STAT 574 Linear and Nonlinear Mixed Models

Lecture 5: Meta-analysis Models

Chencheng Cai

Washington State University

Meta-analysis Model

- Observation: A common treatment effect that synthesizes the results of several studies.
- Model: Build a super-model (the model of models) over several different studies using random-effect models.
- Methodology: Pulling results from studies on similar objectives to build a powerful overall model.
- Key assumptions:
  - Estimations from each study are assumed to be Gaussian (because of CLT)
  - Random effects may be heavy-tailed.

Simple Meta-analysis Model

- n studies estimate a common parameter of interest β by y_i with its variance σ_i^2.
- How to generate an overall estimate for β?

Simple Meta-analysis Model

- n studies estimate a common parameter of interest β by y_i with its variance σ_i^2.
- How to generate an overall estimate for β?
- **Naive Average**:
  - $\hat{\beta} = n^{-1} \sum_{i=1}^{n} y_i$
  - Problem: variance not optimal.
- **Weighted Average**:
  - $\hat{\beta} = \sum_{i=1}^{n} w_i y_i$ with $w_i \propto 1/\sigma_i^2$ and $\sum_{i=1}^{n} w_i = 1$.
  - Problem: result dominated by studies with small σ_i^2's.

Simple Meta-analysis Model

- n studies estimate a common parameter of interest β by y_i with its variance σ_i^2.
- Consider a random-effect approach:

$$y_i = \beta + b_i + \epsilon_i$$

  where b_i is a random effect with **unknown** variance σ^2 and ϵ_i is an error term with **known** variance σ_i^2
- or, equivalently,

$$y_i \sim \mathcal{N}(\beta, \sigma^2 + \sigma_i^2)$$

- **Is this a LME model?**

Simple Meta-analysis Model

- n studies estimate a common parameter of interest β by y_i with its variance σ_i^2.
- Consider a random-effect approach:

$$y_i = \beta + b_i + \epsilon_i$$

  where b_i is a random effect with **unknown** variance σ^2 and ϵ_i is an error term with **known** variance σ_i^2
- or, equivalently,

$$y_i \sim \mathcal{N}(\beta, \sigma^2 + \sigma_i^2)$$

- **Is this a LME model?** Not exactly because σ_i's are known.

Simple Meta-analysis Model

$$y_i = \beta + b_i + \epsilon_i$$

- y_i: study-specific treatments
- β: common treatment effect
- σ^2: heterogeneity parameter (variance of the random effect)

Remark:

- We do not specify how each y_i and σ_i^2 are obtained.
- The model is built on many other individual studies.

Simple Meta-analysis Model

If σ^2 is **known**, the best linear unbiased estimator (BLUE) for β is

$$\hat{\beta} = \frac{\sum_{i=1}^{n} y_i(\sigma^2 + \sigma_i^2)^{-1}}{\sum_{i=1}^{n}(\sigma^2 + \sigma_i^2)^{-1}}.$$

Why? Gauss-Markov theorem.

- If $\sigma^2 = 0$, $\hat{\beta}$ is the weighted average.
- If $\sigma^2 = \infty$, $\hat{\beta}$ is the naive average.

Simple Meta-analysis Model — Estimation

Write it as an intermedia step in LME model:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i \Longleftarrow y_i = \beta + b_i + \epsilon_i$$

when $\boldsymbol{X}_i = \boldsymbol{Z}_i = 1$ and when σ_i^2 are known.
Recall what we have from random effects estimation:

$$\min_{\boldsymbol{\beta},\boldsymbol{b}_1,\ldots,\boldsymbol{b}_N} \sum_{i=1}^{N} \left(\|\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b}_i\|^2 + \boldsymbol{b}_i^T\boldsymbol{D}^{-1}\boldsymbol{b}_i \right)$$

Now, it turns to

$$\min_{\beta,b_1,\ldots,b_n} \sum_{i=1}^{n} \left(\frac{(y_i - \beta - b_i)^2}{\sigma_i^2} + \frac{b_i^2}{\sigma^2} \right)$$

Simple Meta-analysis Model — Estimation

$$\min_{\beta, b_1, \ldots, b_n} \sum_{i=1}^{n} \left(\frac{(y_i - \beta - b_i)^2}{\sigma_i^2} + \frac{b_i^2}{\sigma^2} \right)$$

When we have estimates $\hat{\beta}$ and $\hat{\sigma}^2$, the solution is

$$\hat{b}_i = \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \sigma_i^2}(y_i - \hat{\beta})$$

Verify:

- When $\hat{\sigma}^2 = 0$, $\hat{b}_i = 0$.
- When $\hat{\sigma}^2 = \infty$, $\hat{b}_i = y_i - \hat{\beta}$.

Simple Meta-analysis Model — Estimation

To estimate β and σ^2, under normal assumption for y_i's, the log-likelihood function is

$$\ell(\beta, \sigma^2) = -\frac{1}{2} \sum_{i=1}^{n} \left[\log(\sigma^2 + \sigma_i^2) + \frac{(y_i - \beta)^2}{\sigma^2 + \sigma_i^2} \right]$$

Solution:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} y_i (\hat{\sigma}^2 + \sigma_i^2)^{-1}}{\sum_{i=1}^{n} (\hat{\sigma}^2 + \sigma_i^2)^{-1}}.$$

When $\sigma_i^2 = \sigma_1^2$ for all i,

$$\hat{\beta} = \bar{y}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2 - \sigma_1^2$$

Simple Meta-analysis Model — Estimation

Other estimators for variance:

- Variance least squares:

$$\min_{\sigma^2} \sum_{i=1}^{n} \left[(y_i - \hat{\beta})^2 - \sigma_i^2 - \sigma^2 \right]^2$$

  The solution is:

$$\hat{\sigma}^2_{VLS} = \frac{1}{n} \sum_{i=1}^{n} \left((y_i - \hat{\beta})^2 - \sigma_i^2 \right)$$

- Method of Moments:

$$\bar{y} = \beta$$

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 = \frac{n-1}{n} (n\sigma^2 + \sum_{i=1}^{n} \sigma_i^2)$$

Simple Meta-analysis Model — Testing

Test $H_0 : \sigma = 0$ (no random effect)

- under H_0:

$$Q = \sum_{i=1}^{n} \sigma_i^{-2} (y_i - \hat{\beta}_0)^2 \sim \chi^2_{n-1},$$

- $\hat{\beta}_0$ is the weighted mean.
- Same as the F-test we mentioned in LME.

Simple Meta-analysis Model — Testing

Test $H_0 : \sigma = 0$ (no random effect)

- under H_0:

$$Q = \sum_{i=1}^{n} \sigma_i^{-2}(y_i - \hat{\beta}_0)^2 \sim \chi^2_{n-1},$$

- $\hat{\beta}_0$ is the weighted mean.
- Same as the F-test we mentioned in LME.

Test $H_0 : \beta = 0$ (zero populational mean)

- Wald CI
- PL CI

Meta-analysis Model with Covariates

Consider more covariates in the meta-analysis model:

$$y_i = \boldsymbol{\beta}^T \boldsymbol{x}_i + b_i + \epsilon_i, i = 1, \ldots, n,$$

where $b_i \sim \mathcal{N}(0, \sigma_i)$.
When ϵ_i is Gaussian, we have

$$y_i \sim \mathcal{N}(\boldsymbol{\beta}^T \boldsymbol{x}_i, \sigma^2 + \sigma_i^2)$$

- Example: a multi-center study with different characteristics for the centers.

Meta-analysis Model with Covariates

When σ^2 is known, we have

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} (\sigma^2 + \sigma_i^2)^{-1} \boldsymbol{x}_i \boldsymbol{x}_i^T \right)^{-1} \left(\sum_{i=1}^{n} (\sigma^2 + \sigma_i^2)^{-1} \boldsymbol{x}_i y_i \right)$$

- When $\sigma^2 = 0$, we have

$$\hat{\boldsymbol{\beta}}_0 = \left(\sum_{i=1}^{n} \sigma_i^{-2} \boldsymbol{x}_i \boldsymbol{x}_i^T \right)^{-1} \left(\sum_{i=1}^{n} \sigma_i^{-2} \boldsymbol{x}_i y_i \right)$$

- When $\sigma^2 \to \infty$, we have

$$\hat{\boldsymbol{\beta}}_{OLS} = \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^T \right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{x}_i y_i \right)$$

Meta-analysis Model with Covariates — MLE

The MLE maximizes the loglikelihood function:

$$\ell(\boldsymbol{\beta}, \sigma^2) = -\frac{1}{2} \sum_{i=1}^{n} \left[\log(\sigma^2 + \sigma_i^2) + \frac{(y_i - \boldsymbol{\beta}^T \boldsymbol{x}_i)^2}{\sigma_i^2 + \sigma^2} \right]$$

When $\sigma_i^2 = \sigma_1^2$ for all i, we have the closed form solution:

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{OLS}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{\boldsymbol{\beta}}^T \boldsymbol{x}_i)^2 - \sigma_1^2.$$

Meta-analysis Model with Covariates — Testing

Consider test $H_0 : \sigma^2 = 0$.
The test statistics:

$$Q = \sum_{i=1}^{n} \sigma_i^{-2} (y_i - \hat{\boldsymbol{\beta}}_0^T \boldsymbol{x}_i)^2$$

The distribution:

$$Q \sim \chi^2_{n-m},$$

where m is the number of covariates (length of $\boldsymbol{\beta}$).

Multivariate Meta-analysis Model with Covariates

When each study estimates a vector parameter (e.g. primary effect and secondary effect), we observe a vector from each study with known covariance.

$$\boldsymbol{y}_i = \boldsymbol{\beta}^T \boldsymbol{x}_i + \boldsymbol{b}_i + \boldsymbol{\epsilon}_i,$$

where

$$\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{C}_i)$$

with **unknown** $\boldsymbol{D}$ and known $\boldsymbol{C}_i$.

What if different studies report different numbers of outcomes?

Multivariate Meta-analysis Model

- Suppose we are interested in three parameters (y, s, t): the primary, secondary, and tertiary outcomes.
- Now we have three different studies that report different outcomes:

$$\boldsymbol{y}_1 = \begin{bmatrix} y_1 \end{bmatrix}, \quad \boldsymbol{y}_2 = \begin{bmatrix} y_2 \\ s_2 \end{bmatrix}, \quad \boldsymbol{y}_3 = \begin{bmatrix} y_3 \\ s_3 \\ t_3 \end{bmatrix}$$

Multivariate Meta-analysis Model

- Suppose we are interested in three parameters (y, s, t): the primary, secondary, and tertiary outcomes.
- Now we have three different studies that report different outcomes:

$$\boldsymbol{y}_1 = \begin{bmatrix} y_1 \end{bmatrix}, \quad \boldsymbol{y}_2 = \begin{bmatrix} y_2 \\ s_2 \end{bmatrix}, \quad \boldsymbol{y}_3 = \begin{bmatrix} y_3 \\ s_3 \\ t_3 \end{bmatrix}$$

- **Does it fit into our model?**

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

Multivariate Meta-analysis Model

- Suppose we are interested in three parameters (y, s, t): the primary, secondary, and tertiary outcomes.
- Now we have three different studies that report different outcomes:

$$\boldsymbol{y}_1 = \begin{bmatrix} y_1 \end{bmatrix}, \quad \boldsymbol{y}_2 = \begin{bmatrix} y_2 \\ s_2 \end{bmatrix}, \quad \boldsymbol{y}_3 = \begin{bmatrix} y_3 \\ s_3 \\ t_3 \end{bmatrix}$$

- **Does it fit into our model?**

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

- If $\boldsymbol{y}_i$ has p_i elements, we can set $\boldsymbol{X}_i$ of dimension $p_i \times m$, and $\boldsymbol{b}_i$ of size $p_i \times 1$.
- What about covariance of $\boldsymbol{b}_i$?

Multivariate Meta-analysis Model

- Define three full observation vectors:

$$\boldsymbol{y}_1' = \begin{bmatrix} y_1 \\ s_1 \\ t_1 \end{bmatrix}, \quad \boldsymbol{y}_2' = \begin{bmatrix} y_2 \\ s_2 \\ t_2 \end{bmatrix}, \quad \boldsymbol{y}_3' = \begin{bmatrix} y_3 \\ s_3 \\ t_3 \end{bmatrix}$$

- Define three truncation matrices:

$$\boldsymbol{H}_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{H}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \boldsymbol{H}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Then we have

$$\boldsymbol{y}_i = \boldsymbol{H}_i \boldsymbol{y}_i', \quad \text{for } i = 1, 2, 3$$

Multivariate Meta-analysis Model

- Since $\boldsymbol{y}_i$'s are aligned, we can assume the following model:

$$\boldsymbol{y}_i' = \boldsymbol{X}_i'\boldsymbol{\beta} + \boldsymbol{b}_i' + \boldsymbol{\epsilon}_i'$$

  with $\boldsymbol{b}_i' \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ and $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{C}_i')$

- Using $\boldsymbol{y}_i = \boldsymbol{H}_i\boldsymbol{y}_i'$, we have

$$\boldsymbol{y}_i = \boldsymbol{H}_i\boldsymbol{X}_i'\boldsymbol{\beta} + \boldsymbol{H}_i\boldsymbol{b}_i' + \boldsymbol{H}_i\boldsymbol{\epsilon}_i'$$

- Or simply

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

  with $\boldsymbol{X}_i = \boldsymbol{H}_i\boldsymbol{X}_i'$, $\boldsymbol{b}_i = \boldsymbol{H}_i\boldsymbol{b}_i'$, $\boldsymbol{\epsilon}_i = \boldsymbol{H}_i\boldsymbol{\epsilon}_i'$

- Covariance structure:

$$\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{H}_i\boldsymbol{\Omega}\boldsymbol{H}_i^T), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{H}_i\boldsymbol{C}_i'\boldsymbol{H}_i^T) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{C}_i)$$

Multivariate Meta-analysis Model

Multivariate Meta-analysis Model

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{b}_i + \boldsymbol{\epsilon}_i,$$

with

$$\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{H}_i \boldsymbol{\Omega} \boldsymbol{H}_i^T), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{C}_i).$$

- Known: $\boldsymbol{X}_i$, $\boldsymbol{H}_i$, $\boldsymbol{C}_i$
- Unknown parameters: $\boldsymbol{\beta}$, $\boldsymbol{\Omega}$
- Dimensions:
  - $\boldsymbol{y}_i : p_i \times 1$, $\boldsymbol{X}_i : p_i \times m$, $\boldsymbol{\beta} : m \times 1$
  - $\boldsymbol{b}_i : p_i \times 1$, $\boldsymbol{H}_i : p_i \times k$, $\boldsymbol{\Omega} : k \times k$

Multivariate Meta-analysis Model

If $\boldsymbol{\Omega}$ is known:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{y}_i\right)$$

with $\boldsymbol{V}_i = \boldsymbol{C}_i + \boldsymbol{H}_i \boldsymbol{\Omega} \boldsymbol{H}_i^T$.

- When $\boldsymbol{\Omega} = \mathbf{0}$,

$$\hat{\boldsymbol{\beta}}_0 = \left(\sum_{i=1}^{n} \boldsymbol{X}_i^T \boldsymbol{C}_i^{-1} \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{X}_i^T \boldsymbol{C}_i^{-1} \boldsymbol{y}_i\right)$$

- When $\boldsymbol{\Omega} = \sigma^2 \boldsymbol{I}$ with $\sigma^2 \to \infty$,

$$\hat{\boldsymbol{\beta}}_{OLS} = \left(\sum_{i=1}^{n} \boldsymbol{X}_i^T (\boldsymbol{H}_i \boldsymbol{H}_i^T)^{-1} \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{X}_i^T (\boldsymbol{H}_i \boldsymbol{H}_i^T)^{-1} \boldsymbol{y}_i\right)$$

Multivariate Meta-analysis Model — MLE

- Because $\boldsymbol{y}_i \sim \mathcal{N}(\boldsymbol{X}_i \boldsymbol{\beta}, \boldsymbol{V}_i)$
- the log-likelihood function is

$$\ell(\boldsymbol{\beta}, \boldsymbol{\Omega}) = -\frac{1}{2} \sum_{i=1}^{n} \left(\log |\boldsymbol{V}_i| + (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^T \boldsymbol{V}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) \right)$$

- the restricted log-likelihood function is

$$\ell_R(\boldsymbol{\beta}, \boldsymbol{\Omega}) = \ell(\boldsymbol{\beta}, \boldsymbol{\Omega}) - \frac{1}{2} \log \left| \sum_{i=1}^{n} \boldsymbol{X}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{X}_i \right|$$

Multivariate Meta-analysis Model — Testing

- Consider test $H_0 : \boldsymbol{\Omega} = \mathbf{0}$.
- Test statistics:

$$Q = \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}_0)^T \boldsymbol{C}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \hat{\boldsymbol{\beta}}_0)$$

- distribution:

$$Q \sim \chi^2_d$$

with $d = \sum_i p_i - m$