ARTICLE IN PRESS

Journal of Mathematical Psychology ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Evaluating the model fit of diffusion models with the root mean square error of approximation

Anna-Lena Schubert *, Dirk Hagemann, Andreas Voss, Katharina Bergmann

University of Heidelberg, Institute of Psychology, Hauptstrasse 47-51, D-69117 Heidelberg, Germany

HIGHLIGHTS

- We evaluated the fit of diffusion models with the root mean square error of approximation.
- The RMSEA correctly distinguished between good- and bad-fitting models.
- The RMSEA was largely invariant to trial numbers.
- Decisions based on the RMSEA were superior to decisions based on statistical tests.
- RMSEA values were comparable to values in structural equation modeling.

ARTICLE INFO

Article history:
Received 21 May 2015
Received in revised form
3 August 2016
Available online xxxx

Keywords:
Reaction times
Diffusion model
Model fit
Root mean square error of approximation

ABSTRACT

The statistical evaluation of model fit is one of the greatest challenges in the application of diffusion modeling in research on individual differences. Relative model fit indices such as the AIC and BIC are often used for model comparison, but they provide no information about absolute model fit. Statistical and graphical tests can be used to identify individuals whose data cannot be accounted for by the diffusion model, but they become overly sensitive when trial numbers are large, and are subjective and time-consuming. We propose that the evaluation of model fit may be supplemented with the root mean square error of approximation (RMSEA; Steiger & Lind, 1980), which is one of the most popular goodness-of-fit indices in structural equation modeling. It is largely invariant to trial numbers, and allows identifying cases with poor model fit, calculating confidence intervals, and conducting power analyses. In two simulation studies, we evaluated whether the RMSEA correctly rejects badly-fitting models irrespective of trial numbers. Moreover, we evaluated how variation in the number of trials, the degree of measurement noise, the presence of contaminant outliers, and the number of estimated parameters affect RMSEA values. The RMSEA correctly distinguished between well- and badly-fitting models unless trial numbers were very small. Moreover, RMSEA values were in a value range expected from structural equation modeling. Finally, we computed cut-off values as heuristics for model acceptance or rejection. In a third simulation study we assessed how the RMSEA performs in model selection in comparison to the AIC and BIC. The RMSEA correctly identified the generating model in the majority of cases, but was outperformed by the AIC and BIC. All in all, we showed that the RMSEA can be of great value in the evaluation of absolute model fit, but that it should only be used in addition to other fit indices in model selection scenarios.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

In recent years, the diffusion model for binary responses (Ratcliff, 1978) has seen a huge rise of popularity in a wide area of research fields (Voss, Nagler, & Lerche, 2013). Within the diffusion model framework, response time distributions can be decomposed in terms of different parameters associated with specific cognitive processes. While traditionally researchers interested in mental chronometry drew inferences based on mean (and sometimes SDs of) response times and could therefore only infer *whether* response times differ between experimental conditions and/or individuals, they are now able to infer which processing components may be responsible for the observed response time differences.

The diffusion model has been successfully applied in the context of social cognitive research (e.g., Germar, Schlemmer,

* Corresponding author. Fax: +49 6221 547325.
E-mail addresses: anna-lena.schubert@psychologie.uni-heidelberg.de (A.-L. Schubert), dirk.hagemann@psychologie.uni-heidelberg.de (D. Hagemann), andreas.voss@psychologie.uni-heidelberg.de (A. Voss), katharina.bergmann@psychologie.uni-heidelberg.de (K. Bergmann).

http://dx.doi.org/10.1016/j.jmp.2016.08.004
0022-2496/© 2016 Elsevier Inc. All rights reserved.

Please cite this article in press as: Schubert, A.-L., et al., Evaluating the model fit of diffusion models with the root mean square error of approximation. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.08.004

Krug, Voss, & Mojzisch, 2014; Klauer, Stahl, & Voss, 2011; Voss, Rothermund, & Brandtstädter, 2008; Voss & Schwieren, 2015), prospective memory (e.g., Boywitt & Rummel, 2012), aging (e.g., McKoon & Ratcliff, 2013; Spaniol, Madden, & Voss, 2006), individual differences (e.g., Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007; Schubert, Hagemann, Voss, Schankin, & Bergmann, 2015), and in many more areas of research. It is, however, not always a priori known whether the diffusion model is an adequate process model for the cognitive processes that resulted in a specific distribution of response times and response frequencies in an experimental paradigm. A necessary (but not sufficient) precondition for the interpretation of diffusion model results is the model fit, that is, the degree of match between predicted and observed response time data. Even if it is undisputed that a specific task is principally suited for diffusion modeling, researchers still have to make an informed decision about the specific implementation of the model (e.g., the coding of responses or the number of estimated parameters). Therefore, it is crucial to evaluate how well the estimated model parameters can account for the actual accuracy and response time data, and to identify cases in which the diffusion model fails to describe the data.

Currently, there is no universally accepted gold standard for evaluating model fit in the diffusion model framework. Model fit indices such as the Akaike Information Criterion (AIC; Akaike, 1973) or the Bayes Information Criterion (BIC; Schwarz, 1978) are often used for model comparison and selection purposes, but they provide no information about absolute model fits. Therefore, in diffusion modeling these criteria only convey information about which of several models accounts for the empirical data best, but they do not help to decide whether the models should be accepted or rejected. Moreover, such relative fit indices do not allow identifying and possibly excluding individuals whose data cannot be accounted for within the diffusion model framework, as general suggestions when a model should be rejected cannot be given. In order to identify individuals with badly-fitting model parameters, different strategies are usually pursued. Statistical tests of model fit, such as the χ^2 test, are very common, but sensitivity of these tests is closely tied to the amount of data that is available. For small data sets this leads to a power problem as the test power may be too small to reject the model, and for larger data the test will become overly sensitive. Statistical tests are also biased in favor of more complex models, as a model with higher degrees of freedom that most often provides a better account for the data is not punished in comparison to a more parsimonious model.

To overcome the problems associated with null-hypothesis testing of model fit, Clauset, Shalizi, and Newman (2009), and Voss, Nagler et al. (2013) suggested simulating a large number of synthetic data sets based on the estimated model parameters and deriving critical p-values from these subsequently re-fitted data sets. The 5% or 10% quantile of the distribution of p-values can then be taken as a critical value for the evaluation of the empirical models. This method overcomes some of the problems associated with the statistical testing of model fits, but does not specifically consider trial numbers and the parsimony of a model. Moreover, models get accepted with an unknown error probability, as this procedure does not offer a method to estimate the statistical power of the test.

Graphical methods provide an alternative approach to the evaluation of model fit. For this approach the deviation of the predicted response times from the empirical response times is displayed either individually, or for a complete sample (e.g., Schmitz & Voss, 2014; Voss, Rothermund, Gast, & Wentura, 2013). Decisions based on graphical tests are, however, subjective and may therefore lead to spurious conclusions (D'Agostino, 1986). Moreover, extensive graphical model tests for each individual can quickly become time-consuming in large samples.

An ideal goodness-of-fit (GOF) index that can be used to identify data sets that the diffusion model is not able to account well for should have the properties of the very popular AIC and BIC (i.e., reward parsimonious models and not be strongly affected by variations in trial numbers), but would be an absolute measure of model fit, not a relative one. As such, it would presume that a perfectly fitting model has a fit value of zero and that a deviation from zero indicates how far the model deviates from perfect fit. Then, this deviation from perfect model fit could be compared across different models as well as between cases, and standards for acceptable model-fit could be defined. Moreover, conventions for cut-off values could be suggested indicating when a model should be rejected, and these cut-off values would be invariant across different applications (and therefore, different trial numbers and different degrees of parsimony) of the model.

Another field of research that has been concerned with the performance of GOF indices is structural equations modeling (SEM), which is a statistical technique for testing the structural relations within a multivariate data space. To maximize model fit, the discrepancy between the empirical covariance matrix of all measured variables and the covariance matrix implied by the model specifications gets minimized. As in diffusion modeling, this minimization process does not yield a GOF value that can reasonably be used to decide about model acceptance due to the same problems as listed above. Because participant numbers are typically very large in SEM studies (>200 participants), model predictions often deviate significantly from the empirical data, although the model fit is actually quite good. Therefore, there have been several suggestions on how to evaluate model fit in the SEM framework (see Jackson, Gillaspy, & Purc-Stephenson, 2009, for a review). Several of these GOF indices are not easily transferable to the diffusion model framework, because they compare the performance of the estimated model to a baseline model (in which all variables are presumed to be uncorrelated). Within the diffusion model framework, no such baseline model could be easily specified without further debatable assumptions, because it is entirely unclear which configuration of parameter values might reflect an appropriate baseline model.

One very popular absolute GOF index, the root mean square error of approximation (RMSEA; Steiger & Lind, 1980), however, does not require the assumption of a baseline model, but is based on the noncentrality parameter of the χ^2 distribution. The RMSEA is relatively unaffected by variations in sample size and rewards parsimonious models. Moreover, as the RMSEA is an absolute fit index with a minimum of zero, conventions for cut-off values have been suggested and are frequently used in the SEM framework. Because the RMSEA is based on the noncentrality parameter of the χ^2 distribution, it could be easily reported and used as an evaluation criterion in addition to the χ^2 statistic and its corresponding p-value in the context of diffusion modeling. In the present paper, we propose to use the RMSEA as an index of absolute model fit within the diffusion model framework and discuss its benefits in comparison to standard methods of model evaluation.

1.1. The diffusion model

The diffusion model makes the basic assumption that during a decision process with two alternatives, information is accumulated continuously until the diffusion process reaches one of two thresholds. Specifically, this information accumulation process consists of a constant systematic component, the *drift*, and normally distributed random noise. The basic diffusion model estimates four parameters from empirical response time distributions: The drift rate (v), which describes the strength and direction of the systematic influence on the diffusion process, the threshold separation (a), which maps the amount of information that is used for a decision,

Please cite this article in press as: Schubert, A.-L., et al., Evaluating the model fit of diffusion models with the root mean square error of approximation. Journal of Mathematical Psychology (2016), http://dx.doi.org/10.1016/j.jmp.2016.08.004

Download English Version:

https://daneshyari.com/en/article/4931829

Download Persian Version:

https://daneshyari.com/article/4931829

Daneshyari.com