Lecture 10: Linear Mixed Models
(Linear Models with Random Effects)

Claudia Czado

TU München

Overview

West, Welch, and Galecki (2007)
Fahrmeir, Kneib, and Lang (2007) (Kapitel 6)

- Introduction
- Likelihood Inference for Linear Mixed Models
  - Parameter Estimation for known Covariance Structure
  - Parameter Estimation for unknown Covariance Structure
  - Confidence Intervals and Hypothesis Tests

Introduction

So far: independent response variables, but often

- Clustered Data
  - response is measured for each subject
  - each subject belongs to a group of subjects (cluster)

  Ex.:

  - math scores of student grouped by classrooms (class room forms cluster)
  - birth weigths of rats grouped by litter (litter forms cluster)

- Longitudinal Data
  - response is measured at several time points
  - number of time points is not too large (in contrast to time series)

  Ex.: sales of a product at each month in a year (12 measurements)

Fixed and Random Factors/Effects

How can we extend the linear model to allow for such dependent data structures?

fixed factor	=	**qualitative** covariate (e.g. gender, agegroup)
fixed effect	=	**quantitative** covariate (e.g. age)
random factor	=	**qualitative** variable whose levels are randomly sampled from a population of levels being studied

Ex.: 20 supermarkets were selected and their number of cashiers were reported

$$\left.\begin{array}{l}\text{10 supermarkets with 2 cashiers}\\ \text{5 supermarkets with 1 cashier}\\ \text{5 supermarkets with 5 cashiers}\end{array}\right\}\ \begin{array}{l}\text{observed levels of random factor}\\ \text{“number of cashiers”}\end{array}$$

random effect	=	**quantitative** variable whose levels are randomly sampled from a population of levels being studied

Ex.: 20 supermarkets were selected and their size reported. These size values are random samples from the population of size values of all supermarkets.

Modeling Clustered Data

Y_{ij}	$=$	response of j-th member of cluster i, $i = 1, \ldots, m, j = 1, \ldots, n_i$
m	$=$	number of clusters
n_i	$=$	size of cluster i
$\boldsymbol{x_{ij}}$	$=$	covariate vector of j-th member of cluster i for fixed effects, $\in \mathbb{R}^p$
$\boldsymbol{\beta}$	$=$	fixed effects parameter, $\in \mathbb{R}^p$
$\boldsymbol{u_{ij}}$	$=$	covariate vector of j-th member of cluster i for random effects, $\in \mathbb{R}^q$
$\boldsymbol{\gamma_i}$	$=$	random effect parameter, $\in \mathbb{R}^q$

Model:

$$\boxed{\begin{array}{l} Y_{ij} = \underbrace{\boldsymbol{x_{ij}}^t \boldsymbol{\beta}}_{\text{fixed}} + \underbrace{\boldsymbol{u_{ij}}^t \boldsymbol{\gamma_i}}_{\text{random}} + \underbrace{\epsilon_{ij}}_{\text{random}} \\ \\ i = 1, \ldots, m; j = 1, \ldots, n_i \end{array}}$$

Mixed Linear Model (LMM) I

Assumptions:

$$\boldsymbol{\gamma_i} \sim N_q(\mathbf{0}, D), \qquad D \in \mathbb{R}^{q \times q}$$

$$\boldsymbol{\epsilon_i} := \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{in_i} \end{pmatrix} \sim N_{n_i}(\mathbf{0}, \Sigma_i), \qquad \Sigma_i \in \mathbb{R}^{n_i \times n_i}$$

$\boldsymbol{\gamma_1}, \ldots, \boldsymbol{\gamma_m}, \boldsymbol{\epsilon_1}, \ldots, \boldsymbol{\epsilon_m}$ independent

$D =$ covariance matrix of random effects $\boldsymbol{\gamma_i}$

$\Sigma_i =$ covariance matrix of error vector $\boldsymbol{\epsilon_i}$ in cluster i

Mixed Linear Model (LMM) II

Matrix Notation:

$$X_i := \begin{pmatrix} \boldsymbol{x}_{i1}^t \\ \vdots \\ \boldsymbol{x}_{in_i}^t \end{pmatrix} \in \mathbb{R}^{n_i \times p}, \quad U_i := \begin{pmatrix} \boldsymbol{u}_{i1}^t \\ \vdots \\ \boldsymbol{u}_{in_i}^t \end{pmatrix} \in \mathbb{R}^{n_i \times q}, \quad \boldsymbol{Y_i} := \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in_i} \end{pmatrix} \in \mathbb{R}^{n_i}$$

$$\Rightarrow \boxed{\begin{array}{ll} \boldsymbol{Y_i} = X_i\boldsymbol{\beta} + U_i\boldsymbol{\gamma_i} + \boldsymbol{\epsilon_i} \quad i = 1, \ldots, m & \\ \boldsymbol{\gamma_i} \sim N_q(\boldsymbol{0}, D) & \boldsymbol{\gamma_1}, \ldots, \boldsymbol{\gamma_m}, \boldsymbol{\epsilon_1}, \ldots, \boldsymbol{\epsilon_m} \text{ independent} \\ \boldsymbol{\epsilon_i} \sim N_{n_i}(\boldsymbol{0}, \Sigma_i) & \end{array}} \qquad (1)$$

Modeling Longitudinal Data

Y_{ij} = response of subject i at j-th measurement, $i = 1, \ldots, m, j = 1, \ldots, n_i$

n_i = number of measurements for subject i

m = number of objects

$\boldsymbol{x_{ij}}$ = covariate vector of i-th subject at j-th measurement for fixed effects $\boldsymbol{\beta} \in \mathbb{R}^p$

$\boldsymbol{u_{ij}}$ = covariate vector of i-th subject at j-th measurement for random effects $\boldsymbol{\gamma_i} \in \mathbb{R}^q$

$\overset{\text{matrix notation}}{\Longrightarrow}$

$$\boxed{\begin{array}{ll} \boldsymbol{Y_i} = X_i\boldsymbol{\beta} + U_i\boldsymbol{\gamma_i} + \boldsymbol{\epsilon_i} & \\ \boldsymbol{\gamma_i} \sim N_q(\boldsymbol{0}, D) & \boldsymbol{\gamma_1}, \ldots, \boldsymbol{\gamma_m}, \boldsymbol{\epsilon_1}, \ldots, \boldsymbol{\epsilon_m} \text{ independent} \\ \boldsymbol{\epsilon_i} \sim N_{n_i}(\boldsymbol{0}, \Sigma_i) & \end{array}}$$

Remark: The general form of the mixed linear model is the same for clustered and longitudinal observations.

Matrix Formulation of the Linear Mixed Model

$$\boldsymbol{Y} := \begin{pmatrix} \boldsymbol{Y_1} \\ \vdots \\ \boldsymbol{Y_m} \end{pmatrix} \in \mathbb{R}^n, \text{where } n := \sum_{i=1}^{m} n_i$$

$$X := \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \in \mathbb{R}^{n \times p}, \qquad \boldsymbol{\beta} \in \mathbb{R}^p \qquad \mathcal{G} := \begin{pmatrix} D & & \\ & \ddots & \\ & & D \end{pmatrix} \in \mathbb{R}^{mq \times mq}$$

$$U := \begin{pmatrix} U_1 & 0_{n_1 \times q} & \cdots & 0_{n_1 \times q} \\ 0_{n_2 \times q} & U_2 & & \\ \vdots & & \ddots & \\ 0_{n_m \times q} & & & U_m \end{pmatrix} \in \mathbb{R}^{n \times (m \cdot q)}, \quad 0_{n_i \times q} := \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \\ 0 & \cdots & 0 \end{pmatrix} \in \mathbb{R}^{n_i \times q}$$

$$\boldsymbol{\gamma} := \begin{pmatrix} \boldsymbol{\gamma_1} \\ \vdots \\ \boldsymbol{\gamma_m} \end{pmatrix} \in \mathbb{R}^{m \cdot q}, \qquad \boldsymbol{\epsilon} := \begin{pmatrix} \boldsymbol{\epsilon_1} \\ \vdots \\ \boldsymbol{\epsilon_m}, \end{pmatrix} \qquad R := \begin{pmatrix} \Sigma_1 & & 0 \\ & \ddots & \\ 0 & & \Sigma_m \end{pmatrix} \in \mathbb{R}^{n \times n}$$

Linear Mixed Model (LMM) in matrix formulation

With this, the linear mixed model (1) can be rewritten as

$$\boldsymbol{Y} = X\boldsymbol{\beta} + U\boldsymbol{\gamma} + \boldsymbol{\epsilon} \qquad (2)$$

$$\text{where } \begin{pmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\epsilon} \end{pmatrix} \sim N_{mq+n}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathcal{G} & 0_{mq \times n} \\ 0_{n \times mq} & R \end{pmatrix} \right)$$

Remarks:

- LMM (2) can be rewritten as two level hierarchical model

$$\begin{aligned} \boldsymbol{Y}|\boldsymbol{\gamma} &\sim N_n(X\boldsymbol{\beta} + U\boldsymbol{\gamma}, R) && (3) \\ \boldsymbol{\gamma} &\sim N_{mq}(\mathbf{0}, R) && (4) \end{aligned}$$

- Let $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}^*$, where $\boldsymbol{\epsilon}^* := U\boldsymbol{\gamma} + \boldsymbol{\epsilon} = \underbrace{\begin{pmatrix} U & I_{n\times n} \end{pmatrix}}_{A} \begin{pmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\epsilon} \end{pmatrix}$

$\overset{(2)}{\Rightarrow} \boldsymbol{\epsilon}^* \sim N_n(\mathbf{0}, V)$, where

$$\boxed{V} = A \begin{pmatrix} \mathcal{G} & \mathbf{0} \\ \mathbf{0} & R \end{pmatrix} A^t = \begin{pmatrix} U & I_{n\times n} \end{pmatrix} \begin{pmatrix} \mathcal{G} & \mathbf{0} \\ \mathbf{0} & R \end{pmatrix} \begin{pmatrix} U^t \\ I_{n\times n} \end{pmatrix}$$

$$= \begin{pmatrix} U\mathcal{G} & R \end{pmatrix} \begin{pmatrix} U^t \\ I_{n\times n} \end{pmatrix} = \boxed{U\mathcal{G}U^t + R}$$

Therefore (2) implies $\left.\begin{array}{l} \boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}^* \\ \boldsymbol{\epsilon}^* \sim N_n(\mathbf{0}, V) \end{array}\right\}$ (5) marginal model

- (2) or (3)+(4) implies (5), however (5) does not imply (3)+(4)
  $\Rightarrow$ If one is only interested in estimating $\boldsymbol{\beta}$ one can use the ordinary linear model (5)
  If one is interested in estimating $\boldsymbol{\beta}$ and γ one has to use model (3)+(4)

Likelihood Inference for LMM:
1) Estimation of β and γ for known $\mathcal{G}$ and R

Estimation of β: Using (5), we have as MLE or weighted LSE of $\boldsymbol{\beta}$

$$\boxed{\tilde{\boldsymbol{\beta}} := \left(X^t V^{-1} X\right)^{-1} X^t V^{-1} \mathbf{Y}} \qquad (6)$$

Recall: $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \Sigma), \quad \Sigma$ known, $\quad \Sigma = \Sigma^{1/2}\left(\Sigma^{1/2}\right)^t$

$$\Rightarrow \Sigma^{-1/2}\boldsymbol{Y} = \Sigma^{-1/2} X\boldsymbol{\beta} + \underbrace{\Sigma^{-1/2}\boldsymbol{\epsilon}} \qquad (7)$$

$$\sim N_n(\mathbf{0}, \underbrace{\Sigma^{-1/2}\Sigma\Sigma^{-1/2^t}}_{=I_n})$$

$\Rightarrow$ LSE of $\boldsymbol{\beta}$ in (7) :

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(X^t \Sigma^{-1/2^t} \Sigma^{-1/2} X\right)^{-1} X \Sigma^{-1/2^t} \Sigma^{-1/2} \boldsymbol{Y} \\ &= \left(X^t \Sigma^{-1} X\right)^{-1} X^t \Sigma^{-1} \boldsymbol{Y} \qquad (8) \end{aligned}$$

This estimate is called the weighted LSE

Exercise: Show that (8) is the MLE in $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \Sigma)$

Estimation of γ:

From (3) and (4) it follows that $\quad \boldsymbol{Y} \sim N_n(X\boldsymbol{\beta}, V) \qquad \boldsymbol{\gamma} \sim N_{mq}(\boldsymbol{0}, \mathcal{G})$

$$
\begin{aligned}
Cov(\boldsymbol{Y}, \boldsymbol{\gamma}) &= Cov(X\boldsymbol{\beta} + U\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \boldsymbol{\gamma}) \\
&= \underbrace{Cov(X\boldsymbol{\beta}, \boldsymbol{\gamma})}_{=0} + U \underbrace{Var(\boldsymbol{\gamma}, \boldsymbol{\gamma})}_{\mathcal{G}} + \underbrace{Cov(\boldsymbol{\epsilon}, \boldsymbol{\gamma})}_{=0} = U\mathcal{G}
\end{aligned}
$$

$$
\Rightarrow \begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{\gamma} \end{pmatrix} \sim N_{n+mq}\left(\begin{pmatrix} X\boldsymbol{\beta} \\ \boldsymbol{0} \end{pmatrix}, \begin{pmatrix} V & U\mathcal{G} \\ \mathcal{G}U^t & \mathcal{G} \end{pmatrix} \right)
$$

$$
\begin{aligned}
\text{Recall: } \boldsymbol{X} &= \begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{Z} \end{pmatrix} \sim N_p\left(\begin{pmatrix} \boldsymbol{\mu_Y} \\ \boldsymbol{\mu_Z} \end{pmatrix}, \begin{pmatrix} \Sigma_Y & \Sigma_{YZ} \\ \Sigma ZY & \Sigma_Z \end{pmatrix} \right) \\
\Rightarrow \boldsymbol{Z}|\boldsymbol{Y} &\sim N\left(\boldsymbol{\mu_{Z|Y}}, \Sigma_{Y|Z}\right) \text{ with} \\
\boldsymbol{\mu_{Z|Y}} &= \boldsymbol{\mu_Z} + \Sigma_{ZY}\Sigma_Y^{-1}\left(\boldsymbol{Y} - \boldsymbol{\mu_Y}\right), \Sigma_{Z|Y} = \Sigma_Z - \Sigma_{ZY}\Sigma_Y^{-1}\Sigma_{YZ}
\end{aligned}
$$

$$
\boxed{E(\boldsymbol{\gamma}|\boldsymbol{Y})} = \boldsymbol{0} + \mathcal{G}U^tV^{-10}(\boldsymbol{Y} - X\boldsymbol{\beta}) = \boxed{\mathcal{G}U^tV^{-1}(\boldsymbol{Y} - X\boldsymbol{\beta})} \tag{9}
$$

is the best linear unbiased predictor of γ (BLUP)

Therefore $\tilde{\gamma} := \mathcal{G}U^tV^{-1}(\boldsymbol{Y} - X\tilde{\boldsymbol{\beta}})$ is the empirical BLUP (EBLUP)

Joint maximization of log likelihood of $(\boldsymbol{Y}^t, \boldsymbol{\gamma}^t)^t$ with respect to $(\boldsymbol{\beta}, \boldsymbol{\gamma})$

$$
\begin{aligned}
f(\boldsymbol{y}, \boldsymbol{\gamma}) &= f(\boldsymbol{y}|\boldsymbol{\gamma}) \cdot f(\boldsymbol{\gamma}) \\
&\overset{(3)+(4)}{\propto} \exp\{-\tfrac{1}{2}(\boldsymbol{y} - X\boldsymbol{\beta} - U\boldsymbol{\gamma})^t R^{-1}(\boldsymbol{y} - X\boldsymbol{\beta} - U\boldsymbol{\gamma})\} \\
&\quad \exp\{-\tfrac{1}{2}\boldsymbol{\gamma}^t \mathcal{G}^{-1}\boldsymbol{\gamma}\} \\
\Rightarrow \ln f(\boldsymbol{y}, \boldsymbol{\gamma}) &= -\tfrac{1}{2}(\boldsymbol{y} - X\boldsymbol{\beta} - U\boldsymbol{\gamma})^t R^{-1}(\boldsymbol{y} - X\boldsymbol{\beta} - U\boldsymbol{\gamma}) \\
&\quad -\tfrac{1}{2} \underbrace{\boldsymbol{\gamma}^t \mathcal{G}^{-1}\boldsymbol{\gamma}}_{\text{penalty term for } \boldsymbol{\gamma}} + \text{constants ind. of } (\boldsymbol{\beta}, \boldsymbol{\gamma})
\end{aligned}
$$

So it is enough to minimize

$$
\begin{aligned}
Q(\boldsymbol{\beta}, \boldsymbol{\gamma}) &:= (\boldsymbol{y} - X\boldsymbol{\beta} - U\boldsymbol{\gamma})^t R^{-1}(\boldsymbol{y} - X\boldsymbol{\beta} - U\boldsymbol{\gamma}) - \boldsymbol{\gamma}^t \mathcal{G}^{-1}\boldsymbol{\gamma} \\
&= \boldsymbol{\gamma}^t R^{-1}\boldsymbol{\gamma} - 2\boldsymbol{\beta}^t X^t R^{-1}\boldsymbol{y} + 2\boldsymbol{\beta}^t X^t R^{-1}U\boldsymbol{\gamma} - 2\boldsymbol{\gamma}^t U^t R^{-1}\boldsymbol{y} \\
&\quad +\boldsymbol{\beta}^t X^t R^{-1}X\boldsymbol{\beta} + \boldsymbol{\gamma}^t U^t R^{-1}U\boldsymbol{\gamma} + \boldsymbol{\gamma}^t \mathcal{G}^{-1}\boldsymbol{\gamma}
\end{aligned}
$$

Recall:

$$
\begin{aligned}
f(\boldsymbol{\alpha}) &:= \boldsymbol{\alpha}^{\boldsymbol{t}}\boldsymbol{b} = \sum_{j=1}^{n} \alpha_j b_j \\
\frac{\partial}{\partial \alpha_i} f(\boldsymbol{\alpha}) &= b_j, \\
\frac{\partial}{\partial \boldsymbol{\alpha}} f(\boldsymbol{\alpha}) &= \boldsymbol{b}
\end{aligned}
$$

$$
\begin{aligned}
g(\boldsymbol{\alpha}) &:= \boldsymbol{\alpha}^{\boldsymbol{t}} A \boldsymbol{\alpha} = \sum_i \sum_j \alpha_i \alpha_j a_{ij} \\
\frac{\partial}{\partial \alpha_i} g(\boldsymbol{\alpha}) &= 2\alpha_i a_{ii} + \sum_{j=1, j\neq i}^{n} \alpha_j a_{ij} + \sum_{j=1, j\neq i}^{n} \alpha_j a_{ji} = 2\sum_{j=1}^{n} \alpha_j a_{ij} = 2A_i^t \boldsymbol{\alpha} \\
\frac{\partial}{\partial \boldsymbol{\alpha}} g(\boldsymbol{\alpha}) &= 2 \begin{pmatrix} A_1^t \\ \vdots \\ A_n^t \end{pmatrix} = 2A\boldsymbol{\alpha} \qquad A_i^t \text{ is ith row of A}
\end{aligned}
$$

Mixed Model Equation

$$\frac{\partial}{\partial \boldsymbol{\beta}} Q(\boldsymbol{\beta}, \boldsymbol{\gamma}) = -2X^t R^{-1} \boldsymbol{y} + 2X^t R^{-1} U \boldsymbol{\gamma} + 2X^t R^{-1} X \boldsymbol{\beta} \overset{\text{Set}}{=} 0$$

$$\frac{\partial}{\partial \boldsymbol{\gamma}} Q(\boldsymbol{\beta}, \boldsymbol{\gamma}) = -2U^t R^{-1} X \boldsymbol{\beta} - 2U^t R^{-1} \boldsymbol{y} + 2U^t R^{-1} U \boldsymbol{\gamma} + 2\mathcal{G}^{-1} \boldsymbol{\gamma} \overset{\text{Set}}{=} 0$$

$$\Leftrightarrow \quad \begin{array}{l} X^t R^{-1} X \widetilde{\boldsymbol{\beta}} + X^t R^{-1} U \widetilde{\boldsymbol{\gamma}} = X^t R^{-1} \boldsymbol{y} \\ U^t R^{-1} X \widetilde{\boldsymbol{\beta}} + (U^t R^{-1} U + \mathcal{G}^{-1}) \widetilde{\boldsymbol{\gamma}} = U^t R^{-1} \boldsymbol{y} \end{array}$$

$$\Leftrightarrow \quad \boxed{\begin{pmatrix} X^t R^{-1} X & X^t R^{-1} U \\ U^t R^{-1} U & U^t R^{-1} R + \mathcal{G}^{-1} \end{pmatrix} \begin{pmatrix} \widetilde{\boldsymbol{\beta}} \\ \widetilde{\boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} X^t R^{-1} \boldsymbol{y} \\ U^t R^{-1} \boldsymbol{y} \end{pmatrix}} \qquad (10)$$

Exercise: Show that $\widetilde{\boldsymbol{\beta}}, \widetilde{\gamma}$ defined by (8) and (9) respectively solve (10).

Define $C := \begin{pmatrix} X & U \end{pmatrix}$, $B := \begin{pmatrix} 0 & 0 \\ 0 & \mathcal{G}^{-1} \end{pmatrix}$

$$\begin{aligned} \Rightarrow C^t R^{-1} C &= \begin{pmatrix} X^t \\ U^t \end{pmatrix} R^{-1} \begin{pmatrix} X & U \end{pmatrix} = \begin{pmatrix} X^t R^{-1} \\ U^t R^{-1} \end{pmatrix} \begin{pmatrix} X & U \end{pmatrix} \\ &= \begin{pmatrix} X^t R^{-1} X & X^t R^{-1} U \\ U^t R^{-1} X & U^t R^{-1} U \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \Rightarrow (10) \quad &\Leftrightarrow \quad (C^t R^{-1} C + B) \begin{pmatrix} \widetilde{\boldsymbol{\beta}} \\ \widetilde{\gamma} \end{pmatrix} = C^t R^{-1} \boldsymbol{y} \\ &\Leftrightarrow \quad \begin{pmatrix} \widetilde{\boldsymbol{\beta}} \\ \widetilde{\gamma} \end{pmatrix} = (C^t R^{-1} C + B)^{-1} C^t R^{-1} \boldsymbol{y} \end{aligned}$$

2) Estimation for unknown covariance structure

We assume now in the marginal model (5)

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}^*, \boldsymbol{\epsilon}^* \sim N_n(\mathbf{0}, V)$$

with $V = U\mathcal{G}U^t + R$, that $\mathcal{G}$ and R are only known up to the variance parameter $\boldsymbol{\vartheta}$, i.e. we write

$$V(\boldsymbol{\vartheta}) = U\mathcal{G}(\boldsymbol{\vartheta})U^t + R(\boldsymbol{\vartheta})$$

ML Estimation in extended marginal model

$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}^*, \boldsymbol{\epsilon}^* \sim N_n(\mathbf{0}, V(\boldsymbol{\vartheta}))$ with $V(\boldsymbol{\vartheta}) = U\mathcal{G}(\boldsymbol{\vartheta})U^t + R(\boldsymbol{\vartheta})$
loglikelihood for $(\boldsymbol{\beta}, \boldsymbol{\vartheta})$:

$$l(\boldsymbol{\beta}, \boldsymbol{\vartheta}) = -\frac{1}{2}\{\ln|V(\boldsymbol{\vartheta})| + (\boldsymbol{y} - X\boldsymbol{\beta})^t V(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y} - X\boldsymbol{\beta})\} + \text{const. ind. of } \boldsymbol{\beta}, \boldsymbol{\vartheta} \qquad (11)$$

If we maximize (11) for fixed $\boldsymbol{\vartheta}$ with regard to $\boldsymbol{\beta}$, we get

$$\widetilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta}) := (X^t V(\boldsymbol{\vartheta})^{-1} X)^{-1} X^t V(\boldsymbol{\vartheta})^{-1} \boldsymbol{y}$$

Then the profile log likelihood is

$$\begin{aligned} l_p(\boldsymbol{\vartheta}) &:= l(\widetilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta}), \boldsymbol{\vartheta}) \\ &= -\tfrac{1}{2}\{\ln|V(\boldsymbol{\vartheta})| + (\boldsymbol{y} - X\widetilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))^t V(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y} - X\widetilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))\} \end{aligned}$$

Maximizing $l_p(\boldsymbol{\vartheta})$ wrt to $\boldsymbol{\vartheta}$ gives MLE $\hat{\boldsymbol{\vartheta}}_{ML}$. $\hat{\boldsymbol{\vartheta}}_{ML}$ is however biased; this is why one uses often restricted ML estimation (REML)

Restricted ML Estimation in extended marginal model

Here we use for the estimation of $\boldsymbol{\vartheta}$ the marginal log likelihood:

$$l_R(\boldsymbol{\vartheta}) := \ln\left(\int L(\boldsymbol{\beta}, \boldsymbol{\vartheta}) \mathrm{d}\boldsymbol{\beta}\right)$$

$$\int L(\boldsymbol{\beta}, \boldsymbol{\vartheta}) \mathrm{d}\boldsymbol{\beta} = \int \frac{1}{(2\pi)^{n/2}} |V(\boldsymbol{\vartheta})|^{-1/2} + \exp\{-\tfrac{1}{2}(\boldsymbol{y} - X\boldsymbol{\beta})^t V(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y} - X\boldsymbol{\beta})\} \mathrm{d}\boldsymbol{\beta}$$

Consider:

$$(\boldsymbol{y} - X\boldsymbol{\beta})^t V(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y} - X\boldsymbol{\beta}) = \boldsymbol{\beta}^t \underbrace{X^t V(\boldsymbol{\vartheta})^{-1} X}_{A(\boldsymbol{\vartheta})} \boldsymbol{\beta} - 2\boldsymbol{y}^t V(\boldsymbol{\vartheta})^{-1} X\boldsymbol{\beta} + \boldsymbol{y}^t V(\boldsymbol{\vartheta})^{-1}\boldsymbol{y}$$

$$= (\boldsymbol{\beta} - B(\boldsymbol{\vartheta})\boldsymbol{y})^t A(\boldsymbol{\vartheta})(\boldsymbol{\beta} - B(\boldsymbol{\vartheta})\boldsymbol{y}) + \boldsymbol{y}^t V(\boldsymbol{\vartheta})^{-1} - \boldsymbol{y}^t B(\boldsymbol{\vartheta})^t A(\boldsymbol{\vartheta}) B(\boldsymbol{\vartheta})\boldsymbol{y}$$

where $B(\boldsymbol{\vartheta}) := A(\boldsymbol{\vartheta})^{-1} X^t V(\boldsymbol{\vartheta})^{-1}$

(Note that $B(\boldsymbol{\vartheta})^t A(\boldsymbol{\vartheta}) = V(\boldsymbol{\vartheta})^{-1} X A(\boldsymbol{\vartheta})^{-1} A(\boldsymbol{\vartheta}) = V(\boldsymbol{\vartheta})^{-1} X$)

Therefore we have

$$\int L(\boldsymbol{\beta},\boldsymbol{\vartheta})\mathrm{d}\boldsymbol{\beta} = \frac{|V(\boldsymbol{\vartheta})|^{-1/2}}{(2\pi)^{n/2}}\exp\{-\tfrac{1}{2}(\boldsymbol{y^t}[V(\boldsymbol{\vartheta})^{-1}+B(\boldsymbol{\vartheta})^tA(\boldsymbol{\vartheta})B(\boldsymbol{\vartheta})]\boldsymbol{y}\} \cdot \underbrace{\int \exp\{-\frac{1}{2}(\boldsymbol{\beta}-B(\boldsymbol{\vartheta})\boldsymbol{y})^tA(\boldsymbol{\vartheta})(\boldsymbol{\beta}-B(\boldsymbol{\vartheta})\boldsymbol{y})\}\mathrm{d}\boldsymbol{\beta}}_{\frac{(2\pi)^{p/2}}{|A(\boldsymbol{\vartheta})^{-1}|^{-1/2}} \quad \text{(Variance is } A(\boldsymbol{\vartheta})^{-1}\text{!)}} \tag{12}$$

Now

$$\boxed{(\boldsymbol{y}-X\widetilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))^tV(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y}-X\widetilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))}$$

$$= \boldsymbol{y^t}V(\boldsymbol{\vartheta})^{-1}\boldsymbol{y}-2\boldsymbol{y^t}V(\boldsymbol{\vartheta})^{-1}X\widetilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta})+\widetilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta})^t\underbrace{X^tV(\boldsymbol{\vartheta})^{-1}X}_{A(\boldsymbol{\vartheta})}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta})$$

$$= \boldsymbol{y^t}V(\boldsymbol{\vartheta})^{-1}\boldsymbol{y}-2\boldsymbol{y^t}V(\boldsymbol{\vartheta})^{-1}XB(\boldsymbol{\vartheta})\boldsymbol{y}+\boldsymbol{y}^tB(\boldsymbol{\vartheta})^tA(\boldsymbol{\vartheta})B(\boldsymbol{\vartheta})\boldsymbol{y}$$

$$\boxed{= \boldsymbol{y^t}V(\boldsymbol{\vartheta})^{-1}\boldsymbol{y}-\boldsymbol{y}^tB(\boldsymbol{\vartheta})^tA(\boldsymbol{\vartheta})B(\boldsymbol{\vartheta})\boldsymbol{y}}$$

Here we used:

$$\widetilde{\boldsymbol{\beta}} = (X^tV(\boldsymbol{\vartheta})^{-1}X)^{-1}X^tV(\boldsymbol{\vartheta})^{-1}\boldsymbol{y} = A(\boldsymbol{\vartheta})^{-1}X^tV(\boldsymbol{\vartheta})^{-1}\boldsymbol{y} = B(\boldsymbol{\vartheta})\boldsymbol{y}$$

and

$$B(\boldsymbol{\vartheta})^tA(\boldsymbol{\vartheta})B(\boldsymbol{\vartheta}) = V(\boldsymbol{\vartheta})^{-1}XA(\boldsymbol{\vartheta})^{-1}A(\boldsymbol{\vartheta})B(\boldsymbol{\vartheta}) = V(\boldsymbol{\vartheta})^{-1}XB(\boldsymbol{\vartheta})$$

Therefore we can rewrite (12) as

$$
\begin{aligned}
\int L(\boldsymbol{\beta},\boldsymbol{\vartheta})\mathrm{d}\boldsymbol{\beta} &= \frac{|V(\boldsymbol{\vartheta})|^{-1/2}}{(2\pi)^{n/2}}\exp\{-\tfrac{1}{2}(\boldsymbol{y}-X\widetilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))^t V(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y}-X\widetilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))\} \\
& \cdot \frac{(2\pi)^{n/2}}{|A(\boldsymbol{\vartheta})^{-1}|^{-1/2}} \qquad\qquad |A(\boldsymbol{\vartheta})^{-1}| = \tfrac{1}{|A|} \\
\Rightarrow l_R(\boldsymbol{\theta}) &= \ln(\textstyle\int L(\boldsymbol{\beta},\boldsymbol{\vartheta})\mathrm{d}\boldsymbol{\beta}) \\
&= -\tfrac{1}{2}\{\ln|V(\boldsymbol{\vartheta})| + (\boldsymbol{y}-X\widetilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))^t V(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y}-X\widetilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))\} \\
& -\tfrac{1}{2}\ln|A(\boldsymbol{\vartheta})| + C \\
&= l_p(\boldsymbol{\vartheta}) - \tfrac{1}{2}\ln|A(\boldsymbol{\vartheta})| + C
\end{aligned}
$$

Therefore the restricted ML (REML) of $\boldsymbol{\vartheta}$ is given by $\hat{\boldsymbol{\vartheta}}_{\boldsymbol{REML}}$ which maximizes

$$l_R(\boldsymbol{\vartheta}) = l_p(\boldsymbol{\vartheta}) - \frac{1}{2}\ln|X^t V(\boldsymbol{\vartheta})^{-1}X|$$

Summary: Estimation in LMM with unknown cov.

For the linear mixed model

$$\boldsymbol{Y} = X\boldsymbol{\beta} + U\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \begin{pmatrix} \gamma \\ \boldsymbol{\epsilon} \end{pmatrix} \sim N_{mq+n}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathcal{G}(\boldsymbol{\vartheta}) & 0_{mq \times n} \\ 0_{n \times mq} & R(\boldsymbol{\vartheta}) \end{pmatrix} \right)$$

with $V(\boldsymbol{\vartheta}) = U\mathcal{G}(\boldsymbol{\vartheta})U^t + R(\boldsymbol{\vartheta})$

the covariance parameter vector $\boldsymbol{\vartheta}$ is estimated by either

$\hat{\boldsymbol{\vartheta}}_{\boldsymbol{ML}}$ which maximizes

$$l_p(\boldsymbol{\vartheta}) = -\tfrac{1}{2}\{\ln|V(\boldsymbol{\vartheta})| + (\boldsymbol{y} - X\widetilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))^t V(\boldsymbol{\vartheta})^{-1}(\boldsymbol{y} - X\widetilde{\boldsymbol{\beta}}(\boldsymbol{\vartheta}))\}$$

where $\widetilde{\boldsymbol{\beta}} = (X^t V(\boldsymbol{\vartheta})^{-1} X)^{-1} X^t V(\boldsymbol{\vartheta})^{-1}\boldsymbol{Y}$

or by

$\hat{\boldsymbol{\vartheta}}_{\boldsymbol{REML}}$ which maximizes $\quad l_R(\boldsymbol{\vartheta}) = l_p(\boldsymbol{\vartheta}) - \tfrac{1}{2}\ln|X^t V(\boldsymbol{\vartheta})^{-1} X|$

The fixed effects $\boldsymbol{\beta}$ and random effects γ are estimated by

$$\widehat{\boldsymbol{\beta}} = \left(X^t \widehat{V}^{-1} X\right)^{-1} X^t \widehat{V}^{-1}\boldsymbol{Y}$$

$$\widehat{\gamma} = \widehat{\mathcal{G}} U^t \widehat{V}^{-1}(\boldsymbol{Y} - X\widehat{\boldsymbol{\beta}})$$

where $\widehat{V} = V(\widehat{\boldsymbol{\vartheta}}_{\boldsymbol{ML}})$ or $V(\widehat{\boldsymbol{\vartheta}}_{\boldsymbol{REML}})$

Special Case

(Dependence on ϑ is ignored to ease notation)

$$\mathcal{G} = \begin{pmatrix} D & & \\ & \ddots & \\ & & D \end{pmatrix}, U = \begin{pmatrix} U_1 & & \\ & \ddots & \\ & & U_m \end{pmatrix}, R = \begin{pmatrix} \Sigma_1 & & \\ & \ddots & \\ & & \Sigma_m \end{pmatrix},$$

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix}, \boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y_1} \\ \vdots \\ \boldsymbol{Y_m} \end{pmatrix}$$

$$\Rightarrow V = U\mathcal{G}U^t + R = \begin{pmatrix} U_1DU_1^t + \Sigma_1 & & 0 \\ & \ddots & \\ 0 & & U_mDU_m^t + \Sigma_m \end{pmatrix} \quad \text{(blockdiagonal)}$$

$$= \begin{pmatrix} V_1 & & \\ & \ddots & \\ & & V_m \end{pmatrix} \qquad \text{where } V_i := U_iDU_i^t + \Sigma_i$$

Define

$\widehat{V}_i := U_i D(\widehat{\boldsymbol{\vartheta}}) U_i^t + \Sigma_i(\widehat{\boldsymbol{\vartheta}})$, where $\widehat{\boldsymbol{\vartheta}} = \widehat{\boldsymbol{\vartheta}}_{\boldsymbol{ML}}$ or $\widehat{\boldsymbol{\vartheta}}_{\boldsymbol{REML}}$

$$\begin{aligned}
\Rightarrow \boxed{\widehat{\boldsymbol{\beta}}} &= (X^t \widehat{V}^{-1} X)^{-1} X^t \widehat{V}^{-1} \boldsymbol{Y} \\
&= \boxed{\sum_{i=1}^m X_i^t \widehat{V}_i^{-1} X_i)^{-1} (\sum_{i=1}^m X_i^t \widehat{V}_i^{-1} \boldsymbol{Y_i})}
\end{aligned}$$

and

$\widehat{\gamma} = \widehat{\mathcal{G}} U^t \widehat{V}^{-1} (\boldsymbol{Y} - X\widehat{\boldsymbol{\beta}})$ has components

$$\boxed{\widehat{\boldsymbol{\gamma}}_{\boldsymbol{i}} = D(\widehat{\gamma}) U_i \widehat{V}_i^{-1} (\boldsymbol{y_i} - X_i \widehat{\boldsymbol{\beta}})}$$

3) Confidence intervals and hypothesis tests

Since $\boldsymbol{Y} \sim N(X\boldsymbol{\beta}, V(\boldsymbol{\vartheta}))$ holds, an approximation to the covariance of $\widehat{\boldsymbol{\beta}} = \left(X^t V^{-1}(\widehat{\boldsymbol{\vartheta}})X\right)^{-1} X^t V^{-1}(\widehat{\boldsymbol{\vartheta}})\boldsymbol{Y}$ is given by

$$A(\widehat{\boldsymbol{\vartheta}}) := (X^t V^{-1}(\widehat{\boldsymbol{\vartheta}})X)^{-1}$$

Note: here one assumes that $V(\widehat{\boldsymbol{\vartheta}})$ is fixed and does not depend on $\boldsymbol{Y}$.
Therefore $\widehat{\sigma}_j := (X^t V^{-1}(\widehat{\boldsymbol{\vartheta}})X)^{-1}_{jj}$ are considered as estimates of $Var(\widehat{\beta}_j)$.
Therefore

$$\widehat{\beta}_j \pm z_{1-\alpha/2}\sqrt{(X^t V^{-1}(\widehat{\boldsymbol{\vartheta}})X)^{-1}_{jj}}$$

gives an approximate $100(1-\alpha)\%$ CI for β_j.

It is expected that $(X^t V^{-1}(\widehat{\boldsymbol{\vartheta}})X)^{-1}_{jj}$ underestimates $Var(\widehat{\beta}_j)$ since the variation in $\widehat{\boldsymbol{\vartheta}}$ is not taken into account.
A full Bayesian analysis using MCMC methods is preferable to these approximations.

Under the assumption that $\widehat{\boldsymbol{\beta}}$ is asymptotically normal with mean $\boldsymbol{\beta}$ and covariance matrix $A(\boldsymbol{\vartheta})$, then the usal hypothesis tests can be done; i.e. for

- $H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0$

  Reject $H_0 \Leftrightarrow |t_j| = |\frac{\widehat{\beta}_j}{\widehat{\sigma}_j}| > z_{1-\alpha/2}$

- $H_0 : C\boldsymbol{\beta} = \boldsymbol{d}$ versus $H_1 : C\boldsymbol{\beta} \neq \boldsymbol{d}$ rank$(C) = r$

  Reject $H_0 \Leftrightarrow W := (C\widehat{\boldsymbol{\beta}} - \boldsymbol{d})^t (C^t A(\widehat{\boldsymbol{\vartheta}})C)^{-1} (C\widehat{\boldsymbol{\beta}} - \boldsymbol{d}) > \chi^2_{1-\alpha,r}$
  (Wald-Test)

  or

  Reject $H_0 \Leftrightarrow -2[l(\widehat{\boldsymbol{\beta}}, \widehat{\gamma}) - l(\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}, \widehat{\gamma}_{\boldsymbol{R}})] > \chi^2_{1-\alpha,r}$
  where $\widehat{\boldsymbol{\beta}}, \widehat{\gamma}$ estimates in unrestricted model
  $\widehat{\boldsymbol{\beta}}_{\boldsymbol{R}}, \widehat{\gamma}_{\boldsymbol{R}}$ estimates in restricted model $(C\boldsymbol{\beta} = \boldsymbol{d})$
  (Likelihood Ratio Test)

References

Fahrmeir, L., T. Kneib, and S. Lang (2007). *Regression: Modelle, Methoden und Anwendungen*. Berlin: Springer-Verlag.

West, B., K. E. Welch, and A. T. Galecki (2007). *Linear Mixed Models: a practical guide using statistical software*. Boca Raton: Chapman-Hall/CRC.