

Computational Precision of the Power Function for Conditional Tests of Assumptions of the Rasch Model

Clemens Draxler, Jan Philipp Nolte

UMIT—The Health and Life Sciences University, Hall in Tirol, Austria

Email: clemens.draxler@umit.at, jan.nolte@umit.at

How to cite this paper: Draxler, C. and Nolte, J.P. (2018) Computational Precision of the Power Function for Conditional Tests of Assumptions of the Rasch Model. *Open Journal of Statistics*, 8, 873-884. <https://doi.org/10.4236/ojs.2018.86058>

Received: August 28, 2018

Accepted: November 12, 2018

Published: November 15, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Draxler and Zessin [1] derived the power function for a class of conditional tests of assumptions of a psychometric model known as the Rasch model and suggested an MCMC approach developed by Verhelst [2] for the numerical approximation of the power of the tests. In this contribution, the precision of the Verhelst approach is investigated and compared with an exact sampling procedure proposed by Miller and Harrison [3] for which the discrete probability distribution to be sampled from is exactly known. Results show no substantial differences between the two numerical procedures and quite accurate power computations. Regarding the question of computing time the Verhelst approach will have to be considered much more efficient.

Keywords

Conditional Tests, Conditional Probability Distribution, Hypergeometric Distribution, Power Function, Random Sampling, Rasch Model

1. Introduction

Draxler and Zessin [1] derived the power function for conditional tests of assumptions of a psychometric model known as the Rasch model [4] [5]. These tests can be viewed as a generalization of Fisher's exact test for testing independence in contingency tables by considering an extended covariance structure. The exact probability distributions of the test statistics under both the hypothesis to be tested and a deviation from it are obtained from a family of conditional probability mass functions which can be considered as a generalization of the class of multidimensional non-central hypergeometric distributions [6]. These are rather complicated and time-consuming to compute. Hence, in practice, these

distributions as well as the power function of the tests are usually approximated by random sampling processes. Basically, there are three approaches to accomplish this: sequential importance sampling suggested by Chen *et al.* [7] as well as Chen and Small [8], a Markov chain Monte Carlo (MCMC) approach by Verhelst [2], and so-called exact sampling [3].

Verhelst's MCMC technique may be considered as the most promising in terms of handling practically realistic cases in psychometric research (regarding sample sizes and item numbers) and computing times. On the basis of the stationary distribution of the Markov chain the conditional probability distributions of interest can be computed to obtain size, p value, and power of the tests. The stationary distribution of the chain can be arbitrarily well approximated. Unlike the MCMC technique, the exact sampling approach is based on an analytical solution of a combinatorial problem which arises as a consequence of the conditioning involved in the procedure. This solution enables the exact determination of the conditional probability distributions of interest but, nonetheless, computing them in practically relevant cases in psychometrics is still too time-consuming so that one still relies on random sampling from the known exact distributions.

This paper essentially deals with two questions. The first one generally concerns the precision of power computations of conditional tests as introduced by Draxler and Zessin [1] in case of using random sampling procedures for the approximation of the exact power function. Thereby, precision is expressed in terms of dispersion measures observed for the power values (e.g. variances, standard deviations, quantiles, etc.) computed from the random samples drawn. The second question specifically deals with a comparison of Verhelst's MCMC and the exact sampling approach.

2. The Problem, Conditional Tests, and Their Power Function

Consider a typical psychometric problem that a sample of n persons responds to k items. Let $Y_{ij} \in \{0,1\}$ denote the binary response of person $i = 1, \dots, n$ to item $j = 1, \dots, k$ and $x_i \in \{0,1\}$ be a fixed covariate, *i.e.* a known characteristic of the persons like gender. The covariate may be also treated as a random variable. Examples are quoted by Draxler and Zessin [1]. Consider the following exponential family of probability distributions given by

$$P(Y_{ij} = y_{ij} | x_i) \propto \exp\left[y_{ij}(\tau_i + \beta_j + x_i\delta_j)\right], \quad (1)$$

with $\tau_i \in \mathbb{R}$ as a person parameter, $\beta_j \in \mathbb{R}$ as an item parameter, and $\delta_j \in \mathbb{R}$ as the conditional effect of the item given the covariate. Assuming local independence of the Y s the joint distribution of all binary responses is obtained by

$$P(\mathbf{Y} = \mathbf{y} | \mathbf{x}) = \prod_{i=1}^n \prod_{j=1}^k P(Y_{ij} = y_{ij} | x_i), \quad (2)$$

with \mathbf{Y} as an $n \times k$ matrix-valued random variable containing the Y s

arranged in n rows and k columns and with $\mathbf{x}' = (x_1, \dots, x_n)$. Factorizing this product immediately shows that the statistics $R_i = \sum_j Y_{ij}$, $S_j = \sum_i Y_{ij}$, and $T_j = \sum_i x_i Y_{ij}$ are sufficient for τ_i , β_j , and δ_j . Note that the former two sufficient statistics are the row and column sums of the matrix of responses Y . Suppose the interest lies in making inferences about the δ s where the τ s and β s are treated as nuisance parameters. One way of eliminating the influence of nuisance parameters is conditioning on the observed values of their sufficient statistics. Proceeding in this way one obtains the conditional distribution

$$P(T_1 = t_1, \dots, T_{k-1} = t_{k-1} \mid \mathbf{R} = \mathbf{r}, \mathbf{S} = \mathbf{s}, \mathbf{x}) = \frac{\sum_{\mathbf{T}} \exp\left(\sum_{j=1}^k t_j \delta_j\right)}{\sum_{\Omega} \exp\left(\sum_{j=1}^k t_j \delta_j\right)}, \quad (3)$$

with $\mathbf{R}' = (R_1, \dots, R_n)$, $\mathbf{S}' = (S_1, \dots, S_k)$. For identifiability let $\delta_k = 0$. Note that all information needed for making inferences about the δ s is provided by the T statistics because of their sufficiency property. Hence, the original observations, the Y s, can be represented in condensed form. It suffices to consider the joint distribution of the T s as a function of the Y s. Note also that at least one of the T s is not free conditional on $\mathbf{R} = \mathbf{r}, \mathbf{S} = \mathbf{s}$. The denominator on the right side of (3) is a normalizing constant requiring a summation over the set Ω . The latter is the set of all possible $n \times k$ matrices given the condition $\mathbf{R} = \mathbf{r}, \mathbf{S} = \mathbf{s}$. In other words, this is the set of all matrices with given, fixed row and columns sums. The subset $\mathbf{T} \subseteq \Omega$ contains those $n \times k$ matrices additionally satisfying $T_1 = t_1, \dots, T_k = t_k$.

Suppose the interest lies in testing the composite hypothesis $\delta_1 = \dots = \delta_{k-1} = 0$ against the alternative $(\delta_1, \dots, \delta_{k-1})' = \mathbf{c}_{k-1}$, where \mathbf{c}_{k-1} is any $(k-1)$ -dimensional column vector of constants except a $(k-1)$ -dimensional column vector of zeros, *i.e.* at least one δ is different from 0. Note that both hypotheses would be termed simple if the δ s were the only parameters involved in the problem. The restriction on the parameter space of the free δ s given by the hypothesis to be tested yields the Rasch model as a special case which assumes the Y s independent of the covariate. In other words, the hypothesis to be tested is equivalent with the well-known scenario of testing the equality of the item parameters of the Rasch model between two groups of persons. In psychometric literature, such an analysis is known as testing the invariance of the item parameters or, more general, as investigating differential item functioning (DIF). Moreover, if the covariate vector \mathbf{x} divides the sample of persons according to their scores, *i.e.* the row sums of Y , yielding one group of persons with low score and another with high score the hypothesis $\delta_1 = \dots = \delta_{k-1} = 0$ will be equivalent to the assumption of equal item discriminations. This is a basic assumption of the Rasch model which has to be tested in almost every application. Thus, the conditional procedure discussed can be considered to have practical potential.

A most powerful test and its power function are obtained as follows. Let α

denote the probability of the error of the first kind and C the critical region of the test. Consider the $(k-1)$ -dimensional sufficient statistic T_1, \dots, T_{k-1} for $\delta_1, \dots, \delta_{k-1}$ to serve as the test statistic. Denote by P_0 the conditional distribution given by (3) evaluated at $\delta_1 = \dots = \delta_{k-1} = 0$ (the hypothesis to be tested) and by P_1 the respective distribution evaluated at $(\delta_1, \dots, \delta_{k-1})' = \mathbf{c}_{k-1}$ (the alternative). According to the fundamental lemma of Neyman and Pearson [9] one will obtain a most powerful critical region if C is composed of those values of the test statistic T_1, \dots, T_{k-1} which yield the 100α % largest values of P_1/P_0 . Eventually, the power function of a critical region C chosen this way is obtained by $\sum_C P_1$. Note that Fisher's well-known exact test is obtained as a special case by setting $k = 2, \mathbf{R} = \mathbf{1}_n = (R_1 = 1, \dots, R_n = 1)'$. In this case, (3) becomes the one dimensional non-central hypergeometric distribution and, under the hypothesis to be tested, the (central) hypergeometric distribution.

For further conditional tests and their power functions as well as an application to real-world data from educational research one is referred to Draxler and Zessin [1]. Other tests of various assumptions of the Rasch model also based on Ω have already been suggested by Ponocny [10]. A Bayesian approach also based on the conditional distribution (3) is discussed by Draxler [11].

3. Computational Issues

To compute the conditional distribution given by (3) one obviously has to determine the cardinalities of the two sets T and Ω . Counting the total number of matrices in Ω is not an easy task. Miller and Harrison [3] suggest an exact recursive counting algorithm based on graph theory, the Gale-Ryser theorem [12] [13], and dynamic programming which, additionally, enables exact and efficient sampling from Ω . Their solution can be considered as the fastest and most efficient amongst other exact algorithms (excluding approximate solutions) up until now. It is feasible for many real-world applications, primarily in ecological research and also in some cases in psychometrics. It may be also reasonably used for the evaluation of the accuracy of approximate algorithms like the ones mentioned in the introduction. Nonetheless, in most cases in psychometric research the size of Ω (the total number of matrices) will be probably too large for the exact algorithm because of RAM capacity limitations of the usual desktop machines. For practice, one can suggest approximately $n + k < 100$. In the majority of cases in psychometric research the matrices are far larger so that the number of matrices in Ω is usually not determined or counted but random sampling procedures are used to approximate the ratio of cardinalities of T and Ω . Since $T \subseteq \Omega$ this can be accomplished by drawing a random sample of matrices from the discrete uniform distribution of the elements (matrices) of Ω . One simply has to count the number of matrices within the sample which satisfy $T_1 = t_1, \dots, T_k = t_k$, *i.e.* those matrices drawn from the set T within all matrices drawn. The precision or the variance of this relative frequency obviously depends

on the size of the random sample. Hence, to approximate the conditional distribution (3) under both $\delta_1 = \dots = \delta_{k-1} = 0$ and $(\delta_1, \dots, \delta_{k-1})' = \mathbf{c}_{k-1}$ the summations in the numerator and denominator on the right side of (3) have to be taken only over the respective matrices drawn.

Miller and Harrision [3] also suggest an exact sampling algorithm that can be used after counting the exact total number of matrices in Ω . In this case, the discrete uniform distribution over Ω is exactly known and one can directly sample from it ensuring that every matrix has the same probability of being selected. As already remarked, the practicality of this exact procedure is only ensured for smaller matrices or very sparse larger matrices. The program package EXACT, an efficient C routine with R, Python, and Matlab wrappers, can be downloaded from:

<http://jwmi.github.io/software.html><http://jwmi.github.io/software.html>. An alternative for larger matrices as they typically occur in psychometric research are procedures for the approximation of the desired discrete uniform distribution over Ω . These are a sequential importance sampling [8] and an MCMC approach by Verhelst [2]. The former is less appropriate compared to the latter with respect to psychometric problems. Therefore, it is not considered any further. Verhelst's MCMC approach may be considered as the most efficient and fastest sampling algorithm amongst all algorithms (exact and approximate) available. The stationary distribution of the Markov chain is given by the discrete uniform distribution over Ω which can be, in principle, arbitrarily well approximated by using an appropriate burn in phase of the algorithm. This ensures random draws with approximate equal probabilities for every matrix. It is easily accessible as an R package [14] which is practicable for larger matrices up to 4096 rows and 128 columns. No other technique can handle matrices of this size.

4. Study Design

A first natural question continuing, supplementing, and enhancing the work of Draxler and Zessin [1] concerns the precision or accuracy of the power computations and the approximation of the exact power, respectively, in case of drawing random samples from Ω regardless of the particular sampling approach utilized. For this, the MCMC approach by Verhelst [2] shall be chosen since it is practical for larger matrices. Hence, one may get an idea of the precision of power computations in cases as they typically occur in psychometric research. The chosen scenarios consider matrices with row numbers 10, 30, 90, 150, 250, 350, 500 and 25 columns in each row number case. The row sums of each of these matrices, also called person scores, are chosen so that values in the middle of the possible range from 0 to 25 are more frequent than values near 0 and near 25 obtaining best possible symmetric frequency distributions (best possible zero skewness). Concerning the choice of the 25 column sums, also called item scores, low, middle, and high values have been selected more or less equally frequent.

Note that these choices are reasonable with respect to psychometric research since the items of a psychological (or educational) test are usually selected to cover a wide range of difficulties (item parameters) to ensure best possible accurate measurements for all persons, *i.e.* persons from very low to very high ability (person parameter). The covariate vector \mathbf{x} is chosen so that it divides the sample of persons in two groups of equal sizes. Regarding each of the scenarios to be considered, a random sample of 8000 matrices is drawn using Verhelst's MCMC procedure and the power is computed given a choice of c_{k-1} and given $\alpha = 0.05$. This procedure is replicated 3000 times to observe the distribution of power values and thus, to get an idea of the accuracy of the computations. Concerning the choice of c_{k-1} the following different scenarios are selected. In all scenarios only one of the $k-1$ free δ s is selected to take a value different from 0. These values range from -1.75 to 1.75 with a spacing of 0.25 (excluding 0). The rest of the δ s is set equal to 0 . The particular δ parameter that is chosen to differ from 0 is called DIF parameter in the following since it defines a deviation from the hypothesis to be tested. As already noted in Sec. 2, such a deviation is often called DIF in psychometrics. The DIF parameter refers to different items, *i.e.* to items with a score in the middle of the possible range as well as to items with small and high scores.

A second question focuses on potential differences with respect to power computations between the exact sampling [3] and the Verhelst MCMC approach. Since the latter only approximates the discrete uniform distribution over Ω through a Markov chain the power computations may be imprecise (having larger variances) compared to the exact sampling procedure which samples from the exactly determined discrete uniform distribution. These comparisons can be carried out only for smaller matrices of approximate size $n+k < 100$ because of computational limitations of the exact sampling procedure. The selected matrices refer to 4 items and 10, 30, 60, 90, 120 and 150 persons. Note that the small number of items, *i.e.* matrices with only 4 columns, allows for the consideration of person numbers up to 150. These choices are reasonable and practical from the psychometric point of view since one usually has person numbers far exceeding the numbers of items. Thus, a 100×4 matrix is more realistic than a 50×50 , say. Moreover, concerning the exact sampling approach, the latter case is already too much for usual RAM capacities of today's machines, whereas the case 100×4 is quite manageable.

The chosen column sums for each person number condition are illustrated in **Table 1**. The frequency distribution of the person scores, row sums, is shown in **Table 2**, where person scores 0 and 4 are excluded. These are uninformative. The size of the random sample drawn from Ω is set to 3000 . The selected values of the DIF parameter range again from -1.75 to 1.75 but results are presented only for the case 0.6 (in Sec. 5). These results may be viewed as typical. No substantially different results have been observed with respect to all other choices of the DIF parameter as far as the comparison of the two sampling

Table 1. Selected values of the item scores, column sums.

Item	Person number					
	10	30	60	90	120	150
1	8	23	60	73	95	117
2	6	17	28	51	69	89
3	3	12	16	34	41	53
4	2	7	12	15	19	24

Table 2. Frequency distributions of the persons scores or row sums.

Person Score	Person number					
	10	30	60	90	120	150
1	3	8	17	26	38	46
2	5	15	30	45	60	75
3	2	7	13	19	22	29

approaches is concerned. One can observe only the trivial fact of increasing power with increasing absolute value of the DIF parameter regardless of the sampling technique used. Moreover, an absolute value of roughly at least 0.5 or 0.6 of the DIF parameter may be considered meaningful in most practical contexts in psychometric research. For a deeper discussion on the practical meaning of a deviation from the hypothesis to be tested in a broader context of power and sample size issues one is referred to Draxler [15]. The power is computed using both Verhelst's MCMC approach and the exact sampling procedure given $\alpha = 0.05$. This procedure is replicated 1000 times to observe the distribution of the power values with respect to both sampling procedures. Note the reason for decreasing the number of draws from Ω as well as the number of replications compared to Question 1 which is the greater computational effort and computing time needed for the execution of the exact sampler.

5. Results

The observed results with respect to question 1 are summarized as follows.

Table 3 shows summary statistics of the power values for different sample sizes and a constant DIF parameter of 0.6 referring to an item with a score in the middle of the possible range. As can be seen, the standard deviations are quite small so that the power computations may be considered quite stable. An exception is the 90 person number case which yields a considerably larger dispersion and additionally a slightly higher mean power than the 150 person number case which one would not expect. In this case, the number of random draws from Ω may have to be increased to get more accurate results.

Figure 1 shows box plots of the power values concerning two scenarios. The diagram on the left side of **Figure 1** illustrates a scenario with DIF parameter

referring to an item with score in the middle of the range of possible values.

The diagram on the right side concerns the case assuming a difficult item (low item score) affected by DIF. As can be seen, the effect of the DIF parameter on the observed power depends on which item is assumed to be affected by DIF. These results are expected from theory. To explain, consider the diagram on the right side of **Figure 1**. Negative values of the DIF parameter yield smaller effects on the power (smaller power on average) as well as larger standard deviations of the power values and thus less precision than positive values. This is because a negative delta (DIF parameter) shifts the response probabilities of the respective item closer to the boundary 0 for persons with covariate value 1. Since the response probabilities are more extreme for these persons the responses contain less information about the respective δ (DIF) parameter. As a consequence, its effect on the power as well as the precision of the power computations become smaller. **Figure 2** shows scenarios for which the DIF parameter refers to an item

Table 3. Summary statistics of the observed frequency distributions of power values for different sample sizes.

Matrix	Min	2.5% Quan.	25% Quan.	Median	Mean	75% Quan.	97.5% Quan.	Max	SD
10 × 25	0.12	0.13	0.14	0.14	0.14	0.15	0.16	0.17	0.007
30 × 25	0.37	0.39	0.40	0.41	0.41	0.42	0.43	0.47	0.011
90 × 25	0.28	0.31	0.47	0.49	0.47	0.50	0.51	0.53	0.051
150 × 25	0.38	0.42	0.44	0.46	0.46	0.48	0.51	0.59	0.024
250 × 25	0.53	0.56	0.60	0.62	0.62	0.64	0.68	0.76	0.031
350 × 25	0.61	0.68	0.73	0.76	0.76	0.78	0.82	0.97	0.035
500 × 25	0.75	0.80	0.84	0.86	0.86	0.88	0.92	0.97	0.031

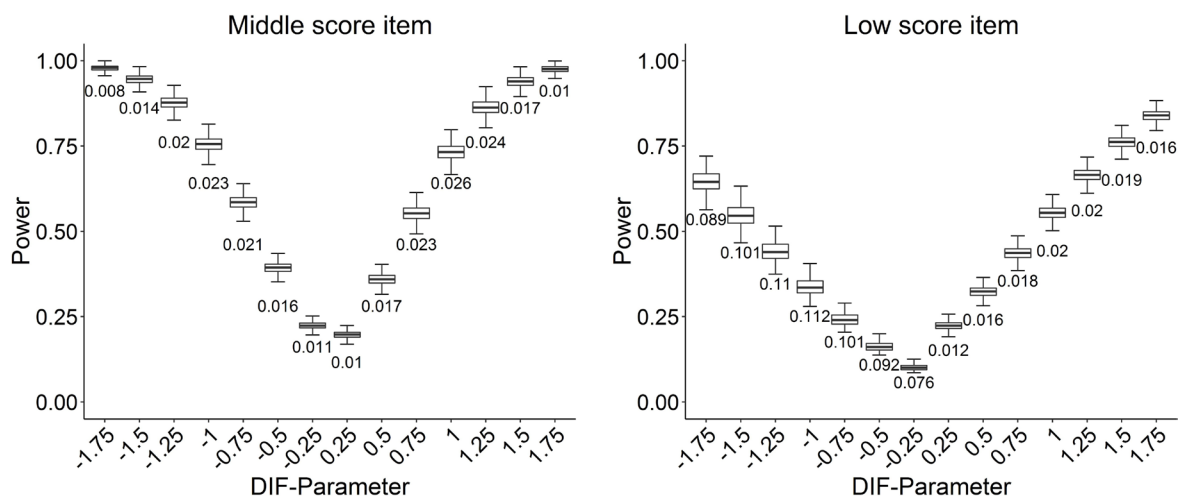


Figure 1. Box plots of observed frequency distributions of power values and standard deviations for two cases differing in the choice of the item affected by DIF. On the left the DIF parameter refers to an item with a score in the middle of the range of possible values, on the right the DIF parameter refers to an item with a low score (difficult item).

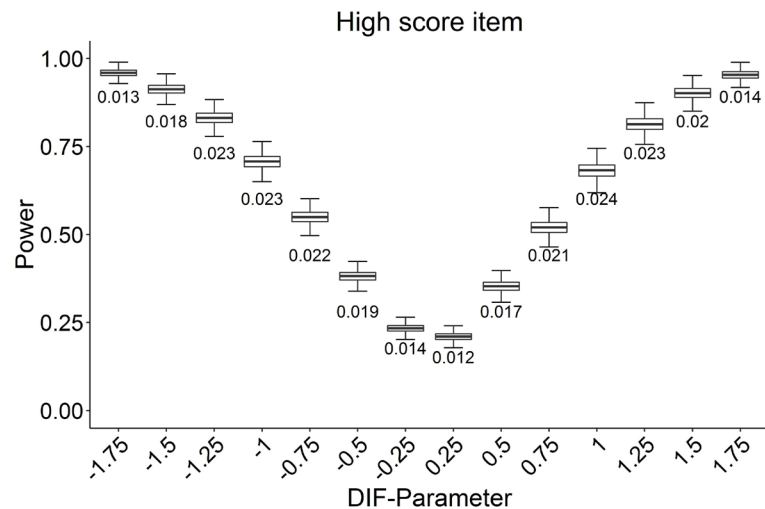


Figure 2. Box plots of observed frequency distributions and standard deviations of power values. The DIF parameter refers to an item with a high score (easy item).

with high score. In these cases, the effect on the power computations is exactly the other way round *i.e.* positive values of the DIF parameter yielding smaller power on average even though not so obvious as on the right side of **Figure 1**. This is just because the score of the item is not so extreme as in the example shown in **Figure 1** (right side).

Generally, as seen in all diagrams the observed standard deviations of the power values are quite small. They are higher for those scenarios yielding a mean power around 0.5 which is also obvious from theory. Thus, the computations may be considered quite stable. Further analyses have been carried out decreasing the number of samples drawn from Ω to 3000 and even to 1000 without considerably increasing the standard deviations of the observed power values.

Moreover, the burn in phase of the MCMC algorithm has been varied from 300 up to 8000 as well as three different values of the so-called step parameter have been considered, *i.e.* 16, 32, and 50. The latter is simply to avoid dependence of the matrices to be drawn (states in the Markov chain), e.g. 16 means that only every 16th matrix is selected. **Figure 3** shows a few exemplary results. It can be summarized that the precision of the computations seems to be barely influenced by the choice of the size of the burn in sample. Thus, the approximation of the discrete uniform distribution over Ω seems to be sufficient already in cases of burn in samples as small as 300. The step parameter, on the contrary, does seem to have a certain impact on the precision of the computations. Increasing it yields decreasing standard deviations.

Finally, **Figure 4** illustrates results concerning Question 2, the comparison of Verhelst's MCMC and the exact sampling procedure. Obviously, no noticeable difference between those two sampling procedures have been observed. Note that the observed average power is slightly larger in the case considering 60 persons than in the one with 90. One would not expect a higher power in case of

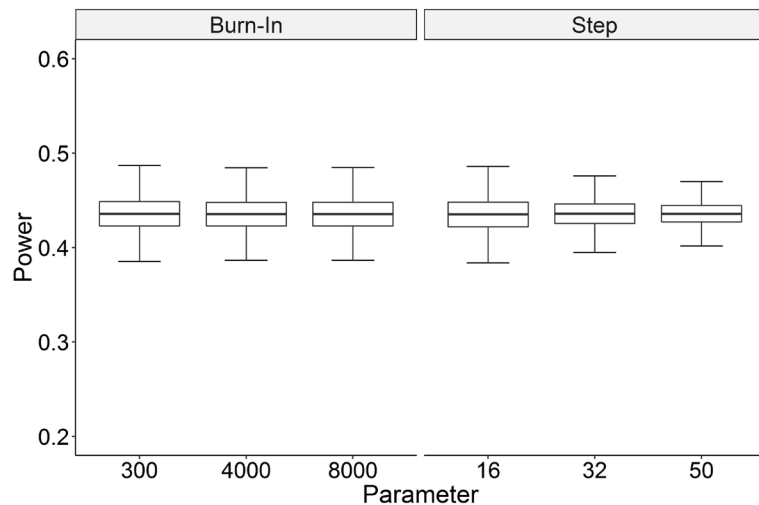


Figure 3. Typical results observed concerning the impact of burn in phase and step parameter of Verhelst’s MCMC procedure.

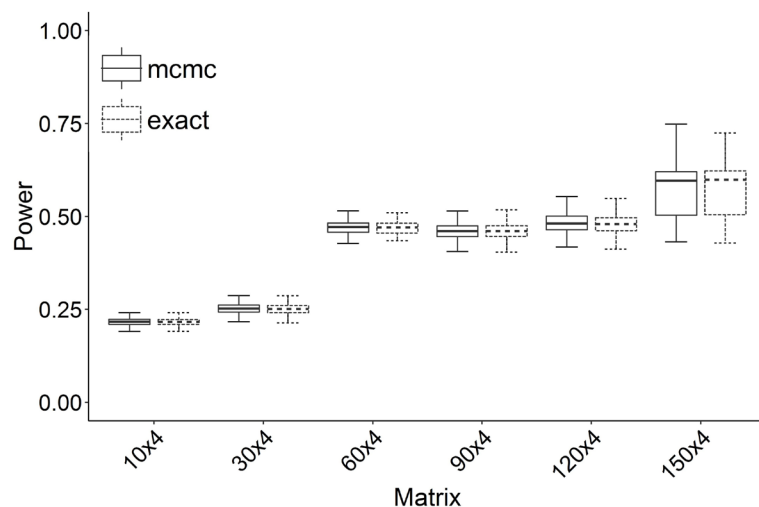


Figure 4. Box plots of frequency distributions of observed power values obtained from computations based on Verhelst’s MCMC and exact sampling approach.

a smaller sample size but, nonetheless, this result is quite understandable. In the presented examples the DIF parameter refers to item 2. In **Table 1** it can be seen that its score is 28 in the 60 person number case, whereas it is 51 in the case of 90 persons. The latter, 51, is more extreme, *i.e.* it is farther from the center of the range of values from 0 to 90. Consequently, the effect of the DIF parameter is smaller and one observes even a slightly lower average power than in the 60 person scenario.

6. Final Remarks

In cases of smaller sample sizes, *i.e.* up to a few hundred, conditional testing of assumptions of the Rasch model [1] is a reasonable alternative to asymptotic or large sample χ^2 tests usually applied in this context (e.g. [15] [16] [17] [18] [19]). In contrast to χ^2 tests the conditional procedure treated in this paper is a

one-sided hypothesis test which generally has higher power than its two sided counterpart.

The power function of the conditional test can be well approximated by numerical procedures and random sampling techniques, respectively. The results of this work hint at quite accurate and stable computations. Nonetheless, many more scenarios could be investigated. Particularly, scenarios assuming more extreme values of the person (row sums) as well as the item scores (column sums) than have been analyzed in this contribution. In such cases higher variances and less precision of the power computations have to be expected.

Probably the most important result of this contribution for the practice of psychometric data analysis is that the exact sampling approach based on an exact counting algorithm [3] for the number of matrices in Ω does not yield substantially more accurate, or rather not even slightly more accurate, power computations than the MCMC approach by Verhelst [2] which only approximates the distribution over Ω . Moreover, the Verhelst approach is practicable for considerably larger matrices and is much more efficient in terms of computing time.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Draxler, C. and Zessin, J. (2015) The Power Function of Conditional Tests of the Rasch Model. *ASTA Advances in Statistical Analysis*, **99**, 367-378. <https://doi.org/10.1007/s10182-015-0249-5>
- [2] Verhelst, N.D. (2008) An Efficient Mcmc Algorithm to Sample Binary Matrices with Fixed Marginals. *Psychometrika*, **73**, 705-728. <https://doi.org/10.1007/s11336-008-9062-3>
- [3] Miller, J.W. and Harrison, M.T. (2013) Exact Sampling and Counting for Fixed-Margin Matrices. *The Annals of Statistics*, **41**, 1569-1592. <https://doi.org/10.1214/13-AOS1131>
- [4] Rasch, G. (1960) Probabilistic Models for Some Intelligence and Attainment Tests. Danish Institute for Educational Research, Copenhagen.
- [5] Fischer, G.H. and Molenaar, I.W. (1995) Rasch Models. Foundations, Recent Developments and Applications. Springer, New York.
- [6] Draxler, C. (2011) A Note on a Discrete Probability Distribution Derived from the Rasch Model. *Advances and Applications in Statistical Sciences*, **6**, 665-673.
- [7] Chen, Y., Diaconis, P., Holmes, S.P. and Liu J.S. (2005) Sequential Monte Carlo Methods for Statistical Analysis of Tables. *Journal of the American Statistical Association*, **100**, 109-120. <https://doi.org/10.1198/016214504000001303>
- [8] Chen, Y. and Small, D. (2005) Exact Tests for the Rasch Model via Sequential Importance Sampling. *Psychometrika*, **70**, 11-30. <https://doi.org/10.1007/s11336-003-1069-1>

- [9] Neyman, J. and Pearson, E.S. (1933) On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philosophical Transactions of the Royal Society A*, **231**, 289-337. <https://doi.org/10.1098/rsta.1933.0009>
- [10] Ponocny, I. (2001) Nonparametric Goodness-of-Fit Tests for the Rasch Model. *Psychometrika*, **66**, 437-459. <https://doi.org/10.1007/BF02294444>
- [11] Draxler, C. (2018) Bayesian Conditional Inference for Rasch Models. *ASTA Advances in Statistical Analysis*, **102**, 245-262. <https://doi.org/10.1007/s10182-017-0303-6>
- [12] Gale, D. (1957) A Theorem on Flows in Networks. *Pacific Journal of Mathematics*, **7**, 1073-1082. <https://doi.org/10.2140/pjm.1957.7.1073>
- [13] Ryser, H.J. (1957) Combinatorial Properties of Matrices of Zeros and Ones. *Canadian Journal of Mathematics*, **9**, 371-377. <https://doi.org/10.4153/CJM-1957-044-3>
- [14] Verhelst, N.D., Hatzinger, R. and Mair, P. (2007) The Rasch Sampler. *Journal of Statistical Software*, **20**, 1-14. <https://doi.org/10.18637/jss.v020.i04>
- [15] Draxler, C. (2010) Sample Size Determination for Rasch Model Tests. *Psychometrika*, **75**, 708-724. <https://doi.org/10.1007/s11336-010-9182-4>
- [16] Andersen, E.B. (1973) A Goodness of Fit Test for the Rasch Model. *Psychometrika*, **38**, 123-140. <https://doi.org/10.1007/BF02291180>
- [17] Glas, C.A. and Verhelst, N.D. (1995) Testing the Rasch Model. *Rasch Models*, Springer, New York, 69-95. https://doi.org/10.1007/978-1-4612-4230-7_5
- [18] Draxler, C. and Alexandrowicz, R.W. (2015) Sample Size Determination within the Scope of Conditional Maximum Likelihood Estimation with Special Focus on Testing the Rasch Model. *Psychometrika*, **80**, 897-919. <https://doi.org/10.1007/s11336-015-9472-y>
- [19] Draxler, C. and Kubinger, K. (2018) Power and Sample Size Considerations in Psychometrics. In: Pilz, J., Rasch, D., Melas, V. and Moder, K., Eds., *Statistics and Simulation*. IWS 2015. Springer Proceedings in Mathematics & Statistics, Vol 231, Springer, Cham.