Chapter 14
A Bayesian Test for the Association of Binary Response Distributions

Rudy Ligtvoet

Abstract   Item response theory models impose testable restrictions on the observed distribution of the response variables. In this chapter, the inequality restrictions are investigated imposed by Mokken's model of monotone homogeneity (MH) for binary item response variables. A Bayesian test for the observable property of variables being associated is proposed for the trivariate distributions of all triplets of items. This test applies to a wide range of item response theory models that extend beyond the MH model assumptions.

List of Abbreviations

MH	Mokken's (1971) model of *monotone homogeneity*.
SPOD	*Strongly positive orthant dependence* (Definition 1).
A	Variables being *associated* (Definition 2).
CI	Assumption of (local or) *conditional independence*.
UD	Assumption of *unidimensionality*.
M	Assumption of (latent) *monotonicity*.
LND	Assumption of *local non-negative dependence*.
DINA	The *deterministic inputs, noisy, AND gate* model.

14.1   Introduction

Mokken scale analysis consists of a collection of diagnostics to assess the assumption underlying the model of *monotone homogeneity* (MH; Mokken, 1971; Mokken & Lewis, 1982; Molenaar & Sijtsma, 2000; Sijtsma, 1988; Van der Ark, 2007; see List of Abbreviations section). The MH model allows for ordinal inferences about

R. Ligtvoet (✉)
University of Cologne, Köln, Germany
e-mail: rligtvoe@uni-koeln.de; https://sites.google.com/site/rligtv/

a latent variable based on the observed responses to the item variables that make up a test. For example, for binary item variables, the MH model implies a stochastic ordering on the latent variable by the sum of the item scores (Grayson 1988; Ghurye & Wallace 1959; Huynh 1994; Ünlü 2008). The assumptions that constitute the MH model are useful in applications when ordinal inferences are required and no further (parametric) assumptions are deemed either necessary or appropriate.

The diagnostic statistics most prominently used in Mokken scale analysis are the scalability coefficients, which are calculated from the bivariate distribution of pairs of item variables (Loevinger 1948; Molenaar 1991; Warrens 2008). The MH model requires that these scalability coefficients are all non-negative, whereas any negative coefficient discredits or invalidates the model assumptions. However, non-negative scalability coefficients are not sufficient for the MH model. A more restrictive demand of the model is the requirement of non-negative partial correlations, for all triplets of items (Ellis 2014). The restrictions imposed on the trivariate item distributions by this latter requirement are part of a wider class of properties of multivariate positive dependence. Examples of these properties are *multivariate total positivity* (Ellis 2015; Karlin & Rinott 1980) and *conditional association* (Holland & Rosenbaum 1986; Rosenbaum 1984), which are also implied (necessary but not sufficient) by the MH model.

A set of assumptions, other than those that define the MH model, was proposed by Holland (1981), with the aim of testing a wider range of item response theory model. These assumptions pertain to perfect scores (all zeros or ones) on subsets of item variables and hold, if and only if any subset of item variables satisfies the property of *strongly positive orthant dependence* (SPOD; Joag-Dev, 1983). Because conditional association implies SPOD (Definition 1 below), it follows that SPOD includes the MH model as a special case (Holland & Rosenbaum 1986).

In this chapter, a Bayesian test is proposed for a wide range of models for binary response data. This test is based on the observable property of *variables being associated* (A; Esary et al., 1967; Walkup, 1968), but applied to all trivariate distributions of triplets of item variables (cf. non-negative partial correlations). In the following section, I introduce some properties of latent variable models and the observable properties that are implied by these models. In Sect. 14.3, the restrictions imposed on the data distribution by the testable properties are expressed in terms of inequality constraints on the log-odds ratios related to the multinomial parameters. This allows for a convenient way of expressing the Bayes factor in favor of property A. The analysis of response data using the Bayes factor is illustrated in Sect. 14.4, followed by a short discussion in Sect. 14.5.

14.2 Preliminaries and Notation

Let $\mathbf{X} = (X_1, \ldots, X_J)$ denote the random vector containing the J binary item response variables. The (real) latent vector $\boldsymbol{\theta}$ is introduced through the law of total probability, $P(\mathbf{X} = \mathbf{x}) = \int P(\mathbf{X} = \mathbf{x}|\boldsymbol{\theta})dG(\boldsymbol{\theta})$, with its distribution function

G further left unspecified. To define a model, several conditions are considered. The first of these conditions is that of local or *conditional independence* (CI): $X_1, \ldots, X_J$ are conditionally independent, given $\boldsymbol{\theta}$. The latent vector is said to be *unidimensional* (UD), whenever $\boldsymbol{\theta} = \theta$ (scalar-valued). The conditions of CI and UD alone, however, are not enough to impose testable restrictions on the distribution of $\mathbf{X}$ (Suppes & Zanotti 1981). The additional condition of *monotonicity* (M) states that the item response functions $P(X_j = 1|\boldsymbol{\theta})$ are (element-wise) non-decreasing in $\boldsymbol{\theta}$, for all $j = 1, \ldots, J$. Holland and Rosenbaum (1986) referred to a model that assumes CI and M as a *monotone latent variable model*, whereby the additional assumption UD defines Mokken's (1971) MH model for $(\mathbf{X}, \theta)$.

14.2.1 Strongly Positive Orthant Dependence

Holland (1981) proposed a set of conditions, with the purpose of providing a test of a wide range of item response theory models. He showed that these conditions hold, if and only if $\mathbf{X}$ satisfies the observable property of SPOD (Joag-Dev 1983), for any selection of variables from $\mathbf{X}$.

Definition 1 *The vector $\mathbf{X}$ is said to satisfy the property of SPOD ($\mathbf{X}$ is SPOD), if for any partition $\mathbf{X} = (\mathbf{Y}, \mathbf{Z})$, the following three inequalities hold:*

$$P(\mathbf{Y} = \mathbf{1})P(\mathbf{Z} = \mathbf{1}) \leq P(\mathbf{X} = \mathbf{1}), \;\; P(\mathbf{Y} = \mathbf{0})P(\mathbf{Z} = \mathbf{0}) \leq P(\mathbf{X} = \mathbf{0}),$$
$$and \; P(\mathbf{Y} = \mathbf{1})P(\mathbf{Z} = \mathbf{0}) \geq P(\mathbf{Y} = \mathbf{1}, \mathbf{Z} = \mathbf{0}). \tag{14.1}$$

For the special case $J = 3$, there are only three distinct (non-empty) partitions of $\mathbf{X}$ to consider, with $\mathbf{Y} = X_i$ and $\mathbf{Z} = (X_j, X_k)$, for $i = 1, 2, 3$ and $j, k \neq i$. This is because, for $J = 3$, the first two inequalities in Definition 1 imply (if and only if) the last inequality for both $\mathbf{Y} = X_i$ and $\mathbf{Y} = (X_j, X_k)$. For example, for $\mathbf{Y} = X_1$ and $p(\mathbf{x}) = P(\mathbf{X} = \mathbf{x})$, with $\mathbf{u} = (0, 1, 1)$, the first inequality implies that

$$P(X_1 = 1)(p(\mathbf{u}) + p(\mathbf{1})) \leq p(\mathbf{1}) \Leftrightarrow p(\mathbf{u})/p(\mathbf{1}) \leq 1/P(X_1 = 1) - 1$$
$$\Leftrightarrow p(\mathbf{1})/p(\mathbf{u}) \geq 1/P(X_1 = 0) - 1 \Leftrightarrow P(X_1 = 0)(p(\mathbf{1}) + p(\mathbf{u})) \geq p(\mathbf{u}), \tag{14.2}$$

where the last expression corresponds to the third inequality in Definition 1, for $\mathbf{Y} = (X_2, X_3)$.

For the general case (any J), the condition of *local non-negative dependence* (LND) is obtained from Definition 1, by conditioning each term on θ. The following result by Holland (1981, Theorem 2) shows that SPOD provides a characterization of a wide class of latent variable models for binary response variables.

Theorem 1 (Holland, 1981) *The binary random vector* **X** *is SPOD, if and only if UD and LND hold, and for any partition* **X** $= ($**Y**$, $**Z**$)$, *both*

$$P(\mathbf{Y} = \mathbf{1}|\theta) \text{ is non-decreasing in } \theta \text{ and}$$
$$P(\mathbf{Y} = \mathbf{0}|\theta) \text{ is non-increasing in } \theta. \tag{14.3}$$

The set of conditions listed in Theorem 1 contain the MH model as a special case (Holland & Rosenbaum 1986), where Holland (1981) referred to Eq. 14.3 as *monotonicity of the subtest characteristic curves*.

14.2.2 Variables Being Associated

Holland (1981) generalized the conditions that define the MH model (i.e., UD, CI, and M) by relaxing the CI condition and replacing M by Eq. 14.3. Alternatively, one may replace the UD restriction by less restrictive constraints on the multidimensional vector $\boldsymbol{\theta}$, to obtain a multivariate version of the MH model for $(\mathbf{X}, \boldsymbol{\theta})$. Here, one such relaxation is considered, namely, that $\boldsymbol{\theta}$ is A.

Definition 2 (Esary et al., 1967) *The random vector* $\boldsymbol{\theta}$ *is said to be associated with* ($\boldsymbol{\theta}$ *is A), whenever the covariance between* $\phi(\boldsymbol{\theta})$ *and* $\varphi(\boldsymbol{\theta})$ *is non-negative, for any (element-wise) non-decreasing functions* ϕ *and* φ, *for which the involved expected values are defined.*

If $\boldsymbol{\theta}$ is A, then any selection of variables from $\boldsymbol{\theta}$ is also A, which follows by taking ϕ and φ to pertain only to the selected variables. Assuming CI and M, the following testable result is obtained (Holland & Rosenbaum, 1986, Theorem 8, referring for the proof to Jogdeo, 1978).

Theorem 2 (Jogdeo, 1978) *If* $\boldsymbol{\theta}$ *is A, CI and M imply that* **X** *is also A.*

Proof Together, M and CI imply that $E[\phi(\mathbf{x})|\boldsymbol{\theta}]$ is non-decreasing in $\boldsymbol{\theta}$ for any non-decreasing function ϕ (e.g., Holland & Rosenbaum, 1986, Lemma 2). Also, **X**$|\boldsymbol{\theta}$ is A, because of CI (Esary et al. 1967, Theorem 2.1). Then, by the Theorem in Jogdeo (1978, p. 234), $(\mathbf{X}, \boldsymbol{\theta})$ is A, and **X** is also A, because any subset of associated random variables satisfies A.   □

The MH model is a special case of the conditions in Theorem 2, which in turn are a special case of the conditions in Theorem 1, as property A implies property SPOD (e.g., Holland & Rosenbaum, 1986, p. 1536).

Another example of a model that satisfies the conditions in Theorem 2 can be obtained by considering $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$ to be a binary random vector of latent attributes. The DINA model (Doignon & Falmagne 2012; Tatsuoka 1995) is a response model for cognitive assessment, with a successful outcome expected on an item, if all the relevant attributes are possessed. The relevance of the attributes for item j is determined by the binary vector $(q_{j1}, \ldots, q_{jK})$, which is usually fixed in advance for all items. For the response functions, let $P(X_j = 1|\boldsymbol{\alpha}) = P(X_j = $

$1|\xi_j = 0)P(X_j = 1|\xi_j = 1)$, with $\ln(\xi_j) = q_{j1}\ln(\alpha_1) + \ldots + q_{jK}\ln(\alpha_K)$. Then, the DINA model implies condition M, if and only if $P(X_j = 1|\xi_j = 0) \leq P(X_j = 1|\xi_j = 1)$ (Junker & Sijtsma 2001).

Proposition 1 *Assuming CI and M, the DINA model implies that* $\mathbf{X}$ *is A, if* $(\boldsymbol{\alpha}, \eta)$ *satisfies the MH model, with* η *denoting a second-order latent variable.*

Proof The MH model for $(\boldsymbol{\alpha}, \eta)$ implies that $\boldsymbol{\alpha}$ is A, so that $(\mathbf{X}, \boldsymbol{\alpha})$ satisfies the conditions of Theorem 2 (replacing $\boldsymbol{\theta}$ by $\boldsymbol{\alpha}$). □

The second-order latent variable in Proposition 1 can be thought of the cognitive growth that stimulates the development of the attributes, with η positively related to the total number of attributes under the MH model. The purpose of Proposition 1 is, however, not to propose another cognitive diagnostic model, but rather to illustrate the generality of the conditions in Theorem 2.

14.3   Restrictions on the Log-Odds Ratios

Both the property of SPOD and A impose a number of inequality restrictions on the distribution of $\mathbf{X}$. In order to test these restrictions, it is convenient to denote by $\mathbf{p}$ the vector containing the elements $p(\mathbf{x}) = P(\mathbf{X} = \mathbf{x})$, arranged in lexicographical order of $\mathbf{x}$ (with elements on the right running faster from 0 to 1). Also assume that $\mathbf{p} > \mathbf{0}$. Then, the restrictions imposed by either of the properties can be concisely expressed in terms of inequality restrictions on the log-odds ratios, as

$$\mathbf{K}\ln(\mathbf{Mp}) \geq \mathbf{0}, \tag{14.4}$$

(cf. Bartolucci & Forcina, 2005) with $\mathbf{K} = \mathbf{I}_v \otimes (1, -1, -1, 1)$ and $\mathbf{I}_v$ is the identity matrix of dimensions v equal to the number of restriction. The matrix $\mathbf{M}$ is a binary design matrix which can be adapted to pertain to the restrictions of either A or SPOD. For example, for $J = 2$, take $v = 1$ and $\mathbf{M} = \mathbf{I}_4$, so that Eq. 14.4 yields $\ln p(0, 0) - \ln p(0, 1) - \ln p(1, 0) + \ln p(1, 1) \geq 0$, which corresponds to $\mathrm{Cov}(X_1, X_2) \geq 0$.

Walkup (1968) listed the set of all pairs of binary non-decreasing functions that characterize property A, for up to four items. For $J = 3$, there are nine such pairs of function. One example of such a pair corresponds to $\mathrm{Cov}(X_2, X_3) \geq 0$. It can be verified that this restriction is obtained from Eq. 14.4 using $\mathbf{M} = (1, 1) \otimes \mathbf{I}_4$. Let p_k denote the kth element of $\mathbf{p}$, with $p_1 = p(\mathbf{0})$, $p_2 = p(0, 0, 1), \ldots, p_8 = p(\mathbf{1})$. Another example imposes the restriction $\mathrm{Cov}(1 - (1 - X_1)(1 - X_2), X_3) \geq 0$, which corresponds to the restriction $\ln p_1 - \ln(p_3 + p_5 + p_7) - \ln p_2 + \ln(p_4 + p_6 + p_8) \geq 0$, obtained from Eq. 14.4, using "$\otimes$" for the Kronecker product, as $\mathbf{M} = (\mathbf{I}_2 \otimes (1, 0)', \mathbf{I}_2 \otimes (0, 1)', (1, 1) \otimes \mathbf{I}_2 \otimes (0, 1)')$. By going though all $v = 9$ pairs of functions listed by Walkup (1968), and stacking on top of one another all the corresponding design matrices, we find that property A holds for $J = 3$, if Eq. 14.4

holds, with design matrix

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \mathbf{M}_3 \\ \mathbf{M}_4 \\ \mathbf{M}_5 \\ \mathbf{M}_6 \\ \mathbf{M}_7 \\ \mathbf{M}_8 \\ \mathbf{M}_9 \end{bmatrix} = \begin{bmatrix} (1, 1) \otimes \mathbf{I}_4 \\ \mathbf{I}_2 \otimes (1, 1) \otimes \mathbf{I}_2 \\ \mathbf{I}_4 \otimes (1, 1) \\ \mathbf{I}_2 \otimes ((1, 0)' \otimes (1, 1)) \otimes \mathbf{I}_2 \\ \mathbf{I}_2 \otimes (\mathbf{I}_2, (0, 1)' \otimes (1, 1)) \\ (\mathbf{I}_2 \otimes (1, 0)' \otimes (1, 1), \mathbf{I}_4) \\ (\mathbf{I}_4, \mathbf{I}_2 \otimes (0, 1)' \otimes (1, 1)) \\ ((1, 1) \otimes \mathbf{I}_2 \otimes (1, 0)', \mathbf{I}_2 \otimes (1, 0)', \mathbf{I}_2 \otimes (0, 1)') \\ (\mathbf{I}_2 \otimes (1, 0)', \mathbf{I}_2 \otimes (0, 1)', (1, 1) \otimes \mathbf{I}_2 \otimes (0, 1)') \end{bmatrix}. \qquad (14.5)$$

The matrix $\mathbf{M}$ in Eq. 14.5 consists of $v = 9$ stacked matrices $\mathbf{M}_1, \ldots, \mathbf{M}_9$, each of dimensions 4×8.

The following result shows that SPOD and A coincide in case $J = 3$.

Theorem 3 *For $J = 3$ binary variables and $\mathbf{p} > \mathbf{0}$, property A is satisfied if and only if property SPOD is satisfied for all subsets of variables.*

Proof For any subset of two variables from $\mathbf{X} = (X_1, X_2, X_3)$, SPOD implies that the covariance between the two variables are non-negative. This corresponds to $\mathbf{M}_1$, $\mathbf{M}_2$, and $\mathbf{M}_3$ in Eq. 14.5 for the three distinct subsets (X_2, X_3), (X_1, X_3), and (X_1, X_3), respectively. The remainder of the proof consists of going through the process of exhaustively listing all restrictions imposed by SPOD for $J = 3$ and expressing these in terms of the log-odds ratios. It can then be shown that $\mathbf{M}_4, \ldots, \mathbf{M}_9$ of the design matrix $\mathbf{M}$ in Eq. 14.5 match one-to-one with those obtained for property SPOD. As an example, consider the inequality $P(\mathbf{Y} = \mathbf{1}) P(\mathbf{Z} = \mathbf{0}) \geq P(\mathbf{Y} = \mathbf{1}, \mathbf{Z} = \mathbf{0})$ from Definition 1, which reduces for $\mathbf{Y} = (X_1, X_2)$ and $\mathbf{Z} = X_3$ to $(p_7 + p_8)(p_1 + p_3 + p_5 + p_7) \geq p_7$ and yields $\ln p_8 - \ln(p_2 + p_4 + p_6) - \ln p_7 + \ln(p_1 + p_3 + p_5) \geq 0$. The last inequality is obtained from Eq. 14.4 using $\mathbf{M}_8$ in Eq. 14.5. The remaining five inequalities can be obtained similarly. $\qquad \square$

One problem when testing either the properties A or SPOD using Eq. 14.4 is that the number of constraints grows fast as J increases to a more realistic size. For example, for $J = 4$, Walkup (1968) listed 99 restrictions imposed by property A. Furthermore, many of the restrictions pertain to outcomes for which observations may be sparse as these restrictions involve ever higher-order interactions between the variables in $\mathbf{X}$. A solution to both these problems is to consider testing the property A for all triplets of item response variables from $\mathbf{X}$ only. By considering the trivariate distributions, the hope is to have a test that is more powerful than a test that involves only the bivariate distribution while at the same time being broad enough to target a wide range of response models.

14.3.1 Trivariate Associated Distributions

Considering all the triplets of item variables of a test of length $J \geq 3$. With property A imposing $J(J-1)/2$ inequality restrictions on the bivariate distributions, and six restrictions involving three items for each trivariate distribution, the total number of restrictions $v = J(J-1)(J-3/2)$.

The design matrix $\mathbf{M}$ for assessing the v inequality restrictions is obtained as follows. First, let

$$\mathbf{B}_{jk} = \mathbf{B}_{jk1} \otimes \cdots \otimes \mathbf{B}_{jkJ}, \text{ with } \mathbf{B}_{jkl} = \begin{cases} \mathbf{I}_2 & \text{if either } j = l \text{ or } k = l \\ (1, 1) & \text{otherwise,} \end{cases} \quad (14.6)$$

and let the matrix $\mathbf{B}$ be obtained by stacking on top of one another all matrices $\mathbf{B}_{ij}$, which contains all the restrictions imposed on the bivariate distributions. Second, for the trivariate distributions, let $\mathbf{R}$ be the 8 by 3 matrix with in its rows all binary vectors of length 3, in lexicographical order. Likewise, let $\mathbf{S}$ be the 2^J by J matrix with in its rows all binary vectors of length J, and let $\mathbf{T}$ denote the matrix in Eq. 14.5, but without $\mathbf{M}_1, \mathbf{M}_2$, and $\mathbf{M}_3$. Third, let $\mathbf{C}_{jkl}$ be a matrix of dimensions 24 by 2^J. Matrix $\mathbf{C}_{jkl}$ is assigned to its ath column the same values as $\mathbf{T}$ has in its bth column, whenever $(s_{aj}, s_{ak}, s_{al}) = (r_{b1}, r_{b2}, r_{b3})$, for $1 \leq a \leq 2^J$ and $1 \leq b \leq 8$. Finally, matrix $\mathbf{C}$ is obtained by stacking all matrices $\mathbf{C}_{jkl}$ on top of each other.

The goal is to test the hypothesis H of trivariate A, for all triplets of response variables, with H obtained as the set of vectors $\mathbf{p}$, which satisfy $\mathbf{p} > \mathbf{0}$ and $\mathbf{1}'\mathbf{p} = 1$ (multinomial model), and Eq. 14.4, with the matrix $\mathbf{M}$ obtained from stacking matrix $\mathbf{B}$ on top of $\mathbf{C}$. A maximum likelihood procedure for testing inequality restrictions requires the estimation of $\mathbf{p}$ and produces test statistics which asymptotic sampling distributions are difficult to obtain (e.g., Bartolucci & Forcina, 2000, 2005; Vermunt, 1999). Here, a Bayesian approach is considered instead, which requires a prior density to be assigned to $\mathbf{p}$, but has the advantage that it allows for hypothesis H to be tested against its complement of at least one violation of Eq. 14.4. As a prior for $\mathbf{p}$, a flat (uniform) Dirichlet distribution is chosen, where the influence of this particular choice is expected to be small as long as all observations of $\mathbf{X} = \mathbf{x}$ have enough support.

14.3.2 Bayes Factor for Trivariate Associated Distribution

The Bayes factor in support of hypothesis H is expressed in terms of the prior and posterior probabilities that the restrictions imposed by H are satisfied (Klugkist & Hoijtink 2007; Tijmstra et al. 2015). The prior probability of H is estimated by sampling a large number of vectors $\mathbf{p}$ from the flat Dirichlet distribution and calculating the proportion c that satisfies Eq. 14.4. Let $\mathbf{n}$ denote the vector containing the observed frequencies of $\mathbf{X} = \mathbf{x}$, arranged as $\mathbf{p}$. Also, let d denote the proportion

of samples that satisfy Eq. 14.4, with the samples obtained from a Dirichlet distribution, with the hyper-parameters equal to $\mathbf{n} + \mathbf{1}$. The ratio d/c provides an estimate of the Bayes factor in favor of H over the multinomial model. The Bayes factor for the evidence in favor of H over its complement (at least one violation) then becomes

$$L = (1/c - 1)/(1/d - 1), \tag{14.7}$$

where a value $L > 1$ expresses support in favor of H, whereas a value $L < 1$ expresses support for the hypothesis that there is at least one violation of H (Lavine & Schervish 1999; Kass & Raftery 1995).

The sampling procedure for obtaining the proportions c and d can be made more efficient by "activating" the restrictions one by one. Let c_k denote the conditional proportion of samples that satisfy the kth restrictions, given that all previous $k - 1$ restrictions are satisfied. Then, $c = c_1 \cdots c_v$ and similarly for d (Mulder et al. 2009; Tijmstra & Bolsinova 2019), where a sample of vectors $\mathbf{p}$ under the first $k - 1$ restrictions can be obtained using a Gibbs sampler similar to Hoijtink and Molenaar (1997; Ligtvoet & Vermunt, 2012).

For illustration of the Gibbs sampler, let $J = 3$, and consider sampling p_3 from the prior distribution constrained by Eq. 14.4 using for $\mathbf{M}$ in Eq. 14.5 only $\mathbf{M}_1$, $\mathbf{M}_2$, and $\mathbf{M}_3$. Let

$$a = \min((\tilde{p}_1 + \tilde{p}_5)(\tilde{p}_4 + \tilde{p}_8)/(\tilde{p}_2 + \tilde{p}_6) - \tilde{p}_7, (\tilde{p}_1 + \tilde{p}_2)(\tilde{p}_7 + \tilde{p}_8)/(\tilde{p}_5 + \tilde{p}_6) - \tilde{p}_4)$$

$$\text{and } b = (\tilde{p}_2 + \tilde{p}_4)(\tilde{p}_5 + \tilde{p}_7)/(\tilde{p}_6 + \tilde{p}_8) - \tilde{p}_1, \tag{14.8}$$

with $\tilde{p}_k$ denoting the values sampled at the previous iteration. The newly sampled value $\tilde{p}_3$ obtained from the gamma distribution truncated between a and b then yields $\tilde{\mathbf{p}}/\mathbf{1}'\tilde{\mathbf{p}}$ as a single sample from the prior distribution constrained by the restrictions imposed by matrices $\mathbf{M}_1$, $\mathbf{M}_2$, and $\mathbf{M}_3$. This sampling procedure is repeated (for both the prior and posterior) for each p_k many times over, gradually adding the restrictions by extending $\mathbf{M}_1, \mathbf{M}_2, \ldots,$ up to $\mathbf{M}_v$. An R program for implementing this algorithm and calculating the Bayes factor is available from the author's website.

14.4 Application

As a small application, consider the transitive reasoning data (Verweij et al. 1996, for details), which are available from the **mokken** package in R (Van der Ark 2007). These data consist of the binary responses of $N = 425$ children to transitive reasoning tasks, where we limit the analyses to those tasks that relate to the task property Length ($J = 4$) and Weight ($J = 5$).

For the $J = 4$ items related to the task property Length, hypothesis H imposes $v = 30$ inequality restrictions. Considering only the restrictions on the bivariate distributions, we get a Bayes factor of $L = 1.2936$, indicating no clear evidence in favor or against property A. Including the remaining 24 restrictions imposed by hypothesis H yields $1/L = 4.6555$, which indicates substantial evidence against the variable of the task property Length being A. The result thus discredits any model that is a special case of the general conditions listed in Theorem 2, including the MH model. This result, however, was not obvious when only the information contained in the bivariate distributions was considered.

For the $J = 5$ items related to the task property Weight, $L = 13.1766$, indicating strong evidence in favor of hypothesis H.

14.5   Discussion

The Bayes factor for the hypothesis that the observable property A holds for all trivariate distributions of triplets of item variables (hypothesis H) provides a convenient way of summarizing evidence in favor of the many restrictions the property imposes on the observed binary data distribution. The application illustrated that the restrictions imposed by the property on the trivariate distributions, in addition to the restrictions on the bivariate distributions, cannot generally be ignored. A test of property A for all J item variables becomes practically infeasible due to the large numbers of restrictions. Hence, the proposed test strikes a balance between the power of the test and what is practically feasibility. However, the procedure for computing the Bayes factor is still computationally very intensive and is no longer feasible for more than seven items. The procedure would thus benefit from alternative ways of estimating or approximating the prior and posterior probabilities for the Bayes factor.

One easy way of alleviating the computational burden when assessing trivariate A is to consider calculating the Bayes factors separately for each of the triplets of item variables, rather than combining these same restrictions into a single test. Note that the number of restrictions in both cases is the same, but the Gibbs sampler runs faster many times on smaller problems than on a single run across the entire 2^J multinomial outcome space. However, the challenge then is to combine the $J(J-1)(J-2)/6$ Bayes factors to come to a judgment about the validity of the assumptions being tested for subsets of response variables. The use of a single global test has a clear advantage here.

Theorem 2 shows that property A for all trivariate distributions is implied by any model for binary response data that assumes the conditions CI and M to hold and additionally assumes that the random vector of (multidimensional) latent variables satisfies property A. These conditions include those that were proposed by Holland (1981) for trivariate distributions, Mokken (1971)s MH model, and a special multilevel version of the DINA model (Proposition 1). Whereas specific tests can be designed for each of the special instances of the conditions listed in

Theorem 2, the test proposed here is aimed at assessing whether the pursuit of any of such models is worth the effort at all.

References

Bartolucci, F., & Forcina, A. (2000). A likelihood ratio test for MTP2 within binary variables. *The Annals of Statistics, 28*(4), 1206–1218. https://www.jstor.org/stable/2673960.

Bartolucci, F., & Forcina, A. (2005). Likelihood inference on the underlying structure of IRT models. *Psychometrika, 70*(1), 31–43. https://doi.org/10.1007/s11336-001-0934-z.

Doignon, J. P., & Falmagne, J. C. (2012). *Knowledge spaces*. Springer Science & Business Media.

Ellis, J. L. (2014). An inequality for correlations in unidimensional monotone latent variable models for binary variables. *Psychometrika, 79*(2), 303–316. https://doi.org/10.1007/s11336-013-9341-5.

Ellis, J. L. (2015). MTP2 and partial correlations in monotone higher-order factor models. In R. E. Millsap, D. M. Bolt, L. A. Van der Ark, & W. C. Wang (Eds.), *Quantitative psychology research* (pp. 261–272). Springer.

Esary, J. D., Proschan, F., & Walkup, D. W. (1967). Association of random variables, with applications. *The Annals of Mathematical Statistics, 38*(5), 1466–1474. https://doi.org/10.1214/aoms/1177698701.

Ghurye, S. G., & Wallace, D. L. (1959). A convolutive class of monotone likelihood ratio families. *The Annals of Mathematical Statistics, 30*(4), 1158–1164. https://doi.org/10.1214/aoms/1177706101.

Grayson, D. A. (1988). Two-group classification in latent trait theory: Scores with monotone likelihood ratio. *Psychometrika, 53*(3), 383–392. https://doi.org/10.1007/BF02294219.

Hoijtink, H., & Molenaar, I. W. (1997). A multidimensional item response model: Constrained latent class analysis using the Gibbs sampler and posterior predictive checks. *Psychometrika, 62*(2), 171–189. https://doi.org/10.1007/BF02295273.

Holland, P. W. (1981). When are item response models consistent with observed data? *Psychometrika, 46*(1), 79–92. https://doi.org/10.1007/BF02293920.

Holland, P. W., & Rosenbaum, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics, 14*(4), 1523–1543. https://doi.org/10.1214/aos/1176350174.

Huynh, H. (1994). A new proof for monotone likelihood ratio for the sum of independent Bernoulli random variables. *Psychometrika, 59*(1), 77–79. https://doi.org/10.1007/BF02294266.

Joag-Dev, K. (1983). Independence via uncorrelatedness under certain dependence structures. *The Annals of Probability, 11*(4), 1037–1041. https://doi.org/10.1214/aop/1176993452.

Jogdeo, K. (1978). On a probability bound of Marshall and Olkin. *The Annals of Statistics, 6*(1), 232–234. https://doi.org/10.1214/aos/1176344082.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*(3), 258–272. https://doi.org/10.1177/01466210122032064.

Karlin, S., & Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities, I. Multivariate totally positive distributions. *Journal of Multivariate Analysis, 10*(4), 467–498. https://doi.org/10.1016/0047-259X(80)90065-2.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572.

Klugkist, I., & Hoijtink, H. (2007). The Bayes factor for inequality and about equality constrained models. *Computational Statistics & Data Analysis, 51*(12), 6367–6379. https://doi.org/10.1016/j.csda.2007.01.024.

Lavine, M., & Schervish, T. (1999). Bayes factors: What they are and what they are not. *The American Statistician, 53*(2), 119–122. https://doi.org/10.2307/2685729.

Ligtvoet, R., & Vermunt, J. K. (2012). Latent class models for testing monotonicity and invariant item ordering for polytomous items. *British Journal of Mathematical and Statistical Psychology, 65*(2), 237–250. https://doi.org/10.1111/j.2044-8317.2011.02019.x.

Loevinger, J. A. (1948). The technique of homogeneous tests compared with some aspects of scale analysis and factor analysis. *Psychological Bulletin, 45*(6), 507–530.

Mokken, R. J. (1971). *A theory and procedure of scale analysis*. De Gruyter.

Mokken, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous responses. *Applied Psychological Measurement, 6*(4), 417–430. https://doi.org/10.1177/014662168200600404.

Molenaar, I. W. (1991). A weighted Loevinger H-coefficient extending Mokken scaling to multicategory items. *Kwantitatieve Methoden, 12*(37), 97–117.

Molenaar, I. W., & Sijtsma, K. (2000). *User's manual MSP5 for Windows*. [Computer software], IEC ProGAMMA.

Mulder, J., Klugkist, I., Van de Schoot, R., Meeus, W. H. J., Selfhout, M., & Hoijtink, H. (2009). Bayesian model selection of informative hypotheses for repeated measurements. *Journal of Mathematical Psychology, 53*(6), 530–546. https://doi.org/10.1016/j.jmp.2009.09.003

Rosenbaum, P. R. (1984). Testing the conditional independence and monotonicity assumptions of item response theory. *Psychometrika, 49*(3), 425–435. https://doi.org/10.1007/BF02306030.

Sijtsma, K. (1988). *Contributions to Mokken's nonparametric item response theory*. Free University Press.

Suppes, P., & Zanotti, M. (1981). When are probabilistic explanations possible? *Synthese, 48*(2), 191–199. https://doi.org/10.1007/BF01063886.

Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In P. D. Nichols, S. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 327–359). Erlbaum.

Tijmstra, J., & Bolsinova, M. (2019). Bayes factors for evaluating latent monotonicity in polytomous item response theory models. *Psychometrika, 84*(3), 846–869. https://doi.org/10.1007/s11336-019-09661-w.

Tijmstra, J., Hoijtink, H., & Sijtsma, K. (2015). Evaluating manifest monotonicity using Bayes factors. *Psychometrika, 80*(4), 880–896. https://doi.org/10.1007/s11336-015-9475-8.

Ünlü, A. (2008). A note on monotone likelihood ratio of the total score variable in unidimensional item response theory. *British Journal of Mathematical and Statistical Psychology, 61*(1), 179–187. https://doi.org/10.1348/000711007X173391.

Van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software 20*(11), 1–19. https://doi.org/10.18637/jss.v020.i11.

Vermunt, J. K. (1999). A general class of nonparametric models for ordinal categorical data. *Sociological Methodology, 29*(1), 187–223. https://doi.org/10.1111/0081-1750.00064.

Verweij, A. C., Sijtsma, K., & Koops, W. (1996). A Mokken scale for transitive reasoning suited for longitudinal research. *International Journal of Behavioral Development, 19*(1), 219–238. https://doi.org/10.1177/016502549601900115.

Walkup, D. W. (1968). Minimal conditions for association of binary variables. *SIAM Journal on Applied Mathematics, 16*(6), 1394–1403. https://doi.org/10.1137/0116115.

Warrens, M. J. (2008). On association coefficients for 2×2 tables and properties that do not depend on the marginal distributions. *Psychometrika, 73*(4), 777–789. https://doi.org/10.1007/s11336-008-9070-3.