***Specific Guidelines for Coordinated Paper Sessions***

*Proposals must consist of*

1. *a title of no more than 12 words,*
2. *an abstract of no more than 50 words (for inclusion in the final program),*
3. *a summary of research of no more than 800 words, and*
4. *references, tables, and figures as appropriate.*

*References, tables, and figures do not count toward the word limits.*

*The summary should include research questions, methods, and findings.*

*The Program Committee also strongly recommends that authors include the practical implications of their research (see the Evaluation and Review Criteria below).*

**TITLE**

Examining the Robustness and Comparability of PROX estimation with known item parameters.

**ABSTRACT**

OPITION 1: This study examines the robustness and comparability of a non-iterative estimation method (PROX) with more common estimation methods when item parameters are known. A simulation was conducted to manipulate the person ability distribution, item difficulty distribution, and sample size.

OPTION 2: This study investigates the effects of person ability distribution, item difficulty distribution, and sample size on the robustness and comparability of a non-iterative estimation method (PROX) with more common estimation methods when item parameters are known.

*800 of 800 words*

**INTRODUCTION**

Iterative estimation procedures have been universally adopted for Item Response Theory and Rasch Measurement Theory software packages. These iterative methods are used to produce estimates of person ability even in situations where item difficulty is known, such as exams delivered via calibrated item banks. Under a Rasch Measurement paradigm, it may be argued that in situations where all item parameters are fixed, determining person ability is more calculation than estimation.

Cohen (1979) introduced PROX, a non-iterative method for estimating Rasch measures when data are complete and both items and persons are approximately normally distributed. A method for hand calculating PROX is shown in Wright and Stone (1979, CH 2). Although not widely applied as a stand-alone estimation method, non-iterative PROX does provide the initial estimates for Joint Maximum Likelihood Estimation in Winsteps. Linacre (1994) describes iterative PROX estimation equations for missing data. When item parameters are fixed, this iterative version of PROX becomes non-iterative.

This study examines three primary research questions:

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

**METHODS**

Using R, simulated Rasch items and person responses were created that adhere to the following conditions:

1. Normally distributed item difficulties and person thetas with μ = 0 and σ = 1;
2. Normally distributed item difficulties and person thetas with μ = 0 and σ = 2;
3. Normally distributed item difficulties (μ = -1, σ = 1) and person thetas (μ = +1 and σ = 1);
4. Normally distributed item difficulties (μ = -2, σ = 1) and person thetas (μ = +2 and σ = 1);
5. Normally distributed item difficulties (μ = -3, σ = 1) and person thetas (μ = +3 and σ = 1);
6. Normally distributed item difficulties (μ = 0, σ = 1) and bimodally distributed person thetas (μ1 = -1.5, μ2 = +1.5, and σ1 = σ2 = 1).

These different distribution combinations allow us to investigate the effects of distribution changes between item difficulties and person abilities among the different estimation methods.

The number of simulated persons will be one of n = 25, 50, 100, 250, 500, or 1000, and the number of items is fixed at 200. Given that we are simulating using the Rasch model, these numbers of persons reflect both a smallest viable sample (25) and a more than adequate sample (1000) to be confident in the calculation of person ability estimates in typical use cases. The choice of using 200 items for our simulation allows us to mimic the typical test length seen in certification and licensure examinations. Fully crossing the sample size and the distribution conditions leads to 36 simulated conditions, each of which were repeated 100 times for a total of 3600 conditions for analysis.

In addition to the PROX method, we compare four other common IRT estimators as implemented in specific R packages (shown in parentheses). These are as follows:

1. Joint Maximum Likelihood Estimation (JMLE; TAM package)
2. Marginal Maximum Likelihood Estimation (MMLE; TAM package)
3. Conditional Maximum Likelihood Estimation (CMLE; eRm package)
4. Expected A Postiori Estimation (EAP; ltm via the irtoys package)

The implementations of JMLE, MMLE, and EAP allow for fixed item parameters, so in these cases we feed the simulated item parameters into the estimation functions. However, for CMLE, we are unable to do this so item and person parameters are simultaneously estimated in this case.

**RESULTS**

The results of the analysis will compare the estimation methods in terms of the mean bias and RMSE of the estimates of the person parameters. We will also investigate the degree to which each estimation method aligns with others.

Initial analyses using the first distributional condition suggests that the methods recover person parameters with similar accuracy.

**DISCUSSION**

Different estimation methods have their own unique set of pros and cons, which analysts must consider depending on the testing situation. Although the initial rationale for non-iterative PROX is no longer salient, given the increase in computing power over the past 30 years, there are instances when an IT department is unable to incorporate specialty software into an automated scoring system and non-iterative PROX provides a uniquely simple solution as it is easily computed in SQL.

*Limitations*

There are no R packages that estimate CMLE with fixed item parameters, so additional care must be taken with the interpretating the comparison of those estimates with other methods. Also, the decision to fix the number of items at 200 was made to ensure that the scope of the project was reasonable. Future research should consider the impact of test length.

**REFERENCES**

Cohen, L. (1979). Approximate expressions for parameter estimates in the Rasch model. *British Journal of Mathematical and Statistical Psychology*, *32*(1), 113–120. https://doi.org/10.1111/j.2044-8317.1979.tb00756.x

Linacre, J. M. (1994). PROX with missing data, or known item or person measures. *Rasch Measurement Transactions*, *8*(3), 378.

Wright, B. D., & Stone, M. H. (1979). *Best Test Design*. MESA Press.