Regularized Gaussian variational estimation for detecting intersectional DIF: With Multilevel Random Item Effects Model

He Ren[1], Weicong Lyu[1], Chun Wang[1], Gongjun Xu[2]

[1] College of Education, University of Washington, Seattle, WA, USA
[2] Department of Statistics, University of Michigan, Ann Arbor, MI, USA

July 15, 2024

IMPS 2024, Prague, Czech Republic

Motivation

- Assessment Fairness
  - Differential item functioning (DIF): People from different groups differ in the probability of correctly answering an item even after controlling their ability
- Existing DIF studies (Fixed-effect-based models)
  - Ignorance of intersectionality: People's multiple identities are interlinked; Cannot handle the numerous subgroups well
  - Reliance on reference group: Conventional choice of reference group might result in a misleading idea, that is, the privileged groups are the standard and all other groups are deviations

Model – Extension of 2PL Model

- Two-parameter logistic (2PL) model

$$\mathbb{P}(y_{ijs} = 1 \mid \theta_i, a_j, b_j) = \frac{1}{1 + \exp[-(a_j\theta_i + b_j)]}$$

- 2PL with random intercept and slope and multilevel ability (2PL-RisM)

$$\mathbb{P}(y_{ijs} = 1|\theta_{is}, a_{js}, b_{js}) = \frac{1}{1 + \exp[-(a_{js}\theta_{is} + b_{js})]}$$

Multilevel Ability

$$\theta_{is} \sim \mathcal{N}(\alpha_{0s} + \boldsymbol{\alpha}_1^\top \boldsymbol{X}_s, \sigma_\theta^2),$$

$$\alpha_{0s} \sim \mathcal{N}(\alpha_0, \sigma_{\alpha_0}^2),$$

Random item effect

$$b_{js} \sim \mathcal{N}(b_j, \sigma_{b_j}^2),$$

$$a_{js} \sim \mathcal{N}_+(\bar{a}_j, \bar{\sigma}_{a_j}^2).$$

DIF Item: $\sigma_{b_j}^2, \bar{\sigma}_{a_j}^2 \neq 0$;

DIF-free Item: $\sigma_{b_j}^2 = \bar{\sigma}_{a_j}^2 = 0$

j, i, s denotes the item ($j = 1, \dots, J$), person ($i = 1, \dots, N_s$), and group ($s = 1, \dots, S$), respectively.

Model Estimation Algorithm

- Expectation Maximization (EM) algorithm
  - Popular in the 2PL estimation but require high dimension integral for MIRT and random item effect models
- Markov chain Monte Carlo (MCMC) method
  - Flexible but computationally inefficient
- Variational EM method
  - Circumvents the high dimension integral in EM
- Proposed regularized Gaussian variational EM (GVEM)
  - Results in closed-form solutions for almost all parameters in the EM iteration
  - Encourages the sparsity by log penalty

EM - Basic Idea

- EM algorithm

Let $\boldsymbol{Z} = \bigcup_{s=1}^{S} \bigcup_{i=1}^{N_s} \bigcup_{j=1}^{J} \{\alpha_{0s}, \theta_{is}, a_{js}, b_{js}\}$ be set of all latent variables, the log marginal likelihood of responses $\boldsymbol{Y}$ is

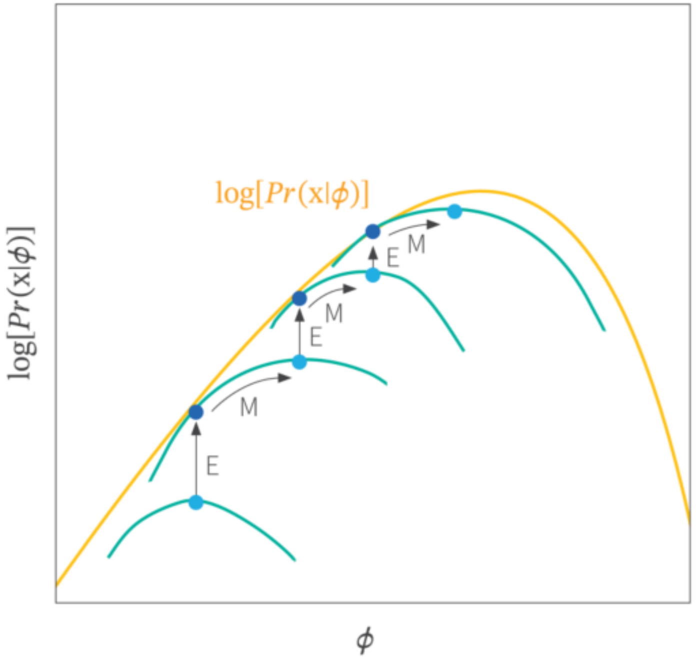
$$\ell(\boldsymbol{Y}) = \log \int \mathbb{P}(\boldsymbol{Y}, \boldsymbol{Z}) \mathrm{d}\boldsymbol{Z} = \log \int \mathbb{P}(\boldsymbol{Y} | \boldsymbol{Z}) p(\boldsymbol{Z}) \mathrm{d}\boldsymbol{Z}$$

For any function satisfying $\int q(\boldsymbol{Z}) \mathrm{d}\boldsymbol{Z} = 1$,

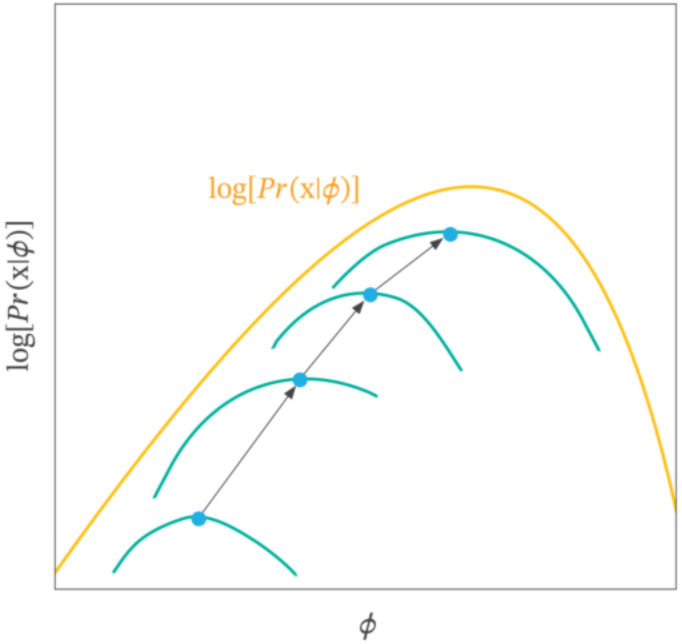
$$\begin{aligned} \ell(\boldsymbol{Y}) &= \int \ell(\boldsymbol{Y}) q(\boldsymbol{Z}) \mathrm{d}\boldsymbol{Z} = \int \log \frac{\mathbb{P}(\boldsymbol{Y}, \boldsymbol{Z})}{\mathbb{P}(\boldsymbol{Z} \mid \boldsymbol{Y})} \times q(\boldsymbol{Z}) \mathrm{d}\boldsymbol{Z} \qquad \mathbf{KL}\big(q(Z) || \mathbb{p}(Z|Y)\big) \geq \mathbf{0} \\ &= \int \log \frac{\mathbb{P}(\boldsymbol{Y}, \boldsymbol{Z})}{q(\boldsymbol{Z})} \times q(\boldsymbol{Z}) \mathrm{d}\boldsymbol{Z} + \boxed{\int \log \frac{q(\boldsymbol{Z})}{\mathbb{P}(\boldsymbol{Z} \mid \boldsymbol{Y})} \times q(\boldsymbol{Z}) \mathrm{d}\boldsymbol{Z}} \\ &\geq \int \log \frac{\mathbb{P}(\boldsymbol{Y}, \boldsymbol{Z})}{q(\boldsymbol{Z})} \times q(\boldsymbol{Z}) \mathrm{d}\boldsymbol{Z} = \underbrace{\boxed{\int \log \mathbb{P}(\boldsymbol{Y}, \boldsymbol{Z}) \times q(\boldsymbol{Z}) \mathrm{d}\boldsymbol{Z}}}_{\textbf{Expectation } Q(Y)} - \text{constant} \end{aligned}$$

EM vs GVEM

- EM
  - $q(\boldsymbol{Z}) = \arg\min \mathrm{KL}(q(\boldsymbol{Z}) || \mathbb{P}(\boldsymbol{Z} \mid \boldsymbol{Y}))$
    $= \mathbb{P}(\boldsymbol{Z} \mid \boldsymbol{Y})$



- GVEM (Cho et al., 2021)
  - $q(\boldsymbol{Z}) = \arg\min_q \mathrm{KL}(q(\boldsymbol{Z}) || \mathbb{P}(\boldsymbol{Z} \mid \boldsymbol{Y}))$ within a distribution family



GVEM

$$
\begin{aligned}
Q(Y) = \sum_{s=1}^{S}\sum_{i=1}^{N_s}\sum_{j=1}^{J}\Bigg\{ & \log\frac{1}{1+\mathrm{e}^{-\xi_{jis}}} + \left(y_{jis}-\frac{1}{2}\right)\left(\mu_{a_{js}}\mu_{\theta_{is}}-\mu_{b_{js}}\right) - \frac{1}{2}\xi_{jis} \\
& - \eta(\xi_{jis})\left[\left(\mu_{a_{js}}^2+\sigma_{a_{js}}^2\right)\left(\mu_{\theta_{is}}^2+\sigma_{\theta_{is}}^2\right) - 2\mu_{a_{js}}\mu_{\theta_{is}}\mu_{b_{js}} + \mu_{b_{js}}^2 + \sigma_{b_{js}}^2 - \xi_{jis}^2\right]\Bigg\} \\
- \frac{1}{2}\sum_{s=1}^{S}\Bigg\{ & (2J+N_s+1)\log 2\pi + \left(\log\sigma_{\alpha 0}^2 + \frac{\mu_{\alpha 0_s}^2+\sigma_{\alpha 0_s}^2}{\sigma_{\alpha 0}^2}\right) \\
& + \sum_{i=1}^{N_s}\left[\log\sigma_\theta^2 + \frac{\sigma_{\theta_{is}}^2+\sigma_{\alpha 0_s}^2+\left(\mu_{\theta_{is}}-\mu_{\alpha 0_s}-\boldsymbol{\alpha}_1^\intercal\boldsymbol{X}_s\right)^2}{\sigma_\theta^2}\right] \\
& + \sum_{j=1}^{J}\left[\boxed{\log\bar{\sigma}_{a_j}^2} + \frac{\sigma_{a_{js}}^2+\left(\mu_{a_{js}}-\bar{a}_j\right)^2}{\bar{\sigma}_{a_j}^2} + 2\log\Phi\left(\frac{\bar{a}_j}{\bar{\sigma}_{a_j}}\right)\right] \\
& + \sum_{j=1}^{J}\left[\boxed{\log\sigma_{b_j}^2} + \frac{\sigma_{b_{js}}^2+\left(\mu_{b_{js}}-b_j\right)^2}{\sigma_{b_j}^2}\right]\Bigg\}.
\end{aligned}
$$

Regularized GVEM

- Regularized GVEM algorithm

$$Q'(\boldsymbol{Y}) = Q(\boldsymbol{Y}) - \lambda \sum_{j=1}^{J} \left(\log \bar{\sigma}_{a_j}^2 + \log \sigma_{b_j}^2 \right)$$

  - The $\lambda \in (0, +\infty)$ is a tuning parameter

- Generalized information criteria (GIC)

$$k c_n - 2\ell(\boldsymbol{Y})$$

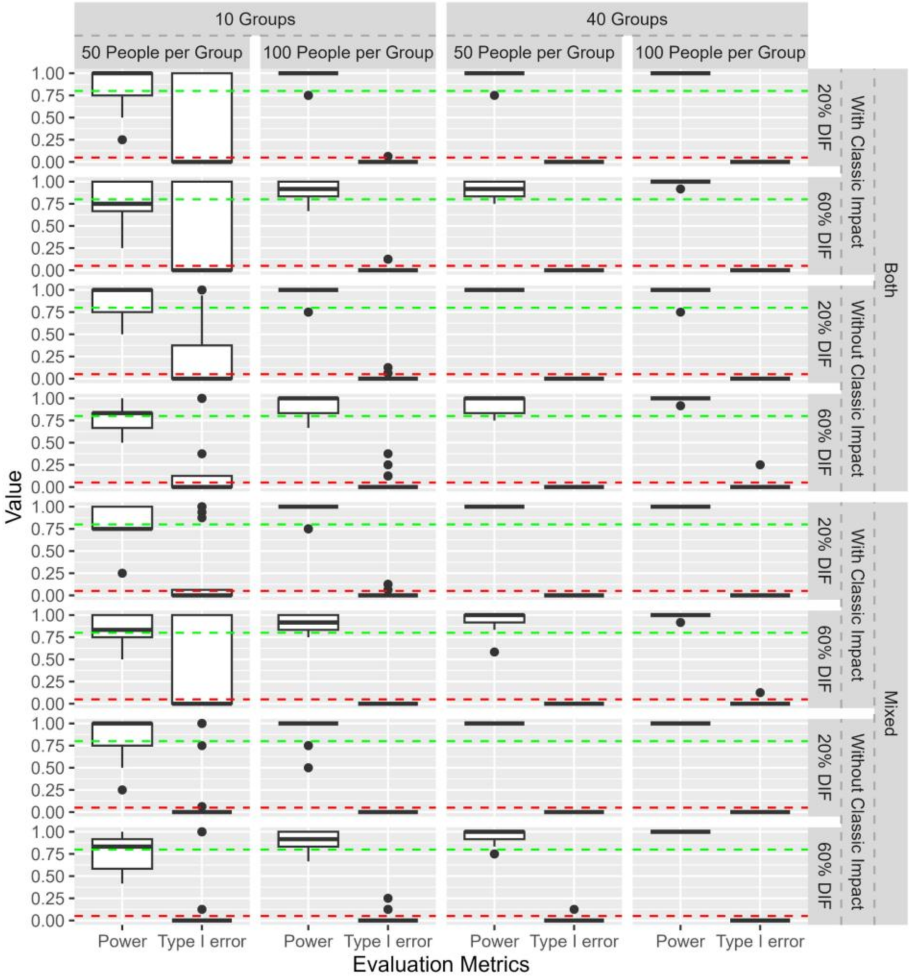
  - The k is the number of parameters, $c_n = c \log(N) \log(\log N)$
  - The c is a tuning parameter, and when $c_n = \log(N)$, GIC reduces to BIC

Simulation Studies - Design

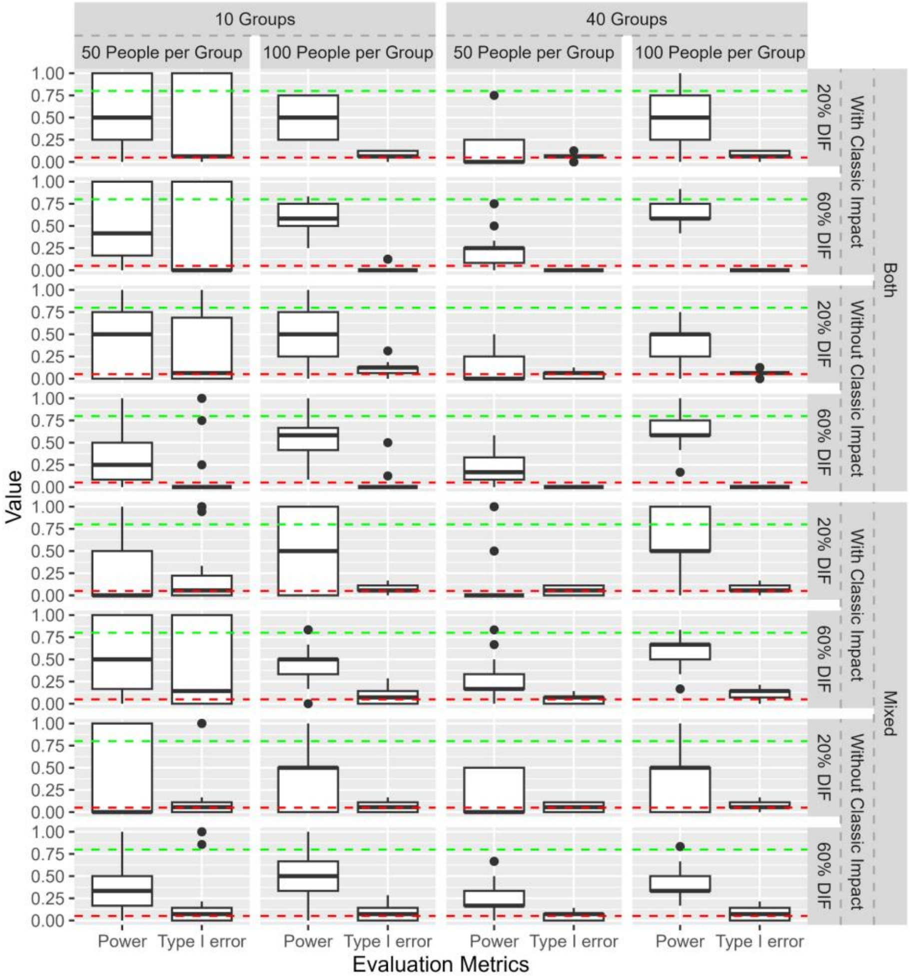
- Test length: 20
- $a_j \sim \text{Lognormal}(0,0.25)$, $b_j \sim \text{Uniform}[-2, 2]$
- Sample size per group: 50; 100
- Number of groups: 10; 40
- Proportion of DIF items: 20%; 60%
- Classic impact (α_1): 0.1; 0
- Multilevel impact ($\sigma^2_{\alpha_0}$): 0.5
- DIF pattern: All DIF items have DIF effect on both slop and intercept, Only half DIF items have DIF effect on both parameters and the other half have the effect only on intercept

Simulation Studies - Results

Intercept



slope



Conclusion

- This study proposes a novel random effect IRT model for detecting intersectional DIF and illustrates the feasibility of a regularized GVEM algorithm in such a context
- The simulation results show that the proposed methods can effectively detect DIF on intercept, although detecting DIF on slope is arguably harder
- The number of groups, group size, and DIF pattern have a substantial impact on the performance

Discussion

- The parameter estimation might be biased as variational method is, in essence, an approximate approach
- Detecting DIF is the first step to helping understand the fairness issue in society, and further investigation into the cause of systematic discrimination is necessary

Thank you!

heren@uw.edu