STAT 574 Linear and Nonlinear Mixed Models

Lecture 8: Nonlinear Mixed Effects Models

Chencheng Cai

Washington State University

Nonlinear Mixed Effects Model

Nonlinear Mixed Effects Model (NLME) is a two-stage model

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\gamma, \boldsymbol{a}_i) + \boldsymbol{\epsilon}_i$$
$$\boldsymbol{a}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{b}_i$$

with $\boldsymbol{\epsilon}_i \sim (0, \sigma^2\boldsymbol{I})$ and $\boldsymbol{b}_i \sim \mathcal{N}(0, \sigma^2\boldsymbol{D})$.

- When γ is absent and $\boldsymbol{f}_i$ is linear, NLME model is the same as the LGC model.
- NLME is not the same as marginal models because $\boldsymbol{a}_i$ is in the argument of the nonlinear functions $\boldsymbol{f}_i$.

Previous Models

- Linear Mixed Effect Model:

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

- Linear Growth Curve Model:

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{Z}_i\boldsymbol{a}_i + \boldsymbol{\epsilon}_i \\ \boldsymbol{a}_i &= \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{b}_i \end{aligned}$$

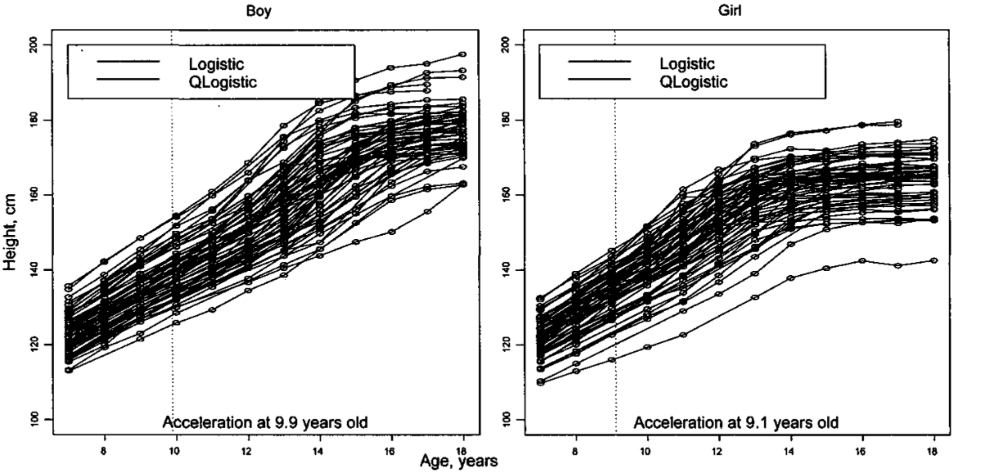
- Nonlinear Marginal Model (Type II):

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{\beta}) + \boldsymbol{Z}_i(\boldsymbol{\beta})\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

- Generalized Linear Mixed Effect Model:

$$\mathbb{P}[y_{ij} = 1] = \mu(\boldsymbol{\beta}^T\boldsymbol{x}_{ij} + \boldsymbol{b}_i^T\boldsymbol{z}_{ij})$$

Example: Height v.s. Age



Example: Height v.s. Age

- LGC model is not appropriate because the curves are not linear.
- Need to consider nonlinear curves with the following parametrization.
  - logistic curve:
  $$f(t) = \frac{a_1}{1 + e^{a_2 - a_3 t}}$$
  - quadratic-logistic curve:
  $$f(t) = \frac{a_1}{1 + e^{a_2 - a_3 t - a_4 t^2}}$$
- In an NLME model, we consider

$$y_{ij} = \frac{a_1 + b_{1i}}{1 + e^{(a_2 + b_{i2}) - (a_3 + b_{i3}) t_{ij} - (a_4 + b_{i4}) t_{ij}^2}} + \epsilon_{ij}$$

where a_1, a_2, a_3, a_4 are fixed effect coefficients, and $b_{i1}, b_{i2}, b_{i3}, b_{i4}$ are random effect coefficients.

Maximum Likelihood Estimation

The log-likelihood function is

$$\ell(\boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{D}) = -\frac{1}{2}\left[N \log|\boldsymbol{D}| + \sum_{i=1}^{N}(n_i + k)\log\sigma^2 + \sum_{i=1}^{N}\log\int g_i(\boldsymbol{\gamma}, \boldsymbol{a}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{D})d\boldsymbol{a}\right]$$

with

$$g_i(\boldsymbol{\gamma}, \boldsymbol{a}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{D}) = \exp\left\{-\frac{1}{2\sigma^2}\left[\|\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\gamma}, \boldsymbol{a})\|^2 + (\boldsymbol{a} - \boldsymbol{A}_i\boldsymbol{\beta})^T\boldsymbol{D}^{-1}(\boldsymbol{a} - \boldsymbol{A}_i\boldsymbol{\beta})\right]\right\}$$

CR Lower Bound for MLE

- Use the following result:
  If $\boldsymbol{X}$ has a distribution parametrized by θ, and $\boldsymbol{Y}$ is independent of θ conditioned on $\boldsymbol{X}$, then
  $$\mathcal{I}_Y \preceq \mathcal{I}_X$$
- Then we have
  $$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{ML}) \geq \sigma^2 \left(\sum_{i=1}^{N} \boldsymbol{A}_i^T \boldsymbol{D}^{-1} \boldsymbol{A}_i \right)^{-1}$$

Two-stage Estimator

If γ is absent,

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{a}_i) + \boldsymbol{\epsilon}_i$$
$$\boldsymbol{a}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{b}_i$$

- First stage: fit the first equation individually for each group.

$$\min_{\boldsymbol{a}_i} \|\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{a}_i)\|^2,$$

  and estimate the covariance
- Second stage: fit the second equation.

Two-stage Estimator — Some Details

- Covariance of $\boldsymbol{a}_i^*$.

$$\mathrm{Cov}(\mathrm{a_i^*} \mid \mathrm{b_i}) = \hat{\sigma}^2(\boldsymbol{R}_i^T \boldsymbol{R}_i)^{-1}$$

  where $\hat{\sigma}^2$ is the variance estimator and $\boldsymbol{R}_i$ is the derivative matrix of $\boldsymbol{f}_i$ at $\boldsymbol{a}_i^*$.
- Estimate $\boldsymbol{\beta}$ from the second equation.
  - Method 1: Assume $\boldsymbol{a}_i^*$ is normal. Use MLE.
  - Method 2: Estimate $\hat{\boldsymbol{D}}$ from MoM. Then use GLS to estimate $\boldsymbol{\beta}$.
- Drawbacks:
  - Require sufficiently large n_i for each group.
  - Outliers in the first stage estimate may ruin the second stage.
- **What if there is γ in the model?**

Two-stage Estimator — Some Details

- Covariance of $\boldsymbol{a}_i^*$.

$$\mathrm{Cov}(\mathrm{a_i^*} \mid \mathrm{b_i}) = \hat{\sigma}^2(\boldsymbol{R}_i^T \boldsymbol{R}_i)^{-1}$$

  where $\hat{\sigma}^2$ is the variance estimator and $\boldsymbol{R}_i$ is the derivative matrix of $\boldsymbol{f}_i$ at $\boldsymbol{a}_i^*$.
- Estimate $\boldsymbol{\beta}$ from the second equation.
  - Method 1: Assume $\boldsymbol{a}_i^*$ is normal. Use MLE.
  - Method 2: Estimate $\hat{\boldsymbol{D}}$ from MoM. Then use GLS to estimate $\boldsymbol{\beta}$.
- Drawbacks:
  - Require sufficiently large n_i for each group.
  - Outliers in the first stage estimate may ruin the second stage.
- **What if there is γ in the model?**

  The first stage becomes

$$\min_{\boldsymbol{\gamma}, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_N} \sum_{i=1}^{N} \|\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\gamma}, \boldsymbol{a}_i)\|^2$$

First-order Approximation

We may transform the problem into a marginal model using the following approximation:

$$\boldsymbol{f}_i(\boldsymbol{\gamma}, \boldsymbol{a}_i) = \boldsymbol{f}_i(\boldsymbol{\gamma}, \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{b}_i) \approx \boldsymbol{f}_i(\boldsymbol{\gamma}, \boldsymbol{A}_i\boldsymbol{\beta}) + \boldsymbol{Z}_i(\boldsymbol{\beta})\boldsymbol{b}_i$$

where

$$\boldsymbol{Z}_i = \frac{\partial \boldsymbol{f}_i(\boldsymbol{\gamma}, \boldsymbol{a}_i)}{\partial \boldsymbol{a}_i}.$$

Now the model becomes a **marginal model**:

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{\gamma}, \boldsymbol{A}_i\boldsymbol{\beta}) + \boldsymbol{Z}_i(\boldsymbol{\beta})\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

One can use GEE or MLE with IRLS.

First-order Approximation — Lindstrom-Bates Version

- The idea: to reduce approximation error, we should expand $\boldsymbol{f}_i$ at a more clear point $\hat{\boldsymbol{b}}_i$ (instead of 0) such that

$$\boldsymbol{f}_i(\boldsymbol{\gamma}, \boldsymbol{a}_i) \approx \boldsymbol{f}_i(\boldsymbol{\gamma}, \boldsymbol{A}_i\boldsymbol{\beta} + \hat{\boldsymbol{b}}_i) + \boldsymbol{R}_i(\boldsymbol{b}_i - \hat{\boldsymbol{b}})$$

  where $\boldsymbol{R}_i = \partial \boldsymbol{f}_i / \partial \boldsymbol{a}_i$ at $\boldsymbol{a}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \hat{\boldsymbol{b}}_i$.
- The method is the actual implementation of `nlme` function in the `nlme` package.

Lindstrom-Bates Estimation

Repeat the following two steps until convergence:

1. Penalized nonlinear least square (PLS): For fixed $\boldsymbol{D}$, minimize the following

$$\min_{\boldsymbol{\gamma},\boldsymbol{\beta},\boldsymbol{b}_1,\dots,\boldsymbol{b}_N} \sum_{i=1}^{N} \left[\|\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\gamma}, \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{b}_i)\|^2 + \boldsymbol{b}_i^T \boldsymbol{D}^{-1} \boldsymbol{b}_i \right]$$

2. Linear mixed effects (LME): Given estimators from step 1, fit the following LME using MLE:

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\hat{\gamma}, \boldsymbol{A}_i\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{b}}_i) + \boldsymbol{R}_i\boldsymbol{A}_i(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \boldsymbol{R}_i(\boldsymbol{b}_i - \hat{\boldsymbol{b}}_i) + \boldsymbol{\epsilon}_i,$$

   with $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I})$ and $\boldsymbol{b}_i \sim \mathcal{N}(0, \sigma^2 \boldsymbol{D})$.
   Get $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}^2$, $\hat{\boldsymbol{D}}$.

Example: Height data

```r
source("./Data/MixedModels/Chapter08/height.dat")
data = height.dat[height.dat$sex==1,]

QLogist = function(a1, a2, a3, a4, x){
return(a1/(1+exp(a2-a3*x-a4*x^2)))
}

library(nlme)
nlme(height~QLogist(a1, a2, a3, a4, x=year),
     fixed=a1+a2+a3+a4~1,
     random=a1+a2+a3+a4~1|id,
     data=data,
     start=c(a1=182, a2=-1, a3=-0.2, a4=0.02))
```

Example: Height data

```
  Nonlinear mixed-effects model fit by maximum likelihood
  Model: height ~ QLogist(a1, a2, a3, a4, x = year)
  Data: data
  Log-likelihood: -1508.839
  Fixed: a1 + a2 + a3 + a4 ~ 1
         a1          a2          a3          a4
165.25699563 -2.80382194 -0.56972462  0.04629372

Random effects:
 Formula: list(a1 ~ 1, a2 ~ 1, a3 ~ 1, a4 ~ 1)
 Level: id
 Structure: General positive-definite, Log-Cholesky parametrization
         StdDev    Corr
a1       6.3110945 a1     a2     a3
a2       0.7977129  0.166
a3       0.2166480  0.128  0.998
a4       0.0161362 -0.109 -0.998 -0.999
Residual 1.2796223

Number of Observations: 702
Number of Groups: 67
```

Likelihood Approximation

Recall the two equations:

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{\gamma}, \boldsymbol{a}_i) + \boldsymbol{\epsilon}_i$$
$$\boldsymbol{a}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{b}_i$$

The (conditional) log-likelihood for the first equation is

$$\ell(\boldsymbol{y}_i \mid \boldsymbol{b}_i, \boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma^2) = -\frac{1}{2}\left\{n_i \log(2\pi) + n_i \log \sigma^2 + \sigma^{-2}\|\boldsymbol{y}_i - \boldsymbol{f}_i\|^2\right\}$$

The (marginal) log-likelihood is therefore

$$\ell(\boldsymbol{y}_i \mid \boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{D}) = \log \int e^{\ell(\boldsymbol{y}_i \mid \boldsymbol{b}_i, \boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma^2)} \phi(\boldsymbol{b} \mid \boldsymbol{D}) d\boldsymbol{b}$$

where

$$\phi(\boldsymbol{b} \mid \boldsymbol{D}) = (2\pi\sigma^2)^{-k/2}\|\boldsymbol{D}\|^{-1/2} \exp\left(-\frac{1}{2\sigma^2}\boldsymbol{b}^T\boldsymbol{D}^{-1}\boldsymbol{b}\right)$$

Likelihood Approximation — Linear Approximation

We approximate the nonlinear function $\boldsymbol{f}_i$ by

$$\boldsymbol{f}_i(\gamma, \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{b}_i) \approx \boldsymbol{f}_i(\gamma, \boldsymbol{A}_i\boldsymbol{\beta}) + \boldsymbol{R}_i\boldsymbol{b}_i$$

where

$$\boldsymbol{R}_i = \left.\frac{\partial \boldsymbol{f}_i(\boldsymbol{\gamma}, \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{b}_i)}{\partial \boldsymbol{b}_i}\right|_{\boldsymbol{b}=\boldsymbol{0}}$$

Then the log-likelihood takes the form:

$$\ell(\boldsymbol{y}_i \mid \boldsymbol{\gamma}, \boldsymbol{\beta}, \sigma^2, \boldsymbol{D}) = \log \int \alpha_0 e^{\alpha_1 + \boldsymbol{\alpha}_2^T \boldsymbol{b} - \boldsymbol{b}^T \boldsymbol{A}\boldsymbol{b}} d\boldsymbol{b}$$

Likelihood Approximation — Penalized Quasi-Likelihood

- Use Laplace approximation of the integral — ℓ_{LA}.
- Consider the penalized quasi-likelihood:

$$\ell_{PQL} = \sum_i \ell_i(\boldsymbol{y}_i; \boldsymbol{b}_i) - \sigma^{-2} \boldsymbol{b}_i^T \boldsymbol{D}^{-1} \boldsymbol{b}_i$$

- PQL method: update the parameters according to ℓ_{LA} and ℓ_{PQL} iteratively.

Example — One-parameter Exponential Family

We consider an exponential family:

$$y_{ij} = e^{a_i} + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2), \quad i = 1, \ldots, N, \; j = 1, \ldots, n$$

with

$$a_i = \beta + b_i, \quad b_i \sim \mathcal{N}(0, \sigma^2 \omega^2)$$

- It is a balanced model.
- $\bar{y}_i = n^{-1} \sum_{j=1}^{n} y_{ij}$ are i.i.d.
- For simplicity, we assume only β is the unknown parameter.
  That is, σ^2 and ω^2 are known.

Example — One-parameter Exponential Family — MLE

The log-likelihood for the i-th group is

$$\ell_i(\boldsymbol{\beta}) = C + \log \int e^{-(2\sigma^2)^{-1}\sum_{j=1}^{n}(y_{ij}-e^a)^2} e^{-(2\sigma^2\omega^2)^{-1}(a-\beta)^2} da$$

The score function is

$$\frac{\partial \ell}{\partial \beta} = -\sum_{i=1}^{N} \frac{\int \left(\frac{e^{2a}-ne^a\bar{y}_i}{\sigma^2} + \frac{a-\beta}{\sigma^2\omega^2}\right) e^{-(2\sigma^2)^{-1}\sum_{j=1}^{n}(y_{ij}-e^{a_i})^2} e^{-(2\sigma^2\omega^2)^{-1}(a-\beta)^2} da}{\int e^{-(2\sigma^2)^{-1}\sum_{j=1}^{n}(y_{ij}-e^a)^2} e^{-(2\sigma^2\omega^2)^{-1}(a-\beta)^2} da}$$

The Fisher's matrix is even more complicated.

Example — One-parameter Exponential Family — First-order Approx

Now we consider the first-order approximation:

$$e^{a_i} = e^{\beta + b_i} \approx e^{\beta} + e^{\beta} b_i$$

The model is now equivalent to

$$\boldsymbol{y}_i = e^{\beta}\mathbf{1} + \boldsymbol{\eta}_i, \quad \boldsymbol{\eta}_i \sim \mathcal{N}(0, \sigma^2(\boldsymbol{I} + e^{2\beta}\omega^2\mathbf{1}\mathbf{1}^T))$$

Let $\boldsymbol{V} = \boldsymbol{I} + e^{2\beta}\omega^2\mathbf{1}\mathbf{1}^T$. We have

$$\boldsymbol{y}_i \sim \mathcal{N}(e^{\beta}\mathbf{1}, \sigma^2\boldsymbol{V})$$

Example — One-parameter Exponential Family — First-order Approx

When $\boldsymbol{V}$ is fixed, we have the GLS solution:

$$\hat{\beta} = \log \frac{\sum_i \mathbf{1}^T \boldsymbol{V}^{-1} \boldsymbol{y}_i}{\sum_i \mathbf{1}^T \boldsymbol{V}^{-1} \mathbf{1}}$$

On the other hand, we have (by Woodbury identity)

$$\boldsymbol{V}^{-1} = (\boldsymbol{I} + e^{2\beta}\omega^2 \mathbf{1}\mathbf{1}^T)^{-1} = \boldsymbol{I} - \frac{e^{2\beta}\omega^2}{1 + ne^{2\beta}\omega^2}\mathbf{1}\mathbf{1}^T$$

Therefore,

$$\mathbf{1}\boldsymbol{V}^{-1}\boldsymbol{y}_i = \frac{1}{1 + ne^{2\beta}\omega^2}\mathbf{1}^T \boldsymbol{y}_i$$
$$\mathbf{1}\boldsymbol{V}^{-1}\boldsymbol{y}_i = \frac{n}{1 + ne^{2\beta}\omega^2}$$

Then we have

$$\hat{\beta} = \log \frac{\sum_i \sum_j y_{ij}}{Nn} = \log \bar{y}$$

Example — One-parameter Exponential Family — First-order Approx

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

Now let $N \to \infty$. Then

$$\hat{\beta} \to \lim_{N\to} \log \left(N^{-1} \sum_{i=1}^{N} \bar{y}_i \right) = \beta + \frac{1}{2}\sigma^2\omega^2.$$

- $\hat{\beta}$ from first-order approximation is not consistent!
- The asymptotic bias is $\frac{1}{2}\sigma^2\omega^2$.
- This bias comes from the approximation error:

$$e^{\beta+b_i} = e^{\beta} + e^{\beta} b_i + e^{2\beta} \cdot \underbrace{\frac{1}{2} b_i}_{\text{expectation: } \frac{1}{2}\sigma^2\omega^2} + O(b_i^3)$$

Example — One-parameter Exponential Family — Two-stage Estimation

First stage: estimate a_i from each group individually.

$$\hat{a}_i = \log \bar{y}_i$$

Second stage: estimate β from the second equation.

$$\hat{\beta} = \frac{1}{N} \sum_{i=1}^{N} \log \bar{y}_i$$

Now let $N \to \infty$, we have

$$\lim_{N \to \infty} \hat{\beta} = \mathbb{E}[\log(e^{\beta + b_i} + \bar{\epsilon}_i)] < \mathbb{E}[log(e^{\beta + b_i})] = \beta$$

The inequality uses Cauchy-Schwartz inequality.

- Two-stage estimator is inconsistent.

Example — One-parameter Exponential Family — L-B Estimation

The penalized least squares (PLS) is

$$\min_{\beta,b_1,\ldots,b_N} \sum_{i=1}^{N}\left(\sum_{j=1}^{n}(y_{ij}-e^{\beta+b_i})^2+\frac{b_i^2}{\omega^2}\right)$$

The estimating equations are

$$\sum_{i=1}^{N}\sum_{j=1}^{n}(y_{ij}-e^{\beta+b_i})e^{\beta+b_i}=0$$

$$\sum_{j=1}^{n}(y_{ij}-e^{\beta+b_i})e^{\beta+b_i}-\frac{b_i}{\omega^2}=0 \quad i=1,\ldots,N.$$

The L-B estimator is the solution to

$$\sum_{i=1}^{N}\hat{b}(\beta,\bar{y}_i)=0,$$

where $\hat{b}(\beta,\bar{y}_i)$ is the solution to the second equation.

Example — One-parameter Exponential Family — L-B Estimation

The L-B estimator satisfies

$$\sum_{i=1}^{N} \hat{b}(\hat{\beta}, \bar{y}_i) = 0$$

When $N \to \infty$, the $\hat{\beta}$ converges to the zero of

$$\mathbb{E}[\hat{b}(\beta, \bar{y})]$$

where $\hat{b}(\beta, \bar{y})$ is the solution to

$$e^{2(\beta+b)} - e^{\beta+b}\bar{y} + \frac{b}{n\omega^2} = 0.$$

- In general, the L-B estimation here is inconsistent.

Equivalence of MLE, TS, and LB Estimators

- In the one-parameter exponential family example, TS and LB are inconsistent when $N \to \infty$.
- But, TS and LB are consistent when $n \to \infty$. (Check!)

Here we provide the general equivalence result.

Theorem

Under mild asymptotic conditions, MLE, TS, and LB estimators have the same limit normal distributions when $N \to \infty$ *and* $\min_i \; n_i \to \infty$.