Linear Mixed Effects Models

STA721: Lecture 24

Merlise Clyde
Duke University

Random Effects Regression

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

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

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

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

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

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

- Fixed effects contribution $\boldsymbol{\beta}$ is constant across groups
- Random effects are $\boldsymbol{\alpha}_j$ 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

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

- dimension of $\mathbf{x}_{ij}$ $p \times 1$
- dimension of $\mathbf{z}_{ij}$ $q \times 1$
- 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 vary across schools but are constant within a school

Marginal Distribution of Data

GLS Estimation

Marginal Model

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

- Define covariance of $\mathbf{Y}$ to be $\mathbf{V} = \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}^T + \sigma^2\mathbf{I}_n$
- Use GLS conditional on $\boldsymbol{\Sigma}, \sigma^2$ to estimate $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}$$

- since $\mathbf{V}$ has unknown parameters, typical practice (non-Bayes) is to use an estimate of $\mathbf{V}$, and replace $\mathbf{V}$ by $\hat{\mathbf{V}}$. (MLE, Methods of Moments, REML)
- frequentist random effects models arose from analysis of variance models so generally some simplification in $\boldsymbol{\Sigma}$!

One Way Anova Random Effects Model

- Consider Balance data so that $n_1 = n_2 = \cdots = n_J = r$ and $n = rJ$
- design matrix $\mathbf{X} = \mathbf{1}_n$
- covariance for random effects is $\boldsymbol{\Sigma} = \sigma^2_\alpha \mathbf{I}_J$
- matrix $\mathbf{Z}$ is $n \times J$

$$\mathbf{Z} = \begin{pmatrix} \mathbf{1}_r & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_r & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_r \end{pmatrix}$$

- Covariance

$$\mathbf{V} = \sigma^2 \mathbf{I}_n + \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}^T = \sigma^2 \mathbf{I}_n + \sigma^2_\alpha \mathbf{Z}\mathbf{Z}^T = \sigma^2 \mathbf{I}_n + \sigma^2_\alpha r \mathbf{P}_{\mathbf{Z}}$$

MLEs for One-Way Random Effects Model

- Model $\mathbf{Y} = \mathbf{1}_n\beta + \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \sim \mathsf{N}(\mathbf{0}, \mathbf{V})$
- Since $C(\mathbf{V}\mathbf{X}) \subset C(\mathbf{X})$, the GLS of $\boldsymbol{\beta}$ is the same as the OLS of β in this case

$$\hat{\beta} = \bar{y}_{..} = \sum_{j=1}^{J} \sum_{i=1}^{r} y_{ij}/n$$

- We need the determinant and inverse of $\mathbf{V}$ to get the MLEs for σ^2 and σ^2_α
- note that $\mathbf{V}$ is block diagonal with blocks $\sigma^2 \mathbf{I}_r + \sigma^2_\alpha r \mathbf{P}_{\mathbf{1}_r}$ (use eigenvalues based on svd of $\mathbf{P}_{\mathbf{1}_r}$ and $\mathbf{I}_r$)
- determinant of $\mathbf{V}$ is the product of determinants of blocks ${\sigma^2}^n(1 + r\sigma^2_\alpha/\sigma^2)^J$
- find inverse of $\mathbf{V}$ via Woodbury identity (or svd of projections/eigenvalues)

$$\mathbf{V}^{-1} = \frac{1}{\sigma^2}\left(\mathbf{I}_n - \frac{r\sigma^2_\alpha}{\sigma^2 + r\sigma^2_\alpha}\mathbf{P}_{\mathbf{Z}}\right)$$

Log likelihood

- plug in $\hat{\beta}$

$$\begin{aligned}
\log L(\sigma^2, \sigma^2_\alpha) &= -\frac{1}{2}\log|V| - \frac{1}{2}(\mathbf{Y} - \mathbf{1}_n\bar{y})^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{1}_n\bar{y}) \\
&= -\frac{1}{2}\log|V| - \frac{1}{2}\mathbf{Y}^T(\mathbf{I} - \mathbf{P}_{\mathbf{1}_n})\mathbf{V}^{-1}(\mathbf{I} - \mathbf{P}_{\mathbf{1}_n})\mathbf{Y} \\
&= -\frac{J(r-1)}{2}\log\sigma^2 - \frac{J}{2}\log(\sigma^2 + r\sigma^2_\alpha) \\
&\quad - \frac{1}{2\sigma^2}\left(\mathbf{Y}^T(\mathbf{I} - \mathbf{P}_{\mathbf{1}_n})(\mathbf{I}_n - \frac{r\sigma^2_\alpha}{\sigma^2 + r\sigma^2_\alpha}\mathbf{P}_\mathbf{Z})(\mathbf{I} - \mathbf{P}_{\mathbf{1}_n})\mathbf{Y}\right) \\
&= -\frac{J(r-1)}{2}\log\sigma^2 - \frac{J}{2}\log(\sigma^2 + r\sigma^2_\alpha) \\
&\quad - \frac{1}{2\sigma^2}\left(\mathbf{Y}^T(\mathbf{I} - \mathbf{P}_{\mathbf{1}_n})\left(\frac{\sigma^2\mathbf{I}_n + r\sigma^2_\alpha(\mathbf{I}_n - \mathbf{P}_\mathbf{Z})}{\sigma^2 + r\sigma^2_\alpha}\right)(\mathbf{I} - \mathbf{P}_{\mathbf{1}_n})\mathbf{Y}\right)
\end{aligned}$$

MLEs

- Simplify using Properties of Projections; ie $(\mathbf{I}_n - \mathbf{P}_{\mathbf{1}_n})(\mathbf{I}_n - \mathbf{P}_{\mathbf{Z}})$ to rewrite in terms of familiar $\textsf{SSE} = \mathbf{Y}^T(\mathbf{I} - \mathbf{P}_{\mathbf{Z}})\mathbf{Y}$ and $\textsf{SST} = \mathbf{Y}^T(\mathbf{P}_{\mathbf{Z}} - \mathbf{P}_{\mathbf{1}_n})\mathbf{Y}$ based on the fixed effects one-way anova model
- take derivatives and solve for MLEs (some alegebra involved!)
- MLE of σ^2 is $\hat{\sigma}^2 = \textsf{MSE} = \textsf{SSE}/(n - J)$
- MLE of σ^2_α is

$$\hat{\sigma}^2_\alpha = \frac{\frac{\textsf{SST}}{J} - \textsf{MSE}}{n}$$

but this is true only if $\textsf{MSE} < \textsf{SST}/J$ otherwise the mode is on the boundary and $\hat{\sigma}^2_\alpha = 0$

Comments

For the One-Way model (and HW) we can find MLEs in closed form - but several approaches to simplify the algebra

- steps outlined here (via the stacked approach - more general)
- treating the response as a matrix and using the matrix normal distribution with the mean function and covariance via Kronecker transformations (lab)
  - extends to other balanced ANOVA models
- simplify the problem based on summary statistics - i.e. the distributions in terms of $\textsf{SSE}$. (Gamma) and the sample means (Normal) and integrate out random effects (Approach in Box & Hill for Bayesian solution)
  - easiest imho for the one-way model

For more general problems we may need iterative methods to find MLEs (alternating between conditional MLE of $\boldsymbol{\beta}$ and MLE of $\boldsymbol{\Sigma}$) (Gauss-Siedel optimization)

Best Linear Prediction

Given a linear model with $\mathsf{E}[Y^*] = \mathbf{X}\boldsymbol{\beta}$ with or without correlation structure, we can *predict* a new observation Y^* at $\mathbf{x}$ as $\mathbf{x}^T\hat{\boldsymbol{\beta}}$ where $\hat{\boldsymbol{\beta}}$ is the OLS or GLS of $\boldsymbol{\beta}$.

- but if Y^* and $\mathbf{Y}$ are correlated we can do better!

Theorem: Christensen 6.3.4; Sec 12.2

Let $\mathbf{Y}$ and Y^* be random variables with the following moments

$$\mathsf{E}[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta} \qquad \mathsf{E}[\mathbf{Y}^*] = \mathbf{x}^T\boldsymbol{\beta}$$
$$\mathsf{Var}[\mathbf{Y}] = \mathbf{V} \qquad \mathsf{Cov}[\mathbf{Y}, Y^*] = \psi$$

Then the best linear predictor of Y^* given $\mathbf{Y}$ is

$$\mathsf{E}[Y^* \mid \mathbf{Y}] = \mathbf{x}^T\hat{\boldsymbol{\beta}} + \delta(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

where $\delta = \mathbf{V}^{-1}\psi$

Best Linear Unbiased Prediction

To go from BLPs to BLUPs we need to estimate the unknown parameters in the model $\boldsymbol{\beta}$

Mixed Model Equations via Bayes Rule

The *mixed model equations* are the normal equations for the mixed effects model and provide both BLUEs and BLUPs

- Consider the model

$$
\begin{aligned}
\mathbf{Y} &\sim \mathsf{N}(\mathbf{W}\boldsymbol{\theta}, \sigma^2 \mathbf{I}_n) \\
\mathbf{W} &= [\mathbf{X}, \mathbf{Z}] \\
\boldsymbol{\theta} &= [\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T]
\end{aligned}
$$

- estimate $\boldsymbol{\theta}$ using Bayes with the prior $\boldsymbol{\theta} \sim \mathsf{N}(\mathbf{0}, \Omega)$ where $\Omega = \begin{pmatrix} \mathbf{I}_p/\kappa & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \end{pmatrix}$

- posterior mean of $\boldsymbol{\theta}$

$$
\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{X}/\sigma^2 + \kappa \mathbf{I}_p & \mathbf{X}^T\mathbf{Z}/\sigma^2 \\ \mathbf{Z}^T\mathbf{X}/\sigma^2 & \mathbf{Z}^T\mathbf{Z} + \boldsymbol{\Sigma}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}^T\mathbf{Y}/\sigma^2 \\ \mathbf{Z}^T\mathbf{Y}/\sigma^2 \end{pmatrix}
$$

BLUEs and BLUPs via Bayes Rule

- take the limiting prior with $\kappa \to 0$ and $\mathbf{\Sigma} \to \mathbf{0}$ to get the mixed model equations
- The BLUE of $\boldsymbol{\beta}$ and BLUP of $\boldsymbol{\alpha}$ satisfy the limiting form of the posterior mean of $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\alpha}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{X}/\sigma^2 & \mathbf{X}^T\mathbf{Z}/\sigma^2 \\ \mathbf{Z}^T\mathbf{X}/\sigma^2 & \mathbf{Z}^T\mathbf{Z} + \mathbf{\Sigma}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}^T\mathbf{Y}/\sigma^2 \\ \mathbf{Z}^T\mathbf{Y}/\sigma^2 \end{pmatrix}$$

- see Christensen Sec 12.3 for details
- the mixed model equations have computational advantages over the usual GLS espression for $\boldsymbol{\beta}$ as it avoids inverting $\mathbf{V}$ $n \times n$ and instead we are inverting $p + q$ matrix!
- related to spatial kriging and Gaussian Process Regression

Other Questions

- How do you decide what is a random effect or fixed effect?
- Design structure is often important
- What if the means are not normal? Extensions to Generalized Linear Models
- what if random effects are not normal? (Mixtures of Normals, Bayes…)
- more examples in Case Studies next semester!
- for more in depth treatment take STA 610