Non**L**inear **M**ixed-**E**ffects Models & Theory

8th Talk

Mixed-Effects Models

Kim, Sung Won

Introduction

So far

- **Linear Mixed-Effects** Models for grouped data
- **Nonlinear Regression** in which where covariates are nonlinear in parameters

- What if, **both situations** are combined?

- i.e. **grouped data** with **nonlinear expectation function** (allow the reg ft to depend nonlinearly on fixed and random effects)

Goal

- NLME Model
  - Two Examples
  - General Model
- Estimation for NLME Models
  - Likelihood function
  - **Approximation Methods*** (Nonlinear)**
- Computational Methods
- An Extended class of NonLinear Regression Model
  - Extended Basic NLME
  - Extended Nonlinear Regression Model

Ex.1 Indomethicin Knetics

- Data

  -Six human Volunteers received (intravenous) Injections

  → **Six different Subjects**

  – **y_{ij}** : Plasma Concentration of Indomethicin (mcg/ml)
  – Time: time at which the sample was drawn

- Interest
  – Estimate the average behavior of **an individual** in the population
  – Estimate the **variability among and within** individuals

Model Ex.1 Indomethicin

Basically,

For Indomethicin data

compartment model ,expressed as a linear combination of (in this case two)exponential terms, is considered

$$y_{ij} = \phi_1 \exp\left[-\exp\left(\phi_2'\right) t_j\right] + \phi_3 \exp\left[-\exp\left(\phi_4'\right) t_j\right] + \epsilon_{ij}$$

$$\phi_2' = \log \phi_2 \text{ and } \phi_4' = \log \phi_4$$

However,

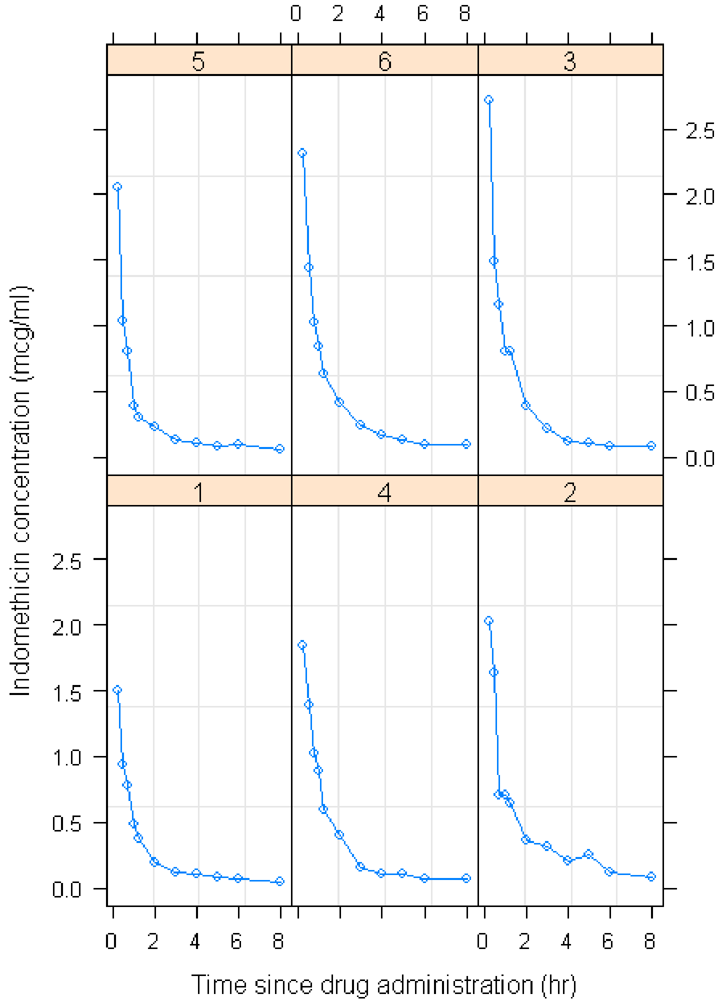
1. Want to **consider Subject Effects** for each individual

And if then

2. Want to know **in which coefficients** Subject Effects should be considered

Observe the data First!!!

Observe the data Ex.1 Indomethicin

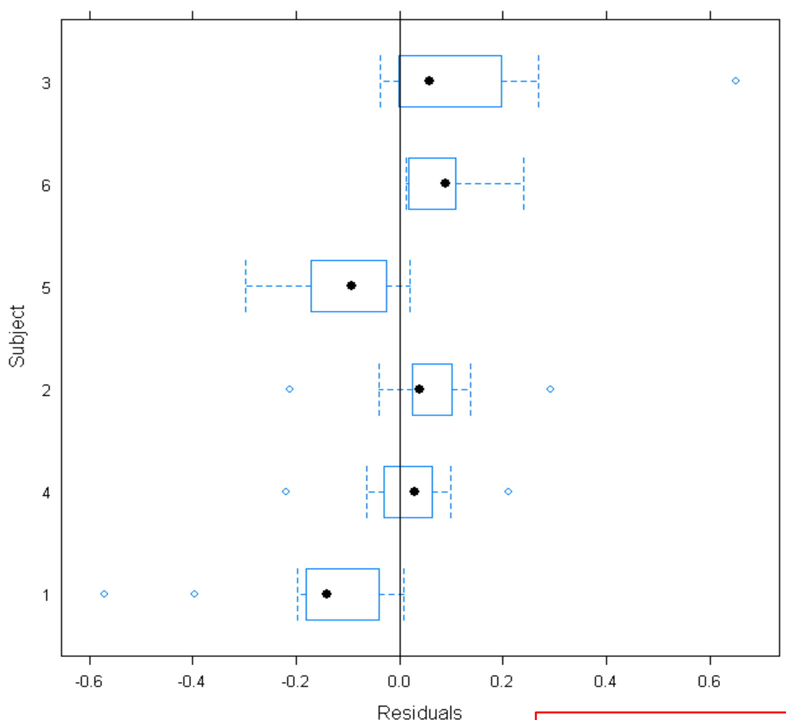


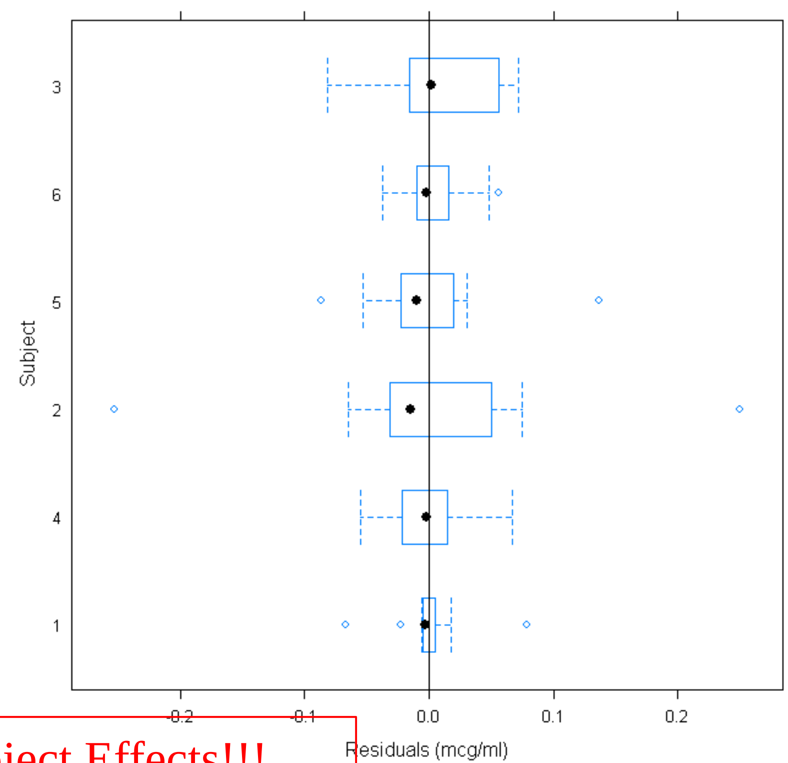
- Nonlinearly decaying
- Similar shape of grouped data
- But differ among six individuals(subjects)

Two Extreme Models Ex.1 Indomethicin

Boxplots of Residuals by subjects for
a NLS
(without considering subjects)
Residual s.e. =0.1747

Boxplots of Residuals by subjects for
a set of **Six Individual NLSs**
(without considering average)
Residual s.e. =0.07555



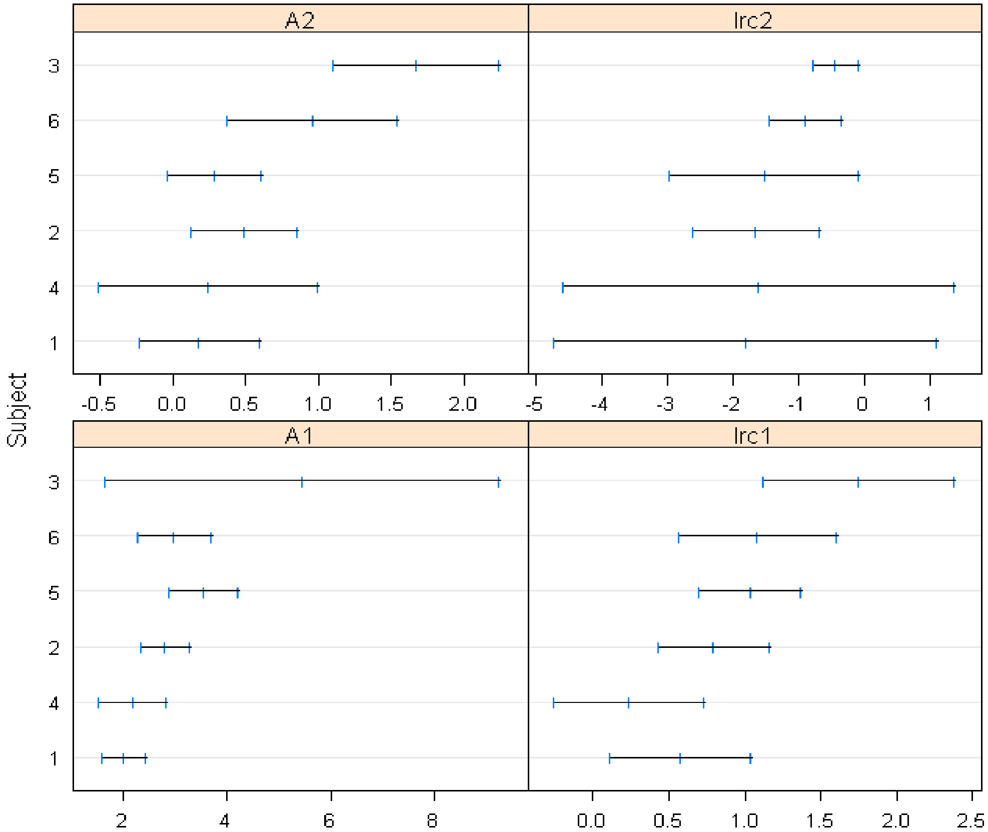


There must be Subject Effects!!!
But in which coefs?

Subject effects**, In which parameters?**

Ex.1 Indomethicin

95% C.I. on model **parameters** for **each Individual**



For some Parameters

There are big variations among indivuals

For which Parameters there is individual(subject) effects?

: Incorporate Random Effects in those Parameters

Model with random effects

Ex.1 Indomethicin

- How are fixed &random effects incorporated in parameters?
- What do they explain for parameters?

$$y_{ij} = \left[\bar{\phi}_1 + \left(\phi_{1i} - \bar{\phi}_1\right)\right] \exp\left\{-\exp\left[\bar{\phi}'_2 + \left(\phi'_{2i} - \bar{\phi}'_2\right)\right] t_j\right\} + \left[\bar{\phi}_3 + \left(\phi_{3i} - \bar{\phi}_3\right)\right] \exp\left\{-\exp\left[\bar{\phi}'_4 + \left(\phi'_{4i} - \bar{\phi}'_4\right)\right] t_j\right\} + \epsilon_{ij}.$$

, or equally in NLME version

$$y_{ij} = (\beta_1 + b_{1i}) \exp\left[-\exp\left(\beta_2 + b_{2i}\right) t_j\right] + (\beta_3 + b_{3i}) \exp\left[-\exp\left(\beta_4 + b_{4i}\right)\right) t_j\right] + \epsilon_{ij}$$

$\beta_1,\ \beta_2,\ \beta_3,$ and β_4	Fixed effects representing the mean values of the parameters
$b_{1i},\ b_{2i},\ b_{3i},$ and b_{4i}	Random effects representing the individual deviations , and being assumed ~N(0, Ψ)

Final Model Ex.1 Indomethicin

We found out there is no random effect for Φ_{4i} i.e. No b_{4i}

$$y_{ij} = \phi_{1i} \exp\left[-\exp\left(\phi'_{2i}\right) t_j\right] + \phi_{3i} \exp\left[-\exp\left(\phi'_{4i}\right) t_j\right] + \epsilon_{ij},$$

$$\underbrace{\begin{bmatrix}\phi_{1i}\\ \phi_{2i}\\ \phi_{3i}\\ \phi_{4i}\end{bmatrix}}_{\boldsymbol{\phi}_{ij}} = \underbrace{\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}}_{\boldsymbol{A}_{ij}} \underbrace{\begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\\ \beta_3\\ \beta_4\end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\0&0&0\end{bmatrix}}_{\boldsymbol{B}_{ij}} \underbrace{\begin{bmatrix}b_{1i}\\ b_{2i}\\ b_{3i}\end{bmatrix}}_{\boldsymbol{b}_i},$$

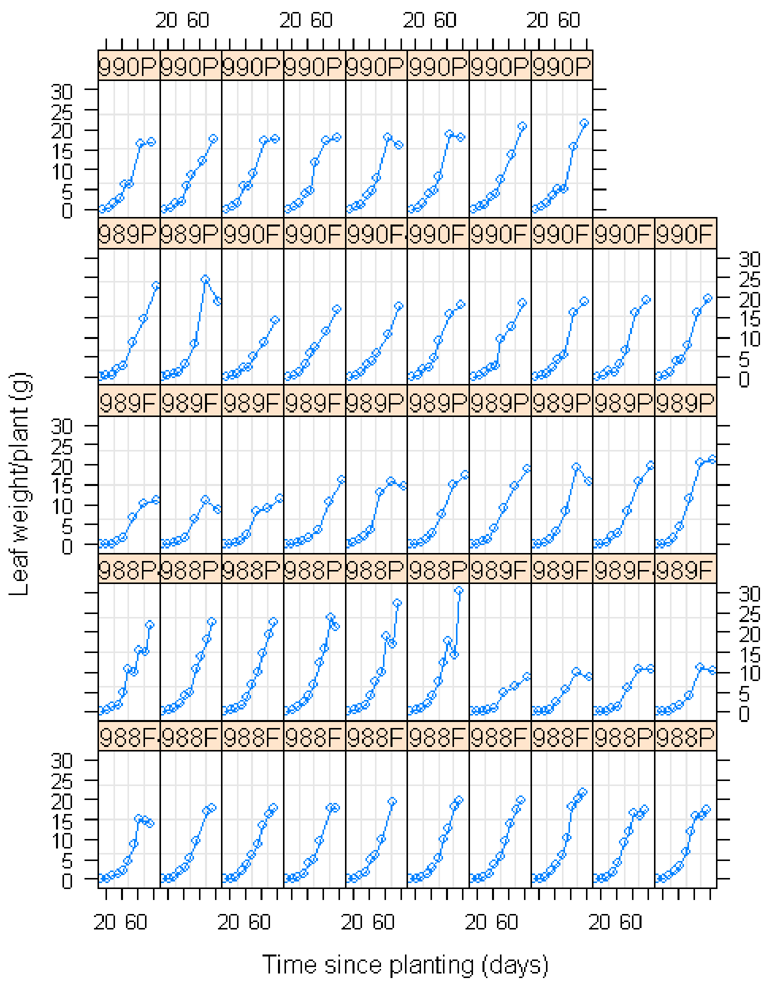
$$\boldsymbol{b}_i \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}\psi_{11}&\psi_{12}&0\\ \psi_{12}&\psi_{22}&0\\ 0&0&\psi_{33}\end{bmatrix}\right), \quad \epsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2\right).$$

Ex.2 Growth of Soybean Plants

- Feature
  - Showing Growth curve data
  - Using covariates to explain between-group variability
- Data
  - $\mathbf{y_{ij}}$: average leaf weight per plant(g)
  - Time: time the sample was taken

    (Experimental Factors)
  - **Varieaty** : Plant Introduction #416937(P), Forrest(F)
  - **Year**: different planting years 1988, 1989, 1990
  - For each category (Variety*Year=6), eight **plots**, fromwhich six plants were averaged, were planted → **48 plots(subjects)**
- **Interest**
  - Possible relationship between **the growth pattern of the Soybean Plants** and exprerimental factors **Variety & Year**

Observe the Data Ex.2 Soybean



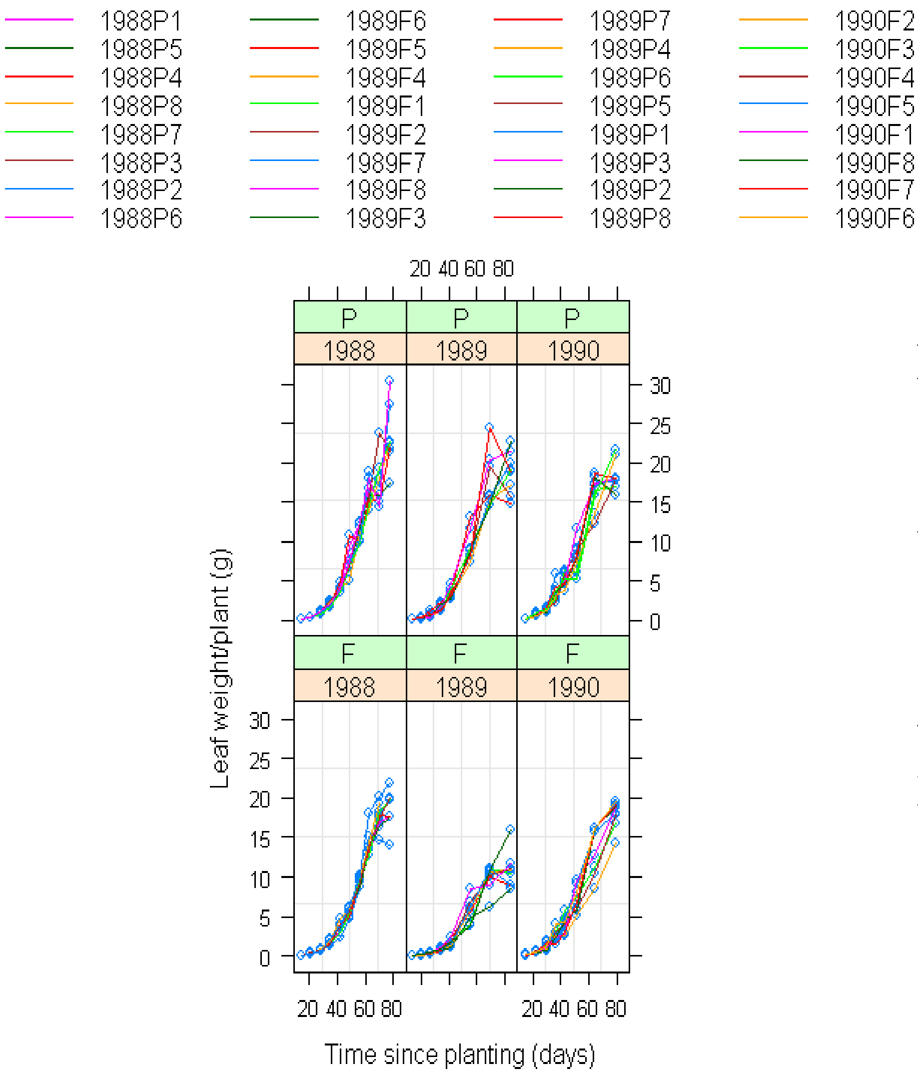
We obeserve all fourty eight plots. All share similar S-shape figure, but there are variations among them?

What are the sources of those variations?

→**Factors**

See the variations with different levels of Factors

Observe the Data Ex.2 Soybean



Average leaf weight per two varieties vs time, over three years

More interesting in Factor Effects (rather than each plot)

➡ Same overall **S-shape**

(nonlinear growth pattern)

But considerable **variation among plots, but more similar within the same Factor levels.**

∴ Significant Factor Effects, Variety and Year, which is what we want to see

Look at the model first!

Model Ex.2 Soybean

Start with

- Accepting the fact that **Nonlinear growth pattern** is well discribed **by three-parameter logistic** model s.t.

$$y_{ij} = \frac{\phi_{1i}}{1 + \exp\left[-\left(t_{ij} - \phi_{2i}\right)/\phi_{3i}\right]} + \epsilon_{ij},$$

$$\boldsymbol{\phi}_i = \begin{bmatrix}\phi_{1i}\\ \phi_{2i}\\ \phi_{3i}\end{bmatrix} = \begin{bmatrix}\beta_1\\ \beta_2\\ \beta_3\end{bmatrix} + \begin{bmatrix}b_{1i}\\ b_{2i}\\ b_{3i}\end{bmatrix} = \boldsymbol{\beta} + \boldsymbol{b}_i,$$

$$\boldsymbol{b}_i \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Psi}\right), \quad \epsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2\right).$$

,whose parameters are $\phi_1 = \texttt{Asym}$, $\phi_2 = \texttt{xmid}$, and $\phi_3 = \texttt{scal}$

Φ1: the asymptotic height(weight),

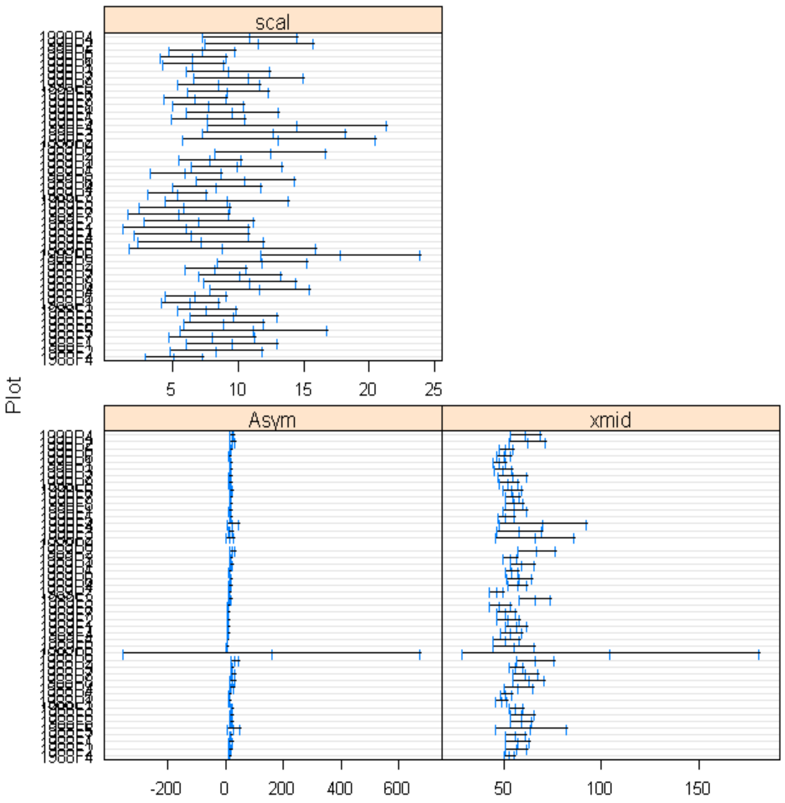
Φ2: the time at which the tree reaches half of ist asymptotic height(weight),

Φ3: the time elapsed between the tree reaching half and 1/{1+exp(-1)}≈¾ of its asymptotic height(weight)

Notice that in this **Nonlinear model** the parameters have physical interpretations
Then there must be some relationship between **Factors**, which seem to effect on Weight, and **Parameters**, which describe growth pattern

Subject(plot) Dependency for Parameters Ex.2
Soybean

95% C.I.s on the model parameters for each individual



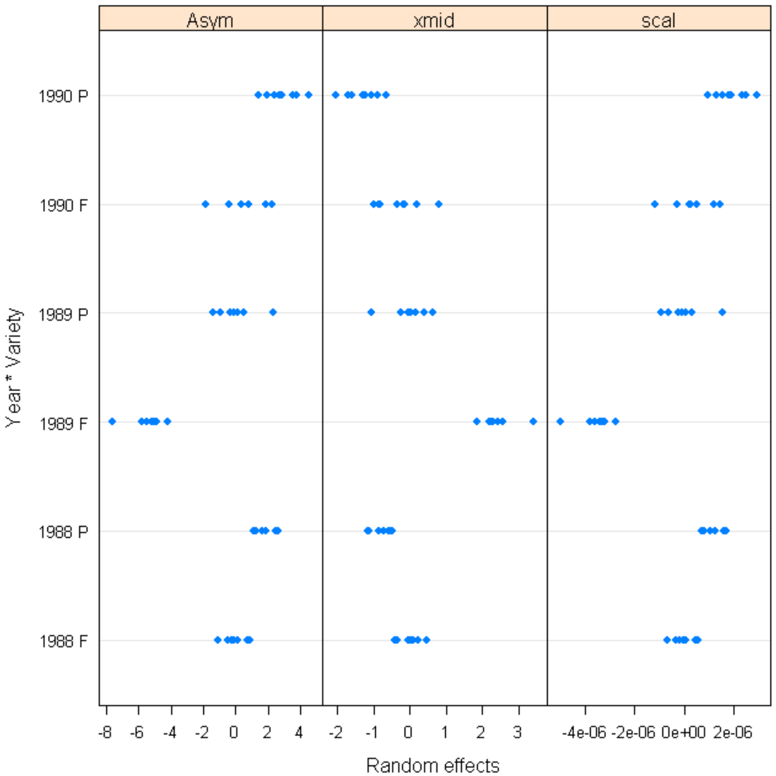
- All the three parameters vary with individuals(subjects)
- All the three parameters need random effects
- However, we are interested in the **relationships** between **growth pattern and factors**
- To find it out we consider linear modeling of parameters wrt coveriates(Variety*Year)
  - ➝ Refer to next slide
- ,which will allow factor effects on parameters to be incorporated in fixed effect

➝ Less needs of random effect terms

➝ Most of between gp variation can be explained by fixed effects

Parameters depend on Variety*Year! Ex.2 Soybean

Estimates of the random effets
Idea: incorporate them in fixed effects using factors ─► more interpretable



- Observe that all three parameters vary according to Variety or/and Year , which look linear
- How are they related?
- After linear fitting on three parameters wrt factors, we obtain proper model such as
- **Fixed-**: Asym~Variety*Year(interaction), xmid~Year+Variety, scal~Year
- After incorporating most of factor effects in fixed effects,
- **Random-** : Asym varies with plots
- ∴ Most of variations can be explained by experimental factors ,which can be incorporated in Fixed effects

Final Model Ex.2 Soybean

Ex. Model for 1990P

$$y_{ij} = \frac{\phi_{1i}}{1 + \exp\left[-\left(t_{ij} - \phi_{2i}\right)/\phi_{3i}\right]} + \epsilon_{ij},$$

$$\underbrace{\begin{bmatrix} \phi_{1i} \\ \phi_{2i} \\ \phi_{3i} \end{bmatrix}}_{\phi_{ij}} = \underbrace{\begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}}_{\boldsymbol{A}_{ij}} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \\ \beta_8 \\ \beta_9 \\ \beta_{10} \\ \beta_{11} \\ \beta_{12} \\ \beta_{13} \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{B}_{ij}} \underbrace{\begin{bmatrix} b_{1i} \end{bmatrix}}_{b_i},$$

$$b_i \sim \mathcal{N}\left(0, \psi\right), \quad \epsilon_{ij} | \boldsymbol{\phi}_i \sim \mathcal{N}\left(0, \sigma^2 \left[\mathrm{E}\left(y_{ij} | \boldsymbol{\phi}_i\right)\right]^{\theta}\right).$$

Generalize the **model** for **NLME!!!**

NLME Model model fomulation

- The model for the *j*th obs on *i*th group is

$$y_{ij} = f(\boldsymbol{\phi}_{ij}, \boldsymbol{v}_{ij}) + \epsilon_{ij}, \qquad i = 1, \ldots, M, \; j = 1, \ldots, n_i,$$

Where M is the #of groups, n_i is the # of obs on the *i*th group, ***f* is a nonlinear,** differentiable ft of **a group-specific parameter vector**, which is modeled as

$$\boldsymbol{\phi}_{ij} = \boldsymbol{A}_{ij}\boldsymbol{\beta} + \boldsymbol{B}_{ij}\boldsymbol{b}_i, \quad \boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Psi})$$

and a covariate vector $\boldsymbol{v}_{ij}$, *and* $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$

Estimation for Parameters $\beta,\ \sigma^2$ and Ψ NLMEmodel

- Idea

  Use Likelihood function

  Parallel to LME until some moment!!!

Likelihood function

$$\mathcal{L}(\beta,\sigma^2,\Delta|y) = \underbrace{p(y|\beta,\sigma^2,\Delta)}_{\text{marginal density of} y} = \int \underbrace{p(y|b,\beta,\sigma^2)}_{\text{conditional density of } y \text{given} b} p(b|\Delta,\sigma^2)db$$

Notice1. Espress **Ψ** in terms of *relative precision factor* ***Δ*** for simplicity s.t.

$$\boldsymbol{\Psi}^{-1} = \sigma^{-2}\boldsymbol{\Delta}^T\boldsymbol{\Delta} \quad \text{,or} \quad \Psi = \sigma^2(\Delta^T\Delta)^{-1}$$

Notice2.

$$y_i|b_i \sim \mathcal{N}(f_i(\beta,b_i),\sigma^2 I) \text{ and } b_i \sim \mathcal{N}(0,\sigma^2(\Delta^T\Delta)^{-1})$$

***Derivation will be on the board!!!

Estimation using Likelihood ft NLMEmodel

- The likelihood ft, **the marginal density of *y*,** is

$$p\left(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\Delta}\right) = \frac{|\boldsymbol{\Delta}|^M}{(2\pi\sigma^2)^{(N+Mq)/2}}\prod_{i=1}^{M}\int \exp\left\{\frac{\|\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta},\boldsymbol{b}_i)\|^2 + \|\boldsymbol{\Delta}\boldsymbol{b}_i\|^2}{-2\sigma^2}\right\} d\boldsymbol{b}_i,$$

,where $\boldsymbol{f}_i(\boldsymbol{\beta},\boldsymbol{b}_i) = \boldsymbol{f}_i\left[\boldsymbol{\phi}_i(\boldsymbol{\beta},\boldsymbol{b}_i), \boldsymbol{v}_i\right]$

- **Notice** that ***f*** is **nonlinear** in the random effects, so the integral is cannot be calculated, which makes **Optimization of Likelihood ft** infeasible
- To make it tractable, three **approximations** to Likelihood ft are proposed

Three Approximations to Likelihood ft NLME

Approximation of Likelihood Function in NLME

Three approximation methods are represented in the following:

1. LME Approximation (Alternating Algorithm)
2. Laplacian Approximation
3. Adaptive Gaussian Approximation

1st Approximation for Likelihood ft :

LME Approximation (of **Alternating algorithm)**

- **Idea**

  approximate likelihood ft **by the likelihood of a linear mixed-effects** model

- It is implemented in “nlme” ft in R

- The Alternating Estimation algorithm alternates two steps

1. Penalized Nonlinear Least Squares(PNLS)
2. **Linear Mixed Effects(LME) *****

1st Approximation for Likelihood ft : **Alternating algorithm**

1st Step **P**enalized **N**onlinear **L**east **S**quares, **PNLS**step

- Goal
  For Fixed current estimate Δ, **Estimate** $\mathbf{b}_i$ **and** $\boldsymbol{\beta}$

- How?
  By minimizing objective ft
  Recall the goal

Penalized **N**onlinear **L**east **S**quares

$$\sum_{i=1}^{M}\left[\|\boldsymbol{y}_i-\boldsymbol{f}_i(\boldsymbol{\beta},\boldsymbol{b}_i)\|^2+\|\boldsymbol{\Delta}\boldsymbol{b}_i\|^2\right]$$

$$p\left(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\Delta}\right)=\frac{|\boldsymbol{\Delta}|^M}{(2\pi\sigma^2)^{(N+Mq)/2}}\prod_{i=1}^{M}\int\exp\left\{\frac{\|\boldsymbol{y}_i-\boldsymbol{f}_i(\boldsymbol{\beta},\boldsymbol{b}_i)\|^2+\|\boldsymbol{\Delta}\boldsymbol{b}_i\|^2}{-2\sigma^2}\right\}d\boldsymbol{b}_i,$$

- Computation for optimization?
  Make it simpler (psedo) and apply **Gauss Newton**
  , which will come later

1st Approximation for Likelihood ft : **Alternating algorithm**
1st Step **PNLS**

Penalized nonlinear Least Squares(PNLS)

We optimize objective function that is equal to

$$\sum_{i=1}^{M} [\| Y_i - f_i(\beta, b_i)\|^2 + \|\Delta b_i\|^2]$$

with Gauss-Newton method.

We will see soon how to apply it !

1st Approximation for Likelihood ft : **Alternating algorithm**

2st step **LME**

„**that is where its name LME comes from !!!**“

- Goal

  Update the estimate of **Δ**

- Idea

  Approximate **loglikelihood** whose **form is identical to that of LME** and use same algorithm as in LME

- How?

  Apply **Taylor Expansion to** $f_i(\beta, b_i)$ around current estimates of β and b_i which gives the identical form to that of a LME ⟹ Same way in LME

  ; **„LME Approximation“**

- Result?

  Obtain the approximate log-likelihood ft to estimate **Δ**

- Plug in current optimal values for $\hat{\beta}(\Delta)$ and $\hat{\sigma}^2(\Delta)$, ft of **Δ .** Then, work with profiled log-likelihood of **Δ** → Optimize **Δ**

1st Approximation for Likelihood ft : **Alternating algorithm**
2st step **LME, Recall the LME**

***Derivation will be on the board
i) Plug in Taylor Expansion
ii) Make it (approximately)linear
iii) Obtain Likelihood ft (usingTransformation)
iv) Calcultae optimal values of β, σ^2 and Profile them on. Esimate Δ!!!

Recall the LME

- the Likelihood ft in LME

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \boldsymbol{y}) = \prod_{i=1}^{M} \frac{\text{abs}\,|\boldsymbol{\Delta}|}{(2\pi\sigma^2)^{n_i/2}} \int \frac{\exp\left[-\left(\|\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b}_i\|^2 + \|\boldsymbol{\Delta}\boldsymbol{b}_i\|^2\right)/2\sigma^2\right]}{(2\pi\sigma^2)^{q/2}}\, d\boldsymbol{b}_i$$

- Reparametrization to describe the density as Normal

$$\boldsymbol{y}_i = \boldsymbol{X}_i\beta + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i = \boldsymbol{X}_i\beta + \boldsymbol{\epsilon}_i^*, \quad i = 1, \ldots, M,$$

Where $\boldsymbol{\epsilon}_i^* = \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$

Sum of Two Indep. MultiNormal

so as to derive the Likelihood ft

$$p(\boldsymbol{y}_i | \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2) = (2\pi\sigma^2)^{-\frac{n_i}{2}} \exp\left(\frac{(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})^T \boldsymbol{\Sigma}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})}{-2\sigma^2}\right) |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}}$$

,where $\boldsymbol{\Sigma}_i = \boldsymbol{I} + \boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i^T/\sigma^2$.

- Now, obtain optimal values for β, σ^2, and then derive profiled likelihood

1st Approximation for Likelihood ft : **Alternating algorithm**
2st step **LME**

- The approximate log-likelihood ft

$$\ell_{\mathrm{LME}}\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Delta} \mid \boldsymbol{y}\right) = -\frac{N}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\sum_{i=1}^{M}\left\{\log\left|\boldsymbol{\Sigma}_i(\boldsymbol{\Delta})\right| + \sigma^{-2}\left[\widehat{\boldsymbol{w}}_i^{(w)} - \widehat{\boldsymbol{X}}_i^{(w)}\boldsymbol{\beta}\right]^T \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\Delta})\left[\widehat{\boldsymbol{w}}_i^{(w)} - \widehat{\boldsymbol{X}}_i^{(w)}\boldsymbol{\beta}\right]\right\}$$

, where $\boldsymbol{\Sigma}_i(\boldsymbol{\Delta}) = \boldsymbol{I} + \widehat{\boldsymbol{Z}}_i^{(w)}\boldsymbol{\Delta}^{-1}\boldsymbol{\Delta}^{-T}\widehat{\boldsymbol{Z}}_i^{(w)^T}$

***Derivation will be on the board

i) Plug in Taylor Expansion
ii) Make it (approximately)linear
iii) Obtain Likelihood ft (usingTransformation)
iv) Calcultae optimal values of β, σ^2 and Profile them on. Esimate Δ!!!

2st Approximation for Likelihood ft

Laplacian Approximation

How

- Use Laplacian approximation to approximate likelihood ft

Idea

- i) Apply second order of **Taylor expansion** to obejective ft

  $\|\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)\|^2 + \|\boldsymbol{\Delta b}_i\|^2$ around $\widehat{\boldsymbol{b}}_i$

  Then integration will be done with Gaussian density
- ii)Modification for simpler calculation
- **Approximate Hessian** bydropping negligible term

Result

- i) Use profiled modified likelihood ft to get MLE
- ii) profile ℓ_{LA} on σ^2 to reduce the dim of optimization problem

  → ft of $\boldsymbol{\beta}, \boldsymbol{\Delta}$

2st Approximation for Likelihood ft : **Laplacian Approximation**

***Derivation will be on the board

Recall the objective ft (Likelihood ft)

$$p\left(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\Delta}\right) = \frac{|\boldsymbol{\Delta}|^M}{(2\pi\sigma^2)^{(N+Mq)/2}} \prod_{i=1}^{M} \int \exp\left\{\frac{\|\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta},\boldsymbol{b}_i)\|^2 + \|\boldsymbol{\Delta}\boldsymbol{b}_i\|^2}{-2\sigma^2}\right\} d\boldsymbol{b}_i,$$

Set $g(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{y}_i,\boldsymbol{b}_i) = \|\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta},\boldsymbol{b}_i)\|^2 + \|\boldsymbol{\Delta}\boldsymbol{b}_i\|^2$

And let $\widehat{\boldsymbol{b}}_i = \widehat{\boldsymbol{b}}_i(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{y}_i) = \arg\min_{\boldsymbol{b}_i} g(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{y}_i,\boldsymbol{b}_i),$

$$g'(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{y}_i,\boldsymbol{b}_i) = \frac{\partial g(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{y}_i,\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i},$$

$$g''(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{y}_i,\boldsymbol{b}_i) = \frac{\partial^2 g(\boldsymbol{\beta},\boldsymbol{\Delta},\boldsymbol{y}_i,\boldsymbol{b}_i)}{\partial \boldsymbol{b}_i \partial \boldsymbol{b}_i^T},$$

A second order of Taylor expansion of g around $\widehat{\boldsymbol{b}}_i$ gives Laplacian approximation defined as

2st Approximation for Likelihood ft : **Laplacian Approximation**

- **The Laplacian Approximation**

$$p\left(y \mid \beta, \sigma^2, \boldsymbol{\Delta}\right) \simeq \left(2\pi\sigma^2\right)^{-\frac{N}{2}} \left|\boldsymbol{\Delta}\right|^M \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{M} g(\boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{y}_i, \widehat{\boldsymbol{b}}_i)\right] \times \prod_{i=1}^{M} \int \left(2\pi\sigma^2\right)^{\frac{q}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\boldsymbol{b}_i - \widehat{\boldsymbol{b}}_i\right]^T g''(\boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{y}_i, \widehat{\boldsymbol{b}}_i)\left[\boldsymbol{b}_i - \widehat{\boldsymbol{b}}_i\right]\right\} d\boldsymbol{b}_i$$

- Furthermore,**the modified Laplacian approximation** to the log-likelihood

$$\ell_{\mathrm{LA}}\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Delta}, \mid \boldsymbol{y}\right) = -\frac{N}{2}\log\left(2\pi\sigma^2\right) + M\log\left|\boldsymbol{\Delta}\right| - \frac{1}{2}\left\{\sum_{i=1}^{M}\log\left|\boldsymbol{G}\left(\boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{y}_i\right)\right| + \sigma^{-2}\sum_{i=1}^{M} g\left(\boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{y}_i, \widehat{\boldsymbol{b}}_i\right)\right\}$$

by approximation Hessian (by dropping second derivatives of *f)*

→ Easier to Compute!

3rd Approximation for Likelihood ft :

Adaptive Gaussian Approximation

- Goal
- Improve Laplacian Apporximation
- (modified Laplacian approximation is the simplest case of Gaussian Approximation)

- Idea
- Use Gaussian quadrature rule
- Gaussian Quadrature Rule?
  - It is used to approximate integrals of fts by a weighted average of the integrand evaluated at predetermined abscissas

- How?
- Apply Gaussian quadrature rule to Laplacian Approximation

3rd Approximation for Likelihood ft : **Adaptive Gaussian Approximation**

Adaptive Gaussian Approximation

To improve Laplacian approximation Gaussian quadrature rules are used, which approximates integrals of functions by a weighted average of the integral evaluated predetermined abscisses such that

$$\int_{-1}^{1} f(x)dx = \sum_{i=1}^{n} w_i f(x_i)$$

Alternating algorithm vs Laplacian approximation

- **(+)Laplacian** approximation generally gives **more accurate estimates** than the alternating algorithm

  reason : it uses an expansion around estimated random effect **only**, while LME approximation in the alternating algorithm uses an expansion around **both** of the estimated fixed and random effects

- **(-) Laplacian** approximation is computationally **intensive** than the alternating algorithm

  reason : it requires solving a different penalized nonlinear least square problem for each group in the data, and its objective function cannot be profiled on the β(profiled log-likelihood is still ft of β and Δ)

$$g(\boldsymbol{\beta}, \boldsymbol{\Delta}, \boldsymbol{y}_i, \boldsymbol{b}_i) = \|\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)\|^2 + \|\boldsymbol{\Delta}\boldsymbol{b}_i\|^2$$

Computational Method

for **Alternating Approximation** on **PNLS** step **to find optimal values of β and b_i**

Computational Methods for Estimating Parameters

For this nonlinear square problem, we use *Gaussian-Newton optimization*. Replacing nonlinear $\tilde{f}$ into Taylor approximation around current estimates gives *Least-Squares problem*.

$$\sum_{i=1}^{M} \left[\|Y_i - f_i(\beta, b_i)\|^2 + \|\Delta b_i\|^2 \right] \simeq \sum_{i=1}^{M} \|\tilde{\omega}_i^{(w)} - \tilde{X}_i^{(w)}\beta - \tilde{Z}_i^{(w)} b_i\|^2$$

Computational Method
for **Alternating Approximation** on **PNLS** step

- Focus on Alternating algorithm

  i) **PNLS step**

  ii) LME Step : same as the LME case

- Situation (Recall!)

  Want **to find optimal values of β and b_i** minimizing Penalized Sum of Square

$$\sum_{i=1}^{M}\left[\|\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)\|^2 + \|\boldsymbol{\Delta b}_i\|^2\right]$$

- How? **

  i) Simplify the objective ft by adding Pseudo data

  → standard **Nonlinear Least-Squares problem** (common method? GN)

  ii) Apply **Gauss-Newton** (Recall!)

Computational Method
for **Alternating Approximation** on **PNLS** step

- i) Simplify the objective ft by adding Pseudo observation

$$\tilde{\boldsymbol{y}}_i = \begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{0} \end{bmatrix} \qquad \tilde{\boldsymbol{f}}_i(\boldsymbol{\beta}, \boldsymbol{b}_i) = \begin{bmatrix} \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i) \\ \boldsymbol{\Delta b}_i \end{bmatrix}$$

Penalized NonLinear Sum of Squares

$$\sum_{i=1}^{M} \left[\|\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)\|^2 + \|\boldsymbol{\Delta b}_i\|^2 \right] \longrightarrow$$

a standard Nonlinear Least-Squares

$$\sum_{i=1}^{M} \|\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{f}}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)\|^2$$

→ Apply **Gauss-Newton** Method (Recall!)

Alternating Approximation on **PNLS** step , recall **Gauss-Newton**

We want to solve $\sum_{i=1}^{M} \|\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{f}}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)\|^2$

To solve **Nonlinear Least-Squares** problems

i) Replace the ft by a **first-order Taylor series approximation** about current estimates

ii) Solve Least-Squares problem $\left\| \left[\boldsymbol{y} - \boldsymbol{f}(\widehat{\boldsymbol{\alpha}}^{(w)})\right] - \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\alpha}^T}\right|_{\widehat{\boldsymbol{\alpha}}^{(w)}} \left(\boldsymbol{\alpha} - \widehat{\boldsymbol{\alpha}}^{(w)}\right) \right\|^2$

So that the soln is the prameter increments $\widehat{\delta}^{(w+1)} = \widehat{\alpha}^{(w+1)} - \widehat{\alpha}^{(w)}$

Then the **new estimate** is $\widehat{\alpha}^{(w)} + \widehat{\delta}^{(w+1)}$, and we can obtain the value of the obj.ft

iv) Iterate algorithm checking Step-halving at each step

*** ***Step-halving*** : to ensure that updated estimate results in a decrease of the objective ft

If no decrease ? Halve the increment and iterate this halving process until the estimate gives decreases value of objective ft cf. $\widehat{\alpha}^{(w)} + \widehat{\delta}^{(w+1)}/2$

Go back to Our PNLS problem!

Computational Method
Alternating Approximation on **PNLS** step

Want to solve $$\sum_{i=1}^{M} \|\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{f}}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)\|^2$$

By Taylor expansion $$\sum_{i=1}^{M} \left\| \left[\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{f}}_i \left(\widehat{\boldsymbol{\beta}}^{(w)}, \widehat{\boldsymbol{b}}_i^{(w)} \right) \right] - \tilde{\boldsymbol{X}}_i^{(w)} \left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}^{(w)} \right) - \tilde{\boldsymbol{Z}}_i^{(w)} \left(\boldsymbol{b}_i - \widehat{\boldsymbol{b}}_i^{(w)} \right) \right\|^2$$

,where derivative matrices are $$\left. \frac{\partial \tilde{\boldsymbol{f}}_i(\boldsymbol{\beta}, \boldsymbol{b}_i | \boldsymbol{\Delta})}{\partial \boldsymbol{\beta}^T} \right|_{\widehat{\boldsymbol{\beta}}^{(w)}, \widehat{\boldsymbol{b}}_i^{(w)}} = \tilde{\boldsymbol{X}}_i^{(w)} = \begin{bmatrix} \widehat{\boldsymbol{X}}_i^{(w)} \\ \boldsymbol{0} \end{bmatrix},$$
$$\left. \frac{\partial \tilde{\boldsymbol{f}}_i(\boldsymbol{\beta}, \boldsymbol{b}_i | \boldsymbol{\Delta})}{\partial \boldsymbol{b}_i^T} \right|_{\widehat{\boldsymbol{\beta}}^{(w)}, \widehat{\boldsymbol{b}}_i^{(w)}} = \tilde{\boldsymbol{Z}}_i^{(w)} = \begin{bmatrix} \widehat{\boldsymbol{Z}}_i^{(w)} \\ \boldsymbol{\Delta} \end{bmatrix},$$

$$\sum_{i=1}^{M} \left\| \tilde{\boldsymbol{w}}_i^{(w)} - \tilde{\boldsymbol{X}}_i^{(w)} \boldsymbol{\beta} - \tilde{\boldsymbol{Z}}_i^{(w)} \boldsymbol{b}_i \right\|^2, \quad \text{where} \quad \tilde{\boldsymbol{w}}_i^{(w)} = \begin{bmatrix} \widehat{\boldsymbol{w}}_i^{(w)} \\ \boldsymbol{0} \end{bmatrix}$$

Find optimal new estimates of β and b_i. And calculate increment! (difference between new estimates and current estimates)

Check if it meets Step-halving rule!

What if, variant &correlated ε ?
(Relax the Assumptions…) **Extending the Basic NLME**

- Want to consider more general case
- Relax the assumption that within group error $\boldsymbol{\varepsilon}_i$; are indepedent $\mathcal{N}(0, \sigma^2 I)$ random vectors

→ Allow them to be
heteroscedastic (having unequal variance) and/or **correlated**

Consider two models

1. **Extended NLME Model**
2. **Extended Nonlinear Regression Model**

What if, variant &correlated ε ?
(Relax the Assumptions…) **1.Extended NLME**

Extended NLME Model

For general case we extend the basic NLME. We allow the within-group errors ε_i to be heteroscedastic or/and correlated. So our model is given by

$$Y_i = f_i(\phi_i, \nu_i) + \varepsilon_i$$

$$\phi_i = A_i\beta + B_i b_i, \quad b_i \sim \mathcal{N}(0, \Psi) \quad \text{and} \quad \boxed{\varepsilon_i \sim \mathcal{N}(0, \sigma^2\Lambda)}$$

What if, variant &correlated ε ?

(Relax the Assumptions…) **1. Extended NLME : Estimation**

- Estimation by Transformation

$$\mathbf{\Lambda}_i = \mathbf{\Lambda}_i^{T/2}\mathbf{\Lambda}_i^{1/2} \quad \text{and} \quad \mathbf{\Lambda}_i^{-1} = \mathbf{\Lambda}_i^{-1/2}\mathbf{\Lambda}_i^{-T/2}$$

$$\begin{aligned} \boldsymbol{y}_i^* &= \mathbf{\Lambda}_i^{-T/2}\boldsymbol{y}_i, \\ \boldsymbol{f}_i^*(\boldsymbol{\phi}_i, \boldsymbol{v}_i) &= \mathbf{\Lambda}_i^{-T/2}\boldsymbol{f}_i(\boldsymbol{\phi}_i, \boldsymbol{v}_i), \\ \boldsymbol{\epsilon}_i^* &= \mathbf{\Lambda}_i^{-T/2}\boldsymbol{\epsilon}_i, \end{aligned}$$

- Transformed model can be **described by basic NLME**

$$\begin{aligned} \boldsymbol{y}_i^* &= \boldsymbol{f}_i^*(\boldsymbol{\phi}_i, \boldsymbol{v}_i) + \boldsymbol{\epsilon}_i^*, \\ \boldsymbol{\phi}_i &= \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i, \\ \boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Psi}), &\quad \boldsymbol{\epsilon}_i^* \sim \mathcal{N}\left(\mathbf{0}, \sigma^2\boldsymbol{I}\right) \end{aligned}$$

- Notice that

$$\boldsymbol{\epsilon}_i^* \sim \mathcal{N}\left[\mathbf{\Lambda}_i^{-T/2}\mathbf{0}, \sigma^2\mathbf{\Lambda}_i^{-T/2}\mathbf{\Lambda}_i\mathbf{\Lambda}_i^{-1/2}\right] = \mathcal{N}\left(\mathbf{0}, \sigma^2\boldsymbol{I}\right)$$

What if, variant &correlated ε ?

(Relax the Assumptions…) **1. Extended NLME : Estimation**

- The **log-likelihood ft** for extended NLME

  since $\boldsymbol{y}_i^* = \boldsymbol{\Lambda}_i^{-T/2}\boldsymbol{y}_i$ and so $d\boldsymbol{y}_i^* = |\boldsymbol{\Lambda}_i|^{-1/2}\, d\boldsymbol{y}_i$

- $$\begin{aligned}\ell\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Delta}, \boldsymbol{\lambda} | \boldsymbol{y}\right) &= \sum_{i=1}^{M} \log p\left(\boldsymbol{y}_i | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\Delta}, \boldsymbol{\lambda}\right) \\ &= \sum_{i=1}^{M} \log p\left(\boldsymbol{y}_i^* | \boldsymbol{\beta}, \sigma^2, \boldsymbol{\Delta}, \boldsymbol{\lambda}\right) - \frac{1}{2}\sum_{i=1}^{M} \log|\boldsymbol{\Lambda}_i| \\ &= \ell\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Delta}, \boldsymbol{\lambda} | \boldsymbol{y}^*\right) - \frac{1}{2}\sum_{i=1}^{M} \log|\boldsymbol{\Lambda}_i| .\end{aligned}$$

Both are density ft, which integral to 1

- **Same approximations** as those of Basic NLME can be applied to **approximate**

$$\ell\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Delta}, \boldsymbol{\lambda} | \boldsymbol{y}^*\right)$$

1. Alternating Algorithm

2. Laplacian and Adaptive Gaussian Approximations

What if, variant &correlated ε ?

(Relax the Assumptions…) **2.Extended Nonlinear Regression: Model**

Extended nonlinear Regression Model

Make Situatin more Simpler!
No Random Effect!
Let all Variations be explained only through „ ε“ structure!
→ Simplfied Version of Extended NLME!

The Extended nonlinear Regression Model is given by

$$Y_i = f_i(\phi_i, \nu_i) + \varepsilon_i$$

$$\boxed{\phi_i = A_i \beta} \quad \text{and} \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2 \Lambda_i)$$

What if, variant &correlated ε ?
(Relax the Assumptions…) **2.Extended Nonlinear Regression:**

Estimation &Inference

- Assuming that Λi matrices are known,

  It is referred to as the **generalized nonlinear least-squares(GNLS)** model

- Using the same transformation as Extended NLME case(cf. slide# 41),

  model is

$$\boldsymbol{y}_i^* = \boldsymbol{f}_i^* \left(\boldsymbol{\phi}_i, \boldsymbol{v}_i\right) + \boldsymbol{\epsilon}_i^*,$$
$$\boldsymbol{\phi}_i = \boldsymbol{A}_i \boldsymbol{\beta}, \quad \boldsymbol{\epsilon}_i^* \sim \Lambda \text{fixed } \boldsymbol{\beta} \text{ and } \boldsymbol{\lambda}$$

- Find MLE using log-likelihood ft

$$\ell\left(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda} \middle| \boldsymbol{y}\right) = -\frac{1}{2}\left\{N \log\left(2\pi\sigma^2\right) + \sum_{i=1}^{M}\left[\frac{\left\|\boldsymbol{y}_i^* - \boldsymbol{f}_i^*\left(\boldsymbol{\beta}\right)\right\|^2}{\sigma^2} + \log\left|\boldsymbol{\Lambda}_i\right|\right]\right\}$$

- Profile it on σ^2 ⟶ MLE of β and λ

Reference

[1] Pinheiro, J.C. and Bates, D.M. *Mixed-Effects Models in S and S-Plus*, Springer 2000

Thanks for your attention!