

AN ISOTONIC PARTIAL CREDIT MODEL FOR ORDERING SUBJECTS ON THE BASIS OF THEIR SUM SCORES

RUDY LIGTVOET

UNIVERSITY OF AMSTERDAM

In practice, the sum of the item scores is often used as a basis for comparing subjects. For items that have more than two ordered score categories, only the partial credit model (PCM) and special cases of this model imply that the subjects are stochastically ordered on the common latent variable. However, the PCM is very restrictive with respect to the constraints that it imposes on the data. In this paper, sufficient conditions for the stochastic ordering of subjects by their sum score are obtained. These conditions define the isotonic (nonparametric) PCM model. The isotonic PCM is more flexible than the PCM, which makes it useful for a wider variety of tests. Also, observable properties of the isotonic PCM are derived in the form of inequality constraints. It is shown how to obtain estimates of the score distribution under these constraints by using the Gibbs sampling algorithm. A small simulation study shows that the Bayesian p -values based on the log-likelihood ratio statistic can be used to assess the fit of the isotonic PCM to the data, where model-data fit can be taken as a justification of the use of the sum score to order subjects.

Key words: item response theory, monotone likelihood ratio, ordinal inferences, partial credit model, polytomously scored items, stochastic ordering, sum score.

1. Introduction

In item response theory (IRT), inferences are made about the ordering of subjects' latent scores on the basis of their responses to multiple test items. Many IRT models have been proposed to describe test data, and these models differ from each other with respect to the restrictions they impose on the data. For example, the Rasch (1960) model allows for consistent estimates of the differences between subjects' latent scores, as the sum of the item scores is a sufficient statistic for the latent variable (Andersen, 1973). This enables different subjects to be compared, irrespective of the particular set of items that were administered to them. But the restrictive parametric requirements the Rasch model imposes on the data also mean that the model might not fit. A nonparametric model like Mokken's (1971) monotone homogeneity model is looser in terms of the requirements it imposes on the data, making it applicable to a wider variety of tests. Unlike the Rasch model, however, it does not allow for the estimation of the differences between subjects, but allows for ordinal inferences only. More specifically, the Mokken model implies a stochastic ordering of subjects' latent scores on the basis of the sum of the item scores (Grayson, 1988; Huynh, 1994; Ünlü, 2008). This measurement property is called the *stochastic ordering of the latent trait by the sum score* (SOL; Hemker, Sijtsma, Molenaar, & Junker, 1996, 1997), which means that a subject with a higher sum score is also expected to have a higher latent score than a subject with a lower sum score. When a test is used with the goal of obtaining an ordering of subjects on the basis of their sum scores, both the Rasch model and the Mokken model are sufficient; but for many tests the Rasch model does not fit the data, and the nonparametric Mokken model might be more appropriate.

The Rasch (1960) model and Mokken's (1971) monotone homogeneity model are examples of IRT models for dichotomously scored items, where the sum score corresponds to the total

number correct for an ability test or the number of times a subject agrees with the presented statement of an attitude test. For many test applications, however, the responses to the items may be assigned more than two scores. For example, the performance on a complex problem may be assigned credits for each step correctly taken in solving the problem, and attitudes are often graded on a Likert scale containing more than two ordered options to choose from. The IRT models for items with more than two ordered scores are called polytomous IRT models.

Masters' (1982) partial credit model (PCM) is a polytomous IRT model which, like the Rasch (1960) model, has the sum scores as a sufficient statistic for the latent scores. However, the PCM imposes constraints on the data that are far more restrictive than necessary for the SOL property to hold. So, if the PCM does not fit the data, the SOL property might still hold. Like the dichotomous IRT models, it is tempting to think that a nonparametric version of the PCM is sufficient for the SOL property. Hemker et al. (1996, 1997) defined such a nonparametric PCM, and showed that Samejima's (1969) graded response model and Muraki's (1992) generalized PCM are special cases of this nonparametric PCM. However, they also found that none of these models imply the SOL property. In fact, they found that, except for the PCM and special cases hereof, none of the popular polytomous IRT models implies the SOL property. Hence, there seems to be a mismatch between the polytomous IRT models and the SOL property.

One way to resolve this mismatch is to define a weaker form of SOL that is implied by the polytomous IRT models. For example, Van der Ark and Bergsma (2010) defined weak SOL for the dichotomized sum scores, and Scheiblechner (2007) defined a monotone likelihood ordering on the basis of concordant and discordant item scores. The weak SOL property subdivides the sum scores into two groups, with sum scores higher than a particular sum score and sum scores lower than that sum score, respectively. These weaker forms of SOL imply that the group with the higher scores is also expected to have (on average) a higher latent value than the group with the lower scores. As such, the weak SOL property justifies the polytomous IRT models for the ordinal measurement of subjects, whereby the dichotomized sum scores yield a stochastic ordering. However, these weaker forms of SOL do not imply a stochastic ordering of subjects' latent values for each individual sum scores; and, as such, they do not justify the individual sum scores as practical measures by which subjects' latent values can be ordered. In contrast to the weaker forms of SOL, the SOL property justifies the used of the sum score for ordinal measurement of subjects, and it allows the sum score to be used as a direct scoring rule for actually performing such measurement.

In this paper, the gap that exists between the polytomous IRT models and the SOL property is bridged by proposing an additional constraint to the nonparametric PCM. It will be shown that the additional constraint is sufficient for the *monotone likelihood ratio* property (MLR; Lehmann, 1959; Hemker et al., 1996, 1997), which implies the SOL property. The nonparametric PCM with the additional constraint is referred to as the isotonic PCM. The isotonic PCM justifies the use of the sum score as a scoring rule for ordinal measurement, without having to assume any parametric distributions. Also, observable properties are derived from the isotonic PCM, allowing the model to be empirically tested. In a simulation study, the merits of the Gibbs sampling algorithm for assessing the model-data fit are investigated for tests consisting of a small numbers of items.

2. Theory

Let A denote the set of subjects taking the test J . The test is assumed to be a finite and fixed set of items. Also, let X_{ai} denote the item score variable for subject $a \in A$ on item $i \in J$. The item scores are assumed to be ordinal and have realizations $x_{ai} \in \{0, \dots, m_i\}$. Here, m_i denotes the maximum score assigned to item i , which is allowed to vary across different items. Notice that for $m_i = 1$, item i is dichotomously scored. The probability of the score $X_{ai} = x_{ai}$ is denoted

by $P_x(a, i)$, which depends on both the subject a and the item i . It is assumed that the $m_i + 1$ probabilities sum to unity for each item–subject pair, and that all probabilities are positive.

In IRT, a latent variable Θ is proposed to explain the associations that exist between the item scores on the test. Two assumptions lie at the heart of almost all IRT models. First, each subject $a \in A$ is characterized by a value θ_a , where the assumption of unidimensionality (UD) states that these values are located on a real-line. In our notation, the subject a is interpreted as an individual subject. However, a can also be considered a subject from an equivalence class of subjects, all with the same latent value θ_a . For example, if Θ is assumed to be discrete, θ_a corresponds to a homogeneous latent class of subjects (e.g., Lazarsfeld & Henry, 1968). Usually, IRT models are expressed by a function that is defined across Θ and relates the latent variable to each item score, allowing the items to be characterized by a small number of parameters, and also allowing inferences to be made about the distribution of the latent variable. However, to emphasize that our goal is to attain an ordering of subjects, we instead focus here on the probabilities $P_x(a, i)$ defined for the set of subjects A .

Second, it is assumed that the item scores are locally (i.e., conditional on $\Theta = \theta_a$) independent (LI), which means that the joint score probability $P(X_{a1} = x_{a1}, X_{a2} = x_{a2}, \dots)$ equals the product of the corresponding marginals $P_x(a, i)$ for each $a \in A$. Together, the assumptions UD and LI allow the latent variable to be interpret as the variable underlying or explaining the item scores.

2.1. Partial Credit Models

Let the local odds corresponding to item score $X_{ai} = x_{ai}$ be defined as

$$O_x(a, i) = \frac{P_x(a, i)}{P_{x-1}(a, i)}$$

(e.g., Agresti, 1990; Douglas, Fienberg, Lee, Sampson, & Whitaker, 1991), where $O_x(a, i) = 1$ for $x_{ai} = 0$, by definition. Assuming UD and LI, Masters’ (1982) PCM can be defined for $x_{ai} > 0$ as

$$O_x(a, i) = \exp(\theta_a - \delta_{ix}), \tag{1}$$

where for each item score δ_{ix} denotes the corresponding difficulty parameter. Notice that in Equation (1) the local odds is exponentially increasing in Θ . Hemker (1996) and Hemker et al. (1996, 1997) defined the nonparametric PCM, where Equation (1) was relaxed by assuming the local odds to be nondecreasing in Θ . The latter assumption was also proposed by Holland and Rosenbaum (1986) under the name of *latent total positivity 2* (LTP₂; cf. Karlin, 1968). Formally, the assumption of LTP₂ states that, for all x_{ai} and x_{bi} (where $i \in J$ and $a, b \in A$), if $\theta_a < \theta_b$ then

$$O_x(a, i) \leq O_x(b, i). \tag{2}$$

Special cases of the nonparametric PCM include Muraki’s (1992) generalized PCM, and Samejima’s (1969) graded response model (Hemker et al., 1997, Theorem 3). However, none of these three models imply the SOL property (Hemker et al., 1996, 1997).

In order to obtain a polytomous IRT model that does imply the SOL property, one additional assumption is made: the assumption of *weak instrumental independence* (WI; cf. Scheiblechner, 1999). Scheiblechner (1999) introduced weak instrumental independence as an axiom to the isotonic probabilistic modeling framework in order to obtain decomposable scales for the subject, the items, and the instrument scales. In another context, Samejima (1972) investigated the orderliness property for graded responses, which corresponds to assumption WI. The difference between both Scheiblechner’s models and the graded response models, and the model presented

here is that in this paper the focus is on the local odds, whereas both Scheiblechner (1995, 1999) and Samejima (1969, 1972) consider cumulative probabilities instead. Neither Scheiblechner’s isotonic models nor the graded response models imply SOL. Here, assumption WI states that for all $1 + x_{ai} = y_{ai} \leq m_i$ (where $a \in A$ and $i \in J$),

$$O_x(a, i) \geq O_y(a, i). \tag{3}$$

In words, assumption WI restricts the different local odds pertaining to the same item to be ordered the same for all subjects, in descending order of the item scores. In terms of Scheiblechner’s models, the nonparametric model with assumption WI is isotonic with respect to θ for both subjects and item scores. The model that is defined by the assumptions of UD, LI, LTP₂ (Equation (2)), and WI (Equation (3)) is therefore called the isotonic PCM. Strictly speaking, the PCM (Equation (1)) does not imply WI, as the difficulty parameters δ_{ix} are formally not restricted to be increasing in x . Notice also that for dichotomously scored items (i.e., $m_i = 1$) WI is implied by definitions as only a single odds exists. Next, the MLR property is discussed.

Let the sum score be denoted by X_{a+} and let $P(X_{a+} = x_{a+})$ denote the probability of subject a obtaining a sum score equal to x_{a+} . The MLR property states that: for all $1 + x_{a+} = y_{a+}$ (where $a, b \in A$), if $\theta_a < \theta_b$ then

$$P(X_{a+} = x_{a+})P(X_{b+} = y_{b+}) \geq P(X_{b+} = x_{b+})P(X_{a+} = y_{a+}). \tag{4}$$

Let $\mathbf{x} = (x_1, x_2, \dots)$ and $\mathbf{y} = (y_1, y_2, \dots)$ be two observed vectors of the item scores, for which $1 + \sum \mathbf{x} = \sum \mathbf{y}$. Assuming LI, the probability of a sum score $X_{a+} = x_{a+}$ can be expressed as

$$\begin{aligned} P(X_{a+} = x_{a+}) &= \sum_{\mathbf{x}} \prod_{i \in J} P_x(a, i) \\ &= \prod_{i \in J} P_0(a, i) \sum_{\mathbf{x}} \left(\prod_{i \in J} \frac{P_x(a, i)}{P_0(a, i)} \right) \\ &= \prod_{i \in J} P_0(a, i) \sum_{\mathbf{x}} \left(\prod_{i \in J} \frac{P_1(a, i)}{P_0(a, i)} \times \frac{P_2(a, i)}{P_1(a, i)} \times \dots \times \frac{P_x(a, i)}{P_{x-1}(a, i)} \right) \\ &= \prod_{i \in J} P_0(a, i) \sum_{\mathbf{x}} \left(\prod_{i \in J} Z_x(a, i) \right) \end{aligned} \tag{5}$$

(cf. Huynh, 1994), with

$$Z_x(a, i) = \prod_{z=0}^x O_z(a, i). \tag{6}$$

Next, inserting Equation (5) in Equation (4) yields

$$\begin{aligned} \sum_{\mathbf{x}} \left(\prod_{i \in J} Z_x(a, i) \right) \sum_{\mathbf{y}} \left(\prod_{i \in J} Z_y(b, i) \right) &\geq \sum_{\mathbf{x}} \left(\prod_{i \in J} Z_x(b, i) \right) \sum_{\mathbf{y}} \left(\prod_{i \in J} Z_y(a, i) \right) \\ \Leftrightarrow \sum_{\mathbf{x}} \sum_{\mathbf{y}} \left(\prod_{i \in J} Z_x(a, i) Z_y(b, i) - \prod_{i \in J} Z_x(b, i) Z_y(a, i) \right) &\geq 0. \end{aligned} \tag{7}$$

The MLR property is satisfied if for all sum scores $1 + x_{a+} = y_{a+}$ (where $a, b \in A$) the left-hand side of Equation (7) is nonnegative.

Hemker et al. (1996) showed that MLR implies (i.e., is sufficient for) the SOL property, but the MLR property is more general than the PCM. With ‘ \Rightarrow ’ denoting a logical implication, this means that

$$\text{PCM} \Rightarrow \text{MLR} \Rightarrow \text{SOL},$$

but the reverse relationships do not hold. Both the PCM and the isotonic PCM are special cases of the nonparametric PCM as they impose additional restrictions on top of the UD, LI and LTP₂ assumptions, but the PCM and the isotonic PCM do not logically imply one another. (The PCM implies the isotonic PCM only if, for each item, $\delta_{i1} \leq \delta_{i2} \leq \dots \leq \delta_{im}$.) Hence,

$$\text{PCM} \Rightarrow \text{nonparametric PCM}$$

and

$$\text{isotonic PCM} \Rightarrow \text{nonparametric PCM},$$

but the reverse relationships do not hold. Further, the theorem below shows that the isotonic PCM implies the MLR property (but not the other way around), and thus also implies the SOL property:

$$\text{isotonic PCM} \Rightarrow \text{MLR} \Rightarrow \text{SOL}.$$

Theorem. *The isotonic PCM is sufficient (but not necessary) for MLR.*

To prove the theorem, the property of MLR is broken down into smaller, more tractable parts. First, element-wise MLR is defined for a single item, for which Equation (7) holds with all the odds of the other items being constant. It is then shown (see lemma below) that element-wise MLR for each item implies the MLR property. Consequently, the proof of the theorem consists of showing that the isotonic PCM implies element-wise MLR for each item. This proof is given after the proof of the lemma.

Define *element-wise MLR* for item i as the case for which Equation (7) holds with all the odds except those of item i being constant on the interval between θ_a and θ_b . Hence, element-wise MLR for item i means that Equation (7) holds with $Z_x(a, k) = Z_x(b, k)$ and $Z_y(a, k) = Z_y(b, k)$ for all $k \neq i$. To break down Equation (7) into smaller and more tractable parts, it will first be shown by induction that Equation (7) is implied if element-wise MLR holds for each item separately.

Lemma. *MLR (Equation (7)) is implied if element-wise MLR holds for each item separately.*

Proof: For two items i and k , Equation (7) yields

$$\sum_x \sum_y (Z_x(a, i)Z_y(b, i)Z_x(a, k)Z_y(b, k) - Z_x(b, i)Z_y(a, i)Z_x(b, k)Z_y(a, k)) \geq 0. \quad (8)$$

We start by showing that element-wise MLR for both items i and k separately imply Equation (8).

Suppose that element-wise MLR holds for item i . Then the odds corresponding to item k are constant on the interval between θ_a and θ_b , which means that the subject indices pertaining to the odds of item k can be arbitrarily assigned the index a , as $Z_x(a, k) = Z_x(b, k)$ and $Z_y(a, k) = Z_y(b, k)$. This yields for Equation (8) the following inequality

$$\sum_x \sum_y (Z_x(a, i)Z_y(b, i)Z_x(a, k)Z_y(a, k) - Z_x(b, i)Z_y(a, i)Z_x(a, k)Z_y(a, k)) \geq 0. \quad (9)$$

Likewise, if we assume element-wise MLR holds for item k , then the indices of the odds of item i can be arbitrarily assigned the index a . However, as the assignment of indices to item i is arbitrary if element-wise MLR holds only for item k , we might just as well conveniently adapt these indices from Equation (9). Adapting the subject indices pertaining to the odds of item i from Equation (9), and assuming element-wise MLR only for item k , yields an inequality with indices that correspond to the situation where element-wise MLR holds for both item i and item k separately. Moreover, this inequality is equal to Equation (8). This shows that if element-wise MLR holds for item i and item k separately, then MLR holds also for item i and item k together. The same line of reasoning can be readily extended to any number of items, which shows that Equation (7) holds if element-wise MLR holds for each item separately. \square

The lemma shows that to prove the theorem it is sufficient to prove that the isotonic PCM implies element-wise MLR for each item $i \in J$, for which Equation (7) reduces to

$$\sum_x \sum_y \left((Z_x(a, i)Z_y(b, i) - Z_x(b, i)Z_y(a, i)) \prod_{\substack{k \in J \\ k \neq i}} Z_x(b, k)Z_y(b, k) \right) \geq 0. \tag{10}$$

In Equation (10), element-wise MLR is defined as a nonnegative summation across the difference in odds between the parentheses. This does not mean that for each pair of patterns \mathbf{x} and \mathbf{y} the difference between the parentheses is nonnegative. Let x_i and y_i refer to the i th element of \mathbf{x} and \mathbf{y} , respectively, and suppose that $x_i < y_i$. Then, Equation (6) shows that

$$Z_y(a, i) = Z_x(a, i) \prod_{z=x+1}^y O_z(a, i),$$

which means that for those pairs of patterns \mathbf{x} and \mathbf{y} in Equation (10) for which $x_i \leq y_i$,

$$\begin{aligned} & (Z_x(a, i)Z_y(b, i) - Z_x(b, i)Z_y(a, i)) \prod_{\substack{k \in J \\ k \neq i}} Z_x(b, k)Z_y(b, k) \\ &= Z_x(a, i)Z_x(b, i) \left(\prod_{z=x+1}^y O_z(b, i) - \prod_{z=x+1}^y O_z(a, i) \right) \prod_{\substack{k \in J \\ k \neq i}} Z_x(b, k)Z_y(b, k) \geq 0. \end{aligned}$$

The last inequality holds for the isotonic PCM because LTP_2 implies that the difference in between the parentheses is nonnegative. However, using the same line of reasoning it can be shown that, for those pairs of patterns \mathbf{x} and \mathbf{y} for which $x_i > y_i$, the difference between the parentheses of Equation (10) can take on negative values. To show that the isotonic PCM implies element-wise MLR, it needs to be shown that the sum of these negative values cannot be larger than the sum of the positive values implied by the isotonic PCM. The proof of the theorem thus consists in showing that the isotonic PCM implies that the negative values associated with the patterns \mathbf{x} and \mathbf{y} for which $x_i > y_i$ are canceled out by the positive values for those patterns for which $x_i < y_i$.

Proof of the Theorem: Define the domain D as a collection of all pairs of patterns \mathbf{x} and \mathbf{y} in Equation (10) for which $x_i > y_i$. Also, consider the codomain C to contain all pairs of patterns for which $x_i < y_i$. The proof of the theorem consists of two parts. In the first part a function is defined that assigns all elements of D (i.e., all pairs of patterns \mathbf{x} and \mathbf{y} for which $x_i > y_i$) to a distinct element in C . In the second part, it is shown that the isotonic PCM implies that

the negative values of the elements in D are canceled out by the positive values of the assigned elements in C . The result is that Equation (10) yields nonnegative values under the assumptions of the isotonic PCM, which by the lemma implies MLR.

Let $\{\mathbf{x}_1^d, \mathbf{y}_1^d\}$ denote an element in D for which $x_i = y_i + u$ for any $u \geq 1$ (i.e., $x_i > y_i$). Also, let item j be referred to as the *operator* if $j \neq i$ and for the item it holds that $x_j < y_j$. For each element from D , an operator item exists, as $1 + \sum \mathbf{x} = \sum \mathbf{y}$ (from the definition of MLR). Further, let \mathbf{x}_{ijk} and \mathbf{y}_{ijk} denote the patterns excluding the scores on the items i, j , and k . Without loss of generality, the pair of patterns \mathbf{x}_1^d and \mathbf{y}_1^d can be expressed as

$$\mathbf{x}_1^d = (x_i, x_j, x_k, \mathbf{x}_{ijk}) \quad \text{and} \quad \mathbf{y}_1^d = (x_i - u, y_j, y_k, \mathbf{y}_{ijk}). \quad (11)$$

Next, consider also the pairs of patterns \mathbf{x}_1^c and \mathbf{y}_1^c for which

$$\mathbf{x}_1^c = (x_i - u, y_j - 1, y_k, \mathbf{y}_{ijk}) \quad \text{and} \quad \mathbf{y}_1^c = (x_i, x_j + 1, x_k, \mathbf{x}_{ijk}). \quad (12)$$

Because $u \geq 1$, the pair $\{\mathbf{x}_1^c, \mathbf{y}_1^c\}$ is an element from the codomain C . Notice that \mathbf{x}_1^c is obtained from \mathbf{y}_1^d by subtracting 1 from the score on the operator j (for which $x_j < y_j$), and likewise, \mathbf{y}_1^c is obtained from \mathbf{x}_1^d by adding 1 to the score on the operator. This operation for obtaining the element $\{\mathbf{x}_1^c, \mathbf{y}_1^c\}$ from $\{\mathbf{x}_1^d, \mathbf{y}_1^d\}$ is denoted by the function $f_j : D \rightarrow C$, which depends on the choice of the operator (in case more than one item exist for which $x_j < y_j$). For a given operator, the function yields distinct elements in C for each element in D . However, if different items are used for the role of operator, then different elements in D can be assigned to the same element in C (i.e., a surjection). To illustrate, consider the pair $\{\mathbf{x}_2^d, \mathbf{y}_2^d\}$ from the D , where

$$\mathbf{x}_2^d = (x_i, x_j + 1, x_k - 1, \mathbf{x}_{ijk}) \quad \text{and} \quad \mathbf{y}_2^d = (x_i - u, y_j - 1, y_k + 1, \mathbf{y}_{ijk}).$$

Suppose now that item k is assigned the role of operator, then $f_k : \{\mathbf{x}_2^d, \mathbf{y}_2^d\}$ yields $\{\mathbf{x}_1^c, \mathbf{y}_1^c\}$; the same element in the codomain as $f_j : \{\mathbf{x}_1^d, \mathbf{y}_1^d\}$. This example illustrates that the function f_j does not guarantee that each element in D can be assigned to a distinct element in C . If an alternative item exists for the role of operator, the surjection can be resolved by assigning the role of operator to another item. However, a problem occurs when for $\{\mathbf{x}_1^d, \mathbf{y}_1^d\}$ only item j can be the operator (i.e., only for item j it holds that $x_j < y_j$) and for $\{\mathbf{x}_2^d, \mathbf{y}_2^d\}$ only item k can be the operator. This situation, where two elements in D yield the same element in C without alternative choices for the operators, is referred to as the *exceptional case*. If for $\{\mathbf{x}_1^d, \mathbf{y}_1^d\}$ only item j can be the operator, then the exceptional case implies that $x_k \geq y_k$ and $\sum \mathbf{x}_{ijk} \geq \sum \mathbf{y}_{ijk}$, which means that $x_j + 1 \leq y_j - u$. In addition, if for $\{\mathbf{x}_2^d, \mathbf{y}_2^d\}$ only item k can take up the role of operator, then $x_j + 1 \geq y_j - 1$, which means that $u = 1$. Hence, if two elements from D are assigned to the same element in C , and both elements have only a single (but different) item to take the role of operator, then it has to be true that $u = 1$. Furthermore, this means that $x_k \leq y_k + 1 - u = y_k$, which holds only if $x_j + 1 = y_j - 1$, $x_k = y_k$, and $\mathbf{x}_{ijk} = \mathbf{y}_{ijk}$. Hence, if $f_j : \{\mathbf{x}_1^d, \mathbf{y}_1^d\}$ and $f_k : \{\mathbf{x}_2^d, \mathbf{y}_2^d\}$ both yield the same element in C , then the vectors patterns $\mathbf{x}_1^d, \mathbf{y}_1^d, \mathbf{x}_2^d$, and \mathbf{y}_2^d can be expressed as

$$\begin{aligned} \mathbf{x}_1^d &= (x_i, x_j, x_k, \mathbf{x}_{ijk}), & \mathbf{y}_1^d &= (x_i - 1, x_j + 2, x_k, \mathbf{x}_{ijk}), \\ \mathbf{x}_2^d &= (x_i, x_j + 1, x_k - 1, \mathbf{x}_{ijk}), & \mathbf{y}_2^d &= (x_i - 1, x_j + 1, x_k + 1, \mathbf{x}_{ijk}). \end{aligned} \quad (13)$$

Also, we obtain that the patterns in Equation (12) become

$$\mathbf{x}_1^c = (x_i - 1, x_j + 1, x_k, \mathbf{x}_{ijk}) \quad \text{and} \quad \mathbf{y}_1^c = (x_i, x_j + 1, x_k, \mathbf{x}_{ijk}). \quad (14)$$

Notice that the existence of the pairs $\{\mathbf{x}_1^d, \mathbf{y}_1^d\}$ and $\{\mathbf{x}_2^d, \mathbf{y}_2^d\}$ implies that $0 \leq x_j < m_j$. This immediately shows that the surjection can be resolved by allowing for the exceptional case item j to take up the role of operator for $\{\mathbf{x}_2^d, \mathbf{y}_2^d\}$ (an exceptional case as $x_j = y_j$). For this exceptional case, $f'_j : \{\mathbf{x}_2^d, \mathbf{y}_2^d\}$ (prime added to indicate the exception) is properly defined and yields the distinct element $\{\mathbf{x}_2^c, \mathbf{y}_2^c\}$ in C . This element $\{\mathbf{x}_2^c, \mathbf{y}_2^c\}$ is distinct as it could not be attained without the exception of an operator for which $x_j = y_j$. Applying function f'_j to $\{\mathbf{x}_2^d, \mathbf{y}_2^d\}$ reveals that

$$\mathbf{x}_2^c = (x_i - 1, x_j, x_k + 1, \mathbf{x}_{ijk}) \quad \text{and} \quad \mathbf{y}_2^c = (x_i, x_j + 2, x_k - 1, \mathbf{x}_{ijk}). \tag{15}$$

Next it is shown that the isotonic PCM implies that the negative values in Equation (10) of the elements in D are canceled out by the positive values of the assigned elements in C . This is first shown for the operator item j for which $x_j < y_j$. Second, it is shown also to be true in the exceptional case for which $x_j = y_j$.

First, for the pair $\{\mathbf{x}_1^d, \mathbf{y}_1^d\}$ in Equation (11) and $\{\mathbf{x}_1^c, \mathbf{y}_1^c\}$ in Equation (12), Equation (10) reduces to

$$\begin{aligned} & (Z_x(a, i)Z_{x-u}(b, i) - Z_x(b, i)Z_{x-u}(a, i))Z_x(b, j)Z_y(b, j) \prod_{\substack{k \in J \\ k \neq i, j}} Z_x(b, k)Z_y(b, k) \\ & + (Z_{x-u}(a, i)Z_x(b, i) - Z_{x-u}(b, i)Z_x(a, i))Z_{y-1}(b, j)Z_{x+1}(b, j) \\ & \times \prod_{\substack{k \in J \\ k \neq i, j}} Z_x(b, k)Z_y(b, k) \geq 0 \\ \Leftrightarrow & (Z_x(b, i)Z_{x-u}(a, i) - Z_x(a, i)Z_{x-u}(b, i)) \\ & \times (Z_{y-1}(b, j)Z_{x+1}(b, j) - Z_x(b, j)Z_y(b, j)) \prod_{\substack{k \in J \\ k \neq i, j}} Z_x(b, k)Z_y(b, k) \geq 0, \tag{16} \end{aligned}$$

where the inequality holds if both terms between the parentheses in Equation (16) are nonnegative. The first term is nonnegative if

$$\begin{aligned} & Z_x(b, i)Z_{x-u}(a, i) \geq Z_x(a, i)Z_{x-u}(b, i) \\ \Leftrightarrow & \prod_{z=x-u+1}^x O_z(b, i) \geq \prod_{z=x-u+1}^x O_z(a, i), \end{aligned}$$

which holds because of assumption LTP₂. The second term between the parentheses in Equation (16) is nonnegative if

$$\begin{aligned} & Z_{y-1}(b, j)Z_{x+1}(b, j) \geq Z_x(b, j)Z_y(b, j) \\ \Leftrightarrow & O_{x+1}(b, j) \geq O_y(b, j). \end{aligned}$$

Because $x_j < y_j$, the second term in Equation (16) is nonnegative because of assumption WI.

Second, for the exceptional case Equation (10) is applied to the four pairs $\{\mathbf{x}_1^d, \mathbf{y}_1^d\}$ and $\{\mathbf{x}_2^d, \mathbf{y}_2^d\}$ in Equation (13), the pair $\{\mathbf{x}_1^c, \mathbf{y}_1^c\}$ in Equation (14), and the pair $\{\mathbf{x}_2^c, \mathbf{y}_2^c\}$ in

Equation (15):

$$\begin{aligned}
 & (Z_x(a, i)Z_{x-1}(b, i) - Z_x(b, i)Z_{x-1}(a, i))Z_x(b, j)Z_{x+2}(b, j)Z_x(b, k)Z_x(b, k) \prod_{\substack{l \in J \\ l \neq i, j, k}} Z_x^2(b, l) \\
 & + (Z_x(a, i)Z_{x-1}(b, i) - Z_x(b, i)Z_{x-1}(a, i))Z_{x+1}(b, j)Z_{x+1}(b, j)Z_{x-1}(b, k)Z_{x+1}(b, k) \\
 & \times \prod_{\substack{l \in J \\ l \neq i, j, k}} Z_x^2(b, l) \\
 & + (Z_{x-1}(a, i)Z_x(b, i) - Z_{x-1}(b, i)Z_x(a, i))Z_{x+1}(b, j)Z_{x+1}(b, j)Z_x(b, k)Z_x(b, k) \\
 & \times \prod_{\substack{l \in J \\ l \neq i, j, k}} Z_x^2(b, l) \\
 & + (Z_{x-1}(a, i)Z_x(b, i) - Z_{x-1}(b, i)Z_x(a, i))Z_x(b, j)Z_{x+2}(b, j)Z_{x+1}(b, k)Z_{x-1}(b, k) \\
 & \times \prod_{\substack{l \in J \\ l \neq i, j, k}} Z_x^2(b, l) \geq 0,
 \end{aligned}$$

for which factorization similar to Equation (16) yields

$$\begin{aligned}
 & (Z_x(b, i)Z_{x-1}(a, i) - Z_x(a, i)Z_{x-1}(b, i))(O_{x+1}(b, j) - O_{x+2}(b, j)) \\
 & \times (O_x(b, k) - O_{x+1}(b, k))Z_x(b, j)Z_{x+1}(b, j)Z_{x-1}(b, k)Z_x(b, k) \prod_{\substack{l \in J \\ l \neq i, j, k}} Z_x^2(b, l) \\
 & \geq 0.
 \end{aligned} \tag{17}$$

Equation (17) holds if all three terms between the parentheses are nonnegative. Similar to Equation (16), the first term is nonnegative because of assumption LTP₂. The second and third terms between the parentheses are nonnegative because of assumption WI. Hence, for any set of elements from D , a distinct set of elements from C exists for which the isotonic PCM implies that Equation (10) is nonnegative. This concludes the proof. \square

2.2. Observable Properties

In order to test the assumptions of the isotonic model, observable consequences of the model need to be obtained. Such observable consequences allow the use of the sum score for a stochastic ordering of subjects to be empirically tested. Holland and Rosenbaum (1986, Theorem 5) showed that the assumptions UD, LI, and LTP₂ imply a property which they called conditional multivariate TP₂ (CMTP₂; cf. Karlin & Rinott, 1980). This means that

$$\text{isotonic PCM} \Rightarrow \text{nonparametric PCM} \Rightarrow \text{CMTP}_2.$$

Consider the item score vector $\mathbf{x} = (x_i, x_j, \mathbf{x}_{ij})$, where \mathbf{x}_{ij} is the vector of item scores excluding the scores on items i and j . Also, consider the scores z_i and z_j , for which $x_i < z_i$ and $x_j < z_j$. The observable property CMTP₂ states that for all $i, j \in J$ and all \mathbf{x}_{ij} ,

$$P(\mathbf{X} = \mathbf{x})P(\mathbf{X} = (z_i, z_j, \mathbf{x}_{ij})) \geq P(\mathbf{X} = (x_i, x_j, \mathbf{x}_{ij}))P(\mathbf{X} = (z_i, x_j, \mathbf{x}_{ij})). \tag{18}$$

In words, for any 2×2 sub-table of the joined score probabilities of items i and j (conditional on the other item scores), CMTP₂ means that the product of the two diagonal elements is larger

than the product of the two off-diagonal elements. Any reversal of the inequality in Equation (18) indicates a violation of the joined assumptions UD, LI, and LTP₂. Hence, with ‘¬’ denoting the logical negation,

$$\neg\text{CMTP}_2 \Rightarrow \neg\text{nonparametric PCM} \Rightarrow \neg\text{isotonic PCM}.$$

Besides the property of CMTP₂, a manifest property for WI can be obtained. Let \mathbf{x}_i denote the item score vector, excluding the score on item i . Let the manifest local odds be expressed as

$$\frac{P(X = (x_i, \mathbf{x}_i))}{P(X = (x_i - 1, \mathbf{x}_i))}.$$

Here, conditional manifest WI (CMWI) is defined as follows: for all $i \in J$, all $0 < x_i < m_i$, and all \mathbf{x}_i ,

$$\frac{P(X = (x_i, \mathbf{x}_i))}{P(X = (x_i - 1, \mathbf{x}_i))} \geq \frac{P(X = (x_i + 1, \mathbf{x}_i))}{P(X = (x_i, \mathbf{x}_i))}. \tag{19}$$

The next proposition shows that the isotonic PCM is sufficient for CMWI; that is,

$$\text{isotonic PCM} \Rightarrow \text{CMWI}.$$

Proposition. *The assumptions UD, LI, and WI imply CMWI.*

Proof: Let $X \setminus X_i$ denote the score vector excluding the i th item score variable. Assuming UD and LI, the manifest local odds can be expressed as

$$\frac{P(X_i = x_i | X \setminus X_i = \mathbf{x}_i)}{P(X_i = x_i - 1 | X \setminus X_i = \mathbf{x}_i)} = E(O_x(a, i) | X \setminus X_i = \mathbf{x}_i)$$

(cf. Junker, 1993, Proposition 4.1a). Inserting $E(O_x(a, i) | X \setminus X_i = \mathbf{x}_i)$ into Equation (19) shows that CMWI is implied by WI, as the inequality holds for all $a \in A$. □

The proposition states that under the assumptions UD, LI, and WI, the manifest local odds of item i are ordered the same for all subjects, in descending order of the item scores (conditional on the rest of the item scores). Any reversal of the inequality in Equation (19) indicates a violation of the joined assumptions UD, LI, and WI. Hence,

$$\neg\text{CMWI} \Rightarrow \neg\text{isotonic PCM}.$$

The observable properties CMTP₂ and CMWI form the basis for our test for the fit of the isotonic PCM to the data. For CMTP₂ the score vector \mathbf{x} is partitioned into disjoint item scores (x_i, x_j) and \mathbf{x}_{ij} , where \mathbf{x}_{ij} corresponds to the remaining item scores. This means that depending on the number of items in the test and numbers m_i of these items, a large number of inequalities in Equation (18) and Equation (19) need to be assessed. To illustrate, for 4 items for which each $m_i = 2$, there are 486 inequalities corresponding to CMTP₂ and 108 corresponding to CMWI. Beside the problem of having to combine these Boolean outcomes into a single conclusion about whether or not the isotonic PCM fits the data, we also need to keep in mind that a large sample size would be needed to obtain reliable estimates for the probabilities in Equation (18) and Equation (19). To deal with sparse data, the remaining item scores can be combined ad hoc to form rest-score groups (Molenaar & Sijtsma, 2000), with each group containing at least a preset number of subjects. However, this method also results in a loss of power of the test as many of the inequalities in Equation (18) and Equation (19) are ignored. Instead of this method of combining the remaining item score into rest-score groups, we explore in the next section the use of

the Gibbs sampling algorithm to simultaneously deal with the large number of inequalities and with sparse data.

3. The Gibbs Sampling Algorithm

The Gibbs sampler allows inequality constraints to be imposed on the probability $P(\mathbf{X} = \mathbf{x})$, for all score vectors \mathbf{x} , by restricting its prior distribution. Hereby, each probability is iteratively sampled from its posterior distribution, conditional on all the other probabilities at the previous iteration. Examples of such applications of the Gibbs sampler include Hoijtink (1998), Hoijtink and Molenaar (1997), Karabatsos (2001), Karabatsos and Sheu (2004), Ligtoet and Vermunt (2012), and Van Onna (2002).

For notational convenience, we consider (without loss of generality) the expected frequency $F(\mathbf{X} = \mathbf{x})$ instead of the probability $P(\mathbf{X} = \mathbf{x})$. Suppose that $x_i < z_i$ and $x_j < z_j$. The observable property CMTP_2 implies that

$$F(\mathbf{X} = \mathbf{x})F(\mathbf{X} = (z_i, z_j, \mathbf{x}_{ij})) \geq F(\mathbf{X} = (x_i, z_j, \mathbf{x}_{ij}))F(\mathbf{X} = (z_i, x_j, \mathbf{x}_{ij})).$$

This means that $F(\mathbf{X} = \mathbf{x})$ is restricted from below by

$$F(\mathbf{X} = (x_i, z_j, \mathbf{x}_{ij}))F(\mathbf{X} = (z_i, x_j, \mathbf{x}_{ij}))/F(\mathbf{X} = (z_i, z_j, \mathbf{x}_{ij})).$$

Likewise, the expected frequency of the score vector $(x_i, z_j, \mathbf{x}_{ij})$ is restricted from above by

$$F(\mathbf{X} = \mathbf{x})F(\mathbf{X} = (z_i, z_j, \mathbf{x}_{ij}))/F(\mathbf{X} = (z_i, x_j, \mathbf{x}_{ij})).$$

For all combinations of item pairs and all \mathbf{x}_{ij} , we can list all those inequalities that involve $F(\mathbf{X} = \mathbf{x})$ and collect all the lower bounds for $F(\mathbf{X} = \mathbf{x})$ in the set L and all the upper bounds in the set U .

Similar to CMTP_2 we can collect all the lower and upper bounds for each expected item score frequency on the basis of the observable property CMWI . For example, CMWI implies that $F(\mathbf{X} = \mathbf{x})$ is restricted from below by

$$\sqrt{F(\mathbf{X} = (x_i, x_i - 1))F(\mathbf{X} = (x_i, x_i + 1))}$$

(see Equation (19)). By listing all the inequalities implied by CMWI that involve $F(\mathbf{X} = \mathbf{x})$, and adding the lower and upper bounds to L and U , respectively, we obtain that any frequency of \mathbf{x} between $\max(L)$ and $\min(U)$ satisfies the constraints of both CMTP_2 and CMWI .

3.1. Estimation

The goal of the following procedure is to obtain estimates for the expected frequencies under the constraints imposed by both CMTP_2 and CMWI . Subsequently, these estimates can be compared to the observed item score frequencies to assess the fit of the isotonic PCM to the data.

With the frequencies $F(\mathbf{X} = \mathbf{x})$ following a multinomial distribution, the posterior distribution of interest (with the conjugated prior) is Dirichlet distributed. Using the Gibbs sampling algorithm, estimates of the expected frequencies are obtained by sampling values from this posterior. This can be achieved iteratively for each \mathbf{x} by fixing all expected frequencies except $F(\mathbf{X} = \mathbf{x})$ to the sampled values from the previous iteration, determining the bounds for $F(\mathbf{X} = \mathbf{x})$ and sampling its value under the appropriate constraints, and then repeating the procedure for each \mathbf{x} . Sampling from the appropriate posterior is problematic here, because the

probabilities $P(X = \mathbf{x})$ are not independent under a Dirichlet distribution and, so, the value of $F(X = \mathbf{x})$ cannot be sampled independently from the other frequencies. To circumvent this problem, each probability is first re-scaled to a Gamma distribution, which is truncated by the upper and lower bounds. Subsequently, a sampled value for $F(X = \mathbf{x})$ can be obtained from the appropriate posterior distribution (Laudy & Hoijtink, 2007; Ligtvoet & Vermunt, 2012). The resulting Gibbs sampling algorithm contains the following steps:

1. Assign to all frequencies initial values $\hat{F}_0(X = \mathbf{x})$ that conform to CMTP₂ and CMWI.
2. Compute at iteration t for the vector \mathbf{x} all the lower bounds and upper bounds imposed upon it by CMTP₂ and CMWI using the estimates of the other frequencies at the previous iteration. Sample a value $\hat{G}_t(X = \mathbf{x})$ from the truncated Gamma distribution,

$$\hat{G}_t(X = \mathbf{x}) \sim \text{Gamma}(F(X = \mathbf{x}) + 1, 1) \times I(\max(L) < \hat{G}_t(X = \mathbf{x}) < \min(U)),$$

where one observation is added to $F(X = \mathbf{x})$ as a prior, the scaling parameter is set equal to unity, and $I(\cdot)$ is the indicator function containing the restrictions. Sampling from the truncated Gamma distribution can be achieved by inverse probability sampling (e.g., Hoijtink & Molenaar, 1997). Return the frequency

$$\hat{F}_t(X = \mathbf{x}) = \frac{N\hat{G}_t(X = \mathbf{x})}{\hat{G}_t(X = \mathbf{x}) + \sum_{y \neq \mathbf{x}} \hat{F}_{t-1}(X = \mathbf{y})},$$

where N is the total sample size and the summation is taken across all previously sampled frequencies.

3. Repeat Step 2 for all \mathbf{x} .
4. Repeat Steps 2 and 3 until a preset criterion of convergence has been reached.

Once the criterion of convergence in Step 4 is reached, subsequent parameter values are sampled as if taken from the posterior distribution of interest. The next step is intended to use these samples to assess model-data fit for the isotonic PCM:

5. After Step 4, repeat Steps 2 and 3 until a sufficiently large sample of values for $\hat{F}(X = \mathbf{x})$ have been obtained.

3.2. Model-Data Fit

Consider two statistics for assessing the model-data fit: Pearson's X^2 and the difference in log-likelihood from the multinomial model. For Pearson's X^2 , the expected frequency of $X = \mathbf{x}$ under the constraints imposed by CMTP₂ and CMWI is estimated by the median value of the samples of $\hat{F}_t(X = \mathbf{x})$ from the Gibbs sampler (after convergence). Notice that the distributions of the two statistics for model-data fit under the constraints imposed by CMTP₂ and CMWI can be obtained by computing the statistics for the samples of $\hat{F}_t(X = \mathbf{x})$. The Bayesian p -values (Gelman, Meng, & Stern, 1996; Meng, 1994) for these statistics are then estimated as the proportion of times the statistic computed for the sampled values exceeds the value of the statistic for the observed score distribution.

3.3. Simulation Study

To assess the use of the Gibbs sampling algorithm to deal with the large number of inequalities and the sparse data, a small simulation study was performed. In this simulation study, the Type I error rate and the power were assessed of two model-data fit procedures that were based on the Bayesian p -values. For the first procedure, the p -values were computed for the Pearson's X^2 statistic. For the second procedure, the p -values were computed for the log-likelihood ratio

statistic. For each of these statistics, the Type I error rate corresponds to the proportion of times the p -value incorrectly indicated misfit of the model assumptions. The power of each statistic corresponds to the proportion of times the p -value correctly indicated misfit of the model assumptions.

3.3.1. Method The independent variables in this study were the number of items, the number of score categories, the sample size, and whether or not the isotonic PCM was valid.

To generate data under a flexible (nonparametric) model, the following procedure was used for each item. First, subjects' latent values were randomly sampled from a Uniform distribution between 1 and 5. Second, a $5 \times m_i$ design matrix \mathbf{D} was generated with random probabilities, where the rows represented $\theta = 1, 2, 3, 4,$ and 5 . These probabilities were subsequently sorted within each column of \mathbf{D} in increasing order. Third, for those data sets in the design for which the isotonic PCM did not hold, the probabilities for adjacent rows in \mathbf{D} were connected with a straight (nondecreasing) line to define a function to relate each subjects' latent score to the probability of attaining each of the m_i item scores. These probabilities were taken as cumulative probabilities; hence, the data sets in the design for which the isotonic PCM did not hold correspond to the nonparametric graded response model (e.g., Molenaar, 1997). This nonparametric graded response model neither implies the assumption LTP_2 nor the assumption WI. For those data sets in the design for which the isotonic PCM did hold, the probabilities in \mathbf{D} correspond to the nonparametric PCM. To facilitate assumption WI, these probabilities were sorted within each column of \mathbf{D} in decreasing order. Fourth, using the formulas in Van der Ark, Hemker, and Sijtsma (2002), the probabilities $P_x(a, i)$ could be obtained for both the nonparametric graded response model and the isotonic PCM. Finally, for each subject an item score vector was randomly sampled from a multinomial distribution using these probabilities.

Van der Ark (2005) found that the severity of the violation of MLR for the nonparametric graded response model depends both on the number of items and the number of score categories. Smaller violations occur for tests consisting of more items and less score categories. Because the severity of the violation of MLR is larger for tests consisting of a few items, we only considered tests of 2, 3, and 4 items. Also, for tests of more than 4 items, the analysis of a single data set took more than 10 minutes, which is no problem when analyzing a single data set, but it became infeasible for the number of replications used in this simulation. For the number of score categories, we fixed all $m_i = m$ and considered three levels: $m + 1 = 3, 5,$ and 7 for a two-item test. For tests consisting of 3 and 4 items, three score categories were considered only. For these five combinations of number of items and number of score categories, sample sizes of $N = 200, 400,$ and 800 were considered for both the nonparametric graded response model and the isotonic PCM, where for each of the resulting 30 cells of the design 50 data sets were generated.

The dependent variables were the Type I error rate and the power of the model-data fit procedures based on the Bayesian p -values for Pearson's X^2 statistic and the log-likelihood ratio statistic. As a decision rule a p -value smaller than 0.10 was taken to indicate misfit of the model assumptions. All 1500 data sets were generated and analyzed in R (R Development Core Team, 2009). The codes provided by Nadarajah and Kotz (2006) were used for sampling from the truncated Gamma distributions. Tests of the Gibbs sampler indicated that for the data sets used in this study a burn-in period of 100 iterations was sufficient for convergence to be reached. This small number might be explained by the large number of constraints that restrict the outcome space. The estimations were based on 400 samples from the posterior after this burn-in period. The R-code for running the Gibbs sampler, which also works for varying numbers of score categories, can be obtained from the author's web site: <http://home.medewerker.uva.nl/r.ligtvoet>.

3.3.2. Results The results are discussed first for the model-data fit procedure based on Pearson's X^2 statistic. Table 1 contains the Type I error rates and power for the different numbers

TABLE 1.

Type I error rate and power based on Pearson's X^2 statistic, for different numbers of items, numbers of score categories, and sample sizes.

Number of categories	Sample size	Number of items					
		2		3		4	
		Type I	Power	Type I	Power	Type I	Power
3	200	0.10	0.92	0.74	1.00	1.00	1.00
	400	0.18	0.96	0.54	1.00	1.00	1.00
	800	0.26	0.96	0.38	1.00	1.00	1.00
5	200	0.80	1.00				
	400	0.86	1.00				
	800	0.74	1.00				
7	200	1.00	1.00				
	400	1.00	1.00				
	800	0.98	1.00				

Cells not part of the design

Note. Proportions are based on 50 replications each.

TABLE 2.

Type I error rate and power based on log-likelihood statistic, for different numbers of items, numbers of score categories, and sample sizes.

Number of categories	Sample size	Number of items					
		2		3		4	
		Type I	Power	Type I	Power	Type I	Power
3	200	0.02	0.62	0.00	0.74	0.08	0.74
	400	0.00	0.78	0.00	0.84	0.00	0.76
	800	0.00	0.80	0.02	0.86	0.00	0.96
5	200	0.20	0.98				
	400	0.04	0.98				
	800	0.08	1.00				
7	200	0.82	0.98				
	400	0.50	1.00				
	800	0.22	1.00				

Cells not part of the design

Note. Proportions are based on 50 replications each.

of items, numbers of score categories, and sample sizes. Table 1 shows that, with the exception of the cells of the design corresponding to 2 items and 3 score categories, the Bayesian p -values for Pearson's X^2 statistic indicate model-data misfit for almost all the data sets. Inspection of the distribution of these p -values across the cells of the design showed a U-shaped distribution, with most p -values close to zero. These results indicate that the Bayesian p -value for Pearson's X^2 statistic is not a good indicator for model-data fit.

The results of the model-data fit procedure based on the log-likelihood statistic are shown in Table 2. Table 2 shows the Type I error rates and power for the different numbers of items, numbers of score categories, and sample sizes. Across the cells of the design, the power was moderate to high (range between 0.62 and 1.00), suggesting that the statistic is sensitive to model violations. In addition, the Type I error rate was small, except for the cells of the design corresponding to 7 score categories. However, for the cells of the design with 7 score categories the Type I error rate decreases for larger sample sizes, which suggest that a sample size of more than 800 subjects is required for data containing 7 score categories. These results suggest that the Bayesian p -value

for the log-likelihood statistic is a good indicator for model-data fit, given that the sample size is large enough.

3.3.3. Conclusion In the simulation study, the Type I error rate and the power were assessed of two model-data fit procedures that were based on the Bayesian p -values. These p -values were computed for the Pearson's X^2 statistic and the log-likelihood ratio statistic. These result indicate that the Bayesian p -value for Pearson's X^2 statistic is not a good indicator for model-data fit, whereas the Bayesian p -value for the log-likelihood statistic did have a high power and a low Type I error rates for fewer than 7 score categories.

4. Discussion

In this paper, sufficient conditions for the stochastic ordering of subjects by their sum score were obtained. These conditions define the isotonic PCM model, which like the PCM justifies the use of the sum score as a basis for ordering subjects. However, the isotonic PCM is more flexible than the PCM in terms of the restrictions it imposes on the test data, which makes it applicable for a wider variety of tests. Also, observable properties of the isotonic PCM were derived in the form of inequality constraints. It was shown how to obtain estimates of the score distribution under these constraints by using the Gibbs sampling algorithm. In a small simulation study, it was shown that the Bayesian p -values based on the log-likelihood ratio statistic can be used to assess the fit of the isotonic PCM to the data.

That the isotonic PCM is more flexible than the PCM does not mean that the isotonic PCM will almost surely hold for a given data set. In fact, the isotonic PCM imposes a great many restrictions onto the data, which illustrate just how restrictive the requirement for using the sum score are. Misfit of the isotonic PCM means that the practical use of the sum score for ordering subjects lacks empirical justification. If the isotonic PCM fits the data, then there is at least some empirical justification for such use of the sum score.

Ligtvoet, Van der Ark, Bergsma, and Sijtsma (2011) studied sufficient conditions for an ordering of the items that is invariant across subjects. For polytomously scored items, they defined several latent scales which yield observable consequences in terms of the restrictions imposed on the observed score distribution. One possible extension of the isotonic PCM is to include such an item ordering, which can be easily tested by including additional inequality restrictions in the Gibbs sampling algorithm.

Acknowledgements

I would like to thank the anonymous reviewers for their careful reading and the suggestions regarding the outline of the proof and the simulation study. Also, I thank Annemarie Zand-Scholten for commenting on an earlier draft of the manuscript.

References

- Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.
- Andersen, E.B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, 38, 123–140.
- Douglas, R., Fienberg, S.E., Lee, M.-L.T., Sampson, A.R., & Whitaker, L.R. (1991). Positive dependence concepts for ordinal contingency tables. In H.W. Block, A.R. Sampson, & T.H. Savits (Eds.), *Topics in statistical dependence* (pp. 189–202). Hayward: Institute of Mathematical Statistics.
- Gelman, A., Meng, X.L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6, 733–807.
- Grayson, D.A. (1988). Two-group classification in latent trait theory: Scores with monotone likelihood ratio. *Psychometrika*, 53, 383–392.

- Hemker, B.T. (1996). *Unidimensional IRT models for polytomous items, with results for Mokken scale analysis*. Unpublished doctoral dissertation, Utrecht University, The Netherlands.
- Hemker, B.T., Sijtsma, K., Molenaar, I.W., & Junker, B.W. (1996). Polytomous IRT models and monotone likelihood ratio of the total score. *Psychometrika*, *61*, 679–693.
- Hemker, B.T., Sijtsma, K., Molenaar, I.W., & Junker, B.W. (1997). Stochastic ordering using the latent trait and the sum score in polytomous IRT models. *Psychometrika*, *62*, 331–347.
- Hojtink, H. (1998). Constrained latent class analysis using the Gibbs sampler and posterior predictive p -values: Applications to educational testing. *Statistica Sinica*, *8*, 691–711.
- Hojtink, H., & Molenaar, I.W. (1997). A multidimensional item response model: Constrained latent class analysis using Gibbs sampler and posterior predictive checks. *Psychometrika*, *62*, 171–189.
- Holland, P.W., & Rosenbaum, P.R. (1986). Conditional association and unidimensionality in monotone latent variable models. *Annals of Statistics*, *14*, 1523–1543.
- Huynh, H. (1994). A new proof for monotone likelihood ratio for the sum of independent Bernoulli random variables. *Psychometrika*, *59*, 77–79.
- Junker, B.W. (1993). Conditional association, essential independence and monotone unidimensional item response models. *Annals of Statistics*, *21*, 1359–1378.
- Karabatsos, G. (2001). The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. *Journal of Applied Measurement*, *2*, 389–423.
- Karabatsos, G., & Sheu, C.-F. (2004). Order-constrained Bayes inference for dichotomous models of unidimensional nonparametric IRT. *Applied Psychological Measurement*, *28*, 110–125.
- Karlin, S. (1968). *Total positivity*. Palo Alto: Stanford University Press.
- Karlin, S., & Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *Journal of Multivariate Analysis*, *10*, 467–498.
- Laudy, O., & Hoijtink, H. (2007). Bayesian methods for the analysis of inequality constrained contingency tables. *Statistical Methods in Medical Research*, *16*, 123–138.
- Lazarsfeld, P.F., & Henry, N.W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Lehmann, E.L. (1959). *Testing statistical hypotheses*. New York: Wiley.
- Ligtvoet, R., Van der Ark, L.A., Bergsma, W.P., & Sijtsma, K. (2011). Polytomous latent scales for the investigation of the ordering of items. *Psychometrika*, *73*, 200–216.
- Ligtvoet, R., & Vermunt, J.K. (2012). Latent class models for testing monotonicity and invariant item ordering for polytomous items. *British Journal of Mathematical & Statistical Psychology*, *65*, 237–250. doi:10.1111/j.2044-8317.2011.02019.x.
- Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149–174.
- Meng, X.L. (1994). Posterior predictive p -values. *Annals of Statistics*, *22*, 1142–1160.
- Mokken, R.J. (1971). *A theory and procedure of scale analysis*. Berlin: De Gruyter.
- Molenaar, I.W. (1997). Nonparametric models for polytomous responses. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 369–380). New York: Springer.
- Molenaar, I.W., & Sijtsma, K. (2000). *User's manual MSP5 for Windows* [Software manual]. Groningen, The Netherlands: IEC ProGAMMA.
- Muraki, E. (1992). A generalized partial credit model: Applications for an EM algorithm. *Applied Psychological Measurement*, *16*, 159–177.
- Nadarajah, S., & Kotz, S. (2006). R Programs for computing truncated distributions. *Journal of Statistical Software*, *16*(2), 1–8.
- R Development Core Team (2009). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org>.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielsen and Lydiche.
- Samejima, F. (1969). Estimation of latent trait ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17.
- Samejima, F. (1972). A general model for free-response data. *Psychometrika Monograph*, No. 18.
- Scheiblechner, H. (1995). Isotonic ordinal probabilistic models (ISOP). *Psychometrika*, *60*, 281–304.
- Scheiblechner, H. (1999). Additive conjoint isotonic probabilistic models (ADISOP). *Psychometrika*, *64*, 295–316.
- Scheiblechner, H. (2007). A unified nonparametric IRT model for d -dimensional psychological test data (d -ISOP). *Psychometrika*, *72*, 43–67.
- Ünlü, A. (2008). A note on monotone likelihood ratio of the total score variable in unidimensional item response theory. *British Journal of Mathematical & Statistical Psychology*, *61*, 179–187.
- Van der Ark, L.A. (2005). Stochastic ordering of the latent trait by the sum score under various polytomous IRT models. *Psychometrika*, *70*, 283–304.
- Van der Ark, L.A., & Bergsma, W.P. (2010). A note on stochastic ordering of the latent trait using the sum of polytomous item scores. *Psychometrika*, *75*, 272–279.
- Van der Ark, L.A., Hemker, B.T., & Sijtsma, K. (2002). Hierarchically related nonparametric IRT models, and practical data analysis methods. In G.A. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models* (pp. 41–62). Mahwah: Erlbaum.
- Van Onna, H.J.M. (2002). Bayesian estimation and model selection in ordered latent class models for polytomous items. *Psychometrika*, *67*, 519–538.

Manuscript Received: 8 APR 2011

Final Version Received: 1 DEC 2011

Published Online Date: 17 MAY 2012