

Designing and Evaluating a Likelihood-Ratio Test for IRT models

Nina van Gerwen (1860852)

13th of October, 2022

- Real life problem: For every type of research, it is important to know whether the model you are using actually correctly illustrates the response process. In other words, it is important to know whether the model fits the data. If the wrong model is used, this can lead to dire consequences such as false conclusions. Furthermore, it also entails that you are making inefficient use of your data. So in order to test whether the model fits the data, there exist goodness-of-fit tests in almost every field of statistics. Most indicators of model fit, such as the AIC/BIC/DIC are relative: they can be used to compare two models on their fit. In Structural Equation Modeling, there also exist fit indices (e.g., SRMR, RMSEA, CFI/TLI) which together with rules of thumb can also indicate model fit. However, for IRT research, there exists no goodness-of-fit LR test that is generally applicable to all IRT models (besides the χ^2 test of a model vs. alternative model). Instead, some models have specific LR tests that tend to suffer from different issues (e.g., Andersen's LR test for all Rasch models, which has been shown to lack power (Krammer, 2018)).
 - Furthermore, goodness-of-fit tests all suffer from specific issues, such as a sensitivity to larger sample sizes. Therefore, there exist fit indices, which can help determine your model fit. Compared to SEM research, IRT models, however, have a lack of fit indices. For a relatively recent overview of fit indices used in IRT, see Nye et al., 2020.
- Research questions: What are the statistical properties (robustness, power, empirical α) of the designed LR test that is applicable to all IRT models?
 - Extra possible research question: something fit index related
- Analytic strategy: To research the statistical properties of our test, a simulation study will be conducted. First, data will be simulated according to certain IRT models. Then, knowing the true model, we can test both the real and other IRT models to the data and see whether our goodness-of-fit test has the ability to determine when we used the right or wrong IRT model.
 - When simulating data, we will for sure vary: the amount of items and sample size. Other factors we can vary in order to increase generalizability are: the parameters used for the IRT model, different types of IRT models, and the amount of group randomization used in the LR test.
- Ethical Consent: considering the fact that the data will be simulated, there should be no issue with the license of the data or ethical consent.

The Likelihood-Ratio test formula:

$$-2\ln\left(\frac{L_{total}}{L_{half1} \cdot L_{half2}}\right) \rightarrow \chi^2(k) \quad (1)$$

References

- Krammer, G. (2018). The andersen likelihood ratio test with a random split criterion lacks power. *Journal of modern applied statistical methods: JMASM*, 17, eP2685. <https://doi.org/10.22237/jmasm/1555594442>
- Nye, C. D., Joo, S.-H., Zhang, B., & Stark, S. (2020). Advancing and evaluating irt model data fit indices in organizational research. *Organizational Research Methods*, 23(3), 457–486. <https://doi.org/10.1177/1094428119833158>