Precision of power computations for pseudo-exact or conditional tests

Jan Philipp Nolte & Clemens Draxler

UMIT – The Health and Life Science University, Austria

International Meeting of the Psychometric Society 2018
Columbia University, NYC, NY, USA, July 11, 2018

Overview

I Pseudo-exact or conditional tests

II Power function of conditional tests illustrated on DIF analysis in the Rasch model

III Computational issues: Verhelst's Rasch Sampler vs. Exact Sampler

IV Results

V Implications for practice

Pseudo-exact or conditional tests

Consider investigating DIF in the Rasch model.

Let

$Y_{ij} \in \{0, 1\}$ be the response of person $i = 1, \ldots, n$ to item $j = 1, \ldots, k$,

$x_i \in \{0, 1\}$ be a binary covariate,

$\tau_i \in \mathbb{R}$ and $\beta_j \in \mathbb{R}$ be person and item parameters,

$\delta_j \in \mathbb{R}$ be the conditional effect of item j given x_i,

$P(Y_{ij} = y_{ij} | x_i) \propto \exp[y_{ij}(\tau_i + \beta_j + x_i \delta_j)]$.

(Draxler & Zessin, 2015)

Pseudo-exact or conditional tests

Joint distribution of all binary responses

$$\prod_{i=1}^{n}\prod_{j=1}^{k} P(Y_{ij} = y_{ij}|x_i).$$

Factorization shows that sufficient statistics are

$R_i = \sum_{j=1}^{k} Y_{ij}$ (row sums) and $S_j = \sum_{i=1}^{n} Y_{ij}$ (column sums) for τ_i and β_j,

$T_j = \sum_{i=1}^{n} x_i Y_{ij}$ for δ_j.

The τs and βs are treated as nuisance, δs are the parameters of interest.

(Draxler & Zessin, 2015)

Pseudo-exact or conditional tests

Elimination of τs and βs by considering conditional distribution

$$P(T_1 = t_1, \ldots, T_{k-1} = t_{k-1} | \boldsymbol{x}, \boldsymbol{R} = \boldsymbol{r}, \boldsymbol{S} = \boldsymbol{s}) = \frac{\sum_{\mathrm{T}} \exp\left(\sum_{j=1}^{k} t_j \delta_j\right)}{\sum_{\Omega} \exp\left(\sum_{j=1}^{k} t_j \delta_j\right)}, \quad (1)$$

$\boldsymbol{x}' = (x_1, \ldots, x_n)$, $\boldsymbol{R}' = (R_1, \ldots, R_n)$, $\boldsymbol{S}' = (S_1, \ldots, S_k)$,

Ω is the set containing all (possible) matrices given $\boldsymbol{R} = \boldsymbol{r}$ and $\boldsymbol{S} = \boldsymbol{s}$,

$\mathrm{T} \subseteq \Omega$ as a subset of binary matrices satisfying $T_1 = t_1, \ldots, T_k = t_k$,

$\delta_k = 0$ for identifiability.

(Draxler & Zessin, 2015)

Power function

Suppose the interest lies in testing the hypothesis

$$\delta_1 = ... = \delta_{k-1} = 0$$

against the alternative that at least one δ is different from 0.

Let C be the critical region of size α (being the probability of the error of the first kind) of the statistical test obtained by

$$\beta(\delta_1, ..., \delta_u) = \sum_C P(T_1 = t_1, \ldots, T_{k-1} = t_{k-1} | \boldsymbol{x}, \boldsymbol{R} = \boldsymbol{r}, \boldsymbol{S} = \boldsymbol{s}).$$

(Draxler & Zessin, 2015)

How to choose the critical region C

Fundamental lemma of Neyman and Pearson (1933):

One will obtain a uniformly most powerful unbiased test if C consists of the 100α % of matrices within Ω yielding the largest values of

$$\frac{P_1(\cdot \mid \cdot)}{P_0(\cdot \mid \cdot)}.$$

Each matrix in Ω is evaluated by (1), first under the assumption that all δs are 0 and second under the assumption that at least one δ is unequal to 0.

Thus, we obtain the probability for each matrix under the null hypotheses and under the alternative.

(Neyman & Pearson, 1933)

Computational issues: Rasch Sampler vs. Exact Sampler

But the complete set Ω (and therefore C) of all $n \times k$ matrices is unknown and not computable under usual circumstances.

Miller & Harrison (2013) solved the combinatorial problem of counting the size of Ω given the marginal sums. Thus, it's possible to draw every matrix from Ω with equal probability. Practically, the procedure is limited to smaller matrices ($n + k \leq 100$).

Verhelst's Rasch Sampler uses Markov Chain Monte Carlo methods to approximate the discrete uniform distribution over Ω. Then it randomly draws with approximately equal probability for each matrix (up to 1024×64).

A random component remains since both samplers only draw a **subset** of Ω.

(Draxler & Zessin, 2015; Verhelst, 2008; Miller & Harrison, 2013)

Research questions

→ What is the precision of the power of conditional tests as suggested by Draxler & Zessin (2015)?

→ How accurate is the Rasch Sampler in approximating the discrete uniform distribution over Ω?

Rasch Sampler vs. Exact Sampler

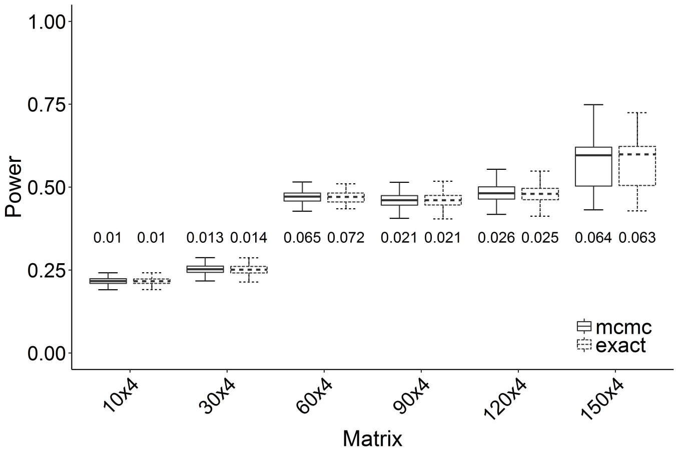

Figure 1. Box plots with standard deviations of the probability distribution of the power values calculated by the Rasch Sampler and Exact Sampler (3000 draws from Ω, repetitions = 1000)

Sample size variation

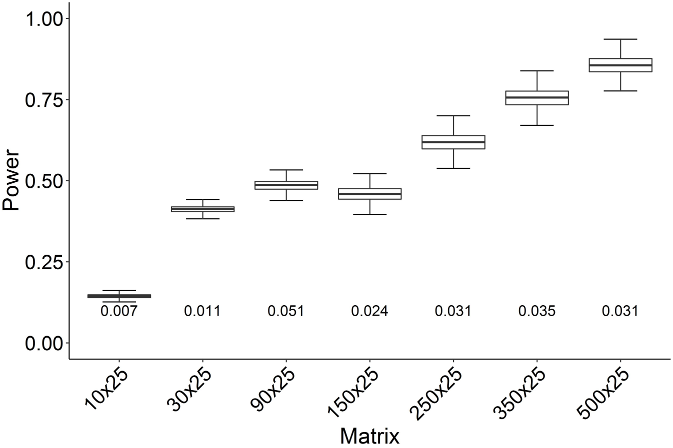

Figure 2. Box plots with standard deviations of the probability distribution of the power values calculated by the Rasch Sampler (8000 draws from Ω, repetitions = 3000)

Moderate item difficulty

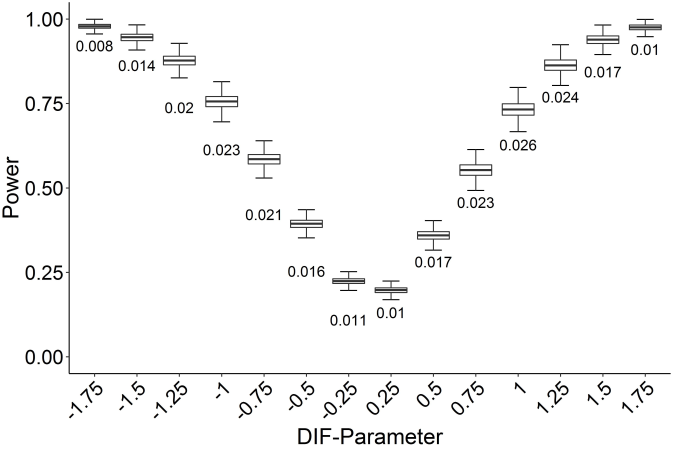

Figure 3. Box plots with standard deviations of the probability distribution of the power values varying the DIF parameters for an item with moderat item difficulty

Low item difficulty

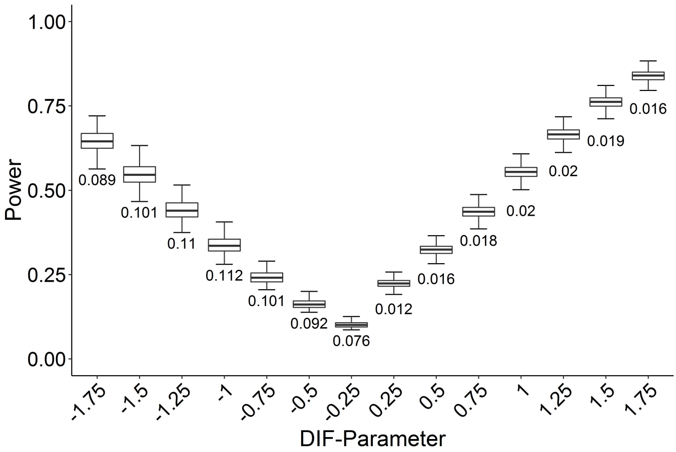

Figure 4. Box plots with standard deviations of the probability distribution of the power values with an easy item chosen as DIF

High item difficulty

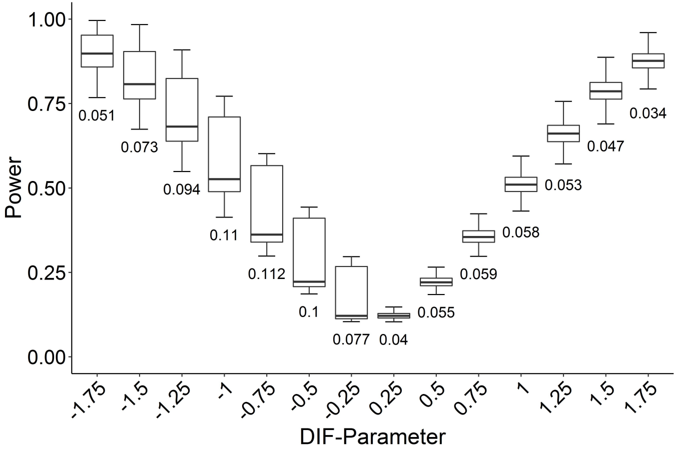

Figure 5. Box plots with standard deviations of the probability distribution of the power values with a difficult item chosen as DIF

High item difficulty: means

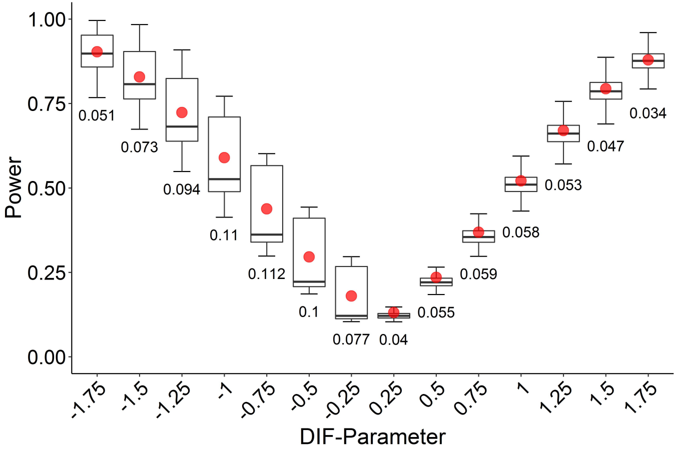

Figure 6. Box plots with **means** and standard deviations of the probability distribution of the power values with a difficult item chosen as DIF

High item difficulty: density

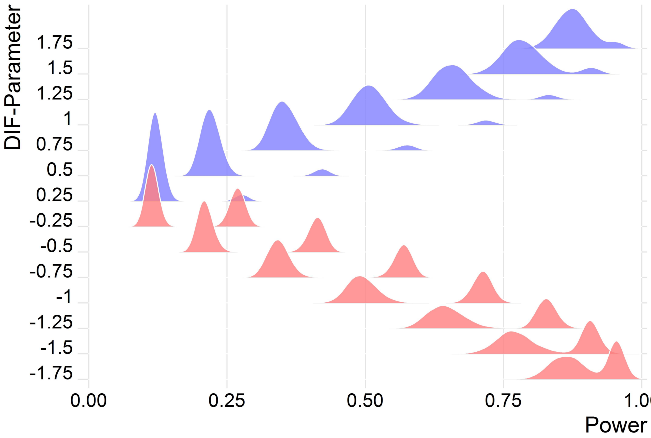

Figure 7. Density functions of the power values with a difficult item chosen as DIF

Moderate item difficulty: density

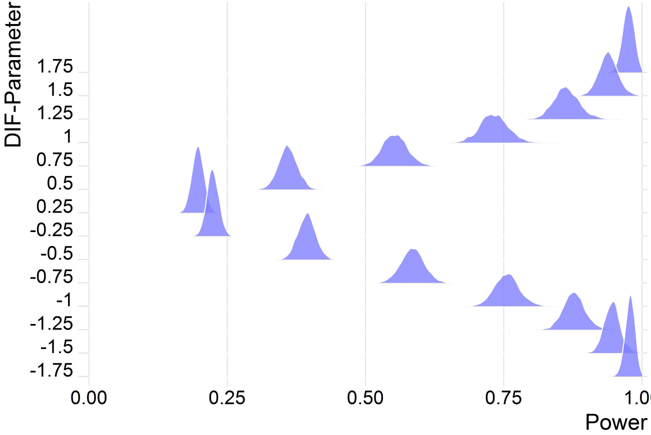

Figure 8. Density functions of the power values with an item of moderate difficulty chosen as DIF

Implications for practice

No need for exact sampling, since Verhelst's Rasch Sampler is as accurate as the Exact Sampler (at least for cases of small matrices that have been investigated in this study).

Conditional tests are powerful.

High accuracy of conditional tests due to small *SD*'s over multiple thousand replications.

For extreme items chosen to be affected by DIF observed distributions of power values are often bimodal with obvious larger *SD*'s.

References

Draxler, C. & Zessin, J. (2015). The power function of conditional tests of the Rasch model. *AStA Advances in Statistical Analysis*, *99*(3), 367–378.

Miller, J. W. & Harrison, M. T. (2013). Exact sampling and counting for fixed-margin matrices. *The Annals of Statistics*, *41*(3), 1569–1592.

Neyman, J. & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. R. Soc. Lond. A*, *231*(694-706), 289–337.

Verhelst, N. D. (2008). An efficient mcmc algorithm to sample binary matrices with fixed marginals. *Psychometrika*, *73*(4), 705–728.