Linear Mixed Effect Models & Generalized Likelihood Ratio Test

- LMEMs in a nut shell
- Estimating LMEMs
- Asymptotic Results for Maximum Likelihood Estimators
- Principles of hypothesis testing
- Generalized Likelihood Ratio Test

- LMEM is a regression model for a random variable Y from the exponential family (*Gaussian*, gamma, Bernoulli, categorial, exponential, beta, ...)
- Extension of a standard linear regression model that allows to model *non-iid variance-covariance patterns*
- Tabular dataset: $[[\mathrm{x}^n, \mathrm{z}^n, y^n]_{n=1}^N]^\top$ where $x \in \mathbb{R}^p$, $z \in \mathbb{R}^q$ and $y \in \mathbb{R}$

General form of LMEM for a single observation

$$y^n = \mathrm{x}^n \boldsymbol{\beta} + \mathrm{z}^n \mathrm{b} + \boldsymbol{\epsilon}^n$$

- $\boldsymbol{\beta}$ fixed effect parameters
- Random effect parameters b and errors ϵ modeled by Gaussian variables
- $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are estimated from the data
- See [McCulloch and Searle, 2001, Demidenko, 2013, Wood, 2017, Riezler and Hagmann, 2021].

Matrix notation for the whole dataset

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

- $\mathrm{y} \in \mathbb{R}^N$, $\mathrm{X} \in \mathbb{R}^{N\times p}$ and $\mathrm{Z} \in \mathbb{R}^{N\times q}$ denote the stacked $y^n/x^n/z^n$
- $\mathrm{b} \sim \mathcal{N}(0, \psi_{\boldsymbol{\theta}})$, $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \Lambda_{\boldsymbol{\theta}})$ where $\psi_{\boldsymbol{\theta}}$ and $\Lambda_{\boldsymbol{\theta}}$ are positive definite
- $\mathbb{E}[\mathrm{y}|\mathrm{X}] = \mathrm{X}\boldsymbol{\beta}$ and $\mathbb{V}[\mathrm{y}] = \mathrm{Z}\psi_{\boldsymbol{\theta}}\mathrm{Z}^\top + \Lambda_{\boldsymbol{\theta}}$

Data distribution for Gaussian Y

$$\mathrm{Y} \sim \mathcal{N}(\mathrm{X}\boldsymbol{\beta}, \mathrm{Z}\psi_{\theta}\mathrm{Z}^\top + \Lambda_{\boldsymbol{\theta}})$$

REMARK I: $f_{\mathrm{y}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\mathrm{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathrm{y}-\boldsymbol{\mu})\right)$
REMARK II: Inversion of $\mathrm{Z}\psi_{\theta}\mathrm{Z}^\top + \Lambda_{\boldsymbol{\theta}}$ is computationally expensive

AIM: Find an expression for f_y that avoids the inversion of $Z\psi_\theta Z^\top + \Lambda_{\boldsymbol{\theta}}$

Elementary facts

$$\text{Marginal Distribution: } f_y = \int f_{y,b}\, db$$
$$\text{Conditional Distribution: } f_{y,b} = f_{y|b} f_b$$

Conditional distribution of Y given b

$$y|b \sim \mathcal{N}(X\boldsymbol{\beta} + Zb, \Lambda_{\boldsymbol{\theta}})$$
$$b \sim \mathcal{N}(0, \psi_{\boldsymbol{\theta}})$$

$\Rightarrow$ Just need to invert Λ_θ and ψ_θ which are typically simple and sparse

Result: Marginal distribution

$$f_{\mathrm{y}}(\boldsymbol{\beta}, \boldsymbol{\theta}) \propto |\mathrm{Z}^\top \Lambda_{\boldsymbol{\theta}}^{-1} \mathrm{Z} + \psi_{\boldsymbol{\theta}}^{-1}|^{-1/2} f_{\mathrm{y}|\hat{\mathrm{b}}}(\boldsymbol{\beta}, \boldsymbol{\theta}) f_{\hat{\mathrm{b}}}(\boldsymbol{\theta})$$

$$\text{where} \quad \hat{\mathrm{b}} := \underset{\mathrm{b} \in \mathbb{R}^q}{\operatorname{argmax}} \log(f_{\mathrm{y},\mathrm{b}}(\boldsymbol{\beta}, \boldsymbol{\theta}))$$

PROOF: $f_y = \int \exp\left(\log(f_{y,\mathrm{b}})\right) d\mathrm{b}$ & Taylor Series of $\log(f_{\mathrm{y},\mathrm{b}})$ around $\hat{\mathrm{b}}$

ML Objective based on the marginal distribution

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathrm{b}) = -(\mathrm{y} - \mathrm{X}\boldsymbol{\beta} + \mathrm{Z}\mathrm{b})^\top \Lambda_{\boldsymbol{\theta}}^{-1} (\mathrm{y} - \mathrm{X}\boldsymbol{\beta} + \mathrm{Z}\mathrm{b}) - \mathrm{b}^\top \psi_{\boldsymbol{\theta}}^{-1} \mathrm{b}$$
$$- \log(|\Lambda_{\boldsymbol{\theta}}|) - \log(|\psi_{\boldsymbol{\theta}}|) - \log(|\mathrm{Z}^\top \Lambda_{\boldsymbol{\theta}}^{-1} \mathrm{Z} + \psi_{\boldsymbol{\theta}}^{-1}|)$$

What happened to $\hat{\mathrm{b}}$?

Assume we know $\boldsymbol{\theta}$: MLE for $\boldsymbol{\beta}$ and b (Henderson equations)

$$\begin{bmatrix} \mathsf{X}^\top \Lambda_{\boldsymbol{\theta}}^{-1}\mathsf{X} & \mathsf{X}^\top \Lambda_{\boldsymbol{\theta}}^{-1}\mathsf{Z} \\ \mathsf{Z}^\top \Lambda_{\boldsymbol{\theta}}^{-1}\mathsf{X} & \mathsf{Z}^\top \Lambda_{\boldsymbol{\theta}}^{-1}\mathsf{Z} + \psi_{\boldsymbol{\theta}}^{-1} \end{bmatrix} \cdot \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathsf{b}} \end{bmatrix} = \begin{bmatrix} \mathsf{X}^\top \Lambda_{\boldsymbol{\theta}}^{-1}\mathsf{y} \\ \mathsf{Z}^\top \Lambda_{\boldsymbol{\theta}}^{-1}\mathsf{y} \end{bmatrix}$$

How to estimate $\boldsymbol{\theta}$

- ML: Find $\hat{\boldsymbol{\theta}}$ by optimizing the profile likelihood $L(\hat{\boldsymbol{\beta}}, \hat{\mathsf{b}}, \boldsymbol{\theta})$
  - No closed form solution
  - Computations can be sped up due to simple and sparse matrices
  - Convergence can be sped up combining EM and Newton-Raphson methods
- REML (Restricted Maximum Likelihood)

Important Large Sample Results for MLE

Consistency [Nie, 2006]

$$\forall \epsilon > 0: \quad P\left(d(\hat{\boldsymbol{\beta}}_N, \boldsymbol{\beta}) > \epsilon\right) \xrightarrow{N \to \infty} 0$$

- $\hat{\boldsymbol{\beta}}$ is asymptotically unbiased
- $\mathbb{V}[\hat{\boldsymbol{\beta}}]$ decreases with N

Distribution of $\hat{\boldsymbol{\beta}}$ given $\boldsymbol{\theta}$

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \overset{app}{\sim} \mathcal{N}(0, (\mathsf{X}^\top(\mathsf{Z}\psi_\theta\mathsf{Z}^\top + \Lambda_{\boldsymbol{\theta}})^{-1}\mathsf{X})^{-1})$$

- Allows statistical inference for $\hat{\boldsymbol{\beta}}$
- Results still holds when $\boldsymbol{\theta}$ is replaced by $\hat{\boldsymbol{\theta}}$

Fundamental goal

Decide between two mutually exclusive and exhaustive sets of hypotheses, called *null hypothesis* H_0 and *alternative hypothesis* H_1 about the data generating probability measures by evidence obtained from observed random samples.

$\Rightarrow$ The test decision is a random event!

Important probabilities of a hypothesis test

$$\begin{aligned} \text{Type-I error probability:} \quad & P(\text{reject } H_0 \text{ while } H_0 \text{ is true}) \\ \text{Power:} \quad & P(\text{reject } H_0 \text{ while } H_1 \text{ is true}) \end{aligned}$$

Conducting a hypothesis test

1. Define a test statistic T which allows to discriminate between H_0 and H_1
2. Assume H_0 is true and derive the distribution of T
3. Set the Type-I error probability to a predefined level α
4. Reject H_0 when $P(|T| > t_{obs}) \leq \alpha$

Notes

- It is important that the actual α_{act} equals the nominal α, otherwise the test either wastes power ($\alpha_{act} < \alpha$) or is not admissible ($\alpha_{act} > \alpha$)
- Usually 2 is the difficult step. If one resorts to resampling based methods one has to be careful to implement an appropriate resampling mechanism [Canty et al., 2006].

Hypothesis

Suppose we have two *nested models* describing the same data $f(\Theta_0)$ and $f(\Theta_1)$ where $\Theta_0 \subseteq \Theta_1$ with $df_0 := \dim(\Theta_0) < \dim(\Theta_1) =: df_1$ are the parameter spaces of the models. We want to test if $m(\Theta_1)$ is more appropriate.

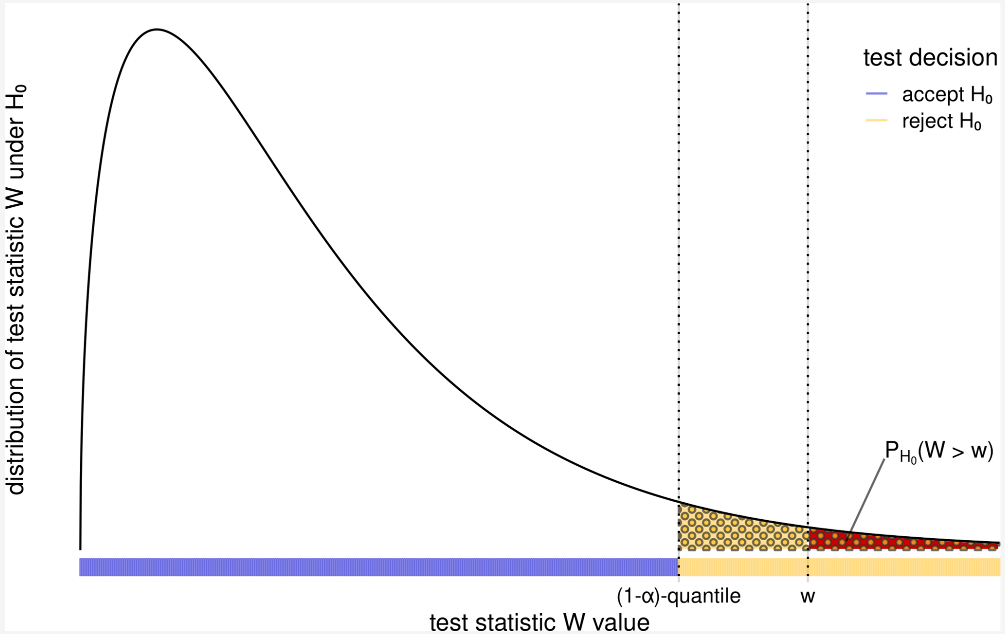
$$H_0\colon \quad \theta \in \Theta_0$$
$$H_1\colon \quad \theta \in \Theta_1 \setminus \Theta_0$$

Likelihood ratio

$$\lambda := \frac{f(\hat{\theta}_0^{ML})}{f(\hat{\theta}_1^{ML})} = \frac{\ell_0^*}{\ell_1^*}$$

Interpretation of $0 < \lambda \leq 1$:

- Values of λ close to 1 suggest that restricted model (H_0) explains the data as well as more complex model (H_1)
- H_0 should be accepted for large values of λ
- Conversely, values close to 0 suggest that the data are not very compatible with the parameter values in the restricted model
- H_0 should be rejected for small values of λ



Test statistic [Wilks, 1938, van der Vaart, 1998]

$$W = -2 \log \lambda = 2(\log \ell_1^* - \log \ell_0^*) \overset{H_0}{\sim} \chi^2_{df_1 - df_0}$$

Reject H_0 if $p := P_{H_0}(W > w_{obs}) \leq \alpha$