ARTICLE IN PRESS

Journal of Mathematical Psychology ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

Identifiability in probabilistic knowledge structures

Jürgen Heller

Department of Psychology, University of Tübingen, Schleichstr. 4, 72076 Tübingen, Germany

HIGHLIGHTS

- The basic local independence model (BLIM) is a probabilistic knowledge structure.
- The paper provides a theoretical treatment of the (local) identifiability of the BLIM.
- It derives conditions giving rise to global non-identifiability, some of them new.
- This provides a full account of the occurring parameter trade-offs.
- Conjectures on the identifiability of the BLIM are shown to be wrong, and are revised.

ARTICLE INFO

Article history:
Received 26 September 2012
Received in revised form
20 July 2016
Available online xxxx

Keywords:
Knowledge spaces
Learning spaces
Basic local independence model
Jacobian matrix
Local identifiability
Global identifiability

ABSTRACT

The so-called basic local independence model (BLIM) constitutes the standard probabilistic model within the theory of knowledge structures. The present paper characterizes local identifiability of the BLIM through the rank of its Jacobian matrix. Within this framework, it reconsiders conditions known to give rise to non-identifiability, and presents some new cases. Together they completely cover the instances cropping up in the collection of BLIMs arising from all the possible knowledge structures on a three-item domain. The derived theoretical results, providing a full account of the trade-offs between parameters that occur in these situations, hold for arbitrary BLIMs, and are not limited to domains of particular cardinality. Moreover, it is shown that previously formulated conjectures on the encountered types of parameter trade-offs need not hold on the whole parameter space, but in general may only be true almost everywhere. If this indeed is the case, remains an open problem.

Knowledge structures provide a highly flexible set-theoretic framework for representing the organization of knowledge elements in a domain. Although they were first introduced by Doignon and Falmagne (1985) within a purely deterministic approach, this perspective was broadened later on by developing a probabilistic framework on top of them (Falmagne & Doignon, 1988a,b). The so-called *basic local independence model* (BLIM; Doignon & Falmagne, 1999) forms the standard probabilistic model within the theory of knowledge structures. Applications of the BLIM raise several questions. How to determine the underlying knowledge structure, and given that, what are the parameter values of the probabilistic model? It is the second question that the current paper addresses. The focus, however, is not on suggesting a particular method for properly estimating the parameters, but to tackle the more fundamental problem of determining whether the estimates obtained with any such method are unique. This property is known as the identifiability of a parametric model. Any such model is considered identifiable when there are no two different sets of parameter values that lead to the same model prediction.

Bamber and van Santen (1985, 2000) formalize these considerations and provide general results concerning the identifiability of a *parametric model*, which they define as a triple $\langle \Theta, f, \Phi \rangle$. Here the set $\Theta \subseteq \mathbb{R}^n$ is considered to form the *parameter space*, the set $\Phi \subseteq \mathbb{R}^m$ is called the *outcome space*, and the *prediction function* $f : \Theta \to \Phi$ is a mapping from the parameter space into the outcome space. A model $\langle \Theta, f, \Phi \rangle$ then is said to be *(globally) identifiable* if its prediction function f is one-to-one (i.e., injective). It is *locally identifiable* at a point $\theta_0 \in \Theta$ if there is an open neighborhood of θ_0 so that f restricted to that neighborhood is one-to-one (cf. Bamber & van Santen, 1985, 2000). There are results showing that the BLIMs induced by knowledge structures of a particular type exhibit non-identifiability (Spoto, Stefanutti, & Vidotto, 2012, 2013). The following goes beyond these results by providing a complete account of the parameter trade-offs that occur in these as well as other cases, which have not been covered yet.

E-mail address: juergen.heller@uni-tuebingen.de.

http://dx.doi.org/10.1016/j.jmp.2016.07.008

1. Probabilistic knowledge structures

Doignon and Falmagne (1985) define a *knowledge structure* as a pair $\langle Q, \mathcal{K}\rangle$ in which Q is a nonempty set (assumed to be finite throughout the paper), and $\mathcal{K}$ is a family of subsets of Q, containing at least Q and the empty set $\emptyset$. The set Q is called the *domain* of the knowledge structure. Its elements are referred to as *items* and the subsets in the family $\mathcal{K}$ are labeled *(knowledge) states*. A knowledge state represents the subset of items in the considered domain that an individual masters.

In practical applications, we may not assume that a person solves a problem if and only if the person masters this problem (i.e. it is an element of the person's knowledge state). There are two kinds of errors that can occur. In case of a careless error the person actually masters an items, but does not solve it, whereas solving an item without actually mastering it is called a lucky guess. Handling these types of errors calls for a probabilistic framework. Within such an approach we can also take into account that knowledge states will not occur with equal probability.

The introduction of a probabilistic framework is based on dissociating the knowledge state K of a person from the actual given response pattern R. Let $\mathcal{R} = 2^Q$ denote the set of all possible response patterns on the domain Q. Given a knowledge structure $\langle Q, \mathcal{K}\rangle$ we have to define an appropriate probability space on the set of outcomes $\mathcal{R} \times \mathcal{K}$ that specifies the joint probability $P(R, K)$ of observing response pattern $R \in \mathcal{R}$ and knowledge state $K \in \mathcal{K}$. Falmagne and Doignon (1988a,b) uniquely determine these joint probabilities by specifying a (marginal) distribution $P_{\mathcal{K}}$ on the states of $\mathcal{K}$, and the conditional probabilities $P(R \mid K)$ for all $R \in \mathcal{R}$ and $K \in \mathcal{K}$. The marginal distribution $P_{\mathcal{R}}$ on $\mathcal{R}$ then is given by

$$P_{\mathcal{R}}(R) = \sum_{K\in\mathcal{K}} P(R \mid K) \cdot P_{\mathcal{K}}(K). \tag{1}$$

The probabilistic model that received most interest satisfies the following additional conditions. First, it is assumed that, given the knowledge state $K \in \mathcal{K}$, the solution behavior is stochastically independent over items (conditional stochastic independence). This may be expressed by

$$P(R \mid K) = \prod_{q\in Q} p_{q,R,K}, \tag{2}$$

with real parameters $0 \leq p_{q,R,K} \leq 1$. Second, it is assumed that the parameters $p_{q,R,K}$ actually do not depend upon the knowledge state K, but only upon item-specific parameters β_q and η_q, which are interpreted as the probabilities of a careless error and a lucky guess on item $q \in Q$, respectively. This means that in (2) we have

$$p_{q,R,K} = \begin{cases} \beta_q & \text{if } q \notin R,\ q \in K \\ 1-\beta_q & \text{if } q \in R,\ q \in K \\ \eta_q & \text{if } q \in R,\ q \notin K \\ 1-\eta_q & \text{if } q \notin R,\ q \notin K \end{cases} \tag{3}$$

with $0 \leq \beta_q < 1$ and $0 \leq \eta_q < 1$. In the sequel, consideration will be confined to the class of models for which (1), (2) and (3) hold, and which are called *basic local independence models* (BLIM).

It is obvious from the above definitions that a BLIM can be considered a latent class model (Andersen, 1982; Dayton, 1998; Goodman, 1978; Lazarsfeld & Henry, 1968; Vermunt & Magidson, 2004) with the knowledge states forming the latent classes. In fact, with (2) and (3) holding, a restricted latent class model is obtained (Schrepp, 2005; Ünlü, 2011). Notice that in contrast to general latent class analysis, the knowledge states impose a structure on the set of latent classes which puts constraints on the conditional probabilities of the response patterns given the latent classes. Within the limited scope of linearly ordered latent classes this idea has been implemented in some particular models (Dayton & Macready, 1976; Lazarsfeld & Henry, 1968; Proctor, 1970) that were intended to provide probabilistic versions of Guttman scaling (Guttman, 1950). For details see Ünlü (2011).

Various approaches to estimating the parameters of a BLIM have been suggested (Heller & Wickelmaier, 2012; Schrepp, 1999; Stefanutti & Robusto, 2009). All of these contributions do not touch upon the problem of identifiability. There is, however, work showing that there are non-identifiable BLIMs (Spoto et al., 2012, 2013), and software to check given BLIMs (of moderate size) for non-identifiability (Stefanutti, Heller, Anselmi, & Robusto, 2012). While this proves their existence, the following for the first time provides a full theoretical account of the parameter trade-offs in those kind of structures as well as others, which have not been considered before. These cases cover all the instances cropping up in the collection of BLIMs arising from all the possible knowledge structures on a three-item domain. Finally, the paper elaborates on results characterizing important classes of knowledge structures, for which the BLIM built upon them is not identifiable, and provides new results on the possible types of parameter trade-offs that can occur.

2. Setting up the framework

Let Q be a domain and $\mathcal{K}$ a knowledge structure on Q. Characterizing the BLIM induced by the knowledge structure $\mathcal{K}$ as a parametric model amounts to specify its parameter and outcome space as well as its prediction function (e.g., Bamber & van Santen, 1985).

2.1. Parameter space

Some of the constraints defining the admissible parameter combinations introduced in this section are quite obvious, while others are more or less tacit assumptions that have not been explicitly addressed in large parts of the literature (but see Heller & Repitsch, 2012).

Let $\beta = (\beta_q)_{q\in Q}$ and $\eta = (\eta_q)_{q\in Q}$ denote the parameter vectors of the item-specific careless error and guessing probabilities, respectively. Concerning the state probabilities, in order to assure independence, all but one of them are included into the corresponding parameter vector π. The natural choice for excluding a state from $\mathcal{K}$ is among those subsets of the domain that are definitely in $\mathcal{K}$, which are the naïve state $\emptyset$ and the state of full mastery Q. The following mainly refers to $\mathcal{K}^* = \mathcal{K} \setminus \{Q\}$ (but see Section 4.2). Thus, let $\pi = (\pi_K)_{K\in\mathcal{K}^*}$ denote the parameter vector of independent state probabilities $\pi_K = P_{\mathcal{K}}(K)$, $K \in \mathcal{K}^*$.

In general, the parameter vectors $\theta = (\beta, \eta, \pi)$ consist of $n = 2 \cdot |Q| + |\mathcal{K}| - 1$ components, with n characterizing the number of free parameters. Several constraints apply that restrict the parameter space to a proper subset of the n-dimensional unit hypercube. First, all parameters are conceived as elements of the open real interval $(0, 1)$. Second, for all $K \in \mathcal{K}^*$ we have

$$\sum_{L\in\mathcal{K}^*} \pi_L < 1. \tag{C1}$$

Third, the theory of knowledge structures heavily relies on natural assumptions on the parameters β_q and η_q, $q \in Q$, that, however, most of the time remain implicit. Consider the parameter restriction captured by inequality

$$\beta_q + \eta_q < 1 \quad \text{for all } q \in Q, \tag{C2}$$

which means nothing else but that a correct response is more likely if the item is mastered than if it is not mastered. This constraint is at the very heart of the idea of a knowledge state, and