Linear Mixed Model-LMM

Emre Karaman

Center for Quantitative Genetics and Genomics
Aarhus University

NADAS course February 2026

What we are going to do

- Convert simple R code for LMM to Julia equivalent
- Get familiarize with forming and working with matrices in Julia
- Apply some of the functions that Bjarke's lecture covered yesterday

For the exercises, please navigate to
`https://github.com/MrodesBook/Examples_R/tree/main`

Simple LMM

Model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Phi} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} \right)$$

- Observation = systematic part + random part
- Systematic part: sex, age, group, year of recording, etc.
- Random part: variation not explained by systematic component

BLUP - Best Linear Unbiased Prediction

Fixed (BLUE) and random effects (BLUP) are solutions to the mixed model equations:

$$\begin{bmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \boldsymbol{\Phi}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{y} \end{bmatrix}$$

Generally:

$$\boldsymbol{\Phi} = \boldsymbol{\Gamma} \sigma_u^2 \qquad \mathbf{D} = \mathbf{I} \sigma_e^2$$

Then:

$$\begin{bmatrix} \mathbf{X}^\top \mathbf{X} & \mathbf{X}^\top \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{X} & \mathbf{Z}^\top \mathbf{Z} + \alpha \boldsymbol{\Gamma}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \mathbf{y} \\ \mathbf{Z}^\top \mathbf{y} \end{bmatrix}$$

$$\alpha = \frac{\sigma_e^2}{\sigma_u^2}$$

Solving the System

- Need to solve

$$\begin{bmatrix} \mathbf{X}^\top\mathbf{X} & \mathbf{X}^\top\mathbf{Z} \\ \mathbf{Z}^\top\mathbf{X} & \mathbf{Z}^\top\mathbf{Z} + \alpha\mathbf{\Gamma}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top\mathbf{y} \\ \mathbf{Z}^\top\mathbf{y} \end{bmatrix}$$

$$\mathbf{C}\boldsymbol{\theta} = \mathbf{r}$$

- Re-arrenging gives

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top\mathbf{X} & \mathbf{X}^\top\mathbf{Z} \\ \mathbf{Z}^\top\mathbf{X} & \mathbf{Z}^\top\mathbf{Z} + \alpha\mathbf{\Gamma}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^\top\mathbf{y} \\ \mathbf{Z}^\top\mathbf{y} \end{bmatrix}$$

$$\boldsymbol{\theta} = \mathbf{C}^{-1}\mathbf{r}$$

Solving the System in Julia

- In Julia, the general recommendation for linear algebra is to solve systems ($\mathbf{C} \backslash \mathbf{r}$) rather than invert matrices (inv($\mathbf{C}$)).
- Using the backslash operator ($\backslash$) is through the LinearAlgebra.jl package
- The ($\backslash$) operator
  - faster
  - more numerically stable
  - more efficient
- You should only use inv($\mathbf{C}$) if you specifically need the $\mathbf{C}^{-1}$

Numerical Example of BLUP

- Suppose we have 3 individuals with one record each:

$$\mathbf{y} = \begin{bmatrix} 10 \\ 12 \\ 9 \end{bmatrix}$$

- Assume:

$$\mathbf{\Phi} = \mathbf{I}\sigma_u^2 \qquad \mathbf{D} = \mathbf{I}\sigma_e^2$$

and

$$\sigma_u^2 = 4, \qquad \sigma_e^2 = 6$$

- Thus:

$$\alpha = \frac{\sigma_e^2}{\sigma_u^2} = \frac{6}{4} = 1.5$$

Model Matrices

- Design matrices:

$$\mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \qquad \mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Mixed model equations:

$$\begin{bmatrix} 3 & 1 & 1 & 1 \\ 1 & 1+\alpha & 0 & 0 \\ 1 & 0 & 1+\alpha & 0 \\ 1 & 0 & 0 & 1+\alpha \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \end{bmatrix} = \begin{bmatrix} 31 \\ 10 \\ 12 \\ 9 \end{bmatrix}$$

$$1 + \alpha = 2.5$$

Solving the System

- Solving the mixed model equations gives:

$$\hat{\mu} = 10.33$$

- Random effects:

$$\hat{u}_1 = -0.13$$

$$\hat{u}_2 = 0.67$$

$$\hat{u}_3 = -0.53$$

Let's solve this example in Julia!

- The system:

$$\begin{bmatrix} \mathbf{X}^\top \mathbf{X} & \mathbf{X}^\top \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{X} & \mathbf{Z}^\top \mathbf{Z} + \alpha \boldsymbol{\Gamma}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \mathbf{y} \\ \mathbf{Z}^\top \mathbf{y} \end{bmatrix}$$

- What do we have:
  - $\mathbf{y}$, α, $\boldsymbol{\Gamma}^{-1} = \mathbf{I}$
- What do we need to form:
  - $\mathbf{X}$, $\mathbf{Z}$ (which are relatively simple!)

Animal Model (Also a LMM!)

Model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

$$\begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_a^2 \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \sigma_e^2 \mathbf{I} \end{pmatrix} \right)$$

- Phenotype = systematic part + random part
- Systematic part: sex, treatment, age, herds, season, year of birth, year of recording, etc.
- Random part: variation not explained by systematic component (genetic factor: animal)
- Genetic effects in **u** are correlated due to pedigree relationships **A**

Pedigree-based BLUP

EBV and fixed effects are solutions to the mixed model equations (MME):

$$\begin{bmatrix} \mathbf{X}^\top\mathbf{X} & \mathbf{X}^\top\mathbf{Z} \\ \mathbf{Z}^\top\mathbf{X} & \mathbf{Z}^\top\mathbf{Z} + \alpha\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top\mathbf{y} \\ \mathbf{Z}^\top\mathbf{y} \end{bmatrix}$$

$$\alpha = \frac{\sigma_e^2}{\sigma_u^2}$$

- EBV for all animals (also those without records)
- $\mathbf{A}^{-1}$ is simple to compute for all animals in the pedigree

An example in Julia

- The system:

$$\begin{bmatrix} \mathbf{X}^\top\mathbf{X} & \mathbf{X}^\top\mathbf{Z} \\ \mathbf{Z}^\top\mathbf{X} & \mathbf{Z}^\top\mathbf{Z} + \alpha\mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top\mathbf{y} \\ \mathbf{Z}^\top\mathbf{y} \end{bmatrix}$$

- What do we have:
  - $\mathbf{y}$, α
- What do we need to form:
  - $\mathbf{X}$, $\mathbf{Z}$ (now they are little tricky!) and $\mathbf{A}$

SNP-BLUP

- Model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \sum_j z_j g_j + \mathbf{e}$$

- Compact notation:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{g} + \mathbf{e}$$

- g_j are SNP effects
- Centering marker codes:

$$\mathbf{z}_j = \mathbf{m}_j - 2p_j\mathbf{1}$$

- SNP effects assumed normally distributed and independent

Mixed Model Equations for SNP-BLUP

$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \mathbf{I}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{bmatrix}$$

$$\alpha = \frac{\sigma_e^2}{\sigma_g^2}$$

- Provides BLUP of SNP effects $\hat{\mathbf{g}}$
- GEBV for individual i:

$$\hat{u}_i = \sum_j z_{ij}\hat{g}_j$$

An example in Julia

- The system:

$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \mathbf{I}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{bmatrix}$$

- What do we have:
  - $\mathbf{y}$, α
- What do we need to form:
  - $\mathbf{X}$, $\mathbf{Z}$

Equivalent Model – GBLUP

Genetic effects:

$$\mathbf{u} = \mathbf{Z}\mathbf{g}$$

Variance:

$$\mathrm{Var}(\mathbf{u}) = \mathbf{Z}\mathrm{Var}(\mathbf{g})\mathbf{Z}^T = \sigma_g^2 \mathbf{Z}\mathbf{Z}^T$$

Model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

$$\mathrm{Var}(\mathbf{a}) = \sigma_u^2 \mathbf{G}$$

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}^T}{\sum_j 2p_j(1-p_j)}$$

$$\sigma_u^2 = \sigma_g^2 \sum_j 2p_j(1-p_j)$$

Mixed Model Equations for GBLUP

$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \mathbf{G}^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{bmatrix}$$

$$\alpha = \frac{\sigma_e^2}{\sigma_u^2}$$

An example in Julia

- The system:

$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \mathbf{G}^{-1}\alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{bmatrix}$$

- What do we have:
  - $\mathbf{y}$, α
- What do we need to form:
  - $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{G}$

Assignment until next week (the 24th)

Make a function or set of functions:

- that can create design matrices for fixed effects automatically for the PBLUP example
- that takes the data frame(s) for BLUP, and variance components as well as analysis type (PBLUP or GBLUP) as its arguments, and returns breeding values for selected individuals. The function should not use the data sets in the current environment as input, but read them from a file