Journal of Mathematical Psychology 86 (2018) 51–63

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

Consensus theory for mixed response formats

André Aßfalg

Department of Psychology, Albert-Ludwigs University Freiburg, Engelbergerstr. 41, 79106 Freiburg, Germany

HIGHLIGHTS

- Estimation of a latent trait when the correct responses are unknown.
- The model accounts for continuous, categorical, or mixed response formats.
- Focus on applications with complex interview data (e.g., eyewitness testimony).
- Minimal requirements for additional response distributions.

ARTICLE INFO

Article history:
Received 9 October 2017
Received in revised form 5 July 2018

Keywords:
Consensus theory
Item response theory
Latent-trait model

ABSTRACT

Measuring shared beliefs, expert consensus, or the details of a crime in eyewitness testimony represents a psychometric challenge. In expert interviews, for example, the correct responses representing the expert consensus (i.e., the answer key) are initially unknown and experts may differ in their contribution to this consensus. I propose the variable-response model, an extension of latent-trait models. The model allows the estimation of the answer key and the latent trait for continuous, categorical, or mixed responses. I describe some minimal requirements for the addition of new response formats to the model. I further propose a Markov chain Monte Carlo algorithm to estimate the model parameters. The results of a simulation study demonstrate that the algorithm accurately recovers the data-generating parameters. I also present an application of the variable-response model to the empirical data of a Geography test. In this application, the parameter estimates correspond well with the true answer key.

1. Consensus theory for mixed response formats

Consensus theory is an extension of latent-trait models that assumes that the correct responses to a set of questions – the answer key – are unknown (e.g., Anders & Batchelder, 2012; Batchelder & Romney, 1988; Romney, Weller, & Batchelder, 1986). Consensus theory includes the answer key as a set of model parameters which are estimated along with the latent trait. Originally, the method was proposed for applications in Anthropology (e.g., Romney et al., 1986): Researchers who question people about an unknown culture initially do not know the answer key. In such an application, estimating the answer key – that is, details about the unknown culture – is typically the primary goal. Another application includes the eyewitness testimony of multiple witnesses. Consensus theory offers an estimate of what happened during a crime – the answer key – while accounting for inter-individual differences in the witnesses' ability to recall the crime (Waubert de Puiseau, Aßfalg, Erdfelder, & Bernstein, 2012; Waubert de Puiseau, Greving, Aßfalg, & Musch, 2017).[1]

Unfortunately, applications of consensus theory are limited by the lack of models that account for mixed response formats within a test. Extant models account for dichotomous (e.g., yes vs. no), multiple choice, ordinal, and continuous response formats (Anders & Batchelder, 2015; Anders, Oravecz, & Batchelder, 2014; Batchelder & Anders, 2012; Batchelder & Romney, 1988; Batchelder, Strashny, & Romney, 2010; Karabatsos & Batchelder, 2003). However, currently, it is not possible to mix these response formats and estimate the latent trait based on the combined responses. Consider again the example of an eyewitness interview. Extant models can be inadequate for eyewitness testimony: yes–no questions, for example, are a poor match for the complexity of eyewitness recollections (e.g., Waubert de Puiseau et al., 2012).

Mixed response formats are well known in latent-trait analysis (Moustaki & Knott, 2000; Moustaki & Papageorgiou, 2005; Sammel,

E-mail address: andre.assfalg@psychologie.uni-freiburg.de.

[1] Strictly speaking, consensus theory provides an estimate of the consensus between witnesses. This consensus typically overlaps with but is not necessarily identical to the ground truth in mock crimes (Waubert de Puiseau et al., 2012, 2017).

https://doi.org/10.1016/j.jmp.2018.08.005
0022-2496/

Ryan, & Legler, 1997). Generalized latent trait models allow a mixture of response formats within the same test. Here, I propose the *variable-response model* (VRM), an analogous approach for consensus theory in which the answer key is unknown. Unfortunately, generalized latent trait models are not readily adapted to consensus theory. Instead, I propose an approach for the VRM that is based on well-known latent-trait models for dichotomously scored items (Birnbaum, 1968; Rasch, 1961). In these models, the *item characteristic curve* describes the probability of a correct response as a function of the model parameters. The VRM extends the concept of the item characteristic curve to all permissible responses of an item. This approach includes the answer key in the model as a latent parameter.

In the following sections, I begin with a brief introduction of the 2-parameter logistic model for dichotomously scored items (Birnbaum, 1968). This model forms the core of the VRM. By extending the 2-parameter logistic model to all permissible responses – independent of the hypothesized response distribution – the model core can be flexibly combined with any response probability distribution that meets some minimal requirements. I formulate the requirements for these distributions and introduce two examples. For the parameter estimation, I further propose a Markov chain Monte Carlo (MCMC) procedure along with model selection and model check indices. Finally, I present the results of a simulation study and the application of the VRM to the empirical data of a Geography test.

2. The variable response model

Arguably, the most well-known latent-trait models account for items that are scored dichotomously as either correct or incorrect (Birnbaum, 1968; Rasch, 1961). These models assume unidimensionality—that is, a single latent trait underlies the responses of a respondent. In consensus theory, a similar concept is referred to as the *common truth* or *single culture* assumption and requires that there is a single answer key for all respondents (e.g., Batchelder & Romney, 1988; Karabatsos & Batchelder, 2003). For present purposes, I assume unidimensionality and a single culture for the VRM as well.

Common latent-trait models for dichotomously scored items assume that the probability of a correct response is a nondecreasing function of the latent trait. Let X_{ik} be the response of respondent i to item k and Z_k the answer key to item k. Further let $\theta_i \in (-\infty, \infty)$ be the latent trait of respondent i. The 2-parameter logistic model (Birnbaum, 1968) defines the probability of a correct response of respondent i to item k as

$$P\left(X_{ik} = Z_k | \varphi\right) = \frac{1}{1 + \exp\left(-a_k\left[\theta_i - d_k\right]\right)} \tag{1}$$

with model parameters $\varphi = \{\theta_i, d_k, a_k\}$. In Eq. (1), $d_k \in (-\infty, \infty)$ and $a_k \in (0, \infty)$ are the item difficulty and item discrimination parameters of item k, respectively. The item difficulty and discrimination parameters determine the location and slope of the logistic function in Eq. (1), respectively. For any probability of a correct response, a higher (lower) item difficulty requires a higher (lower) latent trait to achieve the same probability of a correct response. Conversely, for any change in the probability of a correct response, a relatively small (large) change in the latent trait is necessary if the item discrimination is high (low). Eq. (1) is also known as the *item characteristic curve* of item k. Although I rely on the 2-parameter logistic model in Eq. (1), all following considerations are readily adapted to alternatives such as one- or two-parameter logistic or normal-ogive models.

2.1. The response characteristic curve

The VRM extends the concept of the item characteristic curve in Eq. (1) to that of the *response characteristic curve*—that is, the probability of any permissible response given the model parameters. Like other consensus-theory models, but unlike latent-trait models, the VRM accounts for responses, not item scores (e.g. correct vs. incorrect).

Specifically, in Eq. (1), the probability of the correct response $P\left(X_{ik} = Z_k|\varphi\right)$ is a nondecreasing function of the latent trait. For an item with only two response alternatives, Eq. (1) implies that the probability of the incorrect response is $1 - P\left(X_{ik} = Z_k|\varphi\right)$, a nonincreasing function of the latent trait. The VRM applies this principle to two or more response alternatives, including continuous responses. Specifically, the response characteristic curve of a response X_{ik} is defined as

$$P\left(X_{ik}|\varphi\right) = \begin{cases} P_-\left(X_{ik}\right) + \dfrac{P_+\left(X_{ik}\right) - P_-\left(X_{ik}\right)}{1 + \exp\left(-a_k\left[\theta_i - d_k\right]\right)} \mathit{if}\ P_+\left(X_{ik}\right) \geq P_-\left(X_{ik}\right) \\ P_+\left(X_{ik}\right) + \dfrac{P_-\left(X_{ik}\right) - P_+\left(X_{ik}\right)}{1 + \exp\left(a_k\left[\theta_i - d_k\right]\right)} \mathit{if}\ P_+\left(X_{ik}\right) < P_-\left(X_{ik}\right). \end{cases} \tag{2}$$

The definition of the response characteristic curve in Eq. (2) adds two important concepts to Eq. (1): P_+ is the probability distribution of the responses to item k as $\theta \to \infty$; conversely, P_- is the probability distribution of the responses to item k as $\theta \to -\infty$. In Eq. (2), P_+ and P_- serve as upper and lower asymptotes (and vice versa) of the logistic function.[2] Consequently, I refer to P_+ and P_- as asymptotic response distributions. If the responses are categorical, P_+ and P_- are probability mass functions. Conversely, if the responses are continuous, P_+ and P_- are probability densities.

Eq. (2) implies that the distribution of the responses for a fixed set of parameters φ follows a mixture of the asymptotic response distributions given as

$$Q\left(X; \varphi\right) = P_+\left(X\right) P\left(\varphi\right) + P_-\left(X\right)\left[1 - P\left(\varphi\right)\right], \tag{3}$$

where $P\left(\varphi\right)$ denotes the 2-parameter logistic model in Eq. (1).

Proof. For the sake of brevity, let $P\left(X|\varphi\right)$ be the response characteristic as defined by Eq. (2) for an arbitrary combination of a response X and a set of parameters φ, let $Q(X; \varphi)$ be the mixture distribution in Eq. (3) for this pair, and let $P_+(X)$ and $P_-(X)$ be the value of the asymptotic response distributions for the response X. Now, consider the three mutually exclusive and jointly exhaustive cases: $P_+(X) = P_-(X)$, $P_+(X) > P_-(X)$, and $P_+(X) < P_-(X)$. In all three cases, $P\left(X|\varphi\right)$ in Eq. (2) is equivalent to the finite mixture $Q(X; \varphi)$ in Eq. (3).

Case 1, $P_+(X) = P_-(X)$: According to Eq. (2), $P\left(X|\varphi\right) = P_-(X)$. Because $P_+(X) = P_-(X)$, $P_+(X)$ can be replaced with $P_-(X)$ in Eq. (3) which gives $Q(X; \varphi) = P_-(X) = P\left(X|\varphi\right)$.

Case 2, $P_+(X) > P_-(X)$: In this case, Eq. (2) yields $P\left(X|\varphi\right) = P_-\left(X\right) + \left[P_+\left(X\right) - P_-\left(X\right)\right] P\left(\varphi\right)$ and applying elementary algebra one can see that $P\left(X|\varphi\right) = P_+(X)P\left(\varphi\right) + P_-(X)\left[1 - P\left(\varphi\right)\right] = Q(X; \varphi)$.

Case 3, $P_+(X) < P_-(X)$: By noting that $1 - P\left(\varphi\right) = 1/[1 + \exp(a_k[\theta_i - d_k])]$, Eq. (2) yields $P\left(X|\varphi\right) = P_+\left(X\right) + \left[P_-\left(X\right) - P_+\left(X\right)\right]\left[1 - P\left(\varphi\right)\right]$. Again, applying elementary algebra

[2] Barton and Lord (1981) tested a similar approach to Eq. (2) with upper and lower asymptotes for the logistic model Eq. (1). However, the logistic model only accounts for the probability of the correct response which requires knowledge of the answer key. Conversely, Eq. (2) describes the probability of any response given the model parameters.