Nonparametric item response theory axioms and properties under nonlinearity and their exemplification with knowledge space theory

Ali Ünlü

Ali Ünlü

**Nonparametric Item Response Theory Axioms and Properties Under
Nonlinearity and their Exemplification with Knowledge Space Theory**

NONPARAMETRIC ITEM RESPONSE THEORY AXIOMS AND PROPERTIES UNDER NONLINEARITY AND THEIR EXEMPLIFICATION WITH KNOWLEDGE SPACE THEORY

ALI ÜNLÜ

DEPARTMENT OF MATHEMATICS

UNIVERSITY OF AUGSBURG

UNIVERSITÄTSSTRASSE 14

D-86159 AUGSBURG, GERMANY

July 17, 2007

Correspondence should be sent to

E-mail: ali.uenlue@math.uni-augsburg.de (Ali Ünlü)

Phone: +49 821 598 2236

Fax: +49 821 598 2200

Website: http://www.math.uni-augsburg.de/~uenlueal/

NONPARAMETRIC ITEM RESPONSE THEORY AXIOMS AND PROPERTIES
UNDER NONLINEARITY AND THEIR EXEMPLIFICATION WITH KNOWLEDGE
SPACE THEORY

Abstract

This paper investigates the dichotomous Mokken nonparametric item response theory (IRT) axioms and properties under incomparabilities among latent trait values and items. Generalized equivalents of the unidimensional nonparametric IRT axioms and properties are formulated for nonlinear (quasi-ordered) person and indicator spaces. It is shown that monotone likelihood ratio (MLR) for the total score variable and nonlinear latent trait implies stochastic ordering (SO) of the total score variable, but may fail to imply SO of the nonlinear latent trait. The reason for this and conditions under which the implication holds are specified, based on a new, simpler proof of the fact that in the unidimensional case MLR implies SO. The approach is applied in knowledge space theory (KST), a combinatorial test theory. This leads to a (tentative) Mokken-type nonparametric axiomatization in the currently parametric theory of knowledge spaces. The nonparametric axiomatization is compared with the assumptions of the parametric basic local independence model which is fundamental in KST. It is concluded that this paper may provide a first step toward a basis for a possible fusion of the two split directions of psychological test theories IRT and KST.

Key words: nonparametric item response theory, quasi-ordered person space, quasi-ordered indicator space, invariant item ordering, monotone likelihood ratio, stochastic ordering, generalized

monotone-likelihood-ratio-stochastic-ordering implication, parametric knowledge space theory, basic local independence model, nonparametric knowledge space theory.

Contents

1. Introduction

1.1. Preliminaries

More than 50 years ago, Louis Guttman introduced his scalogram technique (Guttman, 1944, 1950). This deterministic model allows for 'linear' orderings ('chains') of persons (e.g., regarding their abilities) and items (e.g., regarding their difficulties). Since then, the Guttman model has been generalized into, *at least*, two directions.[1] On the one hand (in a probabilistic, statistical direction), based on the Rasch (1960) model and generalized by Mokken's (1971) monotone homogeneity model, a family of 'linear' probabilistic models (*item response theory*, IRT; e.g., Fischer & Molenaar, 1995; Van der Linden & Hambleton, 1997) has emerged, taking over 'linearity' of person and item orderings. On the other hand (in a deterministic, order-theoretic direction), starting with Airasian and Bart (1973), Bart and Airasian (1974), and Bart and Krus (1973), a family of 'nonlinear' deterministic models (*knowledge space theory*, KST; e.g., Doignon & Falmagne, 1985, 1999) has been developed, weakening 'linearity' of person and item orderings to allow for incomparabilities among persons and items, respectively.[2]

---

[1]In the following, I mention item response theory and knowledge space theory. By doing so, I do not claim that the latter theory is as popular as the former one, nor that it is as popular as other theories such as classical test theory or generalizability theory.

[2]In KST, persons are represented by collections of items (of a domain) they are capable of mastering. In this spirit, persons can be incomparable, with respect to set-inclusion. Items, in turn, are assumed to be ordered, for instance, with respect to a hierarchy of mastery dependencies. In this spirit, items can be incomparable, with respect to that hierarchy. Examples illustrating what it means when persons and

Although there exist scattered approaches to creating specific probabilistic versions of the latter 'nonlinear' deterministic models (e.g., Doignon & Falmagne, 1999; Ünlü, 2006), respectively, 'nonlinear' generalizations of the former 'linear' probabilistic models (e.g., Junker, 2001; Van der Linden & Hambleton, 1997), such activities have not arisen from the idea of conflating these two split directions of psychological test theories (however, see Stefanutti, 2006).

The main motivation for the research reported in this paper is to provide a *first step toward a basis* for a possible fusion of these theories. In IRT literature, the Mokken (1971) unidimensional nonparametric models have been satisfactorily applied. Based on such applications, as an important first step, this paper investigates the scope of the unidimensional Mokken axioms in a nonlinear framework allowing for incomparabilities among latent trait values and items, and introduces a first application of nonparametric modeling in the theory of knowledge spaces. This should not, however, be misconceived as claiming ability to achieve that basis already with the present paper. The latter definitely requires future research, as a necessary prerequisite for the development of a superior probabilistic test theory (with corresponding statistical inference methodology) that could include (most of) the existing models as special cases.

1.2. Nonparametric Approach

This paper pursues a nonparametric, as opposed to a parametric, (probabilistic) approach. A nonparametric perspective may, for instance, offer the following

items are incomparable are given in Subsections 3.1 and 5.2. In IRT, on the other hand, persons and items are, for instance, represented by single real numbers, ability and difficulty parameters, respectively. In this spirit, persons and items are linearly ordered, with respect to the natural ordering of the real numbers. Conceptually speaking, KST may be viewed as a more 'qualitative, behavioral' approach, unlike IRT, as a 'quantitative, statistical' approach. In the following, I will talk about IRT and KST without having introduced the corresponding models technically. This will be done in Sections 2 and 3, respectively. Reading these sections, it should be clear what the corresponding models, their similarities and dissimilarities are about. Also, Subsection 1.3 discusses what the KST models can do that the IRT models cannot do, and vice versa.

advantages:[3]

1. So far, a nonparametric approach in KST is lacking. Because KST at present is merely formulated parametrically—see Subsection 3.3; the parametric (probabilistic) *basic local independence model*, reviewed in Definition 4, is fundamental in KST, in the sense that most of the KST probabilistic models are special cases of it—, a nonparametric approach in KST would be an important contribution.

2. In realistic contexts, the parametric models in KST generally contain too many model parameters (e.g., a prohibitively large number of knowledge state probabilities) to be estimated from the data (cf. Subsection 3.3). For instance, in an experiment by Kambouri (1991), the number of knowledge states ranges from several hundreds to several thousands (for 50 items). In such cases, without any restrictions, it may be infeasible to obtain reliable estimates of the model parameters. On the other hand, setting constraints to reduce the number of independent model parameters necessitates the justification of such restrictions. In turn, assumptions not justified by the data can lead to biases in the estimates. Finally, justified restrictions may require data from a special, costly experimental design. Given these observations, a nonparametric approach in KST may help in developing a nonparametric axiomatic framework, with corresponding feasible new statistical inference methodologies for parameter estimation and (separately) testing KST model assumptions.

3. Nonparametric IRT includes a broad range of parametric IRT models. Treating the more general nonparametric cases thus (implicitly) implies the treatment of various parametric special cases. To some degree, this also applies to the nonparametric KST axioms introduced in this paper. For instance, for realistic ranges of parameter values, the basic local independence model satisfies the axiom of isotonicity (see Theorem 6).

Nevertheless, parametric IRT-type approaches in KST, of course, are important directions for future research (for a recent logistic approach, see Stefanutti, 2006; for a

---

[3]Basic deterministic and probabilistic concepts of KST are reviewed in Section 3.

sort of generalized normal ogive approach, see Ünlü, 2006).

This paper is also interesting from an IRT point of view. It provides new structural insights into the long-standing study of monotone likelihood ratio and stochastic ordering in IRT (e.g., Hemker, Sijtsma, Molenaar, & Junker, 1996, 1997; Hemker, Van der Ark, Sijtsma, 2001; Sijtsma, 1998; Van der Ark, 2001, 2005). So far, IRT research has only concentrated on the unidimensional case. In this paper, at an order-theoretic structural level, monotone likelihood ratio and stochastic ordering are discussed for more general *nonlinear* (quasi-ordered) latent trait spaces. I show that the generalized version of the fundamental Ghurye-Wallace-Grayson-Huynh result (Ghurye & Wallace, 1959; Grayson, 1988; Huynh, 1994) on monotone likelihood ratio of the total score variable in the nonlinear latent trait does *not* in general imply stochastic ordering of the nonlinear latent trait by the total score variable. Even more (see Table 1), if the latent trait space is not linearly ordered (in particular, if it is not unidimensional), and if the manifest variable assumes values in a linearly ordered set (e.g., if it is scalar-valued such as the total score variable), monotone likelihood ratio implies stochastic ordering of the manifest variable, but may fail to imply stochastic ordering of the latent trait (though the property of monotone likelihood ratio is still symmetric in its arguments). I quantify the reason for this fact (order-theoretic completeness assumption)[4] and give conditions under which the implication holds.

1.3. Arguments for Fusion of IRT and KST

[5] What could be gained by fusing (nonparametric) IRT and KST? In other words, why should one be interested in trying to unify IRT and KST? Generally speaking, what one ideally would like to have is a unified framework keeping the strengths and at the same time avoiding the drawbacks of both theories. But what can KST do for a test constructor and a test user that IRT cannot deliver, and vice versa? The following lists

[4]The assumption of completeness is defined in Footnote 14.

[5]This subsection was motivated by critical comments made by a reviewer. For selected topics such as surmise systems and adaptive testing in KST, see the references listed in Section 3.

some arguments supporting the importance of a possible fusion of IRT and KST.

Statistical inference methodologies. An IRT-type modeling in KST could provide feasible new statistical inference methodologies (cf. Subsection 1.2). For IRT, unlike KST, has plenty of sophisticated statistical methods that could be suited to and applied in KST. For instance, the random effects approach discussed in Ünlü (2006) allows for the estimation of the response error probabilities of the basic local independence model, without having to estimate them simultaneously along with the knowledge state probabilities (cf. Subsection 3.3). The latter, however, is crucial, because the number of knowledge states generally tends to be quite large in real data applications. This random effects approach represents a straightforward, nonetheless important, application of (parametric) IRT-type modeling with corresponding statistical inference methodology in KST.

Restrictivity. It seems that IRT models that simultaneously imply a person ordering and an item ordering (e.g., *double monotonicity model*, Subsection 2.1) are restrictive models with respect to real data. In general, they will not fit many empirical data sets. A unified test theory combining IRT and KST could positively contribute to and improve on this observation. For, a strength of KST is that it implies very general combinatorial structures, both at the levels of persons and items, contrary to IRT, implying more restrictive linear orderings. KST further provides mathematical theorems on the linkage between these levels, offering flexibility in the choice of a representation. Hence a unified approach could deliver as general as possible probabilistic models that could (a) imply both a person ordering and an item ordering, (b) extend linear orderings to more general and flexible surmise relations or even surmise systems, (c) allow for flexibility in representation by supplying possible theorems on the linkage between the person and item levels, (d) encompass (most of) the existing IRT and KST models as special cases, and thus (e) fit far more data sets in practice.

Adaptive testing. A unified test theory combining IRT and KST could also positively

contribute to the problem of adaptive testing in *nonparametric* IRT using only ordinal measurement information. Developments in this direction have been rather moderate (e.g., Huisman & Molenaar, 2001; Laros & Tellegen, 1991). Adaptive testing, however, is a major strength of (parametric) KST.

Qualitative derivation of hierarchies among items. KST provides a number of 'a priori' qualitative, psychological theory driven methods for the derivation of hierarchies among items (e.g., based on mastery dependencies, Subsection 3.1). Such methods are presented, for example, in the book edited by Albert and Lukas (1999); for instance, Albert and Held (1999), Held (1999), and Korossy (1999). These methods utilize psychological theories and principles based on content– and process–analysis for the qualitative derivation of hierarchies among items. In IRT, however, orderings of items are obtained 'a posteriori' by utilizing quantitative, statistical methods (e.g., by estimating the difficulty parameter of each item). A unified framework could offer (a) qualitative, theory driven (KST), (b) quantitative, statistical (IRT), and (c) hybrid derivation methods.

Software. For a unified test theory combining IRT and KST a unified software environment could be developed. Such a comprehensive environment could encompass many existing IRT and KST models and software, under a single graphical user interface (GUI) with common look and feel. This would imply easy access to and use of software for the practical application of the unified test theory, IRT, and KST models to empirical data, user-friendly unified 'under one umbrella'.

1.4. Structure of Paper

This paper is structured as follows. The unidimensional nonparametric IRT axioms and properties are reviewed (Section 2). In particular, a new, simpler proof of the fact that in the unidimensional case monotone likelihood ratio implies stochastic ordering is presented. Basic deterministic and probabilistic KST concepts are reviewed (Section 3).

Generalized equivalents of the unidimensional nonparametric IRT axioms and properties for nonlinear person and indicator spaces are discussed (Section 4). In particular, based on the new proof, (a) it is seen that the generalized Ghurye-Wallace-Grayson-Huynh result may not imply stochastic ordering of the nonlinear latent trait, and (b) the reason for this fact and conditions under which the implication holds are specified. An application of the nonlinear nonparametrics in KST is presented (Section 5). KST equivalents of the generalized nonparametric IRT concepts are formulated for the knowledge structure, surmise relation, and quasi-ordinal knowledge space models. The nonparametric formulation is compared with the assumptions of the parametric basic local independence model. This paper ends with a discussion containing a summary, concluding remarks, and some suggestions for further research (Section 6).

2. Nonparametric IRT: Axioms and Properties

This section briefly reviews the axioms of unidimensionality, local independence, monotonicity, and invariant item ordering underlying Mokken's (1971) nonparametric IRT models of monotone homogeneity and double monotonicity for dichotomous items (see also Mokken, 1997; Mokken & Lewis, 1982; Sijtsma, 1998; Sijtsma & Molenaar, 2002).[6] The properties of monotone likelihood ratio and stochastic ordering justifying the use of Mokken's models as measurement models for persons are also reviewed.

2.1. Axioms

Let X_l with realization $x_l \in \{0, 1\}$ denote the *item score variable* for item I_l $(1 \leq l \leq m,\ m \in \mathbb{N}_{\geq 2})$,[7] and let $X_+ := \sum_{l=1}^{m} X_l$ with realization $x_+ \in \{0, 1, \ldots, m\}$ be the *total score variable.* A function $f : \{0, 1, \ldots, m\} \to \mathbb{R}$ is *nondecreasing* if and only if (iff)

$$\forall x, y \in \{0, 1, \ldots, m\}, x \leq y : f(x) \leq f(y).$$

[6]Throughout this paper, only dichotomous items are considered.

[7]Let $\mathbb{N} := \{1, 2, 3, \ldots\}$, and let $\mathbb{N}_{\geq 2} := \{n \in \mathbb{N} : n \geq 2\}$.

Let the *latent trait* be denoted by θ, $\theta \in \Theta \subseteq \mathbb{R}$; this is referred to as the axiom of *unidimensionality*. The latent trait θ is assumed to have a sampling distribution in the examinee population under reference. A function $f : \Theta \to \mathbb{R}$ is *nondecreasing* iff

$$\forall \theta_1, \theta_2 \in \Theta, \theta_1 \leq \theta_2 : f(\theta_1) \leq f(\theta_2).$$

Let the conditional positive response probability $P(X_l = 1|\theta)$ as a function of $\theta \in \Theta$ be the *item response function* (IRF) of an item I_l ($1 \leq l \leq m$). The axiom of *local independence* states that

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m|\theta) = \prod_{l=1}^{m} P(X_l = x_l|\theta)$$

for any $x_l \in \{0, 1\}$ ($1 \leq l \leq m$) and $\theta \in \Theta$. The axiom of *monotonicity* holds iff any IRF $P(X_l = 1|.)$ ($1 \leq l \leq m$) is nondecreasing. The axiom of *invariant item ordering* states that the IRFs $P(X_l = 1|.)$ ($1 \leq l \leq m$) can be ordered such that

$$\forall \theta \in \Theta : P(X_{l_1} = 1|\theta) \leq P(X_{l_2} = 1|\theta) \leq \cdots \leq P(X_{l_m} = 1|\theta)$$

where $1 \leq l_i \leq m$ ($1 \leq i \leq m$). (For applications and methods of investigation of the axiom of invariant item ordering, see, for the dichotomous case, Sijtsma and Junker (1996), and for the polytomous case, Sijtsma and Hemker (1998).)

Mokken's *monotone homogeneity model* (MHM) is based on the axioms of unidimensionality, local independence, and monotonicity. His *double monotonicity model* (DMM) further adds the axiom of invariant item ordering.

2.2. Formulation of Properties

Monotone likelihood ratio for the total score variable and latent trait plays an important role in IRT. It implies stochastic ordering (SO) properties that can be conveniently interpreted in an IRT context (see Subsection 2.4): Stochastic ordering of the total score variable by the latent trait, and stochastic ordering of the latent trait by the total score variable.

The total score variable X_+ has *monotone likelihood ratio* (MLR) in θ iff, for any $0 \le x_{+,1} \le x_{+,2} \le m,$[8]

$$\frac{P(X_+ = x_{+,2}|\theta)}{P(X_+ = x_{+,1}|\theta)}$$

is a nondecreasing function of (unidimensional) $\theta \in \Theta$. Similarly, the latent trait θ has *MLR* in X_+ iff, for any $\theta_1, \theta_2 \in \Theta$, $\theta_1 \le \theta_2,$

$$\frac{P(\theta_2|X_+ = x_+)}{P(\theta_1|X_+ = x_+)}$$

is a nondecreasing function of $0 \le x_+ \le m$. The fundamental Ghurye-Wallace-Grayson-Huynh result states that under mild conditions the total score variable has MLR in the unidimensional latent trait (see Theorem 1).

The property of MLR is rather technical and implies the following two important stochastic ordering properties that are easier to interpret in IRT. The property of MLR implies that X_+ is stochastically ordered by θ (see Part 1 of Theorem 2). The *stochastic ordering of the manifest variable X_+ by θ* (SOM) means that, for any $0 \le x_+ \le m$,

$$P(X_+ \ge x_+|\theta)$$

is a nondecreasing function of (unidimensional) $\theta \in \Theta$. Note that this property takes the latent trait as a starting point. In practice, however, the total score variable is observed and inferences about the latent trait are required. This is accommodated by the next property. The MLR property also implies that θ is stochastically ordered by X_+ (see Part 2 of Theorem 2). The *stochastic ordering of the latent trait θ by X_+* (SOL) means that, for any $\theta_0 \in \Theta$,

$$P(\theta \ge \theta_0|X_+ = x_+)$$

is a nondecreasing function of $0 \le x_+ \le m$. Note that the property of SOL is very important for practical measurement, because it justifies the use of the total score variable to estimate the ordering of subjects on the latent trait. This is the key result that justifies the use of the MHM and DMM as measurement models for persons.

---

[8]Throughout this paper, mathematical expressions are assumed to be defined whenever they are written in the text.

2.3. Ghurye-Wallace-Grayson-Huynh Result

The fundamental *Ghurye-Wallace-Grayson-Huynh* (GWGH) *result* on monotone likelihood ratio of the total score variable in unidimensional IRT (for dichotomous items) is as follows (Ghurye & Wallace, 1959; Grayson, 1988; Huynh, 1994).

Theorem 1. (GWGH result) Under the axioms of unidimensionality, local independence, and monotonicity, and the requirement that each item response function assumes values strictly between zero and one, the total score variable has monotone likelihood ratio in the (unidimensional) latent trait.

Proof. For a recent proof of this result correcting flaws in Huynh's (1994) argument, see Ünlü (2007a). □

Note that the requirement of having IRFs assuming values strictly between zero and one is not that restrictive in practice, because 'boundary value' IRFs (assuming the values 0 and/or 1) may be closely approximated by IRFs that do meet this requirement. Also note that the MHM and hence parametric special cases such as the Rasch (1960) and Birnbaum (1968) models possess the MLR property. Mokken's models, by definition, satisfy the axioms of unidimensionality, local independence, and monotonicity.

2.4. Monotone-Likelihood-Ratio-Stochastic-Ordering Implication

The MLR property is *symmetric* in its arguments, that is, MLR of X_+ in θ is equivalent to MLR of θ in X_+ (Bayes' theorem). Another well-known fact is that the property of MLR implies the SOM and SOL properties (e.g., Hemker, Sijtsma, Molenaar, & Junker, 1996, 1997; Hemker, Van der Ark, Sijtsma, 2001; Sijtsma, 1998; Van der Ark, 2001, 2005).

This fact has been established in the statistical literature by means of various mathematical proofs. However, all these proofs have one thing in common. They are

indirect, in the sense that this (in fact, elementary) result is obtained as a corollary of unnecessarily complicated, general statements.

The proof by Lehmann (1986, pp. 85–86, Lemma 2) establishes stochastic ordering (Part (ii) of Lemma 2) as a special result of a more general implication of MLR (Part (i) of Lemma 2): invariance of the property of nondecreasing monotonicity for (indexed) expectations of nondecreasing monotone transformations. Lehmann's lemma is a special case of an even more general theorem by Karlin (1957, 1968). Karlin's theorem relates the number of sign changes of the (indexed) expectation of a nondecreasing transformation to those of the transformation when the involved (indexed) densities are totally positive. The proof by Junker (1993, pp. 1371–1372, Lemma 4.1) unnecessarily introduces a 'conditional' random variable to rely on Esary, Proschan, and Walkup (1967, Property (P3)). In Shaked and Shanthikumar (1994, pp. 28–29, Subsection 1.C.3; see also Ross, 1996), the likelihood ratio order $\leq_{\mathrm{lr}}$ implying the usual stochastic order $\leq_{\mathrm{st}}$ is established via the hazard (respectively, reversed hazard) rate order $\leq_{\mathrm{hr}}$ (respectively, $\leq_{\mathrm{rh}}$): Theorem 1.C.1 states that $\leq_{\mathrm{lr}} \subseteq \leq_{\mathrm{hr}} \bigcap \leq_{\mathrm{rh}}$, and Theorem 1.B.1 (respectively, Theorem 1.B.21) states that $\leq_{\mathrm{hr}} \subseteq \leq_{\mathrm{st}}$ (respectively, $\leq_{\mathrm{rh}} \subseteq \leq_{\mathrm{st}}$). The MLR property implying the stochastic ordering properties can also be obtained from a general theorem by Douglas, Fienberg, Lee, Sampson, and Whitaker (1990, Theorem 3.1), and from the characterizations of stochastic ordering in terms of a maximal invariant with respect to the group of monotone transformations by Lehmann and Rojo (1992).

In the following, a new, simpler proof of this basic (in IRT, important) fact is presented. The proof is direct and elementary based on few purely algebraic, straightforward manipulations. It is self-contained requiring only basic knowledge of probability theory. It seems that this proof is the first of such a kind (direct, elementary, and self-contained) in the statistical literature on this matter. More precisely, the proof follows the line of reasoning by Junker (1993). First, simple sum (respectively, integral) representations are given for the left- and right-hand sides of the required stochastic ordering inequality. Then, however, instead of adequately expressing

the representations further based on a 'conditional' random variable (e.g., to rely on Esary et al., 1967), appropriate terms are simply summarized by means of straightforward algebraic manipulations.

Theorem 2. (MLR-SO implication) It holds:

1. MLR of (unidimensional) θ in X_+ implies the SOM property.

2. MLR of X_+ in (unidimensional) θ implies the SOL property.

Proof. 1. Let $0 < x_+ \leq m$, and $\theta_1, \theta_2 \in \Theta$, $\theta_1 \leq \theta_2$. It holds (cf. Junker, 1993, pp. 1371–1372, Lemma 4.1),

$$P(X_+ \geq x_+|\theta_1) = \frac{\sum_{y=x_+}^m P(\theta_1|X_+ = y)P(X_+ = y)}{\sum_{y'=0}^m P(\theta_1|X_+ = y')P(X_+ = y')},$$

$$P(X_+ \geq x_+|\theta_2) = \frac{\sum_{y''=x_+}^m P(\theta_1|X_+ = y'')\frac{P(\theta_2|X_+=y'')}{P(\theta_1|X_+=y'')}P(X_+ = y'')}{\sum_{y'''=0}^m P(\theta_1|X_+ = y''')\frac{P(\theta_2|X_+=y''')}{P(\theta_1|X_+=y''')}P(X_+ = y''')}.$$

Separating sums '$\sum_{y'=0}^m \cdots$' and '$\sum_{y'''=0}^m \cdots$' into two parts '$\sum_{y'=0}^{x_+-1} \cdots + \sum_{y'=x_+}^m \cdots$' and '$\sum_{y'''=0}^{x_+-1} \cdots + \sum_{y'''=x_+}^m \cdots$', respectively, and summarizing appropriate terms, the required inequality $P(X_+ \geq x_+|\theta_1) \leq P(X_+ \geq x_+|\theta_2)$ is seen to be equivalent to

$$\sum_{y=x_+}^m P(\theta_1|X_+ = y)P(X_+ = y) \cdot \sum_{y'''=0}^{x_+-1} P(\theta_1|X_+ = y''')\frac{P(\theta_2|X_+ = y''')}{P(\theta_1|X_+ = y''')}P(X_+ = y''')$$

$$\leq \sum_{y''=x_+}^m P(\theta_1|X_+ = y'')\frac{P(\theta_2|X_+ = y'')}{P(\theta_1|X_+ = y'')}P(X_+ = y'') \cdot \sum_{y'=0}^{x_+-1} P(\theta_1|X_+ = y')P(X_+ = y').$$

The latter inequality holds because ((1) and (2) indicate the use of MLR)

$$\sum_{y=x_+}^m P(\theta_1|X_+ = y)P(X_+ = y) \cdot \sum_{y'''=0}^{x_+-1} P(\theta_1|X_+ = y''')\frac{P(\theta_2|X_+ = y''')}{P(\theta_1|X_+ = y''')}P(X_+ = y''')$$

$$\overset{(1)}{\leq} \sum_{y=x_+}^m P(\theta_1|X_+ = y)P(X_+ = y) \cdot \sum_{y'''=0}^{x_+-1} P(\theta_1|X_+ = y''')\frac{P(\theta_2|X_+ = x_+)}{P(\theta_1|X_+ = x_+)}P(X_+ = y''')$$

$$= \sum_{y=x_+}^m P(\theta_1|X_+ = y)\frac{P(\theta_2|X_+ = x_+)}{P(\theta_1|X_+ = x_+)}P(X_+ = y) \cdot \sum_{y'''=0}^{x_+-1} P(\theta_1|X_+ = y''')P(X_+ = y''')$$

$$\overset{(2)}{\leq} \sum_{y=x_+}^m P(\theta_1|X_+ = y)\frac{P(\theta_2|X_+ = y)}{P(\theta_1|X_+ = y)}P(X_+ = y) \cdot \sum_{y'''=0}^{x_+-1} P(\theta_1|X_+ = y''')P(X_+ = y''').$$

2. The arguments are the same once sum is replaced by integral. □

Why should one be interested in this new elementary proof? The stochastic ordering of the latent trait by the total score variable (SOL) is very important for practical measurement in IRT. It justifies the use of the (observable) total score variable to estimate the ordering of subjects on the (unobservable) latent trait. The SOL property is the key result that justifies the use of Mokken's (1971) models of monotone homogeneity and double monotonicity as measurement models for persons. In other words, even though elementary from a theoretical-statistical point of view, this result is very important from an application point of view. The elementary proof may thus be of interest for reasons of simplicity and clarity. The main reason, however, is that this proof helps in investigating the MLR, SOM, and SOL properties for more general nonlinear latent trait spaces. Because it is based on 'just necessary' mathematical terms, this is possible in a straightforward and easy manner (see Subsection 4.4).

Two remarks are in order at this point.

1. *Reading* the implications described in Theorem 2 from the pictures in the paper by Douglas et al. (1990) does not constitute a technical-mathematical proof thereof. The pictures only serve as visualizations of mathematically proved (general) statements.

2. The intention here is not to give a *short* indirect and general, but rather direct, elementary, and self-contained proof. A shorter proof, for instance, may be based on the well-known fact that the likelihood ratio order implying the usual stochastic order can be, indirectly, established via the hazard (respectively, reversed hazard) rate order (e.g., Shaked & Shanthikumar, 1994; as already mentioned at the beginning of this subsection). In fact, the proof of Theorem 2 even is a *new* proof of this statistical fact: $\leq_{lr}$ implying $\leq_{st}$. It establishes the latter in a direct and very basic manner.

3. KST: Basic Concepts

In 1985, Jean-Paul Doignon and Jean-Claude Falmagne introduced *knowledge space theory* (KST; Doignon & Falmagne, 1985). Most of the theory of knowledge spaces is

presented in a monograph entitled 'Knowledge Spaces' by Doignon and Falmagne (1999); see also Doignon and Falmagne (1987), Falmagne (1989), and Falmagne, Koppen, Villano, Doignon, and Johannesen (1990). A comprehensive bibliography on KST (including a lot of references on applications of KST) by C. Hockemeyer (University of Graz, Austria) can be retrieved from `http://wundt.kfunigraz.ac.at/kst.php`. For concrete application examples, see in particular Albert and Lukas (1999). The theory of knowledge spaces has been successfully applied for the computerized, adaptive assessment and training of knowledge; for instance, see the ALEKS (Assessment and LEarning in Knowledge Spaces) system, a fully automated math tutor on the Internet: `http://www.aleks.com`.

This section starts with a motivating small example which is taken from Falmagne, Doignon, Cosyn, and Thiéry (2003), and then briefly reviews some of the basic deterministic and probabilistic concepts of KST. For details, the reader is referred to the afore mentioned references.

3.1. Example: Elementary Algebra

A natural starting point for a theory of *knowledge assessment* and *training* stems from the observation that some pieces of knowledge may imply other pieces of knowledge. In the context of this subsection, the mastery of some algebra problem may imply the mastery of other problems. Such implications between pieces of knowledge may be used to design efficient computer-based, adaptive knowledge assessment and training procedures.

Consider the following six (dichotomous) problems in Elementary Algebra:

a. A car travels on the freeway at an average speed of 52 miles per hour. How many miles does it travel in 5 hours and 30 minutes?

b. Using the pencil, mark the point at the coordinates $(1, 3)$.

c. Perform the following multiplication:

$$4x^4y^4 \cdot 2x \cdot 5y^2$$

and simplify your answer as much as possible.

d. Find the greatest common factor of the expressions $14t^6y$ and $4tu^5y^8$. Simplify your answer as much as possible.

e. Graph the line with slope -7 passing through the point $(-3, -2)$.

f. Write an equation for the line that passes through the point $(-5, 3)$ and is perpendicular to the line $8x + 5y = 11$.

A plausible *prerequisite diagram* of mastery dependencies for the six Elementary Algebra problems may look like in Figure 1 (Falmagne et al., 2003). The mastery of Problem b is, for instance, a prerequisite for the mastery of Problem e. In other words, the mastery of Problem e implies that of Problem b.

[Figure 1]

The prerequisite diagram in Figure 1 completely specifies the feasible *knowledge states*. A respondent can certainly master just Problem a. This does not imply mastery of any other problem. In that case, the knowledge state is $\{a\}$. However, if the respondent masters e, for instance, then a, b, and c must also be mastered. This gives the knowledge state $\{a, b, c, e\}$. In this way, one obtains exactly 10 knowledge states *consistent* with the prerequisite diagram:

$$\mathcal{K} = \Big\{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, c\}, \{a, b, c\},$$
$$\{a, b, c, d\}, \{a, b, c, e\}, \{a, b, c, d, e\}, \{a, b, c, d, e, f\}\Big\}.$$

This set $\mathcal{K}$ of all possible knowledge states is called *knowledge structure*. These notions are next formalized mathematically in the following subsection.

3.2. Basic Deterministic Concepts

A general concept is that of a *knowledge structure*.

Definition 1. (Knowledge structure) A *knowledge structure* is a pair $(Q, \mathcal{K})$, with Q a non-empty, finite set, and $\mathcal{K}$ a family of subsets of Q containing at least (the empty

set) $\emptyset$ and Q. The set Q is called the *domain* of the knowledge structure. The elements $q \in Q$ and $K \in \mathcal{K}$ are referred to as *(test) items* and *(knowledge) states*, respectively. One also says that $\mathcal{K}$ is a knowledge structure on Q.

The general definition of a knowledge structure allows for infinite item sets as well. Throughout this paper, however, Q is assumed to be finite.

The set Q is assumed to be a set of dichotomous items. In this paper, Q is interpreted as a set of questions/problems that can either *be solved* or *not be solved*. This stands for the *observed* responses of a subject (*manifest level*), and has to be distinguished from a subject's true, *unobservable* knowledge of the solution to an item (*latent level*). In the latter case, the subject is said to be *capable of mastering* or *not capable of mastering* the item. Let 2^Q denote the *power-set* of Q. The observed responses of a subject are represented by the subset $R \subseteq Q$ containing exactly the items solved by the subject. This subset R is called the *response pattern* of the subject. Similarly, the true latent state of knowledge of a subject is represented by the subset $K \subseteq Q$ containing exactly the items the subject is capable of mastering. This subset K is called the *knowledge state* of the subject, representing her/his 'ability'. Given a knowledge structure $\mathcal{K}$, the only states of knowledge possible are assumed to be the ones in $\mathcal{K}$. In this spirit, $\mathcal{K}$ captures the organization of knowledge in the domain and population of reference. Idealized, if no response errors would be committed, the only response patterns possible would be the knowledge states in $\mathcal{K}$.

As an example knowledge structure consider the one described in Subsection 3.1, on the domain $Q := \{a, b, c, d, e, f\}$ of the six Elementary Algebra problems.

Note that this example knowledge structure is closed under union and intersection.

Definition 2. ((Quasi-ordinal) Knowledge space) A knowledge structure $(Q, \mathcal{K})$ is called a *knowledge space* iff $\mathcal{K}$ is closed under union—that is, for all $\mathcal{F} \subseteq \mathcal{K}$, $\bigcup \mathcal{F} \in \mathcal{K}$. If a knowledge space $(Q, \mathcal{K})$ is closed under intersection—that is, for all $\mathcal{F} \subseteq \mathcal{K}$, $\bigcap \mathcal{F} \in \mathcal{K}$—, it is called a *quasi-ordinal* knowledge space.

The notions of a knowledge structure and (quasi-ordinal) knowledge space are at the level of persons (representing collections of knowledge states of individuals). There is another important notion, that of a *surmise relation*, which is at the level of items (representing collections of mastery dependencies between items).

Definition 3. (Surmise relation) Let Q be a non-empty, finite item set. Any quasi-order—that is, reflexive and transitive binary relation—on Q is called a *surmise relation* (on Q).

A surmise relation $\mathcal{S}$ on Q may model a latent hierarchy among the items based on mastery dependencies of the following type: a subject capable of mastering item $J \in Q$ is also capable of mastering item $I \in Q$ (i.e., $I \mathcal{S} J$).

As an example surmise relation consider the surmise relation $\mathcal{S}$ corresponding to the prerequisite diagram of mastery dependencies in Figure 1:

$$\mathcal{S} = \triangle \bigcup \Big\{ (a,c), (a,d), (a,e), (a,f), (b,d), (b,e), (b,f),$$
$$(c,d), (c,e), (c,f), (d,f), (e,f) \Big\},$$

where $\triangle$ denotes the diagonal in $\{a,b,c,d,e,f\} \times \{a,b,c,d,e,f\}$.

Birkhoff's (1937) theorem (applied in KST) provides a linkage between quasi-ordinal knowledge spaces and surmise relations on an item set.

Theorem 3. (Birkhoff's theorem) There exists a one-to-one correspondence between the family of all quasi-ordinal knowledge spaces $\mathcal{K}$ on a domain Q, and the family of all surmise relations $\mathcal{S}$ on Q. Such a correspondence is defined through the two equivalences ($p, q \in Q$, $K \subseteq Q$):

$$p \mathcal{S} q \quad :\Longleftrightarrow \quad \left[\forall K \in \mathcal{K} : \Big\{ q \in K \Longrightarrow p \in K \Big\} \right],$$
$$K \in \mathcal{K} \quad :\Longleftrightarrow \quad \left[\forall (p \mathcal{S} q) : \Big\{ q \in K \Longrightarrow p \in K \Big\} \right].$$

Proof. See (e.g.) Doignon and Falmagne (1999, pp. 39–40, Theorem 1.49). $\square$

This theorem is important from a practical point of view. Though the quasi-ordinal knowledge space and surmise relation models are empirically interpreted at two different levels, at the levels of persons and items respectively, they are connected with each other, through Birkhoff's theorem, on a solid mathematical basis. Roughly speaking, it mathematically links two different levels of empirical interpretations.

In the example in Subsection 3.1, the 10 knowledge states *consistent* with the prerequisite diagram are obtained by applying the second equivalence of Birkhoff's theorem.

3.3. Basic Probabilistic Concepts

Examinees are randomly drawn from a population of reference. Let $N \in \mathbb{N}$ be the sample size. The data are the observed absolute counts $N(R) \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ of response patterns $R \in 2^Q$, that is, $\mathbf{x} := (N(R))_{R \in 2^Q}$. The data are assumed to be the realization of a random vector, $\mathbf{X} := (X_R)_{R \in 2^Q}$, which is *multinomially* distributed over the power-set 2^Q. That is,

$$P(\mathbf{X} = \mathbf{x}) = P(X_\emptyset = N(\emptyset), \ldots, X_Q = N(Q))$$
$$= \frac{N!}{\prod_{R \in 2^Q} N(R)!} \prod_{R \in 2^Q} \rho(R)^{N(R)},$$

where $\rho(R) > 0$ is the (unknown) true probability of occurrence of a response pattern $R \in 2^Q$, with $\sum_{R \in 2^Q} \rho(R) = 1$, and $0 \leq N(R) \leq N$ for any $R \in 2^Q$, with $\sum_{R \in 2^Q} N(R) = N$.

The *basic local independence model* is fundamental in KST, in the sense that most of the KST probabilistic models are special cases of it.[9]

Definition 4. (Basic local independence model) A quadruple $(Q, \mathcal{K}, p, r)$ is called a *basic local independence model* (BLIM) iff

---

[9]The idea expressed in the definition of the basic local independence model is not a new one and goes back to traditional latent class 'measurement' models such as the Proctor (1970) model, the Dayton and Macready (1976) intrusion–omission model, and more generally, the Lazarsfeld and Henry (1968) latent distance model (see Ünlü, 2006).

1. $(Q, \mathcal{K})$ is a knowledge structure;

2. p is a *probability distribution* on $\mathcal{K}$—that is, $p : \mathcal{K} \to \,]0, 1[$, $K \mapsto p(K)$, with $p(K) > 0$ for any $K \in \mathcal{K}$, and $\sum_{K \in \mathcal{K}} p(K) = 1$;

3. r is a *response function* for $(Q, \mathcal{K}, p)$—that is, $r : 2^Q \times \mathcal{K} \to [0, 1]$,

   $(R, K) \mapsto r(R, K)$, with $r(R, K) \geq 0$ for any $R \in 2^Q$ and $K \in \mathcal{K}$, and

   $\sum_{R \in 2^Q} r(R, K) = 1$ for any $K \in \mathcal{K}$;

4. r satisfies *local independence*—that is, for any $R \in 2^Q$ and $K \in \mathcal{K}$,[10]

$$r(R, K) = \left\{ \left[\prod_{q \in K \setminus R} \beta_q \right] \cdot \left[\prod_{q \in K \cap R} (1 - \beta_q) \right] \right.$$
$$\left. \cdot \left[\prod_{q \in R \setminus K} \eta_q \right] \cdot \left[\prod_{q \in Q \setminus (R \cup K)} (1 - \eta_q) \right] \right\},$$

with constants $\beta_q, \eta_q \in [0, 1[$ for each $q \in Q$, respectively called *careless error probability* and *lucky guess probability* at q.

To each knowledge state $K \in \mathcal{K}$ is attached a probability $p(K)$ measuring the likelihood that a randomly sampled subject is in state K (Part 2). For $R \in 2^Q$ and $K \in \mathcal{K}$, $r(R, K)$ specifies the conditional probability of response pattern R for an examinee in state K (Part 3). The item responses of an examinee are assumed to be independent given the knowledge state of the examinee, and the response error probabilities β_q, η_q $(q \in Q)$ are attached to the items (item-specific) and do not vary from state to state (state-independent) (Part 4).

The BLIM assumes that the manifest multinomial probability distribution on the set of response patterns is governed by the latent knowledge state proportions and

---

[10]Note that for any $R \in 2^Q$ and $K \in \mathcal{K}$, $K \setminus R := \{q \in Q : q \in K \text{ and } q \notin R\}$, $K \cap R := \{q \in Q : q \in K \text{ and } q \in R\}$, $R \setminus K := \{q \in Q : q \in R \text{ and } q \notin K\}$, and $Q \setminus (R \cup K) := \{q \in Q : q \notin R \text{ and } q \notin K\}$ form a partition of Q—that is, $Q = (K \setminus R) + (K \cap R) + (R \setminus K) + (Q \setminus (R \cup K))$. Roughly speaking, items in $K \setminus R$, $K \cap R$, $R \setminus K$, and $Q \setminus (R \cup K)$ are mastered but not solved (careless error), mastered and solved (no careless error), solved but not mastered (lucky guess), and not solved and not mastered (no lucky guess), respectively.

response error rates.

Corollary 1. (Multinomial model) Under the BLIM, the occurrence probabilities $\rho(R)$ of response patterns $R \in 2^Q$ are parameterized as

$$\rho(R) = \sum_{K \in \mathcal{K}} \left\{ \left[\prod_{q \in K \setminus R} \beta_q \right] \cdot \left[\prod_{q \in K \cap R} (1 - \beta_q) \right] \right.$$
$$\left. \cdot \left[\prod_{q \in R \setminus K} \eta_q \right] \cdot \left[\prod_{q \in Q \setminus (R \cup K)} (1 - \eta_q) \right] \right\} p(K). \qquad \square$$

The model parameters of the BLIM are $p(K)$ ($K \in \mathcal{K}$) and β_q, η_q ($q \in Q$). The number of independent model parameters is $2|Q| + (|\mathcal{K}| - 1)$.[11] Parameter estimation and model testing, if possible at all, may be based on classical maximum likelihood methodology. As mentioned in Subsection 1.2, however, the size of $\mathcal{K}$ generally tends to be prohibitively large in practice, emphasizing the importance of a nonparametric approach in KST.

In Section 5, nonparametric IRT-type axioms and properties are discussed for the knowledge structure, surmise relation, and quasi-ordinal knowledge space models. Two sets of general nonparametric KST axioms result, which correspond to Mokken's MHM and DMM. The new axioms are compared with the assumptions underlying the BLIM. Ordinal restrictions on the parameters of the BLIM are identified under which this parametric model satisfies the nonparametric axioms. The restrictions even turn out to provide complete characterizations, in the sense that the BLIM satisfies the nonparametric axioms if, *and only if*, the restrictions are satisfied. The properties of MLR, SOM, and SOL are discussed for the BLIM as well.

4. Nonparametric IRT Axioms and Properties: Nonlinear Generalizations

The basic idea of the extensions discussed next is as follows. In the formulation of Mokken nonparametric IRT, the unidimensional latent trait space and 'linear' ordering

[11]For a set X, let $|X|$ denote the size (cardinality) of X.

of items are replaced by abstract quasi-ordered person and indicator spaces, respectively. These general spaces allow for incomparabilities among latent trait values and items. Based on such spaces, generalized equivalents of the nonparametric axioms and properties are formulated in the most intuitive and obvious way: for pairs of latent trait values and items in 'quasi-order relation'.

This section generalizes the discussion of the unidimensional nonparametric IRT axioms and properties to abstract quasi-ordered person and indicator spaces. In particular, the generalized counterpart of the monotone-likelihood-ratio-stochastic-ordering implication is studied.

4.1. Generalized Axioms

Let $(\Theta, \preceq)$ be a *quasi-ordered* latent trait space, that is, $\preceq$ is a reflexive and transitive binary relation on the set Θ. In this paper, a quasi-ordered latent trait (person) space is also called *nonlinear*; this is referred to as the axiom of *nonlinear dimensionality*. Examples are: (a) in KST, *set-theoretic discrete dimensionality* represented by a knowledge structure $\Theta := \mathcal{K}$ which is partially ordered with respect to set-inclusion; and (b) in multidimensional IRT, *Euclidean multidimensionality* represented by a Euclidean subset $\Theta \subseteq \mathbb{R}^n$ ($n \in \mathbb{N}_{\geq 2}$) which is partially ordered with respect to, for instance, coordinate-wise vector-ordering.

The latent trait $\theta \in \Theta$ is assumed to have a sampling distribution in the examinee population under reference. A function $f : \Theta \to \mathbb{R}$ is *isotonic* iff

$$\forall \theta_1, \theta_2 \in \Theta, \theta_1 \preceq \theta_2 : f(\theta_1) \leq f(\theta_2).^{12}$$

Let the conditional positive response probability $P(X_l = 1|\theta)$ as a function of $\theta \in \Theta$ be the *item response function* (IRF) of an item I_l ($1 \leq l \leq m$). The axiom of *local*

---

[12]Note that for $\preceq$-*incomparable* latent trait values $\theta_1, \theta_2 \in \Theta$, that is, $\theta_1 \not\preceq \theta_2$ and $\theta_2 \not\preceq \theta_1$, no restrictions are imposed on the relationship of the function values $f(\theta_1)$ and $f(\theta_2)$ to each other.

independence states that

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m | \theta) = \prod_{l=1}^{m} P(X_l = x_l | \theta)$$

for any $x_l \in \{0, 1\}$ $(1 \le l \le m)$ and $\theta \in \Theta$. The axiom of *isotonicity* holds iff any IRF $P(X_l = 1|.)$ $(1 \le l \le m)$ is isotonic.

The axiom of *invariant item ordering* states that there exists a quasi-order $\mathcal{S}$ on the item set $Q := \{I_l : 1 \le l \le m\}$ (cf. Subsection 3.2) such that the IRFs $P(X_l = 1|.)$ $(1 \le l \le m)$ can be ordered to yield

$$\forall \theta \in \Theta : P(X_{l_2} = 1|\theta) \le P(X_{l_1} = 1|\theta)$$

for any $(I_{l_1}, I_{l_2}) \in \mathcal{S}$ $(1 \le l_1, l_2 \le m)$.[13] In this paper, a quasi-ordered item (indicator) space $(Q, \mathcal{S})$ is also called *nonlinear*. (Note that the quasi-order $\mathcal{S}$ on the item set Q has to be distinguished from the quasi-order $\preceq$ on the latent trait space Θ. The former refers to the ordering of items, the latter to the ordering of persons.)

4.2. Formulation of Generalized Properties

The total score variable X_+ has *monotone likelihood ratio* (MLR) in θ iff, for any $0 \le x_{+,1} \le x_{+,2} \le m$,

$$\frac{P(X_+ = x_{+,2}|\theta)}{P(X_+ = x_{+,1}|\theta)}$$

is an isotonic function of (nonlinear) $\theta \in \Theta$. Similarly, the latent trait θ has *MLR* in X_+ iff, for any $\theta_1, \theta_2 \in \Theta$, $\theta_1 \preceq \theta_2$,

$$\frac{P(\theta_2|X_+ = x_+)}{P(\theta_1|X_+ = x_+)}$$

is a nondecreasing function of $0 \le x_+ \le m$. The generalized variant of the fundamental GWGH result states that under the mild conditions the total score variable still has MLR in the *nonlinear* latent trait (see Theorem 4).

---

[13]Again, note that for $\mathcal{S}$-*incomparable* items $I_{l_1}, I_{l_2} \in Q$, that is, $(I_{l_1}, I_{l_2}) \notin \mathcal{S}$ and $(I_{l_2}, I_{l_1}) \notin \mathcal{S}$, no restrictions are imposed on the relationship of the IRFs $P(X_{l_1} = 1|.)$ and $P(X_{l_2} = 1|.)$ to each other.

The property of MLR still implies that X_+ is stochastically ordered by θ (see Part 1 of Theorem 5). The *stochastic ordering of the manifest variable X_+ by θ* (SOM) means that, for any $0 \le x_+ \le m$,

$$P(X_+ \ge x_+|\theta)$$

is an isotonic function of (nonlinear) $\theta \in \Theta$. The MLR property, however, may *not* imply that θ is stochastically ordered by X_+ (see Part 2 of Theorem 5). The *stochastic ordering of the latent trait θ by X_+* (SOL) means that, for any $\theta_0 \in \Theta$,

$$P(\theta \succeq \theta_0|X_+ = x_+)$$

is a nondecreasing function of $0 \le x_+ \le m$. Unlike unidimensional IRT, where the MLR property implies both the SOM and SOL properties (for dichotomous as well as polytomous items), under the axiom of nonlinear dimensionality, the MLR property implies the SOM property, but may fail to imply the SOL property.

4.3. Generalized Ghurye-Wallace-Grayson-Huynh Result

As presented in Ünlü (2007a), there is a natural generalization of the fundamental GWGH result (Theorem 1) to abstract nonlinear latent trait spaces (for dichotomous items).

Theorem 4. (Generalized GWGH result) Under the axioms of nonlinear dimensionality, local independence, and isotonicity, and the requirement that each item response function assumes values strictly between zero and one, the total score variable has monotone likelihood ratio in the (nonlinear) latent trait.

Proof. See Ünlü (2007a). □

4.4. Generalized Monotone-Likelihood-Ratio-Stochastic-Ordering Implication

The MLR property is still symmetric in its arguments. This is because only Bayes' theorem is applied. The crucial question is whether the implications 'MLR $\Rightarrow$ SOM' and 'MLR $\Rightarrow$ SOL' still hold in the nonlinear case.

Theorem 5. (Generalized MLR-SO implication) It holds:

1. MLR of (nonlinear) θ in X_+ implies the SOM property.

2. MLR of X_+ in (nonlinear) θ does not imply the SOL property, in general.

Proof. 1. The proof of Theorem 2 carries over. For $0 < x_+ \leq m$, and $\theta_1, \theta_2 \in \Theta$, $\theta_1 \preceq \theta_2$, the required inequality $P(X_+ \geq x_+ | \theta_1) \leq P(X_+ \geq x_+ | \theta_2)$ is seen to be equivalent to

$$\sum_{y=x_+}^{m} P(\theta_1 | X_+ = y) P(X_+ = y) \cdot \sum_{y'''=0}^{x_+ - 1} P(\theta_1 | X_+ = y''') \frac{P(\theta_2 | X_+ = y''')}{P(\theta_1 | X_+ = y''')} P(X_+ = y''')$$

$$\leq \sum_{y''=x_+}^{m} P(\theta_1 | X_+ = y'') \frac{P(\theta_2 | X_+ = y'')}{P(\theta_1 | X_+ = y'')} P(X_+ = y'') \cdot \sum_{y'=0}^{x_+ - 1} P(\theta_1 | X_+ = y') P(X_+ = y').$$

The latter inequality holds because of MLR of (nonlinear) θ in X_+.

2. A counterexample based on the BLIM in KST is presented in Subsection 5.3 (however, see below). $\qquad\qquad\square$

It is instructive to see where the proof of Theorem 2 breaks down when transferring the argument to Part 2 of Theorem 5. For the sake of simplicity, let Θ be finite, equipped with a probability distribution $P(\theta) > 0$ $(\theta \in \Theta)$, $\sum_{\theta \in \Theta} P(\theta) = 1$. For $\theta_0 \in \Theta$ (without restriction, assume there is an $\theta \in \Theta$, $\theta \not\succeq \theta_0$), and $0 \leq x_{+,1} \leq x_{+,2} \leq m$, the inequality $P(\theta \succeq \theta_0 | X_+ = x_{+,1}) \leq P(\theta \succeq \theta_0 | X_+ = x_{+,2})$ is equivalent to (cf. proof of Theorem 2)

$$\sum_{\theta \succeq \theta_0} P(X_+ = x_{+,1} | \theta) P(\theta) \cdot \sum_{\theta''' \not\succeq \theta_0} P(X_+ = x_{+,1} | \theta''') \frac{P(X_+ = x_{+,2} | \theta''')}{P(X_+ = x_{+,1} | \theta''')} P(\theta''')$$

$$\leq \sum_{\theta'' \succeq \theta_0} P(X_+ = x_{+,1} | \theta'') \frac{P(X_+ = x_{+,2} | \theta'')}{P(X_+ = x_{+,1} | \theta'')} P(\theta'') \cdot \sum_{\theta' \not\succeq \theta_0} P(X_+ = x_{+,1} | \theta') P(\theta').$$

At this point the proof breaks down. The reason for this fact is that, unlike the real-valued latent trait in unidimensional IRT which assumes values in a linearly ordered subset of the real numbers, the nonlinear latent trait takes values in a

quasi-ordered space. Because of possible $\preceq$-incomparabilities among latent trait values, there in general no longer exists an $\preceq$-*upper-bound* θ_u of $A := \{\theta \in \Theta : \theta \not\succeq \theta_0\}$ in the quasi-ordered set $(\Theta, \preceq)$—that is, $\theta_u \in \Theta$, with $\theta \preceq \theta_u$ for any $\theta \in A$—, such that for an $\preceq$-*lower-bound* θ_l of $B := \{\theta \in \Theta : \theta \succeq \theta_0\}$ (e.g., $\theta_l := \theta_0$) in $(\Theta, \preceq)$—that is, $\theta_l \in \Theta$, with $\theta_l \preceq \theta$ for any $\theta \in B$—, θ_u is $\preceq$-*majorized* by θ_l—that is, $\theta_u \preceq \theta_l$.[14] (An example illustrating the problem of incomparabilities is given in Subsection 5.2.)

Nevertheless, the proof of Theorem 2 can be used to specify conditions under which the MLR property implies the property of SOL (for a nonlinear latent trait). (For any $\theta_0 \in \Theta$, let $[\theta_0) := \{\theta \in \Theta : \theta \succeq \theta_0\}$ and $\overline{[\theta_0)} := \Theta \setminus [\theta_0) = \{\theta \in \Theta : \theta \not\succeq \theta_0\}$. For $\theta \in \Theta$ and $0 \le x_{+,1} \le x_{+,2} \le m$, let $g(\theta; x_{+,1}, x_{+,2}) := P(X_+ = x_{+,2}|\theta)/P(X_+ = x_{+,1}|\theta)$.)

Proposition 1. (Conditions for MLR implying SOL) The property of MLR implies the SOL property under any of the following conditions:

1. For any $\theta_0 \in \Theta$, $\overline{[\theta_0)} \ne \emptyset$, and $0 \le x_{+,1} \le x_{+,2} \le m$, for an element $\widehat{\theta} = \widehat{\theta}(\theta_0, x_{+,1}, x_{+,2})$ in $\overline{[\theta_0)}$ with

$$g(\widehat{\theta}; x_{+,1}, x_{+,2}) = \max_{\theta \in \overline{[\theta_0)}} g(\theta; x_{+,1}, x_{+,2}),$$

it holds

$$\sum_{\theta \in [\theta_0)} P(X_+ = x_{+,1}|\theta)g(\widehat{\theta}; x_{+,1}, x_{+,2})P(\theta) \le \sum_{\theta \in [\theta_0)} P(X_+ = x_{+,1}|\theta)g(\theta; x_{+,1}, x_{+,2})P(\theta).$$

This, for instance, holds if for an $\widehat{\theta} = \widehat{\theta}(\theta_0, x_{+,1}, x_{+,2}) \in \overline{[\theta_0)}$ with $\widehat{\theta} \in \arg\max_{\theta \in \overline{[\theta_0)}} g(\theta; x_{+,1}, x_{+,2})$, one has $\widehat{\theta} \preceq \theta_0$.[15]

2. For any $\theta_0 \in \Theta$, $\overline{[\theta_0)} \ne \emptyset$, and $0 \le x_{+,1} \le x_{+,2} \le m$, for an element $\widehat{\theta} = \widehat{\theta}(\theta_0, x_{+,1}, x_{+,2})$

---

[14]Note that if $(\Theta, \preceq)$ is *complete*—that is, for all $\theta_1, \theta_2 \in \Theta$, $\theta_1 \preceq \theta_2$ or $\theta_2 \preceq \theta_1$—, this argument can be used to establish that the property of MLR implies the SOL property. In case of completeness, for both finite and infinite Θ, an $\preceq$-upper-bound of $A := \{\theta \in \Theta : \theta \not\succeq \theta_0\} = \{\theta \in \Theta : \theta \preceq \theta_0, \theta \ne \theta_0\}$ is $\theta_u := \theta_0$, which is $\preceq$-majorized by the $\preceq$-lower-bound $\theta_l := \theta_0$ of $B := \{\theta \in \Theta : \theta \succeq \theta_0\}$.

[15]In case of infinite Θ, the supremum $\sup_{\theta \in \overline{[\theta_0)}} g(\theta; x_{+,1}, x_{+,2})$ is assumed to exist (may or may not be

in $\overline{[\theta_0)}$ with

$$g(\widehat{\theta}; x_{+,1}, x_{+,2}) = \max_{\theta \in \overline{[\theta_0)}} g(\theta; x_{+,1}, x_{+,2}),$$

it holds

$$g(\widehat{\theta}; x_{+,1}, x_{+,2}) \leq g(\theta_0; x_{+,1}, x_{+,2}).$$

This, for instance, is the case if for an $\widehat{\theta} = \widehat{\theta}(\theta_0, x_{+,1}, x_{+,2}) \in \overline{[\theta_0)}$ with $\widehat{\theta} \in \arg\max_{\theta \in \overline{[\theta_0)}} g(\theta; x_{+,1}, x_{+,2})$, it holds $\widehat{\theta} \preceq \theta_0$.[16]

3. This condition applies only to finite Θ.[17] For any $\theta_0 \in \Theta$, $\overline{[\theta_0)} \neq \emptyset$, and $0 \leq x_{+,1} \leq x_{+,2} \leq m$, for every $\preceq$-*maximal element* $\widehat{\theta} = \widehat{\theta}(\theta_0) \in \overline{[\theta_0)}$ *of* $\overline{[\theta_0)}$—that is, for any $\theta \in \overline{[\theta_0)}$, $\widehat{\theta} \preceq \theta \Rightarrow \theta \preceq \widehat{\theta}$—, it holds

$$g(\widehat{\theta}; x_{+,1}, x_{+,2}) \leq g(\theta_0; x_{+,1}, x_{+,2}).$$

This, for instance, holds if $\widehat{\theta} \preceq \theta_0$ for every $\preceq$-maximal element $\widehat{\theta} = \widehat{\theta}(\theta_0)$ of $\overline{[\theta_0)}$.

4. For any $\theta_0 \in \Theta$, $\overline{[\theta_0)} \neq \emptyset$, and $0 \leq x_{+,1} \leq x_{+,2} \leq m$, there exists an $\preceq$-*largest element* $\widehat{\theta} = \widehat{\theta}(\theta_0) \in \overline{[\theta_0)}$ *of* $\overline{[\theta_0)}$—that is, for any $\theta \in \overline{[\theta_0)}$, $\theta \preceq \widehat{\theta}$—, with

$$g(\widehat{\theta}; x_{+,1}, x_{+,2}) \leq g(\theta_0; x_{+,1}, x_{+,2}).$$

This, for instance, holds if there exists an $\preceq$-largest element $\widehat{\theta} = \widehat{\theta}(\theta_0)$ of $\overline{[\theta_0)}$ with $\widehat{\theta} \preceq \theta_0$. If this is the case for any $\theta_0 \in \Theta$, $\overline{[\theta_0)} \neq \emptyset$, then $(\Theta, \preceq)$ is even a *weakly ordered* space, that is, a quasi-ordered space which is complete.[18]

a maximum) and to satisfy the corresponding inequality. This, for instance, holds if

$$\sup_{\theta \in \overline{[\theta_0)}} g(\theta; x_{+,1}, x_{+,2}) = g(\widehat{\theta}; x_{+,1}, x_{+,2})$$

for an $\widehat{\theta} \in \Theta$ with $\widehat{\theta} \preceq \theta_0$.

[16] A corresponding formulation of this condition based on 'sup' applies to infinite Θ (cf. Footnote 15).

[17] In case of infinite Θ, there may be an infinite *chain* C in $\overline{[\theta_0)}$—that is, a linearly ordered subset $C \subseteq \overline{[\theta_0)}$—which is (a) *not bounded above in* $\overline{[\theta_0)}$—that is, there exists no θ_b in $\overline{[\theta_0)}$ with $c \preceq \theta_b$ for all $c \in C$—, and for which (b) there is an $c \in C$ such that $c \npreceq \theta_0$.

[18] This condition applies to both finite and infinite person spaces. In the application of this condition in KST (Subsection 5.1), the latter special instance means that the knowledge structure is linearly ordered (with respect to set-inclusion). This implies that the items form a Guttman scale (see Footnote 20).

Proof. 1. Let $\theta_0 \in \Theta$, $\overline{[\theta_0)} \neq \emptyset$, and $0 \leq x_{+,1} \leq x_{+,2} \leq m$. The inequality

$$\sum_{\theta \succeq \theta_0} P(X_+ = x_{+,1}|\theta)P(\theta) \cdot \sum_{\theta''' \not\succeq \theta_0} P(X_+ = x_{+,1}|\theta''')\frac{P(X_+ = x_{+,2}|\theta''')}{P(X_+ = x_{+,1}|\theta''')}P(\theta''')$$

$$\leq \sum_{\theta'' \succeq \theta_0} P(X_+ = x_{+,1}|\theta'')\frac{P(X_+ = x_{+,2}|\theta'')}{P(X_+ = x_{+,1}|\theta'')}P(\theta'') \cdot \sum_{\theta' \not\succeq \theta_0} P(X_+ = x_{+,1}|\theta')P(\theta'),$$

which is equivalent to $P(\theta \succeq \theta_0|X_+ = x_{+,1}) \leq P(\theta \succeq \theta_0|X_+ = x_{+,2})$, is implied by

$$\sum_{\theta \in [\theta_0)} P(X_+ = x_{+,1}|\theta)g(\widehat{\theta}; x_{+,1}, x_{+,2})P(\theta) \leq \sum_{\theta \in [\theta_0)} P(X_+ = x_{+,1}|\theta)g(\theta; x_{+,1}, x_{+,2})P(\theta).$$

This is easily seen based on the line of reasoning in the proof of Theorem 2: A counterpart of the inequality marked by (1) holds using $g(\widehat{\theta}; x_{+,1}, x_{+,2})$; a counterpart of the inequality marked by (2) holds applying the inequality in Condition 1. This proves that under Condition 1 the property of MLR implies the SOL property. (Note that here the property of MLR is not required for the implication between the two aforementioned inequalities to hold.)

2.–4. Because for finite Θ, it holds '4 $\Rightarrow$ 3 $\Rightarrow$ 2 $\Rightarrow$ 1', and for infinite Θ, one has '4 $\Rightarrow$ 2 $\Rightarrow$ 1', the property of MLR implies the SOL property under any of the Conditions 2, 3, and 4. (Note that here the property of MLR is required for the implications between Conditions 1–4 to hold.)

For the last part of Condition 4, let $\theta_1, \theta_2 \in \Theta$. If θ_1 and θ_2 would be $\preceq$-incomparable, one would have $\theta_1 \in \overline{[\theta_2)} \neq \emptyset$. Then, by assumption, there would be an $\preceq$-largest element $\widehat{\theta} = \widehat{\theta}(\theta_2)$ of $\overline{[\theta_2)}$ with $\theta_1 \preceq \widehat{\theta} \preceq \theta_2$. By transitivity of $\preceq$, one would obtain $\theta_1 \preceq \theta_2$, a contradiction to $\theta_1 \not\preceq \theta_2$. $\square$

Some remarks are in order with respect to Proposition 1.

1. In Conditions 1 and 2, an 'optimal' $\widehat{\theta}$ generally depends on both the (fixed) latent trait value $\theta_0 \in \Theta$ and (fixed) total score (variable) values $0 \leq x_{+,1} \leq x_{+,2} \leq m$; that is, $\widehat{\theta} = \widehat{\theta}(\theta_0, x_{+,1}, x_{+,2})$. Conditions 1 and 2 require determining $\widehat{\theta}$ by optimization ($\widehat{\theta} \in \arg\max_{\theta \in \overline{[\theta_0)}} g(\theta; x_{+,1}, x_{+,2})$, or $\widehat{\theta} \in \arg\sup_{\theta \in \overline{[\theta_0)}} g(\theta; x_{+,1}, x_{+,2})$, respectively), separately for any two total score values $0 \leq x_{+,1} \leq x_{+,2} \leq m$, and the mentioned

special instance $\widehat{\theta} \preceq \theta_0$ does not uniformly apply to all pairs of total score values $0 \leq x_{+,1} \leq x_{+,2} \leq m$. In Conditions 3 and 4, however, $\widehat{\theta}$ generally depends only on the (fixed) latent trait value $\theta_0 \in \Theta$; that is, $\widehat{\theta} = \widehat{\theta}(\theta_0)$. Conditions 3 and 4 require determining $\widehat{\theta}$ solely based on the quasi-order $\preceq$ (no 'max' or 'sup' performed, respectively), and the special instance $\widehat{\theta} \preceq \theta_0$ uniformly applies to all pairs of total score values $0 \leq x_{+,1} \leq x_{+,2} \leq m$.

2. For different (regions of) values of the latent trait and total score variable, different conditions stated in Proposition 1 can be applied to verify whether the SOL property is 'locally' satisfied, that is, for specific (latent trait and total score) value combinations. For instance, for a collection of values $\theta_0' \in \Theta$ and $0 \leq x_{+,1}' \leq x_{+,2}' \leq m$, one may apply the special instance in Condition 4 (MLR assumed). If there exists an $\preceq$-largest element $\widehat{\theta} = \widehat{\theta}(\theta_0')$ of $\overline{[\theta_0')}$ with $\widehat{\theta} \preceq \theta_0'$, the SOL property is satisfied for $\theta_0' \in \Theta$ and $0 \leq x_{+,1}' \leq x_{+,2}' \leq m$ (even more, for any combination of θ_0' with total score values $0 \leq x_{+,1} \leq x_{+,2} \leq m$; cf. previous Remark 1). Stated differently, depending on which condition (special instance) may be easier to adopt, one could individually check the SOL property for specific (regions of) values of the latent trait and total score variable, using combinations of the criteria described in Proposition 1.

This may be useful in practical applications, in case the SOL property does not 'globally' hold for all latent trait and total score values. In practice, local violations of the SOL property may not be empirically severe. In the BLIM counterexample in Subsection 5.3, for instance, the violation of the SOL property is based on unrealistic values for some of the BLIM response error parameters. For a discussion of this issue in polytomous unidimensional IRT, refer to Van der Ark (2005); see also Subsection 6.2.

3. In Condition 1 (MLR assumed),

$$\sum_{\theta \in [\theta_0)} P(X_+ = x_{+,1}|\theta)g(\widehat{\theta}; x_{+,1}, x_{+,2})P(\theta) \leq \sum_{\theta \in [\theta_0)} P(X_+ = x_{+,1}|\theta)g(\theta; x_{+,1}, x_{+,2})P(\theta)$$

is implied by

$$\sum_{\theta \in [\theta_0) \bigcap \overline{[\widehat{\theta})}} P(X_+ = x_{+,1}|\theta)g(\widehat{\theta}; x_{+,1}, x_{+,2})P(\theta) \leq \sum_{\theta \in [\theta_0) \bigcap \overline{[\widehat{\theta})}} P(X_+ = x_{+,1}|\theta)g(\theta; x_{+,1}, x_{+,2})P(\theta),$$

both for finite—$\widehat{\theta} \in \arg\max_{\theta \in \overline{[\theta_0)}} g(\theta; x_{+,1}, x_{+,2})$—and

infinite—$\widehat{\theta} \in \arg\sup_{\theta \in \overline{[\theta_0)}} g(\theta; x_{+,1}, x_{+,2})$—$\Theta$. In other words, the sums in Condition 1 can be only taken over the 'smaller-equal' set of latent trait values $[\theta_0) \cap \overline{[\widehat{\theta})} \subseteq [\theta_0)$, as another criterion under which the property of MLR implies the SOL property. In the finite case, $[\theta_0) \cap \overline{[\widehat{\theta})} = \emptyset$ if $\widehat{\theta} \preceq \theta_0$, or $[\theta_0) \cap \overline{[\widehat{\theta})} = \{\theta \in \Theta : \theta \succeq \theta_0, \text{ and } \theta \not\preceq \widehat{\theta}, \theta \not\succeq \widehat{\theta}\}$ if θ_0 and $\widehat{\theta}$ are $\preceq$-incomparable. The case $\theta_0 \preceq \widehat{\theta}$ is not possible because $\widehat{\theta} \in \arg\max_{\theta \in \overline{[\theta_0)}} g(\theta; x_{+,1}, x_{+,2})$. In the infinite case, one can additionally have $\theta_0 \preceq \widehat{\theta}$ because $\widehat{\theta} \in \arg\sup_{\theta \in \overline{[\theta_0)}} g(\theta; x_{+,1}, x_{+,2})$, taken in $(\Theta, \preceq)$. Since for $\widehat{\theta} \preceq \theta_0$ (and also for $\theta_0 \not\preceq \widehat{\theta}, \widehat{\theta} \not\preceq \theta_0$) the same as before applies, one is left with $\theta_0 \preceq \widehat{\theta}, \widehat{\theta} \not\preceq \theta_0$, in which case the set $[\theta_0) \cap \overline{[\widehat{\theta})}$ cannot be further reduced.

5. Nonlinear Nonparametrics: Application in KST

This section presents an application of the generalized nonparametric IRT concepts in the theory of knowledge spaces.

5.1. KST Variants of Generalized Axioms and Properties

Let $Q := \{I_l : 1 \leq l \leq m\}$. Let $\mathcal{K}$ be a knowledge structure on Q. Obviously, $(\mathcal{K}, \subseteq)$ is a *partially ordered* set (i.e., $\subseteq$ is an antisymmetric quasi-order on $\mathcal{K}$). This is called the 'latent trait (person) space' in KST and referred to as the axiom of (*set-theoretic*) *discrete dimensionality*. Because $(\mathcal{K}, \subseteq)$ is quasi-ordered, all the discussion in terms of an abstract quasi-ordered person space applies to KST. For instance, a function $f : \mathcal{K} \to \mathbb{R}$ is *isotonic* iff

$$\forall K_1, K_2 \in \mathcal{K}, K_1 \subseteq K_2 : f(K_1) \leq f(K_2).$$

The *item response function* (IRF) of an item $I_l \in Q$ $(1 \leq l \leq m)$ is defined as

$$P(X_l = 1|.) : \mathcal{K} \to [0,1], K \mapsto P(X_l = 1|K).$$

The axioms of *local independence* and *isotonicity* can be obviously defined.

In the formulation of the axiom of invariant item ordering, the person space is taken to be the knowledge structure corresponding to the surmise relation on the item set. More precisely, let $\mathcal{S}$ be a surmise relation on Q. Let $\mathcal{K}_\mathcal{S}$ be the quasi-ordinal knowledge space derived from it according to Birkhoff's theorem (Theorem 3). The axiom of *invariant item ordering* states that the IRFs $P(X_l = 1|.)$ $(1 \leq l \leq m)$ can be ordered such that

$$\forall K \in \mathcal{K}_\mathcal{S} : P(X_{l_2} = 1|K) \leq P(X_{l_1} = 1|K)$$

for any $(I_{l_1}, I_{l_2}) \in \mathcal{S}$ $(1 \leq l_1, l_2 \leq m)$. (Note that here the indicator space $(Q, \mathcal{S})$ and person space $(\mathcal{K}_\mathcal{S}, \subseteq)$ are mathematically linked with each other using Theorem 3 (one-to-one correspondence). In the general discussion (Section 4), the spaces may not be structurally connected to each other. Also note that here the indicator space is quasi-ordered with respect to an arbitrary quasi-order, whereas the person space is partially ordered with respect to set-inclusion.)

A *tentative* (definitely open to future research) Mokken-type nonparametric axiomatization in KST is obtained as follows. An axiom set corresponding to the MHM consists of the axioms of discrete dimensionality, local independence, and isotonicity, and is based on an arbitrary knowledge structure. An axiom set corresponding to the DMM further adds the axiom of invariant item ordering, and is based on a surmise relation and its corresponding quasi-ordinal knowledge space. To avoid misunderstandings at this point, note that this paper does not aim at providing a fully elaborated nonparametric KST with corresponding statistical inference methodology. This paper rather investigates the scope of the unidimensional Mokken axioms in a nonlinear framework, and introduces a first application of nonparametric modeling in KST (cf. Subsection 1.1).

The properties of MLR, SOM, and SOL can be immediately formulated for an arbitrary KST latent trait space. For instance, the *stochastic ordering of the latent trait $K \in \mathcal{K}$ by X_+* (SOL) means that, for any $K_0 \in \mathcal{K}$,

$$P(K \supseteq K_0|X_+ = x_+)$$

is a nondecreasing function of $0 \leq x_+ \leq m$.

The generalized GWGH result's (Theorem 4) KST variant (for dichotomous items) takes the following form.

Corollary 2. (KST variant of generalized GWGH result) Under the axioms of discrete dimensionality—that is, for a knowledge structure $(\mathcal{K}, \subseteq)$—, local independence, and isotonicity, and the requirement that $0 < P(X_l = 1|K) < 1$ for any $K \in \mathcal{K}$ and $1 \leq l \leq m$, the total score variable has monotone likelihood ratio in the (discrete-dimensional) latent trait $K \in \mathcal{K}$. $\square$

As a corollary of Theorem 5,[19] the MLR property implies the SOM property, but may fail to imply the SOL property.

Corollary 3. (KST variant of generalized MLR-SO implication) It holds:

1. MLR of (discrete-dimensional) $K \in \mathcal{K}$ in X_+ implies the SOM property.
2. MLR of X_+ in (discrete-dimensional) $K \in \mathcal{K}$ does not imply the SOL property, in general. $\square$

Conditions 1–4 stated in Proposition 1, and Remarks 1–3 following Proposition 1, can also be immediately exemplified in the context of KST (all the discussion applies to the special choice $\Theta := \mathcal{K}$ and $\preceq := \subseteq$). In the light of Corollaries 2 and 3, one can thus specify conditions under which the property of MLR implies the SOL property, in the framework of the nonlinear, Mokken-type nonparametric KST formulation proposed in this paper.[20]

---

[19]Note that a counterexample illustrating Part 2 of Theorem 5 is based on the present nonparametric KST formulation (see Subsection 5.3).

[20]As mentioned in Footnote 18, the last part of Condition 4 (Proposition 1) means that the KST latent trait space is linearly ordered ($\subseteq$ is antisymmetric). Because the indicator space $(Q, \mathcal{S})$ is linked with the person space $(\mathcal{K}_{\mathcal{S}}, \subseteq)$ through Birkhoff's theorem, $\mathcal{S}$ is a weak order (according to Theorem 3, for any

5.2. Example: Problem of Incomparabilities

An example helps to illustrate the crucial problem of incomparabilities. This example will be continued in Subsection 5.3, in order to provide a counterexample which shows that the property of MLR does not imply the SOL property in general. Let $Q := \{I_1, I_2, I_3\}$ be a set of three dichotomous items. Consider the knowledge structure (quasi-ordinal knowledge space) $\mathcal{K} := \{\emptyset, \{I_1\}, \{I_1, I_2\}, \{I_1, I_3\}, Q\}$. The surmise relation corresponding to $\mathcal{K}$ according to Theorem 3 is $\mathcal{S} := \{(I_1, I_1), (I_2, I_2), (I_3, I_3), (I_1, I_2), (I_1, I_3)\}$. Note that $\mathcal{S}$ is even a partial order (antisymmetric). The Hasse diagrams of the KST person space $(\mathcal{K}, \subseteq)$ and corresponding KST indicator space $(Q, \mathcal{S})$ are depicted in Figure 2.

[Figure 2]

To exemplify the discussion in Subsection 4.4, consider $K_0 \in \mathcal{K}$, $K_0 := \{I_1, I_2\}$. In this case, $A := \{K \in \mathcal{K} : K \not\supseteq K_0\} = \{\emptyset, \{I_1\}, \{I_1, I_3\}\}$, and $B := \{K \in \mathcal{K} : K \supseteq K_0\} = \{\{I_1, I_2\}, Q\}$. The set of all $\subseteq$-upper-bounds of A in $(\mathcal{K}, \subseteq)$ is $U_A = \{\{I_1, I_3\}, Q\}$, whereas the set of all $\subseteq$-lower-bounds of B in $(\mathcal{K}, \subseteq)$ is $L_B = \{\emptyset, \{I_1\}, \{I_1, I_2\}\}$. There is no $K_u \in U_A$ which is $\subseteq$-majorized by an $K_l \in L_B$. This is shown in Figure 3. Thus, an argument similar to that for the unidimensional implication 'MLR $\Rightarrow$ SOL' (Theorem 2) does not work.

[Figure 3]

5.3. BLIM and KST Variants of Generalized Axioms and Properties

Next the nonparametric KST axioms and properties are compared with the assumptions underlying the fundamental parametric BLIM (see Subsection 3.3). The BLIM satisfies the axioms of discrete dimensionality and local independence by definition (Parts 1 and 4 of the definition, respectively). Because the BLIM assumes item-specific, state-independent careless error and lucky guess probabilities, respectively,

$I, J \in Q$, $I \mathcal{S} J \Leftrightarrow I \in \bigcap\{K \in \mathcal{K}_\mathcal{S} : J \in K\}$). In other words, except for possible 'ties' (i.e., for $I, J \in Q$, $I \mathcal{S} J$ and $J \mathcal{S} I$), the items form a Guttman (1944, 1950) scale.

β_l and η_l, at any item $I_l \in Q$ $(1 \leq l \leq m)$ (Part 4 of the definition), the IRF of an item $I_l \in Q$ $(1 \leq l \leq m)$ under the BLIM is given by (as a function of $K \in \mathcal{K}$)

$$P(X_l = 1|K) = \begin{cases} 1 - \beta_l & : \quad \text{if } I_l \in K, \\ \eta_l & : \quad \text{if } I_l \notin K. \end{cases}$$

A characterization of the axiom of isotonicity under the BLIM is as follows.

Theorem 6. (BLIM characterization of isotonicity) Let $\mathcal{K}$ be a knowledge structure on Q. In general, a set Q of BLIM IRFs does not satisfy the axiom of isotonicity. A set Q of BLIM IRFs satisfies the axiom of isotonicity if, and only if, $\eta_l \leq 1 - \beta_l$ for any $I_l \in Q$ $(1 \leq l \leq m)$.

Proof. Counterexample: Let $Q := \{I_1, I_2, I_3\}$, and $\mathcal{K} := \{\emptyset, \{I_1\}, \{I_1, I_2\}, Q\}$. Choose $K_1 := \{I_1\}$, $K_2 := \{I_1, I_2\}$, (unrealistic) $\eta_2 := 0.35$, and (unrealistic) $\beta_2 := 0.70$. Then, it holds $P(X_2 = 1|K_1) = \eta_2 > 1 - \beta_2 = P(X_2 = 1|K_2)$.

'$\Leftarrow$': Let $I_l \in Q$ $(1 \leq l \leq m)$, and $K_1, K_2 \in \mathcal{K}$, $K_1 \subseteq K_2$. The only critical case to consider is $I_l \in K_2$ and $I_l \notin K_1$. In this case, one has
$P(X_l = 1|K_1) = \eta_l \leq 1 - \beta_l = P(X_l = 1|K_2)$.

'$\Rightarrow$': Because $\emptyset \in \mathcal{K}$ and $Q \in \mathcal{K}$, for any item $I_l \in Q$ $(1 \leq l \leq m)$, it holds
$\eta_l = P(X_l = 1|\emptyset) \leq P(X_l = 1|Q) = 1 - \beta_l$. $\qquad\square$

Note that, from a practical point of view, the *order constraints* $\eta_l \leq 1 - \beta_l$ for any item $I_l \in Q$ $(1 \leq l \leq m)$ on the BLIM response error parameters are not restrictive in general: A 'psychologically plausible' *maximum* careless error rate may be around 15 to 20 percent (i.e., $1 - \beta_l \geq 0.80$), whereas lucky guess rates can be nearly eliminated by appropriate (e.g., open format, no multiple choice) item formulation (i.e., $\eta_l \approx 0$).

The axiom of invariant item ordering can be characterized as follows.

Theorem 7. (BLIM characterization of invariant item ordering) Let $\mathcal{S}$ be a surmise relation on Q, and let $\mathcal{K}$ be the corresponding quasi ordinal knowledge space. In general, a set Q of BLIM IRFs does not satisfy the axiom of invariant item ordering. A

set Q of BLIM IRFs exhibits an invariant item ordering if, and only if, for any $(I_{l_1}, I_{l_2}) \in \mathcal{S}$ $(1 \leq l_1, l_2 \leq m)$,

$$\beta_{l_1} \leq \beta_{l_2},$$

$$\eta_{l_1} \geq \eta_{l_2},$$

and if $(I_{l_2}, I_{l_1}) \notin \mathcal{S}$, in addition,

$$1 - \beta_{l_1} \geq \eta_{l_2}.$$

Proof. Counterexample: For $(I_2, I_3) \in \mathcal{S}$ ($\mathcal{S}$, the surmise relation which $\mathcal{K}$ in the proof of Theorem 6 corresponds to), (realistic) $\eta_2 := 0.04$, and (realistic) $\eta_3 := 0.07$, it holds $P(X_3 = 1|\emptyset) = \eta_3 > \eta_2 = P(X_2 = 1|\emptyset)$.

'$\Leftarrow$': Let $(I_{l_1}, I_{l_2}) \in \mathcal{S}$ $(1 \leq l_1, l_2 \leq m)$, and $K \in \mathcal{K}$. *Case 1.* If $I_{l_2} \in K$, then $I_{l_1} \in K$ (according to Theorem 3, for any $I, J \in Q$, $I \mathcal{S} J \Leftrightarrow I \in \bigcap\{K \in \mathcal{K}_\mathcal{S} : J \in K\}$). In this case, it holds $P(X_{l_2} = 1|K) = 1 - \beta_{l_2} \leq 1 - \beta_{l_1} = P(X_{l_1} = 1|K)$. *Case 2.* If $I_{l_2} \notin K$ and $I_{l_1} \notin K$, then $P(X_{l_2} = 1|K) = \eta_{l_2} \leq \eta_{l_1} = P(X_{l_1} = 1|K)$. *Case 3.* If $I_{l_2} \notin K$ and $I_{l_1} \in K$—this is possible only if $(I_{l_2}, I_{l_1}) \notin \mathcal{S}$—, then $P(X_{l_2} = 1|K) = \eta_{l_2} \leq 1 - \beta_{l_1} = P(X_{l_1} = 1|K)$.

'$\Rightarrow$': Let $(I_{l_1}, I_{l_2}) \in \mathcal{S}$ $(1 \leq l_1, l_2 \leq m)$. It holds $(\emptyset, Q \in \mathcal{K})$

$$1 - \beta_{l_2} = P(X_{l_2} = 1|Q) \leq P(X_{l_1} = 1|Q) = 1 - \beta_{l_1},$$

and

$$\eta_{l_2} = P(X_{l_2} = 1|\emptyset) \leq P(X_{l_1} = 1|\emptyset) = \eta_{l_1}.$$

If, additionally, $(I_{l_2}, I_{l_1}) \notin \mathcal{S}$, there is an $\widehat{K} \in \mathcal{K}$ with $I_{l_1} \in \widehat{K}$ and $I_{l_2} \notin \widehat{K}$, and

$$\eta_{l_2} = P(X_{l_2} = 1|\widehat{K}) \leq P(X_{l_1} = 1|\widehat{K}) = 1 - \beta_{l_1}. \qquad \square$$

Note that violations of the axiom of invariant item ordering can occur for *realistic* values for the BLIM response error parameters (unlike, in general, for the axiom of isotonicity).

Finally, the properties of MLR, SOM, and SOL are discussed. According to Corollary 2, for a set Q of BLIM IRFs (a) satisfying the axiom of isotonicity, that is, for any item $I_l \in Q$ $(1 \leq l \leq m)$, $\eta_l \leq 1 - \beta_l$ (Theorem 6), and (b) having response error probabilities $0 < \beta_l, \eta_l < 1$ for *all* items $I_l \in Q$ $(1 \leq l \leq m)$, the total score variable has MLR in the (discrete-dimensional) latent trait $K \in \mathcal{K}$. In particular (Part 1 of Corollary 3), the property of SOM holds. Note that, from a practical point of view, the *order constraints* $0 < \beta_l, \eta_l < 1$ for all items $I_l \in Q$ $(1 \leq l \leq m)$ on the BLIM response error parameters are not that restrictive. 'Boundary value' BLIM IRFs may be closely approximated by BLIM IRFs that do meet these order constraints.

The following counterexample shows that a set of BLIM IRFs may not possess the property of SOL. This example also illustrates what was asserted in Theorem 5 and Corollary 3: Under nonlinearity, the property of MLR does not in general imply the SOL property; even in case of a restrictive set of parametric BLIM IRFs satisfying the axioms of discrete dimensionality, local independence, and isotonicity.

SOL-counterexample based on BLIM. Let $Q := \{I_1, I_2, I_3\}$, and $\mathcal{K} := \{\emptyset, \{I_1\}, \{I_1, I_2\}, \{I_1, I_3\}, Q\}$ (the example in Subsection 5.2). One can specify values for the BLIM parameters such that (a) the BLIM IRFs are isotonic, assuming values strictly between zero and one, therefore (b) the property of MLR holds, and (c) for $K_0 := \{I_1, I_2\}$, $x_{+,1} := 1$, and $x_{+,2} := 2$ the SOL property is violated. Such a choice of values for the BLIM parameters is given by $p(K) := 1/5$ for any $K \in \mathcal{K}$ (uniform distribution), and $\beta_1 := 0.75$ / $\eta_1 := 0.10$, $\beta_2 := 0.82$ / $\eta_2 := 0.16$, and $\beta_3 := 0.71$ / $\eta_3 := 0.07$. This specification yields $P(K \supseteq K_0 | X_+ = x_{+,1}) \approx 0.532 > 0.477 \approx P(K \supseteq K_0 | X_+ = x_{+,2})$.

Note that the violation of the SOL property reported above is based on unrealistic values for some of the BLIM response error parameters (e.g., $\beta_1 := 0.75$). Extensive simulation studies systematically conducted in Ünlü (2007b) demonstrate that violations of the SOL property occur only for extreme values for some of the BLIM parameters; for non-extreme and thus practical parameter vectors the BLIM seems to satisfy the property of SOL. Therefore, if at all of interest, it is necessary to check for

the SOL property in any fitted BLIM. In fact, the issue of (local) violation of the SOL property parallels the situation in polytomous unidimensional IRT, where violations of the SOL property are rather based on extreme parameter values as well (e.g., Van der Ark, 2005). This common ground may provide interesting directions for further research on this issue in KST applications (cf. Subsection 6.2).

Of course, one can use Proposition 1 (MLR assumed) to derive conditions under which the BLIM satisfies the SOL property, 'globally' (for all value combinations) as well as 'locally' (for specific value combinations); see also Remarks 1–3 following Proposition 1.

6. Discussion

6.1. Summary

At an order-theoretic structural level, this paper has investigated the dichotomous Mokken nonparametric IRT axioms and properties under incomparabilities among latent trait values and items. A new, simpler proof of the fact that in the unidimensional case monotone likelihood ratio implies stochastic ordering has been presented (Subsection 2.4). Based on this proof, (a) it has been seen that the generalized Ghurye-Wallace-Grayson-Huynh (GWGH) result implies stochastic ordering of the total score variable, but may fail to imply stochastic ordering of the nonlinear latent trait, and (b) the reason for this fact and conditions under which the implication holds have been specified (Subsection 4.4). This paper then has exemplified the generalized nonparametric IRT concepts with the combinatorial theory of knowledge spaces. KST variants of the generalized nonparametric IRT axioms and properties, the generalized GWGH result, and the generalized monotone-likelihood-ratio-stochastic-ordering implication have been formulated (Subsection 5.1). Under the parametric BLIM (fundamental in KST), characterizations (based on order constraints) of the axioms of isotonicity and invariant item ordering have been provided, and the monotone likelihood ratio and stochastic ordering properties have been discussed (Subsection 5.3).

6.2. Total Score Variable and Nonlinearity of Latent Trait

Intuitively, it seems obvious that the total score variable may not be a reasonable statistic for estimating the ordering of subjects on a nonlinear latent trait: The total score variable assumes values in a linearly ordered set, whereas the nonlinear latent trait takes values in a not necessarily complete set. It is precisely this aspect that has been investigated in this paper, mathematically.[21] The structural investigation has shown that though the total score variable has monotone likelihood ratio in the nonlinear latent trait (generalized GWGH result), the crucial implication 'MLR $\Rightarrow$ SOL' which guarantees the SOL property breaks down (though the MLR property still implies the SOM property and is symmetric in its arguments).

Nevertheless, it is interesting to see that even for a nonlinear latent trait, the total score variable may, at least in a sense, be a reasonable statistic in certain cases. On the one hand, under the conditions stated in Proposition 1, the MLR property implies the SOL property. On the other hand, simulations using the BLIM demonstrate that violations of the SOL property occur only for extreme (unrealistic) parameter vectors (Ünlü, 2007b). The latter parallels the situation in polytomous unidimensional IRT (see Sijtsma & Van der Ark, 2001; Van der Ark, 2005). Similarly to research there, one can also consider investigating here (a) how often the SOL property is violated, (b) whether violations of the SOL property constitute serious problems for practical purposes, and (c) what factors are related to violations of the SOL property. As well, a reliance on the total score variable may be of practical relevance, as this statistic is quick and simple, and easy to communicate (cf. Hemker et al., 2001, p. 493). Interestingly, that practical relevance can also be evidenced in the literature on (so-called) *ecological rationality* which uses the *tallying* score, the (unweighted) total score, of cue values as a *fast* and *frugal* heuristic in paired comparisons (e.g., Katsikopoulos & Martignon, 2006; Martignon & Hoffrage, 2002).

---

[21]In particular, the driving force for this paper is not to have the SOL property hold imperatively. The focus rather is on a thorough mathematical analysis of the dichotomous Mokken nonparametric IRT axioms and properties under incomparabilities among latent trait values and items.

6.3. Order-Theoretic Structural Remarks

It is worth mentioning that the whole discussion about the nonparametric IRT axioms and properties is solely based on order-theoretic assumptions, not algebraic. For instance, the unidimensional and generalized GWGH results in Theorems 1 and 4, respectively, can be formulated and proved without resorting to any algebraic operation (e.g., addition in the real numbers). To what extent, if at all, this may also apply to the statistical treatment of the discussed concepts is out of the scope of this paper. The development of appropriate statistical inference methodologies for the generalized formulations and an order-theoretic structural analysis of the latter could be investigated in future research.

Further note that the results concerning a nonlinear latent trait could be, *at least mathematically*, analogously formulated for an abstract nonlinear manifest variable X (generalizing the total score variable); that is, the statistic X assumes values in a quasi-ordered set of possible realizations. Without giving any mathematical proof in what follows, an argument in the manifest variable can be seen to be *dual* to an argument in the latent trait. For instance (cf. Theorem 5), if the manifest variable X is nonlinear and the latent trait unidimensional, MLR of (nonlinear) X in (unidimensional) θ implies the SOL property, but MLR of (unidimensional) θ in (nonlinear) X does not in general imply the SOM property. Again (cf. Proposition 1), *at least mathematically*, one could 'dually' formulate conditions under which the MLR property implies the property of SOM.

This paper has demonstrated that the crucial order-theoretic assumption in the discussion of the generalized monotone-likelihood-ratio-stochastic-ordering implication is the assumption of completeness (for the binary relations underlying the sets of possible values of the latent trait and manifest variable), defined in Footnote 14. In Table 1, a cross-tabulation of the latent trait and manifest variable based on the criterion of completeness summarizes the possible implications.

[Table 1]

6.4. Concluding Resume

Though the present approach has its limitation that it does not consider any statistical inference method and application to simulated or empirical data, I hope to have shown, however, that an order-theoretic analysis of the nonparametric IRT axioms and properties provides new structural insights into the long-standing study of those concepts. In particular, I hope to have emphasized the need for important future research on a fusion of IRT and KST. An endeavor such as described in this paper may prove valuable in the search for appropriate sets of nonparametric probabilistic axioms in the (yet parametric) theory of knowledge spaces; for instance, in order to provide a Mokken-type scale analysis for the surmise relation model. This, indeed, would be an important contribution to the scientific literature; in particular, allowing for a unification of both the currently split directions of psychological test theories IRT and KST.

Acknowledgments

References

Airasian, P.W., & Bart, W.M. (1973). Ordering theory: A new and useful measurement model. *Educational Technology, 13*, 56–60.

Albert, D., & Held, T. (1999). Component based knowledge spaces in problem solving and inductive reasoning. In D. Albert, & J. Lukas (Eds.), *Knowledge spaces: Theories, empirical research, and applications* (pp. 15–40). Mahwah, NJ: Lawrence Erlbaum Associates.

Albert, D., & Lukas, J. (Eds.) (1999). *Knowledge spaces: Theories, empirical research, and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Bart, W.M., & Airasian, P.W. (1974). Determination of the ordering among seven Piagetian tasks by an ordering-theoretic method. *Journal of Educational Psychology, 66*, 277–284.

Bart, W.M., & Krus, D.J. (1973). An ordering-theoretic method to determine hierarchies among items. *Educational and Psychological Measurement, 33*, 291–300.

Birkhoff, G. (1937). Rings of sets. *Duke Mathematical Journal, 3*, 443–454.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord, & M.R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Reading, MA: Addison-Wesley.

Dayton, C.M., & Macready, G.B. (1976). A probabilistic model for the validation of behavioral hierarchies. *Psychometrika, 41*, 189–204.

Doignon, J.-P., & Falmagne, J.-Cl. (1985). Spaces for the assessment of knowledge. *International Journal of Man–Machine Studies, 23*, 175–196.

Doignon, J.-P., & Falmagne, J.-Cl. (1987). Knowledge assessment: A set theoretical framework. In B. Ganter, R. Wille, & K.E. Wolfe (Eds.), *Beiträge zur Begriffsanalyse: Vorträge der Arbeitstagung Begriffsanalyse, Darmstadt 1986* (pp. 129–140). Mannheim: BI Wissenschaftsverlag.

Doignon, J.-P., & Falmagne, J.-Cl. (1999). *Knowledge spaces.* Berlin: Springer.

Douglas, R., Fienberg, S.E., Lee, M.T., Sampson, A.R., & Whitaker, L.R. (1990). 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, CA: IMS.

Esary, J.D., Proschan, F., & Walkup, D.W. (1967). Association of random variables, with applications. *The Annals of Mathematical Statistics, 38*, 1466–1474.

Falmagne, J.-Cl. (1989). Probabilistic knowledge spaces: A review. In F.S. Roberts (Ed.), *Applications of combinatorics and graph theory to the biological and social sciences* (Vol. 17, pp. 283–303). New York: Springer.

Falmagne, J.-Cl., Doignon, J.-P., Cosyn, E., & Thiéry, N. (2003). The assessment of knowledge, in theory and in practice. Institute for Mathematical Behavioral Sciences, University of California, Irvine. Paper 26.

Falmagne, J.-Cl., Koppen, M., Villano, M., Doignon, J.-P., & Johannesen, L. (1990). Introduction to knowledge spaces: How to build, test and search them. *Psychological Review, 97*, 201–224.

Fischer, G.H., & Molenaar, I.W. (Eds.) (1995). *Rasch models: Foundations, recent developments, and applications.* New York: Springer.

Ghurye, S.G., & Wallace, D.L. (1959). A convolutive class of monotone likelihood ratio families. *The Annals of Mathematical Statistics, 30*, 1158–1164.

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

Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review, 9*, 139–150.

Guttman, L. (1950). The basis for scalogram analysis. In S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star, & J.A. Clausen (Eds.), *Measurement and prediction* (pp. 60–90). Princeton, NJ: Princeton University Press.

Held, T. (1999). An integrated approach for constructing, coding, and structuring a body of word problems. In D. Albert, & J. Lukas (Eds.), *Knowledge spaces: Theories, empirical research, and applications* (pp. 67–102). Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Hemker, B.T., Van der Ark, L.A., & Sijtsma, K. (2001). On measurement properties of continuation ratio models. *Psychometrika, 66*, 487–506.

Huisman, M., & Molenaar, I.W. (2001). Imputation of missing scale data with item response models. In A. Boomsma, M.A.J. van Duijn, & T.A.B. Snijders (Eds.), *Essays on item response theory* (pp. 221–244). New York: Springer.

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. *The Annals of Statistics, 21*, 1359–1378.

Junker, B.W. (2001). On the interplay between nonparametric and parametric IRT, with some thoughts about the future. In A. Boomsma, M.A.J. van Duijn, & T.A.B. Snijders (Eds.), *Essays on item response theory* (pp. 247–276). New York: Springer.

Kambouri, M. (1991). *Knowledge assessment: A comparison of human experts and computerized procedures.* Ph.D. dissertation, New York University, New York.

Karlin, S. (1957). Pólya type distributions, II. *The Annals of Mathematical Statistics, 28*, 281–308.

Karlin, S. (1968). *Total positivity.* Stanford, CA: Stanford University Press.

Katsikopoulos, K.V., & Martignon, L. (2006). Naïve heuristics for paired comparisons: Some results on their relative accuracy. *Journal of Mathematical Psychology, 50,* 488–494.

Korossy, K. (1999). Modeling knowledge as competence and performance. In D. Albert, & J. Lukas (Eds.), *Knowledge spaces: Theories, empirical research, and applications* (pp. 103–132). Mahwah, NJ: Lawrence Erlbaum Associates.

Laros, J.A., & Tellegen, P.J. (1991). *Construction and validation of the SON-R 5.5-17, the Snijders–Oomen non-verbal intelligence test.* Groningen: Wolters–Noordhoff.

Lazarsfeld, P.F., & Henry, N.W. (1968). *Latent structure analysis.* Boston: Houghton Mifflin.

Lehmann, E.L. (1986). *Testing statistical hypotheses* (2nd ed.). New York: Wiley.

Lehmann, E.L., & Rojo, J. (1992). Invariant directional orderings. *The Annals of Statistics, 20,* 2100–2110.

Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision, 52,* 29–71.

Mokken, R.J. (1971). *A theory and procedure of scale analysis.* The Hague, Mouton/Berlin: De Gruyter.

Mokken, R.J. (1997). Nonparametric models for dichotomous responses. In W.J. van der Linden, & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 351–367). New York: Springer

Mokken, R.J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement, 6,* 417–430.

Proctor, C.H. (1970). A probabilistic formulation and statistical analysis of Guttman scaling. *Psychometrika, 35,* 73–78.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Nielsen & Lydiche.

Ross, S.M. (1996). *Stochastic processes* (2nd ed.). New York: Wiley.

Shaked, M., & Shanthikumar, J.G. (1994). *Stochastic orders and their applications*. New York: Academic Press.

Sijtsma, K. (1998). Methodology review: Nonparametric IRT approaches to the analysis of dichotomous item scores. *Applied Psychological Measurement, 22*, 3–31.

Sijtsma, K., & Hemker, B.T. (1998). Nonparametric polytomous IRT models for invariant item ordering, with results for parametric models. *Psychometrika, 63*, 183–200.

Sijtsma, K., & Junker, B.W. (1996). A survey of theory and methods of invariant item ordering. *British Journal of Mathematical and Statistical Psychology, 49*, 79–105.

Sijtsma, K., & Molenaar, I.W. (2002). *Introduction to nonparametric item response theory*. Thousand Oaks, CA: Sage.

Sijtsma, K., & Van der Ark, L.A. (2001). Progress in NIRT analysis of polytomous item scores: Dilemmas and practical solutions. In A. Boomsma, M.A.J. van Duijn, & T.A.B. Snijders (Eds.), *Essays on item response theory* (pp. 297–318). New York: Springer.

Stefanutti, L. (2006). A logistic approach to knowledge structures. *Journal of Mathematical Psychology, 50*, 545–561.

Ünlü, A. (2006). Estimation of careless error and lucky guess probabilities for dichotomous test items: A psychometric application of a biometric latent class model with random effects. *Journal of Mathematical Psychology, 50*, 309–328.

Ünlü, A. (2007a). A note on monotone likelihood ratio of the total score variable in unidimensional item response theory. *British Journal of Mathematical and Statistical Psychology*, in press.

Ünlü, A. (2007b). Stochastic ordering of the latent trait by the total score variable under the basic local independence model in knowledge space theory. Manuscript in preparation.

Van der Ark, L.A. (2001). Relationships and properties of polytomous item response theory models. *Applied Psychological Measurement, 25*, 273–282.

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 Linden, W.J., & Hambleton, R.K. (Eds.) (1997). *Handbook of modern item response theory*. New York: Springer.

Figure Captions

Figure 1. Prerequisite diagram of mastery dependencies for the six Elementary Algebra problems a–f. Reflexivity and transitivity are assumed to hold and are not explicitly depicted. The mastery of Problem b is, for instance, a prerequisite for the mastery of Problem e. In other words, the mastery of Problem e implies that of Problem b. Taken from Falmagne et al. (2003).

Figure 2. Hasse diagrams of (a) the KST person space $(\mathcal{K} := \{\emptyset, \{I_1\}, \{I_1, I_2\}, \{I_1, I_3\}, Q\}, \subseteq)$ and (b) the KST indicator space $(Q, \mathcal{S} := \{(I_1, I_1), (I_2, I_2), (I_3, I_3), (I_1, I_2), (I_1, I_3)\})$; for a set $Q := \{I_1, I_2, I_3\}$ of three dichotomous items. The spaces correspond to each other according to Birkhoff's theorem (Theorem 3).

Figure 3. KST person space $(\mathcal{K} := \{\emptyset, \{I_1\}, \{I_1, I_2\}, \{I_1, I_3\}, Q\}, \subseteq)$, knowledge state $K_0 := \{I_1, I_2\}$, and relevant sets $A := \overline{[K_0)}$, $B := [K_0)$, U_A (set of $\subseteq$-upper-bounds of A), and L_B (set of $\subseteq$-lower-bounds of B). The set A consists of all knowledge states below the line marked by '$A := \overline{[K_0)}$' (i.e., $\emptyset$, $\{I_1\}$, and $\{I_1, I_3\}$), the set B consists of all knowledge states above this line (i.e., $\{I_1, I_2\}$ and Q). Similarly, the set U_A consists of all knowledge states above the line marked by 'U_A' (i.e., $\{I_1, I_3\}$ and Q), the set L_B consists of all knowledge states below this line (i.e., $\emptyset$, $\{I_1\}$, and $\{I_1, I_2\}$). There is no $K_u \in U_A$ which is $\subseteq$-majorized by an $K_l \in L_B$.
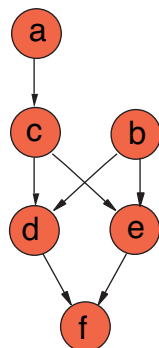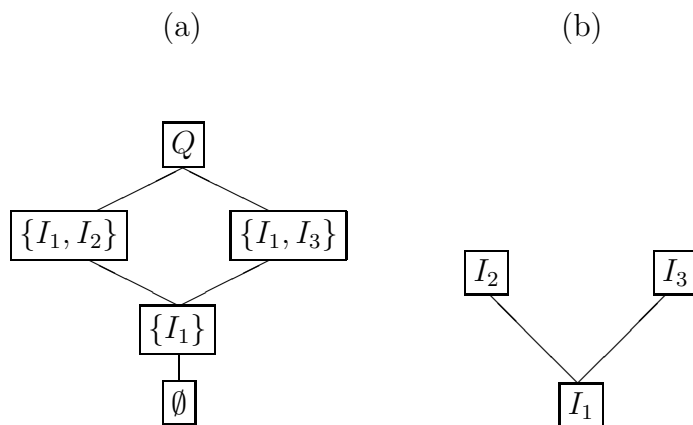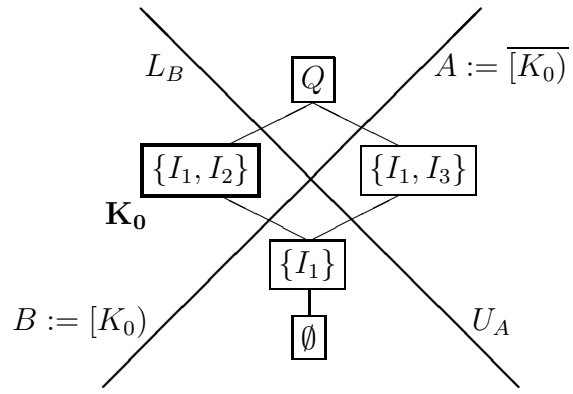
Figures

(a)

(b)

Figure 2.

FIGURE 3.

Tables

Table 1.

Cross-tabulation of implications based on the criterion of completeness (for dichotomous items)

$\theta \setminus X$	complete	not complete
complete	MLR $\Rightarrow$ SOL	MLR $\Rightarrow$ SOL
	MLR $\Rightarrow$ SOM	MLR $\overset{\text{i.g.}}{\not\Rightarrow}$ SOM
not complete	MLR $\overset{\text{i.g.}}{\not\Rightarrow}$ SOL	MLR $\overset{\text{i.g.}}{\not\Rightarrow}$ SOL
	MLR $\Rightarrow$ SOM	MLR $\overset{\text{i.g.}}{\not\Rightarrow}$ SOM

Note: The assumption of completeness for the binary relations underlying the sets of possible values of the latent trait θ (rows) and manifest variable X (columns) is the crucial order-theoretic assumption in the discussion of the generalized monotone-likelihood-ratio-stochastic-ordering implication. The dichotomous unidimensional IRT case using unidimensional θ ($\theta \in \Theta \subseteq \mathbb{R}$) and linear (total score variable) X_+ is represented by the upper-left entry 'complete θ' and 'complete X'. Theorem 5 using nonlinear (quasi-ordered) θ and linear X_+ is represented by the lower-left entry 'not complete θ' and 'complete X'. The case mentioned in Subsection 6.3 using unidimensional θ and nonlinear X is represented by the upper-right entry 'complete θ' and 'not complete X'. Finally, a case using, for instance, nonlinear θ and nonlinear X is represented by the lower-right entry 'not complete θ' and 'not complete X'. Note that parts of this table (e.g., upper-right entry) are just asserted without giving mathematical proofs. Note that 'i.g.' stands for 'in general'.