim} \mathsf{N}(0, \sigma^2) \ oldsymbol{\gamma}_j \sim &\mathsf{N}_p(\mathbf{0}_p, oldsymbol{\Sigma}) \end{aligned}$$

- ullet dimension of $\mathbf{x}_{ij}\,d imes 1$
- ullet dimension of $\mathbf{z}_{ij}\,p imes 1$
- 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})
- ullet features of school j that

4

Likelihoods

• Complete Data Likelihood $(oldsymbol{eta}, \{oldsymbol{\gamma}_j\}, \sigma^2, oldsymbol{\Sigma})$

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

• Marginal likelihood $(oldsymbol{eta}, \{oldsymbol{\gamma}_j\}, \sigma^2, oldsymbol{\Sigma})$

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

- 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):

$$egin{aligned} \mathbf{Y}_j &= \mathbf{X}_j oldsymbol{eta} + \mathbf{Z}_j oldsymbol{\gamma}_j + oldsymbol{\epsilon}_j, & oldsymbol{\epsilon}_j \sim \mathsf{N}(\mathbf{0}_{n_j}, \sigma^2 \mathbf{I}_{n_j}) \ oldsymbol{\gamma}_j &\stackrel{ ext{iid}}{\sim} \mathsf{N}(\mathbf{0}_p, oldsymbol{\Sigma}) \end{aligned}$$

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

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

Priors

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

$$oldsymbol{eta} \sim \mathsf{N}(\mu_0, \Psi_0^{-1}) \ \phi \sim \mathsf{Gamma}(v_0/2, v_o \sigma_0^2/2) \ oldsymbol{\Sigma} \sim \mathrm{IW}_p(\eta_0, oldsymbol{S}_0^{-1})$$

MCMC Sampling

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

$$oldsymbol{eta} \sim \mathsf{N}(\mu_0, \Psi_0^{-1}) \ \phi \sim \mathsf{Gamma}(v_0/2, v_o \sigma_0^2/2) \ oldsymbol{\Sigma} \sim \mathrm{IW}_p(\eta_0, oldsymbol{S}_0^{-1})$$

- Model (2) can be challenging to update the variance components! no conjugacy and need to ensure that MH updates maintain the positive-definiteness of Σ (can reparameterize)
- Is Gibbs always more efficient?
- No because the Gibbs sampler can have high autocorrelation in updating the $\{\gamma_j\}$ 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 $oldsymbol{eta}$ and $oldsymbol{\gamma}$'s as a block! (marginal and conditionals) given the others
- update $oldsymbol{eta}$ using (2) instead of (1) (marginalization so is independent of $oldsymbol{\gamma}_j$'s

```
1. Draw oldsymbol{eta}, \gamma_1, \ldots \gamma_J as a block given \phi, \Sigma by a. Draw oldsymbol{eta} \mid \phi, \Sigma, \mathbf{Y}_1, \ldots \mathbf{Y}_j then b. Draw \gamma_j \mid oldsymbol{eta}, \phi, \Sigma, \mathbf{Y}_j for j=1,\ldots J
2. Draw \Sigma \mid \gamma_1, \ldots \gamma_J, oldsymbol{eta}, \phi, \mathbf{Y}_1, \ldots \mathbf{Y}_j
3. Draw \phi \mid oldsymbol{eta}, \gamma_1, \ldots \gamma_J, \Sigma, \mathbf{Y}_1, \ldots \mathbf{Y}_j
```

• Reduces correlation and improves mixing!

Marginal update for $oldsymbol{eta}$

$$egin{aligned} \mathbf{Y}_j \mid oldsymbol{eta}, oldsymbol{\Sigma}, \sigma^2 \overset{ind}{\sim} \mathsf{N}(\mathbf{X}_joldsymbol{eta}, \mathbf{Z}_joldsymbol{\Sigma}\mathbf{Z}_j^T + \sigma^2\mathbf{I}_{n_j}) \ oldsymbol{eta} \sim \mathsf{N}(\mu_0, \Psi_0^{-1}) \end{aligned}$$

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

$$egin{aligned} \pi(oldsymbol{eta} \mid oldsymbol{\Sigma}, \sigma^2, oldsymbol{Y}) & \propto |\Psi_0|^{1/2} \exp\left\{-rac{1}{2}(oldsymbol{eta} - \mu_0)^T \Psi_0(oldsymbol{eta} - \mu_0)
ight\} \cdot \ & \prod_{j=1}^J |\Phi_j|^{1/2} \exp\left\{-rac{1}{2}(\mathbf{Y}_j - \mathbf{X}_j oldsymbol{eta})^T \Phi_j(\mathbf{Y}_j - \mathbf{X}_j oldsymbol{eta})
ight\} \end{aligned}$$

$$ext{thing} \propto \exp \left\{ -rac{1}{2} \Bigg((oldsymbol{eta} - \mu_0)^T \Psi_0 (oldsymbol{eta} - \mu_0) + \sum_j (\mathbf{Y}_j - \mathbf{X}_j oldsymbol{eta})^T \Phi_j (\mathbf{Y}_j - \mathbf{X}_j oldsymbol{eta}) \Bigg\}
ight.$$

Marginal Posterior for $oldsymbol{eta}$

$$egin{aligned} \pi(oldsymbol{eta} \mid oldsymbol{\Sigma}, \sigma^2, oldsymbol{Y}) \ &\propto \exp\left\{-rac{1}{2}igg((oldsymbol{eta} - \mu_0)^T \Psi_0(oldsymbol{eta} - \mu_0) + \sum_j (\mathbf{Y}_j - \mathbf{X}_j oldsymbol{eta})^T \Phi_j(\mathbf{Y}_j - \mathbf{X}_j oldsymbol{eta})
ight)
ight\} \end{aligned}$$

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

$$oxed{\mu_n = \left(\Psi_0 + \sum_{j=1}^J \mathbf{X}_j^T \Phi_j \mathbf{X}_j
ight)^{-1} \left(\Psi_0 \mu_0 + \sum_{j=1}^J \mathbf{X}_j^T \Phi_j \mathbf{X}_j \hat{oldsymbol{eta}}_j
ight)}$$

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

Full conditional for σ^2 or ϕ

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

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

$$\phi \mid \{Y_j\}, oldsymbol{eta}, \{oldsymbol{\gamma}_j\} \sim \mathsf{Gamma}\left(rac{v_0 + \sum_j n_j}{2}, rac{v_o \sigma_0^2 + \sum_j \|\mathbf{Y}_j - \mathbf{X}_j oldsymbol{eta} - \mathbf{Z}_j oldsymbol{\gamma}_j\|^2}{2}
ight)$$

Conditional posterior for Σ

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

ullet The conditional posterior (full conditional) $oldsymbol{\Sigma} \mid oldsymbol{\gamma}, oldsymbol{Y}$, is then

$$\pi(oldsymbol{\Sigma} \mid oldsymbol{\gamma}, oldsymbol{Y}) \propto \pi(oldsymbol{\Sigma}) \cdot \pi(oldsymbol{\gamma} \mid oldsymbol{\Sigma}) \ \propto |oldsymbol{\Sigma}|^{rac{-(\eta_0 + p + 1)}{2}} \exp\left\{-rac{1}{2} ext{tr}(oldsymbol{S}_0 oldsymbol{\Sigma}^{-1})
ight\} \cdot \underbrace{\prod_{j=1}^{J} |oldsymbol{\Sigma}|^{-rac{1}{2}} \exp\left\{-rac{1}{2} \left[oldsymbol{\gamma}_j^T oldsymbol{\Sigma}^{-1} oldsymbol{\gamma}_j
ight]
ight\}}_{\pi(oldsymbol{\gamma} \mid oldsymbol{\Sigma})}$$

Posterior Continued

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

$$egin{aligned} \pi(oldsymbol{\Sigma} \mid oldsymbol{\gamma}, oldsymbol{Y}) & \propto |oldsymbol{\Sigma}| \stackrel{-(\eta_0+p+1)}{2} \exp\left\{-rac{1}{2} ext{tr}(oldsymbol{S}_0 oldsymbol{\Sigma}^{-1})
ight\} \cdot \prod_{j=1}^J |oldsymbol{\Sigma}|^{-rac{1}{2}} \exp\left\{-rac{1}{2} igl[oldsymbol{\gamma}_j^T oldsymbol{\Sigma}^{-1} oldsymbol{\gamma}_jigr]
ight\} \end{aligned}$$

Full conditional for $\{\gamma_j\}$

$$egin{aligned} \mathbf{Y}_j \mid oldsymbol{eta}, oldsymbol{\gamma}_j, \sigma^2 & \overset{ ext{ind}}{\sim} \mathsf{N}(\mathbf{X}_j oldsymbol{eta} + \mathbf{Z}_j oldsymbol{\gamma}_j, \sigma^2 \mathbf{I}_{n_j}) \ oldsymbol{\gamma}_j \mid oldsymbol{\Sigma} & \overset{iid}{\sim} \mathsf{N}(\mathbf{0}_d, oldsymbol{\Sigma}) \ oldsymbol{\Sigma} & \sim \mathrm{IW}_p(\eta_0, oldsymbol{S}_0^{-1}) \ oldsymbol{eta} & \sim \mathsf{N}(\mu_0, oldsymbol{\Psi}_0^{-1}) \ oldsymbol{\phi} & \sim \mathsf{Gamma}(v_0/2, v_o \sigma_0^2/2) \ \pi(oldsymbol{\gamma}_j \mid oldsymbol{eta}, oldsymbol{\phi}, oldsymbol{\Sigma}) \propto \mathsf{N}(oldsymbol{\gamma}_j; 0, oldsymbol{\Sigma}) \prod_j \mathsf{N}(\mathbf{Y}_j; \mathbf{X}_j oldsymbol{eta} + \mathbf{Z}_j oldsymbol{\gamma}_j, oldsymbol{\phi}^{-1} \mathbf{I}_{n_j})) \end{aligned}$$

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