A Bayesian Approach for Modeling Three-Way Cross-Classified Multilevel Data

Haley E. Yaremych, M.S., Sarah Brown-Schmidt, Ph.D., Sun-Joo Cho, Ph.D., Kristopher J. Preacher, Ph.D.
Department of Psychology & Human Development, Vanderbilt University

INTRODUCTION

- Psychological studies often give rise to **data that are clustered within three distinct units of nesting**.
- This data structure requires fitting multilevel models with three-way cross-classification.
- **Maximum likelihood estimation (MLE) struggles to fit these models; Bayesian estimation may be the better approach.**

METHOD

- Used **empirical data from a psycholinguistics study:** three-way cross-classified with 3 key predictors.
- Identified the "maximal" (most complex) cross-classified model that would converge with MLE.
- Then aimed to fit a more complex (and substantively better) model with Bayesian estimation.

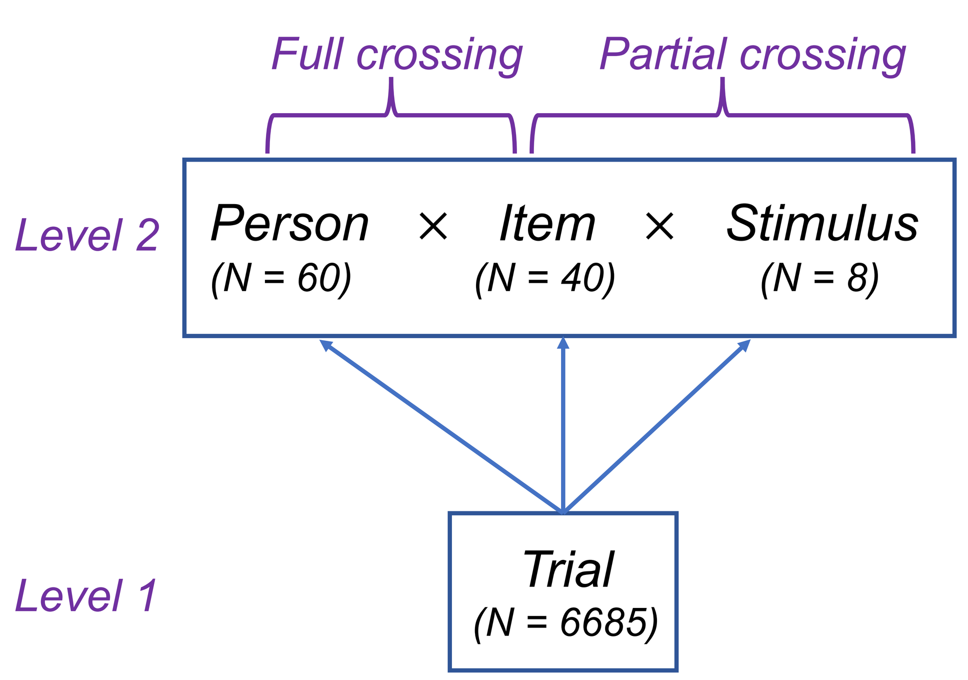

Fig 1: Empirical data nesting structure

RESULTS

- **With Bayesian estimation, a more complex model successfully converged.**
- But for some parameters, highly informative priors were needed to achieve convergence.

DISCUSSION

- Bayesian approaches allow the incorporation of priors, which can stabilize complex models.
- **Priors must be chosen carefully and communicated transparently**, especially if sample size is small.

Compared to maximum likelihood estimation, **Bayesian estimation** led to **successful convergence** of a theoretically better-motivated and **more complex model.**

Components Included in Converged Models Using Empirical Data

	Fixed Effect	Random effect across persons	Random effect across items	Random effect across stimuli
Intercept	MLE B	MLE B	B	MLE B
x_1	MLE B	MLE B		MLE B
x_2	MLE B	MLE B		
x_3	MLE B	B		
$x_1 \times x_2$	MLE B			
$x_1 \times x_3$	MLE B			
$x_2 \times x_3$	MLE B			
$x_1 \times x_2 \times x_3$	MLE B			

MLE *= in converged model with maximum likelihood estimation*
B *= in converged model with Bayesian estimation*

haley.e.yaremych@vanderbilt.edu

DETAILS: PRIOR CHOICES

- For cross-classified models, weakly informative priors are recommended.
- The *Student's t* distribution with 3 < *df* < 7 is the suggested prior for fixed effects when *y* is binary.
- To achieve convergence, a very informative prior was necessary for random effects across the third unit of nesting (stimuli), due to a very small sample size (*N* = 8) for this unit.

Fixed effects:
- Intercept: ***Student's t***(*df* = 6, *M* = 0, *SD* = 10)
- Slopes: ***Student's t***(*df* = 6, *M* = 0, *SD* = 2.5)

Random effects:
- Random intercept and slope SDs across persons and items: ***half Cauchy***(*M* = 0, *SD* = 25)
- Random intercept and slope SDs across stimuli: ***half Cauchy***(*M* = 0, *SD* = 1)
- Correlation matrices for random effects: ***LKJ***(2)

DETAILS: RESULTS

- Across items, random intercept variance was small.
  - This led to convergence problems for MLE, but not Bayesian estimation.
- Bayesian estimation may be especially advantageous when random effect variances are small and/or when level-2 sample size is small.

DETAILS: PLANNED SIMULATION STUDY

Design Factors:
- Sample size for each level-2 unit
- Variance and intraclass correlation of *y*
- Magnitude of random effects

Expected Findings:

Outcome measure	Expected superior estimation method	
	Bayesian	MLE
Computational efficiency		✓
Unbiased random effects estimates	✓	
Standard errors of all estimates	✓	
Convergence despite small level 2 sample size and/or partial crossing	✓	