minque: An R Package for Mixed Linear Model Analyses

Jixiang Wu
Associate Professor of Quantitative Genetics/Biostatistics
Department of Agronomy, Horticulture & Plant ScienceSouth Dakota State University
Brookings, SD 57007
Email: jixiang.wu@sdstate.edu
Phone: 605-688-5947

Outlines

- Overview of linear mixed models (LMM) and analytical approaches
- minque package
- Applications and demonstrations
- Conclusions/discussion

Various Data Types and Structures

- Subject based
  - Animals
  - Crops
  - …
- Location based
  - Field testing
  - Lab work: DNA markers, microarray
  - …
- Structure based
  - Complicated vs simple
  - Balanced vs unbalanced
  - Missing vs no missing

A simplified linear model

- $y = \mu + T + B + X + \varepsilon$
- T component → treatment designs
- B component → blocking designs
- X component→ covariate variable(s)
- Reducible
- Extendable

T Component

- One factor
- Two or more factors
  - Factorial
  - Nested
  - Mixture
  - Interaction effects
  - Nested effects

B component

- Local control of experimental errors
- Technique affects variations and bias
- Selection of uniform experimental units
- Blocking to reduce experimental error/ variation
- Often related to field experimental designs but not limited to

X Component

- Any variables that might contribute to the variation of the response variable y
- Condition of experimental units (EU)
- Soil conditions in field plot

Combining T, B, and X Components

- There will be a number of combinations with T and B components
- There could be a lot of linear models!

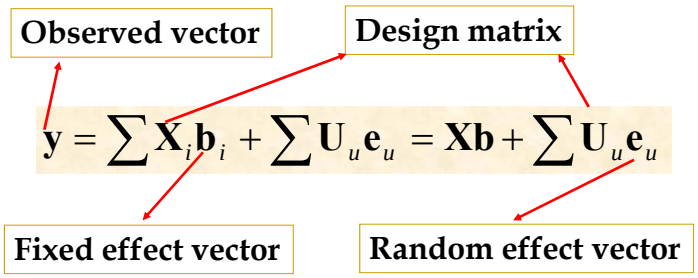
Various Linear Models

- Types of linear models
  - Fixed effect models
  - Random effect models
  - Mixed effect models
- Fixed effect models
  - All effects are fixed except error
  - Regression models
  - Some ANOVA models
- Random effect models
  - Except population mean, all effects are random
  - Some ANOVA models
  - Many genetic models
- Linear mixed models (LMMs)
  - Except population mean, some are random and some are fixed

Linear Mixed Models

- Linear mixed models (LMMs) are a generalization of various linear models
- Fixed effect and random effects are two specific cases
- LMMs can be complex
  - Some genetic models
  - Repeated measurements
    - Spatial patterns
    - Time series
  - More extendable

A Generalized Linear Mixed Model



$$\mathbf{V} = \mathrm{var}(\mathbf{y}) = \sum \sigma_u^2 \mathbf{U}_u \mathbf{U}_u^T$$

Three Components for a LMM Analysis

- Estimate variance components
- Estimate fixed effects
- Predict random effects

LMM approaches

- ML: Maximum Likelihood (Hartley and Rao, 1967)
- REML: Restricted Maximum Likelihood (Patterson & Thompson, 1971)
- MINQUE: Minimum Norm Quadratic Unbiased Estimation (Rao 1971)

Method	Distribution	Data Structure	Iteration	Prior Values
ML	Normal	None	Yes	Yes
REML	Normal	None	Yes	Yes
MINQUE	None	None	No	Yes

Linear Mixed Model Approaches

- Advantages
  - Missing data
  - Complicated data sets
  - Complicated genetic models
- Disadvantage
  - Matrix based: need computers and can be time-consuming
  - Iterations may be needed
  - Computationally intensive

Variance component estimation: MINQUE approach

$$\mathbf{y} = \mathbf{Xb} + \sum_{u=1}^{r} \mathbf{U}_u \mathbf{e}_u$$

Generalized linear mixed model

$$[tr(\mathbf{U}_u^T \mathbf{Q}_\alpha \mathbf{U}_v^T \mathbf{U}_v \mathbf{Q}_\alpha \mathbf{U}_u)][\sigma_u^2] = [\mathbf{y}^T \mathbf{Q}_\alpha \mathbf{U}_u \mathbf{U}_u^T \mathbf{Q}_\alpha \mathbf{y}]$$

MINQUE equations

$$\mathbf{Q}_\alpha = \mathbf{V}_\alpha^{-1} - \mathbf{V}_\alpha^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}_\alpha^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}_\alpha^{-1}$$

$$\mathbf{V}_\alpha = \sum_{u=1}^{r} \alpha_u \mathbf{U}_u \mathbf{U}_u^T$$

MINQUE1 (Zhu, 1989)

Prior values

Random Effect Prediction

$$\hat{\mathbf{e}}_u = \sigma_u^2 \mathbf{U}_u^T \mathbf{Q} \mathbf{y}$$ BLUP

$$\hat{\mathbf{e}}_u = \hat{\sigma}_u^2 \mathbf{U}_u^T \hat{\mathbf{Q}} \mathbf{y}$$ "BLUP"

$$\hat{\mathbf{e}}_u = \alpha_u \mathbf{U}_u^T \mathbf{Q}_\alpha \mathbf{y}$$ LUP

$$\hat{\mathbf{e}}_u = \kappa_u \alpha_u \mathbf{U}_u^T \mathbf{Q}_\alpha \mathbf{y}$$ AUP

Fixed Effect Estimation

- Least square estimation (LSE)
  - $\widehat{\boldsymbol{b}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$
  - **Unbiased but not the best**
- Weighted LSE
  - $\widehat{\boldsymbol{b}} = (\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{V}^{-1}\boldsymbol{y}$
  - **If $\boldsymbol{V}$ *is known***
    - **Best linear unbiased estimation (BLUE)**
  - **If V is unknown**
    - **Not sure if unbiased**
    - **Maybe with smaller variance**

Statistical tests for LMM

- LMM with conventional tests
  - X^2 test for variance components
  - z-test for fixed and random effects
  - Disadvantage: low power

	REML	Power
V(Female)	20.13	0.00
V(Male)	20.57	0.00
V(Rep)	20.84	0.00
V(Female:Male)	20.87	0.35
V(e)	19.84	1.00

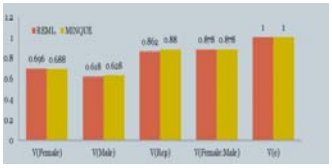
Statistical tests for LMM

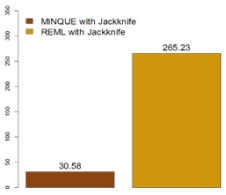
- Issues with X^2-and z- tests
  - Low power
- LMM with resampling techniques
  - Jackknife (*Wu et al., 2008, 2013*)
    - Block based
    - k-fold based
  - Permutation
  - Advantages:
    - Different distributions
    - Different parameters

REML vs MINQUE

- REML and MINQUE perform equally well
- Jackknife can significantly improves power while controlling type I error
- MINQUE performs faster than REML

(*Nan andWu, 2015*)





minque package

- Two components:
  - Random and fixed effects
- Includes two LMM approaches:
  - REML and MINQUE
- Includes two resampling approaches
  - Jackknife and permutation
- Actual data analysis;
  - (Proportional) variance components
  - Fixed effects
  - Random effects
- Simulation:
  - Variance components

Applications

- Classroom teaching
  - Design of experiments
  - Quantitative genetics
- Research
  - Unbalanced/missing genetic data
  - Augmented experimental design
  - Model evaluation
  - Integration of field spatial pattern (current)

Conclusions

- LMM are flexible approaches for various linear mixed models
- Resampling techniques are simple ways for various statistical tests
- minque package is an integration of LMM and above statistical approaches
- minque package can be used in both education and research

Demonstration

- Fixed and random effects
- LMM without and with jackknife
- Actual data analysis
- Simulation

Questions and Comments