Comparing the Robustness of PROX Estimation to Maximum Likelihood Estimation

Anthony Raborn, Psychometrician

What is PROX Estimation?

Also known as Non-iterative Normal Approximation estimation

- Algebraic estimate of person parameters
- Used within a Rasch (or sometimes 1PL) framework
- Can be used with fixed item parameters
- Often used as an *initial* estimate of parameters for other iterative methods

PROX Equation

$$\hat{\theta}_p = \bar{\beta}_i + \ln(\frac{R_p}{N_p - R_p})\sqrt{1 + \frac{\sigma_i}{2.9}}$$

(see Cohen, 1979 for the derivation; Linacre, 1999 for more discussion.)

Purpose

- Iterative methods are *ubiquitous* for applying item response theory and Rasch measurement theory
  - Even with known item parameters!
- These methods, especially with Rasch exams, usually start with PROX
- Given this, how accurate is PROX on its own?
  - Important for practical applications in large-scale testing

Research Questions

1. How robust is (non-iterative) PROX to sample size fluctuations?
2. How robust is PROX to violations of the distributional assumptions for items and persons?
3. How do estimates produced by PROX compare to other common estimation methods under the conditions set in the first two research questions?

Simulation Study Conditions

Person and Item Parameters

1. Standard Normal Parameters
2. Wide Normal Parameters
3. Small Parameter Mismatch
4. Large Parameter Mismatch
5. Extreme Parameter Mismatch
6. Bimodal Person Parameters

Number of Persons and Items

1. Persons: $n_p \in (25, 50, 100, 250, 500, 1000)$
2. Items: $n_i = 200$

Total of **36** data conditions with **100** repetitions for **3600** total simulation iterations.

Estimation Methods

1. PROX Estimation (PROX; manually coded in `R` (R Core Team, 2023))
2. Joint Maximum Likelihood Estimation (JMLE; `TAM` (Robitzsch et al., 2022))
3. Conditional Maximum Likelihood Estimation (CMLE; `eRm` (Mair & Hatzinger, 2007))

- NOTE: did NOT accept fixed item parameters

4. Expected A Posteriori estimation (EAP; `ltm` (Rizopoulos, 2006) via `irtoys`)

Analysis Methods

1. Correlation: $\rho = \frac{\text{cov}(\theta, \hat{\theta})}{\sigma_\theta, \sigma_{\hat{\theta}}}$
2. Mean bias: $\frac{\Sigma_{i=1}^{n}(\hat{\theta}_i - \theta_i)}{n}$
3. Mean absolute difference (MAD): $\frac{\Sigma_{i=1}^{n}|\hat{\theta}_i - \theta_i|}{n-1}$
4. Root mean square error (RMSE): $\sqrt{\frac{\Sigma_{i=1}^{n}(\hat{\theta}_i - \theta_i)^2}{n-1}}$

(Formulas for mean bias, MAD, and RMSE taken from Feinberg & Rubright, 2016)

Results

Correlation

- Average correlation across conditions 1, 2, 3, and 6 for each estimation method were very high (.96 or greater)
- Average correlation for conditions 4 and 5 were high between estimation methods (.97 or greater) but much lower for true values
  - Large Parameter Difference: approximately .88 for each method
  - Extreme Parameter Difference: approximately .68 for each method
- No noticeable variability across sample sizes

Mean Bias

- JMLE: average of -0.08 across all conditions
- PROX: average of 0.14 across all conditions
- EAP: average of -0.34 across all conditions
- CMLE: average of 0.74 across all conditions

MAD

- Best: PROX and JMLE (0.31 and 0.33 across conditions, respectively)
- Next: EAP (0.47 across conditions)
- Worst: CMLE (0.67 across conditions)
- Generally little changes based on number of persons
  - JMLE and EAP showed positive change for worst conditions, while CMLE showed negative change

RMSE

- Best: PROX and JMLE (0.21 and 0.29 across conditions, respectively)
- Next: EAP (0.64 across conditions)
- Worst: CMLE (0.85 across conditions)
- Generally little changes based on number of persons
  - JMLE and EAP showed positive change for worst conditions, while CMLE showed negative change

Conclusions

Limitations

- Implementation of methods may not be perfect
  - Use other programs (e.g., Winsteps) for ML methods?
- Static number of items
  - Operational work implies PROX estimates begin to differ at $n_i = 100$
- Requires known item parameters
  - Limits estimation methods
- Estimates truncated

Acknowledgements

Thanks to Michael Peabody for the initial idea of the study!

Please send any questions to anthony.w.raborn@gmail.com

References

Cohen, L. (1979). Approximate expressions for parameter estimates in the rasch model. *British Journal of Mathematical and Statistical Psychology*, *32*(1), 113–120.

Feinberg, R. A., & Rubright, J. D. (2016). Conducting simulation studies in psychometrics. *Educational Measurement: Issues and Practice*, *35*(2), 36–49. https://doi.org/https://doi.org/10.1111/emip.12111

Linacre, J. M. (1999). Understanding rasch measurement: Estimation methods for rasch measures. *Journal of Outcome Measurement*, *3*, 382–405.

Mair, P., & Hatzinger, R. (2007). Extended rasch modeling: The eRm package for the application of IRT models in r. *Journal of Statistical Software*, *20*. https://doi.org/10.18637/jss.v020.i09

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rizopoulos, D. (2006). Ltm: An r package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, *17*(5), 1–25. https://doi.org/10.18637/jss.v017.i05

Robitzsch, A., Kiefer, T., & Wu, M. (2022). *TAM: Test analysis modules*. https://CRAN.R-project.org/package=TAM