Empirical Likelihood-based Analysis of Variance Component in Linear Mixed-Effects Models

Jingru Zhang, Haochang Shou and Hongze Li
University of Pennsylvania

Introduction

Linear mixed-effects (LME) models are widely used in analyzing repeated measurement and longitudinal data. Although statistical inference of the fixed effects is well studied, inference of the variance component is rarely explored, especially when the random effects and errors often require strong distributional assumptions on the random errors and effects.

Question: How to do distribution-free inference of the variance component in LME models?

Problem setup

- n subjects.
- For the i-th subject, n_i repeated measurements.
- For each repeated measurement, data are collected at time t_i = s_1, s_2, ..., s_m.
- For the i-th subject at time t, we observe a response vector y(t) ∈ R^p, an n_i × p design matrix X_i for the fixed effects β(t) ∈ R^p, and d semi-positive design matrices Φ_{ij} (j = 1, ..., d) for the variance components θ(t) ∈ (R_+ ∪ {0})^d.

LME model

A general setting of the linear mixed-effects model:
y(t) = X_i β(t) + r_i(t), i = 1, ..., n,
where r_i(t) ∈ R^p is a zero-mean random variable with variance H_i(θ(t)).

We consider H_i(θ(t)) with a linear structure, i.e.,
H_i(θ(t)) = \sum_{l=1}^{d} θ_{i}^{l}(t)Φ_{i}^{l},
θ(t) = (θ_1(t), ..., θ_d(t))^T \cong (θ_1(t), θ_2(t))^T.

- In this general setting, we do not specify any distribution for the data.
- The data y(t) are independent over different subjects i, while they are allowed to be non-independent over t.

Testing problems

- Local testing problem H_0: θ_1(t) = θ_1(t) at a given t.
- Global testing problem H_0: θ_1(t) = θ_1(t), t ∈ [t_1, t_2].

Introduction

Linear mixed-effects (LME) models are widely used in analyzing repeated measurement and longitudinal data. Although statistical inference of the fixed effects is well studied, inference of the variance component is rarely explored, especially when the random effects and errors often require strong distributional assumptions on the random errors and effects.

Question: How to do distribution-free inference of the variance component in LME models?

Problem setup

- n subjects.
- For the i-th subject, n_i repeated measurements.
- For each repeated measurement, data are collected at time t_i = s_1, s_2, ..., s_m.
- For the i-th subject at time t, we observe a response vector y(t) ∈ R^p, an n_i × p design matrix X_i for the fixed effects β(t) ∈ R^p, and d semi-positive design matrices Φ_{ij} (j = 1, ..., d) for the variance components θ(t) ∈ (R_+ ∪ {0})^d.

LME model

A general setting of the linear mixed-effects model:
y(t) = X_i β(t) + r_i(t), i = 1, ..., n,
where r_i(t) ∈ R^p is a zero-mean random variable with variance H_i(θ(t)).

We consider H_i(θ(t)) with a linear structure, i.e.,
H_i(θ(t)) = \sum_{l=1}^{d} θ_{i}^{l}(t)Φ_{i}^{l},
θ(t) = (θ_1(t), ..., θ_d(t))^T \cong (θ_1(t), θ_2(t))^T.

- In this general setting, we do not specify any distribution for the data.
- The data y(t) are independent over different subjects i, while they are allowed to be non-independent over t.

Testing problems

- Local testing problem H_0: θ_1(t) = θ_1(t) at a given t.
- Global testing problem H_0: θ_1(t) = θ_1(t), t ∈ [t_1, t_2].