D3mirt: Descriptive Three-Dimensional Multidimensional Item Response Theory for R

30 March 2023

Summary

Descriptive multidimensional item response theory (DMIRT) has been proposed for multidimensional item-level analysis of items that fit the multidimensional extension of the two-parameter logistic model or the two-parameter multidimensional graded response model. However, there is a shortage of software extensions that implement this type of statistical method. In this paper, I introduce the `D3mirt` package for `R` that offers DMIRT analysis restricted to a three-dimensional latent model space. With this, we also introduce new contributions to the DMIRT method: the use of constructs, profile analysis, and the model identification procedure.

Statement of need

Common to most item response theory (IRT) models is the assumption of *unidimensionality*, i.e., that a test or item measures simple structures (Hambleton 1991). There are, however many occasions where this may be improper. Consider a mathematical word problem (Reckase, M. D. 2009; M. D. Reckase 1985; R. L. Reckase M. D. & McKinley 1991). To solve a mathematical word problem, one must often have both verbal and mathematical skills, or abilities (θ) as it is often called in the literature on item response theory. In other words, one's result would be a function of one's ability to read, on the one hand, and one's ability to perform numerical manipulations, on the other. Accordingly, instead of a person's location on a unidimensional latent variable, the mathematical word problem illustrates a situation where it seems more reasonable to assume that a correct response is due to the respondent's location in a multidimensional latent variable space.

Descriptive multidimensional item response theory (DMIRT) (Reckase, M. D. 2009; M. D. Reckase 1985; R. L. Reckase M. D. & McKinley 1991) has been proposed as a data reduction technique for the just mentioned situation. The method is based on using the *compensatory model*, i.e., a type of measurement model that uses linear combinations of θ-values for ability assessment. This type of model assumes that the same sum can be reached by adding different combinations of θ-values. In turn, this implies that items can be unidimensional and *within-multidimensional*, i.e., that the item measure more than one ability.

The `D3mirt` approach is limited to two types of item models, dependent on item type. If dichotomous items are used, the analysis is based on the multidimensional extension of the two-parameter logistic model (M2PL) (McKinley 1983). If polytomous items are used, the analysis is based on the two-parameter multidimensional graded response model (MGRM) (Muraki 1995). The results of the analysis are presented in tables and in interactive three-dimensional devices with vector arrows representing item response functions (see Figure 1). An example how the utility of using the package in an empirical context for item and scale analysis has recently been given in Forsberg (In Print).

Multidimensional item parameters

The theoretical framework for DMIRT rests foremost on three assumptions (M. D. Reckase 1985). Firstly, ability map the probability monotonically, such that a higher level of ability implies a higher probability of

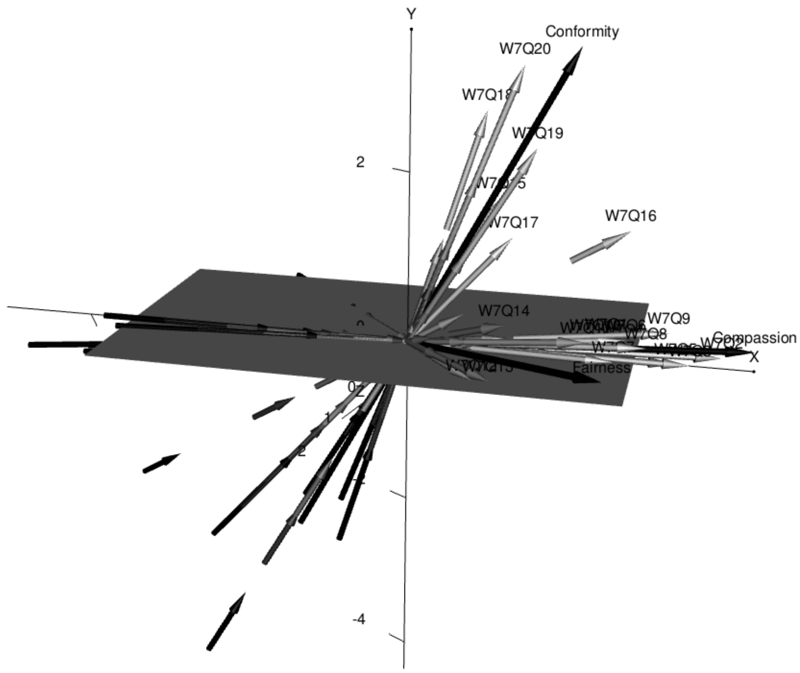

Figure 1: **Figure 1**. A still shot of the graphical output from `D3mirt`. The Figure illustrates a three-dimensional vector plot for all items in the `anes0809offwaves` data set included in the package. The output also includes the three construct vector arrows: Compassion, Fairness, and Conformity (solid black arrows).,

answering an item correctly. Second, we wish to locate an item at a singular point at which it is possible to derive item characteristics for the multidimensional case. Thirdly, an item's maximum level of discrimination, i.e., its highest possible sensitivity score for measuring ability, is the best option for the singular point estimation. The most important parameter equations capturing the just mentioned assumptions in DMIRT will be presented below.

Firstly, by using the discrimination score a_i from the compensatory model we can dervie the multidimensional analog to the unidimensional discrimination parameter, i.e., the multidimensional discrimination index (MDISC), denoted A_i for item i, to highlight the connection to the unidimensional a_i parameter (Reckase, M. D. 2009; R. L. Reckase M. D. & McKinley 1991).

$$MDISC = A_i = \sqrt{\sum_{k=1}^{m} a_{ik}^2} \tag{1}$$

With the slope constant $\frac{1}{4}$ omitted (Reckase, M. D. 2009; R. L. Reckase M. D. & McKinley 1991). Importantly, the MDISC sets the orientation of the item vectors in the multidimensional space (Reckase, M. D. 2009; R. L. Reckase M. D. & McKinley 1991), as follows.

$$\omega_{il} = cos^{-1}\left(\frac{a_{il}}{\sqrt{\sum_{k=1}^{m} a_{ik}^2}}\right) \tag{2}$$

On latent axis l in the model. Note, the ω_{il} is in this solution a characteristic of the item i that tells in what direction i has its highest level of discrimination, assuming a multidimensional latent space (Reckase, M. D. 2009; R. L. Reckase M. D. & McKinley 1991). This gives us the following criteria to use as a rule of thumb. Assume a two-dimensional space, an orientation of 0° with respect to any of the model axes indicates that the item is unidimensional. Such an item describes a singular trait only. In contrast, an orientation of 45° indicated that the item is within-multidimensional. Such an item describes both traits in the two-dimensional model equally well. The same criteria are extended to the three-dimensional case. The MDISC is also used in the graphical output to scale the length of the vector arrows representing the item response functions, e.g., so that longer vector arrows indicate higher discrimination, and shorter arrows lower discrimination in the model, and so on.

Next, to assess multidimensional difficulty, the distance from the origin is calculated suing the multidimensional difficulty (MDIFF) index (M. D. Reckase 1985):

$$MDIFF = B_i = \frac{-d_i}{\sqrt{\sum_{k=1}^{m} a_{ik}^2}} \tag{3}$$

In which d is the d-parameter index from the compensatory model. The MDIFF is denoted B as the DMIRT counterpart to the b-parameter in the unidimensional IRT model. The MDIFF is, therefore, a characteristic of item i such that higher MDIFF values indicate that higher levels of ability are necessary for a correct response (Reckase, M. D. 2009; R. L. Reckase M. D. & McKinley 1991). Observe that the denominator in Equation Equation 3 is the same expression as EquationEquation 1.

Importantly, in DMIRT analysis, the MDISC and MDIFF only apply in the direction set by ω_{il} and Equation Equation 2 (Reckase, M. D. 2009; R. L. Reckase M. D. & McKinley 1991). Thus, we cannot compare these estimates directly across items, as would be the case in the unidimensional model. This is because DMIRT seeks to maximize item discrimination as a global characteristic in a multidimensional environment. To estimate item discrimination as a local characteristic in the multidimensional space, it is, however, possible to select a common direction for the items and then recalculate the discrimination, i.e., to estimate the directional discrimination (DDISC).

$$DDISC = \sum_{k=1}^{m} a_{ik} cos\, \omega_{ik} \tag{4}$$

Since the DDISC is a local characteristic in the model, it is always the case that $DDISC \leq MDISC$.

In `D3mirt`, the DDISC is optional and implemented in `D3mirt`as optional *construct vectors*.

The results, along side the graphical device, include tables for the MDISC and MDIFF estimates as well as spherical coordinates describing the location of the vector arrows. If construct vector are used, the output also include DDISC scores for all items showing the constrained discrimination. The package also include a dedicated function to help identify the DMIRT model and the option of plotting individual scores (i.e., *profile analysis*) in the three-dimensional latent space.

Acknowledgements

I acknowledge support, advice, and suggestions for improvements from my supervisor Dr. Anders Sjöberg, Stockholm University. I also would like to express gratitude to Professor Torun Lindholm Öjmyr and Professor Mats Nilsson, Stockholm University, for their support and professional advice.

References

Forsberg, A., E. & Sjöberg. In Print. "A Validation Study of the Extended Relevance Scale using the D3mirt Package for R." *Measurement: Interdisciplinary Research and Perspectives*, In Print. https://doi.org/10.1080/15366367.2023.2282418.

Hambleton, Swaminathan, R. K. 1991. *Fundamentals of Item Response Theory*. SAGE Publications.

McKinley, M. D., R. L. & Reckase. 1983. "An Extension of the Two-Parameter Logistic Model to the Multidimensional Latent Space." ONR83-2. Iowa City, IA: American College Testing Program.

Muraki, J. E., E. & Carlson. 1995. "Full-Information Factor Analysis for Polytomous Item Responses." *Applied Psychological Measurement* 1 (19): 73–90. https://doi.org/10.1177/014662169501900109.

Reckase, M. D. 1985. "The Difficulty of Test Items That Measure More Than One Ability." *Applied Psychological Measurement* 4 (9): 401–12.

Reckase, M. D. 2009. *Multidimensional Item Response Theory*. Springer-Verlag New York, NY. https://doi.org/10.1007/978-0-387-89976-3.

Reckase, R. L., M. D. & McKinley. 1991. "The Discriminating Power of Items That Measure More Than One Dimension." *Applied Psychological Measurement* 4 (15): 361–73.