Journal of Mathematical Psychology 57 (2013) 249–254

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

Considerations about the identification of forward- and backward-graded knowledge structures

Andrea Spoto [a,*], Luca Stefanutti [b], Giulio Vidotto [a]

[a] Department of General Psychology, University of Padua, Italy
[b] FISPPA, University of Padua, Italy

HIGHLIGHTS

- The BLIM is unidentifiable for FG and BG knowledge structures.
- A relevant number of knowledge structures is FG and/or BG.
- The introduction of equally informative items solves unidentifiability.
- A structure where a tradeoff dimension involves one error parameter is FG or BG.
- A relationship exists between unidentifiability and forward- and backward-gradedness.

ARTICLE INFO

Article history:
Received 19 December 2012
Received in revised form
10 September 2013
Available online 7 October 2013

Keywords:
Identifiability
Forward-graded knowledge structures
Backward-graded knowledge structures
Probabilistic knowledge structures
Learning spaces
Equally informative items

ABSTRACT

The application of the basic local independence model (BLIM) to a knowledge structure $(Q, \mathcal{K})$ that satisfies a particular kind of gradation (namely forward- or backward-gradedness) leads the model to be not identifiable. In the present article, we show that many important types of knowledge structures happen to be either forward- or backward-graded. This means that the application of the BLIM to these structures leads to unidentifiable models. No universal remedy for recovering identifiability is presently known. However, we propose a construction that consists in introducing an equally informative item for each item in Q. We conjecture that the BLIM based on the resulting knowledge structure is always identifiable. This conjecture is proven to be true for knowledge structures on small sets of items.

1. Introduction

In knowledge space theory (KST; Doignon & Falmagne, 1985, 1999; Falmagne & Doignon, 2011) the knowledge state of a student is the collection of all problems that she/he masters in a given field of knowledge, and a knowledge structure is the collection of all possible knowledge states in a given population of students. The basic local independence model (BLIM; Doignon & Falmagne, 1999; Falmagne & Doignon, 1988) is a probabilistic model applied, in KST, for the stochastic assessment of knowledge, and for the empirical validation of knowledge structures.

This model has lately received some attention (de Chiusole, Stefanutti, Anselmi, & Robusto, in press; Heller, under revision; Schrepp, 2005; Stefanutti & Robusto, 2009), particularly concerning its identifiability. At the moment, there is still incomplete knowledge about this fundamental issue of the BLIM. Notwithstanding, a few results have been recently obtained in this direction. For instance, Spoto, Stefanutti, and Vidotto (2012) have shown that the BLIM is not identifiable for two broad classes of knowledge structures named, respectively, forward-graded and backward-graded. Moreover, Stefanutti, Heller, Anselmi, and Robusto (2012) developed a procedure for testing the local identifiability of the BLIM with arbitrary knowledge structures on sets of items having a moderate size. Furthermore, some connections between the BLIM and latent class models, including identifiability issues, have been pointed out by Schrepp (2005) and Ünlü (2011).

While it is known that the identifiability of the BLIM strictly depends on the specific knowledge structure to which it is applied, it is still not clear how to separate all knowledge structures, on a finite set of items, that make the BLIM unidentifiable, from the remaining ones. In other words it is not known, in general, how to read the identifiability of the BLIM, directly from the combinatorial properties of the knowledge structure to which it is applied.

* Correspondence to: Department of General Psychology, Via Venezia 8, 35131 Padua, Italy.
E-mail addresses: andrea.spoto@unipd.it (A. Spoto), luca.stefanutti@unipd.it (L. Stefanutti), giulio.vidotto@unipd.it (G. Vidotto).

0022-2496/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmp.2013.09.002

At the moment this knowledge seems to be restricted to the aforementioned forward- and backward-graded cases. Moreover, what to do when the BLIM is not identifiable for a particular knowledge structure is a sensible question, for which there is no ultimate answer at the present time.

This article can be essentially divided into two main parts: the first one concerns theoretical results on the lack of identifiability of the BLIM for certain classes of knowledge structures that frequently occur in practical applications. This material draws upon theoretical results, obtained in Spoto et al. (2012), on two broad classes of knowledge structures called forward-graded and backward-graded. The second part is a collection of simulation studies whose aim is to provide evidence supporting a particular conjecture on how to restore local identifiability of the BLIM, when it is applied to an arbitrary (not necessarily forward- and/or backward-graded) knowledge structure.

Section 2 introduces basic KST concepts, including the BLIM. Section 3 summarizes the concepts of local identifiability of a model that are needed in this article. Background material about identifiability of the BLIM is presented in Section 4. Section 5 is a collection of theoretical results concerning noticeable instances of forward-graded and backward-graded knowledge structures. This section covers the first part of the article mentioned in the previous paragraph. The second part is covered by Section 6.

2. KST and the basic local independence model

In KST, a *knowledge domain* is the set Q of all items about a specific topic. Throughout the paper we assume that Q is finite. Given Q, a *knowledge state* is the subset $K \subseteq Q$ that a subject can solve. A *knowledge structure* is a pair $(Q, \mathcal{K})$ where $\mathcal{K}$ is a collection of knowledge states containing at least $\emptyset$ and Q itself.

For any item $q \in Q$, let $\mathcal{K}_q = \{K \in \mathcal{K} | q \in K\}$ and $\bar{\mathcal{K}}_q$ be its complement in $\mathcal{K}$. Then two items, p and q, are said to be *equally informative* if $\mathcal{K}_q = \mathcal{K}_p$. Equally informative items form notions: the *notion* corresponding to $q \in Q$ is $q^* = \{p \in Q | \mathcal{K}_q = \mathcal{K}_p\}$. The collection of all the notions forms a partition of Q. Any structure in which every notion contains only one item is called *discriminative*, since every item in such a structure provides unique information.

Whenever a knowledge structure is closed under union, it is a *knowledge space*. The dual $\mathcal{K}^d = \{Q \setminus K : K \in \mathcal{K}\}$ of a knowledge space $\mathcal{K}$ is closed under intersection and it is called *closure space*. A knowledge structure which is closed under both union and intersection is a *quasi-ordinal knowledge space*, and if it is also discriminative, it is an *ordinal knowledge space*.

Going beyond the deterministic case, a *probabilistic knowledge structure* (PKS; Doignon & Falmagne, 1999) is defined as a triple $(Q, \mathcal{K}, \pi)$ where $(Q, \mathcal{K})$ is a knowledge structure and π is a probability distribution on $\mathcal{K}$. A *response pattern* is the subset R of Q consisting of all the items which would receive a correct answer if asked to a student. The probability of an arbitrary response pattern $R \subseteq Q$ is specified (Falmagne & Doignon, 1988) by the following unrestricted latent class model (general concepts about latent class models can be found in, e.g., Lazarsfeld & Neil, 1968):

$$P(R) = \sum_{K \in \mathcal{K}} P(R|K)\pi(K),$$

where $P(R|K)$ is the conditional probability of the response pattern R given the knowledge state K.

The probabilistic model usually applied in KST is the *basic local independence model*, in which the answers to the items are locally independent given the knowledge state of the person. The conditional probability of a pattern given a state is determined by two error parameters of each item q: β_q (*careless error*) and η_q (*lucky guess*) respectively. Eq. (1) displays the connection between $P(R|K)$ and the error parameters:

$$P(R|K) = \left[\prod_{q \in K \setminus R} \beta_q\right]\left[\prod_{q \in K \cap R} (1-\beta_q)\right]\left[\prod_{q \in R \setminus K} \eta_q\right] \times \left[\prod_{q \in \overline{K \cup R}} (1-\eta_q)\right]. \quad (1)$$

Generally speaking, these parameters are expected to be low (see e.g., Stefanutti & Robusto, 2009). More specifically, it can be reasonably assumed that $\beta_q + \eta_q < 1$ for all $q \in Q$. In fact, such an assumption seems to underlie any assessment, because it simply asserts that the probability of observing a wrong response to an item is higher when the student does not master the item than when the student does master it (in formula, $\beta_q < 1 - \eta_q$). Conversely, the probability of a correct answer should be higher when the student masters the item rather than when the student does not master it (i.e. $\eta_q < 1 - \beta_q$).

3. Local identifiability of a model

Whenever a probabilistic model is empirically tested, the identifiability of the model itself is a crucial issue that must be addressed. In general, a model can be regarded as a triple (Θ, f, Φ) where $\Theta \in \mathbb{R}^n$ is the *parameter space* of the model (where n is the number of parameters in the model), $\Phi \in \mathbb{R}^m$ is the *outcome space* of the model (where m is the number of observables), and $f : \Theta \to \Phi$ is a mapping, called the *prediction function*, assigning to each parameter vector $\theta \in \Theta$ a corresponding element $\phi \in \Phi$ of the outcome space. As far as probabilistic models like the BLIM are concerned, a point in the outcome space Φ is a probability mass distribution. Particularly, in the BLIM, it is a distribution on the collection of response patterns. A model is said to be *identifiable* whenever f is injective. If a model is not globally identifiable, it is still possible to test whether it is locally identifiable. A model is said to be *locally identifiable at a point* $\theta \in \Theta$ if the prediction function f is injective when restricted to points with distance less than some $\epsilon > 0$ of θ (Bamber & Van Santen, 2000). Moreover, it is *locally identifiable* if local identification holds true at any point of its parameter space.

4. Identifiability of the BLIM

In the case of the BLIM, a point in the parameter space is a vector containing a pair $\beta_q, \eta_q \in (0, 1)$, such that $\beta_q + \eta_q < 1$, for each item $q \in Q$ and a probability $\pi(K) \in (0, 1)$ for each knowledge state, with the trivial restriction $\sum_{K \in \mathcal{K}} \pi(K) = 1$. Let Θ be the parameter space of the BLIM. Given a parameter vector $\theta \in \Theta$, the prediction function $f(\theta)$ of the BLIM is a probability distribution on the response patterns $R \subseteq Q$. Given a single response pattern R,

$$f_R(\theta) = \sum_{K \in \mathcal{K}} P(R|K)\pi_K.$$

The BLIM is a particular case of latent class models (Lazarsfeld & Neil, 1968). Although identifiability problems are a well-known issue in latent class analysis (McHugh, 1956), the BLIM deserves separate investigation for various reasons. In the first place it introduces the restriction of local independence and the link between conditional probabilities of the response patterns and error rates. Furthermore, as pointed out by Spoto et al. (2012), identification of the BLIM strongly depends on the specific knowledge structure to which the model is applied. This is a specific feature of this model that is not shared by latent class models in general. Given a finite fixed set Q of items, the collection of all knowledge structures on Q could, in principle, be partitioned into two subsets: one containing knowledge structures for which the BLIM is identifiable and