New Computational Approaches to Large/Complex Mixed Effects Models

Norm Matloff
University of California at Davis

JSM 2016

August 1, 2016

http://heather.cs.ucdavis.edu/JSM2016.pdf

A Different Kind of Talk

A Different Kind of Talk

- Methodology for *algebraic* computation, not mainly at the computer stage.

A Different Kind of Talk

- Methodology for *algebraic* computation, not mainly at the computer stage.
- Suggestions on *what* (new) to model, not *how*.

Overview

Overview

- Given, a mixed-effects model.

Overview

- Given, a mixed-effects model.
- Change to a fully random model, treating fixed factors as samples from populations.

Overview

- Given, a mixed-effects model.
- Change to a fully random model, treating fixed factors as samples from populations.
- Note that get (essentially) the same variances, thus the same Method of Moments estimators.

Overview

- Given, a mixed-effects model.
- Change to a fully random model, treating fixed factors as samples from populations.
- Note that get (essentially) the same variances, thus the same Method of Moments estimators.
- Consider the former n_i to be random, i.e. N_i, with their own effects worth studying.

Overview

- Given, a mixed-effects model.
- Change to a fully random model, treating fixed factors as samples from populations.
- Note that get (essentially) the same variances, thus the same Method of Moments estimators.
- Consider the former n_i to be random, i.e. N_i, with their own effects worth studying. (In 2-factor models, N_i are row counts, and have column counts M_i. Etc.)

Simple Example

Simple Example

- Classic 1-random-factor model,

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \;\; i = 1, ..., r, j = 1, ..., n_i \tag{1}$$

Simple Example

- Classic 1-random-factor model,

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \ i = 1, ..., r, j = 1, ..., n_i \tag{1}$$

- Not considered “mixed,” but it really is,

Simple Example

- Classic 1-random-factor model,

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \ \ i = 1, ..., r, j = 1, ..., n_i \tag{1}$$

- Not considered "mixed," but it really is, since the n_i are considered constants.

Simple Example

- Classic 1-random-factor model,

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \;\; i = 1, ..., r, j = 1, ..., n_i \tag{1}$$

- Not considered "mixed," but it really is, since the n_i are considered constants.
- But change n_i to N_i, assume latter are i.i.d.

Simple Example

- Classic 1-random-factor model,

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \ \ i = 1, ..., r, j = 1, ..., n_i \qquad (1)$$

- Not considered "mixed," but it really is, since the n_i are considered constants.
- But change n_i to N_i, assume latter are i.i.d.
- Why do this?

Simple Example

- Classic 1-random-factor model,

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \;\; i = 1, ..., r, j = 1, ..., n_i \tag{1}$$

- Not considered "mixed," but it really is, since the n_i are considered constants.
- But change n_i to N_i, assume latter are i.i.d.
- Why do this?
    - Makes quantities like $Y_{i.} = \sum_{j=1}^{N_i} Y_{ij}/N_i$ i.i.d., thus easy algebra.

Simple Example

- Classic 1-random-factor model,

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \;\; i = 1, ..., r, j = 1, ..., n_i \tag{1}$$

- Not considered "mixed," but it really is, since the n_i are considered constants.
- But change n_i to N_i, assume latter are i.i.d.
- Why do this?
  - Makes quantities like $Y_{i.} = \sum_{j=1}^{N_i} Y_{ij}/N_i$ i.i.d., thus easy algebra.
  - Enables statistical analysis of N!

Investigating Effects of N

Investigating Effects of N

Maybe model as

$$Y_{ij} = c_1 + c_2 N_i + \alpha_i + \epsilon_{ij}, \;\; i = 1, r, j = 1, ..., N_i \qquad (2)$$

Investigating Effects of N

Maybe model as

$$Y_{ij} = c_1 + c_2 N_i + \alpha_i + \epsilon_{ij}, \ \ i = 1, r, j = 1, ..., N_i \qquad (2)$$

- E.g. recommender systems.

Investigating Effects of N

Maybe model as

$$Y_{ij} = c_1 + c_2 N_i + \alpha_i + \epsilon_{ij}, \;\; i = 1, r, j = 1, ..., N_i \tag{2}$$

- E.g. recommender systems. Maybe frequent users (N_i large) become jaded, thus give lower ratings?

Investigating Effects of N

Maybe model as

$$Y_{ij} = c_1 + c_2 N_i + \alpha_i + \epsilon_{ij}, \ \ i = 1, r, j = 1, ..., N_i \qquad (2)$$

- E.g. recommender systems. Maybe frequent users (N_i large) become jaded, thus give lower ratings? (Yes.)

Investigating Effects of N

Maybe model as

$$Y_{ij} = c_1 + c_2 N_i + \alpha_i + \epsilon_{ij}, \ i = 1, r, j = 1, ..., N_i \tag{2}$$

- E.g. recommender systems. Maybe frequent users (N_i large) become jaded, thus give lower ratings? (Yes.)
- Research on family size (N_i), found to be positively related to child longevity (Ahmed *et al*, 2016).

Investigating Effects of N

Maybe model as

$$Y_{ij} = c_1 + c_2 N_i + \alpha_i + \epsilon_{ij}, \;\; i = 1, r, j = 1, ..., N_i \qquad (2)$$

- E.g. recommender systems. Maybe frequent users (N_i large) become jaded, thus give lower ratings? (Yes.)
- Research on family size (N_i), found to be positively related to child longevity (Ahmed *et al*, 2016).
- Research on negative correlation of family size to household income (Berger, 2011).

Streamlining the Algebra

Streamlining the Algebra

Consider again the simple model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \;\; i = 1, ..., r, j = 1, ..., N_i \tag{3}$$

Streamlining the Algebra

Consider again the simple model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \; i = 1, ..., r, j = 1, ..., N_i \tag{3}$$

- We want to estimate σ_α^2, and possibly σ_ϵ^2.

Streamlining the Algebra

Consider again the simple model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \;\; i = 1, ..., r, j = 1, ..., N_i \tag{3}$$

- We want to estimate σ_α^2, and possibly σ_ϵ^2.
- The Method of Moments approach is safer (no normality assumptions).

Streamlining the Algebra

Consider again the simple model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \;\; i = 1, ..., r, j = 1, ..., N_i \tag{3}$$

- We want to estimate σ^2_α, and possibly σ^2_ϵ.
- The Method of Moments approach is safer (no normality assumptions). But deriving the equations is messy.

Streamlining the Algebra

Consider again the simple model

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \;\; i = 1, ..., r, j = 1, ..., N_i \tag{3}$$

- We want to estimate σ_α^2, and possibly σ_ϵ^2.
- The Method of Moments approach is safer (no normality assumptions). But deriving the equations is messy.
- Once we go to two-factor models, add regressors etc., things get even messier, FAST.

General Strategy

General Strategy

- Assume *everything* — including fixed effects, covariates and even the N_i — is random, i.i.d.

General Strategy

- Assume *everything* — including fixed effects, covariates and even the N_i — is random, i.i.d.
- Where possible, use the "Pythagorean Theorem" for variance,

General Strategy

- Assume *everything* — including fixed effects, covariates and even the N_i — is random, i.i.d.
- Where possible, use the "Pythagorean Theorem" for variance,

$$Var(W) = E[Var(W|U)] + Var[E(W|U)] \tag{4}$$

Back to the Simple Example

Back to the Simple Example

Write $Y = \mu + \alpha + \epsilon$ and, motivated by defining

$$S_i = \sum_{j=1}^{N_i} Y_{ij} \approx N\mu + N\alpha \tag{5}$$

also write $S = N\mu + N\alpha + \epsilon_1 + ... + \epsilon_N$

Back to the Simple Example

Write $Y = \mu + \alpha + \epsilon$ and, motivated by defining

$$S_i = \sum_{j=1}^{N_i} Y_{ij} \approx N\mu + N\alpha \tag{5}$$

also write $S = N\mu + N\alpha + \epsilon_1 + ... + \epsilon_N$

Then apply the "Pythagorean Theorem" with $W = S$, $U = N$,

Back to the Simple Example

Write $Y = \mu + \alpha + \epsilon$ and, motivated by defining

$$S_i = \sum_{j=1}^{N_i} Y_{ij} \approx N\mu + N\alpha \qquad (5)$$

also write $S = N\mu + N\alpha + \epsilon_1 + ... + \epsilon_N$

Then apply the "Pythagorean Theorem" with $W = S, \ U = N$, using facts like

$$Var(S|N) = N^2 Var(\alpha) + N\sigma_\epsilon^2 \qquad (6)$$

Back to the Simple Example

Write $Y = \mu + \alpha + \epsilon$ and, motivated by defining

$$S_i = \sum_{j=1}^{N_i} Y_{ij} \approx N\mu + N\alpha \tag{5}$$

also write $S = N\mu + N\alpha + \epsilon_1 + ... + \epsilon_N$

Then apply the "Pythagorean Theorem" with $W = S$, $U = N$, using facts like

$$Var(S|N) = N^2 Var(\alpha) + N\sigma_\epsilon^2 \tag{6}$$

Finally, replace pop. quantities by sample analogs, and solve.

Almost-Simple Example

Almost-Simple Example

Let's add a covariate:

$$Y_{ij} = \beta_0 + X_i \beta_1 + \alpha_i + \epsilon_{ij}, \ \ i = 1, ..., r, j = 1, ..., N_i \tag{7}$$

$$Y = \beta_0 + X \beta_1 + \alpha + \epsilon \tag{8}$$

$$S = N(\beta_0 + X\beta_1 + \alpha) + \epsilon_1 + ... + \epsilon_N \tag{9}$$

and the matrix equation

$$\mathbb{Y} = \mathbb{B}_0 + \mathbb{X}\beta_1 + \mathbb{A} + \mathbb{G} \tag{10}$$

This is already getting messy even in this form, but much better than the standard way, with all the messy sums.

What If the Fixed Effects Really Are Fixed?

What If the Fixed Effects Really Are Fixed?

- E.g. 5 different drugs for hypertension.

What If the Fixed Effects Really Are Fixed?

- E.g. 5 different drugs for hypertension.
- May not feel comfortable with these being sample from a "population" of drugs.

What If the Fixed Effects Really Are Fixed?

- E.g. 5 different drugs for hypertension.
- May not feel comfortable with these being sample from a "population" of drugs.
- But can treat each observation's drug as a sample from the 5, making the N_i random.

What If the Fixed Effects Really Are Fixed?

- E.g. 5 different drugs for hypertension.
- May not feel comfortable with these being sample from a "population" of drugs.
- But can treat each observation's drug as a sample from the 5, making the N_i random.
- Can show that we still get the same answer!

What If the Fixed Effects Really Are Fixed?

- E.g. 5 different drugs for hypertension.
- May not feel comfortable with these being sample from a "population" of drugs.
- But can treat each observation's drug as a sample from the 5, making the N_i random.
- Can show that we still get the same answer!
- Hence my characzieration of the method as an algebraic computational device.

(Computer) Computational Benefits

(Computer) Computational Benefits

- Due to i.i.d. nature, lends to easy parallelization through "software alchemy" (Matloff, 2016 JSS and references therein).

(Computer) Computational Benefits

- Due to i.i.d. nature, lends to easy parallelization through "software alchemy" (Matloff, 2016 JSS and references therein).
- MM is much faster and uses far less memory than MLE.

Summary

Summary

In mixed-effects models, treat the fixed effects, and even the N_i, as random.

Summary

In mixed-effects models, treat the fixed effects, and even the N_i, as random.

- Streamlines algebraic derivations.

Summary

In mixed-effects models, treat the fixed effects, and even the N_i, as random.

- Streamlines algebraic derivations.
- Allows investigation of the effects of the N_i themselves.

Summary

In mixed-effects models, treat the fixed effects, and even the N_i, as random.

- Streamlines algebraic derivations.
- Allows investigation of the effects of the N_i themselves.
- Enables parallelization.