Random-Intercept Linear Mixed Model

Step-by-Step Estimation and Theory

Suaray

February 2026

Overview

This document presents a **slow, mechanical walkthrough** of estimation for the random-intercepts linear mixed model, closely mirroring how modern software implements maximum likelihood and REML estimation.

1 Model and Unknown Parameters

1.1 Index Form

$$y_{ij} = \beta_0 + \mathbf{x}_{ij}^\top \boldsymbol{\beta} + u_j + \varepsilon_{ij},$$

with

$$u_j \sim \mathcal{N}(0, \sigma_u^2), \qquad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2), \qquad u_j \perp \varepsilon_{ij}.$$

1.2 Unknown Parameters

- Fixed effects: $(\beta_0, \boldsymbol{\beta})$
- Variance components: $(\sigma_u^2, \sigma_\varepsilon^2)$

Unlike OLS, estimation requires learning the *covariance structure* induced by the random effects.

2 Matrix Formulation

Stack all N observations:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon},$$

where

$$\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{G}), \quad \mathbf{G} = \sigma_u^2 \mathbf{I}_J,$$
$$\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \quad \mathbf{R} = \sigma_\varepsilon^2 \mathbf{I}_N.$$

Thus

$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}), \qquad \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^\top + \mathbf{R}.$$

For random intercepts, $\mathbf{Z}$ is a group-membership indicator matrix, yielding a block-structured covariance.

3 GLS if V Were Known

If $(\sigma_u^2, \sigma_\varepsilon^2)$ were known, fixed effects are estimated via generalized least squares:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{y}.$$

Key point: conditional on $\mathbf{V}$, estimation of β is a single GLS step.

4 Predicting Random Effects (BLUP)

Random effects are *predicted*, not estimated:

$$\hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}^\top \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

For group j:

$$\hat{u}_j = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2 / n_j} \Big(\bar{y}_j - \overline{\mathbf{x}}_j^\top \hat{\boldsymbol{\beta}} \Big).$$

- Small $n_j \Rightarrow$ strong shrinkage
- Large $n_j \Rightarrow$ weak shrinkage

5 Why Estimation Is Iterative

The covariance matrix

$$\mathbf{V}(\sigma_u^2, \sigma_\varepsilon^2)$$

depends on unknown parameters, so estimation proceeds iteratively.

1. Estimate β via GLS given current $\mathbf{V}$
2. Update variance components using residual information

6 Maximum Likelihood for Variance Components

The Gaussian log-likelihood is

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\tfrac{1}{2} \Big[\log |\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \Big] + C,$$

where $\boldsymbol{\theta} = (\sigma_u^2, \sigma_\varepsilon^2)$.

6.1 Profiling Out Fixed Effects

For each candidate $\boldsymbol{\theta}$:

1. compute $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ via GLS
2. substitute into ℓ
3. maximize over $\boldsymbol{\theta}$

7 REML Estimation

REML maximizes

$$\ell_R(\boldsymbol{\theta}) = -\tfrac{1}{2}\left[\log|\mathbf{V}| + \log|\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X}| + \mathbf{y}^\top\mathbf{P}\mathbf{y}\right] + C,$$

where

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{V}^{-1}.$$

REML uses variation orthogonal to the fixed-effects subspace.

8 Iterative REML Algorithm (Software View)

1. Initialize $(\sigma_u^{2(0)}, \sigma_\varepsilon^{2(0)})$
2. Form $\mathbf{V}^{(k)}$
3. Update fixed effects via GLS
4. Compute BLUPs
5. Update variance components (Fisher scoring / AI)
6. Check convergence

$$\widehat{\mathrm{ICC}} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_\varepsilon^2}.$$

Appendix

A Linear Algebra Facts

A.1 Quadratic Forms

For symmetric A:

$$\frac{\partial}{\partial x}(x^\top A x) = 2Ax.$$

A.2 Matrix Inverses

$$(A^\top)^{-1} = (A^{-1})^\top, \qquad (A^{-1})^\top = (A^\top)^{-1}.$$

A.3 Woodbury Identity

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}.$$

B Multivariate Normal Facts

B.1 Definition

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \iff f(\mathbf{x}) \propto \exp\Big(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\Big).$$

B.2 Conditional Distribution

If

$$\begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix}, \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \right),$$

then

$$\mathbb{E}[\mathbf{a} \mid \mathbf{b}] = \boldsymbol{\mu}_a + \Sigma_{ab}\Sigma_{bb}^{-1}(\mathbf{b} - \boldsymbol{\mu}_b).$$

B.3 Covariance Matrix

$$\mathrm{Var}(\mathbf{x}) = \mathbb{E}[(\mathbf{x} - \mathbb{E}\mathbf{x})(\mathbf{x} - \mathbb{E}\mathbf{x})^\top].$$

C Numerical Methods for REML

C.1 Fisher Scoring

Updates parameters using expected Hessians:

$$\theta^{(k+1)} = \theta^{(k)} + I(\theta^{(k)})^{-1} s(\theta^{(k)}).$$

C.2 Average Information (AI)

Approximates Fisher scoring with half observed + half expected curvature; efficient for large mixed models.

C.3 Convergence Issues

- Flat likelihoods when variance ≈ 0
- Boundary solutions
- Large sparse matrix inversions

Modern implementations exploit sparsity and block structure.