Empirical validation of a knowledge structure for assessment and learning of psychometrics at university level

Luca Stefanutti and Francesca Cristante
University of Padova, Department of Psychology
Via Venezia, 8
35131 Padova, Italy
luca.stefanutti@unipd.it, francesca.cristante@unipd.it

Introduction

In this paper we present a preliminary study aimed at the construction of an intelligent tutoring/assessment system for the courses of statistics, psychometrics and data analysis of the Faculty of Psychology of the University of Padova, Italy. Our intention is to base such a system on some modern psychometric models and methods provided by the theory of knowledge spaces (Doignon & Falmagne, 1985; Albert & Lukas, 1998; Doignon & Falmagne, 1999). According to this theory, a knowledge domain is a set Q of all problems characterizing a given discipline, and the knowledge state of a student is the set $K \subseteq Q$ of problems that this student is capable of solving. A knowledge structure is then the collection $\mathcal{K}$ of all knowledge states that can be observed in a certain population of students. If this collection is closed under union ($K, K' \in \mathcal{K}$ implies $K \cup K' \in \mathcal{K}$) then it is called a knowledge space.

Knowledge spaces are in 1–1 correspondence with a particular kind of binary relations on the set Q of items, called *entail relations* (Koppen & Doignon, 1990). In the simplest case, two items p and q in Q are in an entail relation $\mathcal{E}$ (formally, $p\mathcal{E}q$) if failing item p 'entails' failing item q. In this sense, item p contains prerequisites for solving item q.

If an entail relation — or, equivalently, the corresponding knowledge space — is available for a given set Q of items, then it can be efficiently used in the assessment of the knowledge state of a student. As an example, if $p\mathcal{E}q$ and a student has just provided a wrong response to item p, then the response of that student to q can be obtained by inference. Thus, in general, the whole knowledge state of the student can be recovered presenting her/him with a subset of the items in Q. This subset is usually expected to be much smaller than Q making thus the whole assessment process very efficient especially with large sets of items. Falmagne and Doignon (1988) provide algorithms for the efficient assessment of the knowledge state of a student by means of a knowledge space.

A knowledge space on a set Q of items can be constructed through the computerized query of a set of experts on the corresponding entail relation (see, e.g., Koppen, 1993; Dowling, 1993; Stefanutti & Koppen, 2003). This method was used to build a knowledge space for a set of problems in "data analysis in psychology" (Stefanutti & Cristante, 2001) and this paper focuses on the empirical validation of this structure by means of particular probabilistic models that are introduced in the next section.

Probabilistic models

A knowledge structure (or a knowledge space) is a deterministic model specifying the relationships among a set of problems in terms of prerequisites and background knowledge. This deterministic representation, however, cannot be validated against empirical data directly. In fact, empirical observations are affected by noise and random error and thus it makes sense to shift from a deterministic framework to a non-deterministic (probabilistic) one. Doignon and Falmagne (1999) introduced a probabilistic model which is appropriate for the empirical validation of a knowledge structure (probabilistic knowledge structures). In our application we used a logistic extension of the above-mentioned model proposed by Stefanutti (2003). In this section we provide a short presentation of this model.

We consider a set Q of *items*, a set S of *skills* and a collection $\mathcal{K}$ of *states* and we use the following notation: $Q := \{q_1, q_2, \ldots, q_m\}$, $S := \{s_1, s_2, \ldots, s_p\}$, $\mathcal{K} := \{K_1, K_2, \ldots, K_n\}$, where m, p, and n are the cardinalities of Q, S, and $\mathcal{K}$ respectively. Then we introduce two binary relations: a relation $\mathcal{V} \subseteq Q \times \mathcal{K}$ between items and states, and a relation $\mathcal{W} \subseteq S \times \mathcal{K}$ between skills and states. In an usual interpretation, $\mathcal{W}$ is the membership relation '$\in$', so that $\mathcal{K}$ si a collection of subsets of S, and $s\mathcal{W}K$ iff $s \in K$ for all $s \in S$, and $K \in \mathcal{K}$. On the other hand, if $f : Q \to 2^{\mathcal{K}}$ is a mapping such that, for all $q \in Q$ and $K \in \mathcal{K}$,

$$K \in f(q) \quad \Longleftrightarrow \quad q\mathcal{W}K,$$

then f is a *skill multi-map* in the sense specified by Doignon and Falmagne (1999).

Each skill $s_l \in S$ is characterized by a parameter $\delta_l \in \mathbb{R}$ specifying the *difficulty* to master (or to learn) that skill. Then, we assume a probability distribution on the states in $\mathcal{K}$, and the probability that a student – randomly selected from the population – is in state K_j is given by

$$\pi_j := \frac{\exp(-\xi_j)}{\sum_{v=1}^{n} \exp(-\xi_v)}, \quad \text{where, for all } K_v \in \mathcal{K} \quad \xi_v = \sum_{l=1}^{p} w_{vl}\delta_l \quad \text{for} \quad w_{vl} = \begin{cases} 1 & \text{if } s_l\mathcal{W}K_v, \\ 0 & \text{otherwise,} \end{cases}$$

is the *difficulty* of state K_v.

The *response pattern* of a student is the subset $R \subseteq Q$ of all items that obtained a correct response from that student, and the collection of all response patterns is the power-set 2^Q. Given the probabilities π_j of the states, the probability θ_i of a response pattern $R_i \subseteq Q$ is specified through the local independence model of Doignon and Falmagne (1999):

$$\theta_i := \sum_{j=1}^{n} \rho_{ij}\pi_j,$$

where ρ_{ij} is the conditional probability that a student exhibits pattern R_i given that her/his state is K_j. This conditional probability is specified by means of two parameters of the items $q_k \in Q$: a parameter η_k denoting the probability of a *careless error* for item q_k, and a parameter γ_k denoting the probability of a *lucky guess* for q_k. The details of the derivation of ρ_{ij} from the parameters η_k and γ_k are beyond the scope of the present paper, and they can be found in Doignon and Falmagne (1999). Here, for the sake of completeness, we just give the general equation of the local independence model:

$$\rho_{ij} = \left[\prod_{j=1}^{p} \eta_k^{(1-r_{ik})w_{jk}}\right] \left[\prod_{j=1}^{p} (1-\eta_k)^{r_{ik}w_{jk}}\right] \left[\prod_{j=1}^{p} \gamma_k^{r_{ik}(1-w_{jk})}\right] \left[\prod_{j=1}^{p} (1-\gamma_k)^{(1-r_{ik})(1-w_{jk})}\right],$$

where

$$r_{ik} := \begin{cases} 1 & \text{if } q_k \in R_i, \\ 0 & \text{otherwise.} \end{cases}$$

The parameters that have to be estimated for this model are the difficulties δ_l of the skills and the error probabilities η_k and γ_k of the items. There are thus $2n + p$ parameters to estimate in the whole. Given a suitable random sample of the population of students, the goodness of fit of the model can then be tested through standard statistics like chi-square or the likelihood ratio test, by comparing the observed frequencies F_i of the response patterns R_i in the sample, with the frequencies $N\theta_i$ predicted by the model, where N is the sample size. In both cases, the degrees of freedom are $d.f. = |2^Q| - 2n - p - 1$.

Results

In this preliminary study a subset of 10 items were selected among a large set of problems that are typical in a course of "Techniques of Psychological Research and Data Analysis" of the Faculty of Psychology of the University of Padova, Italy. The items cover a basic and an intermediate level of the course, and their topic is "basic probability theory", "contingency tables" and "log-linear analysis". As an example, the first of them is displayed below:

> "In a group of managers, a percent of 60% speaks two languages. Among all managers, 55% of them works often abroad, and the 20.4% is capable of speaking two languages and is often abroad. If a manager is picked at random, which is the probability that this manager does not work often abroad given that s/he does not speak two languages?"

A knowledge space was obtained for the above-mentioned set of items through a computerized query of five experts of the domain of knowledge. We present in this section the results obtained in the empirical validation of this knowledge space.

A test Q containing the 10 selected items was administered to a group of 400 Italian students attending the course of "Techniques of Psychological Research and Data Analysis". Thus, the data collected from these subjects consisted in a set of 400 response patterns. The knowledge space on the 10 items was then tested empirically through the logistic model presented in the previous section. The parameters of the model were estimated by maximum likelihood, and the fit of the model was tested by means of the likelihood ratio statistic.

The logistic model requires the specification of a set S of skills and the two relations $\mathcal{V}$ and $\mathcal{W}$. In this application we assume the existence of a set S of 10 skills and a bijective mapping $f : S \to Q$ so that there is one skill for each of the 10 items in Q. Moreover, we assume that, for all items $q \in Q$, all states $K \in \mathcal{K}$ and all skills $s \in S$:

$$q\mathcal{V}K \iff q \in K, \qquad \text{and} \qquad s\mathcal{W}K \iff f(s) \in K.$$

Thus $\mathcal{V}$ can be interpreted as the membership relation between Q and $\mathcal{K}$, where $\mathcal{K}$ is a collection of subsets of Q.

The available dataset (400 students) was not very large, if compared with the number of theoretical response patterns: for a set of 10 items there are $2^{10} = 1024$ different response patterns. For this reason it was decided to partition the full set Q of 10 items into two disjoint subsets $Q_1 := \{1, 2, 3, 4, 5\}$ and $Q_2 := \{6, 7, 8, 9, 10\}$ of 5 items each. The corresponding knowledge spaces $\mathcal{K}_1$ and $\mathcal{K}_2$, obtained by

$$\mathcal{K}_1 := \{K \cap Q_1 : K \in \mathcal{K}\}, \quad \text{and} \quad \mathcal{K}_2 := \{K \cap Q_2 : K \in \mathcal{K}\}$$

were then tested separately. The logistic model was thus applied to the following knowledge spaces:

$$\begin{aligned}
\mathcal{K}_1 := \quad & \{\emptyset, \{4\}, \{5\}, \{2,5\}, \{3,4\}, \{4,5\}, \{1,3,4\}, \{2,3,4\}, \{3,4,5\}, \\
& \{1,2,3,4\}, \{1,3,4,5\}, \{2,3,4,5\}, Q_1\}, \\
\mathcal{K}_2 := \quad & \{\emptyset, \{7\}, \{9\}, \{10\}, \{7,10\}, \{9,10\}, \{7,9\}, \{6,7,10\}, \{7,9,10\}, \\
& \{6,7,8,10\}, \{6,7,9,10\}, \{7,8,9,10\}, Q_2\}
\end{aligned}$$

For all the statistical decisions in testing the model a probability value of .05 was used. An application of the logistic model to the knowledge space $\mathcal{K}_1$ gave a likelihood ratio of 23.46 that with $32 - 10 - 5 - 1 = 16$ degrees of freedom has a probability of .102, thus the model is accepted. Table 1 shows the parameter estimates obtained for the items in Q_1 and Q_2. Each row in the table contains the error probabilities and the difficulty parameter of a single item (remember that skills and items are in a 1–1 correspondence).

1. Parameter estimates for the knowledge space on the 10 items: careless error (η), lucky guess (γ), and difficulty (δ).

Items	η	γ	δ	Items	η	γ	δ
1	0.000	0.085	9.335	6	0.000	0.000	−0.529
2	0.000	0.275	−0.697	7	0.000	0.000	1.897
3	0.048	0.416	−2.314	8	0.000	0.780	−1.514
4	0.029	0.001	5.806	9	0.051	0.186	−0.696
5	0.038	0.000	−3.084	10	0.100	0.317	3.197

The model was then applied with the space $\mathcal{K}_2$ and a likelihood ratio of 23.84 was obtained. With 16 degrees of freedom also this value is not significant, suggesting that the model can be accepted. Observe that, although a pretty good fitness was obtained also for this second space, the lucky guess probability γ_8 is very high (0.78), meaning that — if the model reflects the true structure on the items in $\mathcal{K}_2$ — a student provides a correct response to item 8 by lucky guess with a probability of 78%. Taking into account that item 8 was a multiple choice item with 5 different alternatives, this result would suggest that either the model is not correct or the data for this item are unreliable. Further investigations are needed for this second part of the model in order to establish the source of this inconsistency.

REFERENCES

Albert, D., Lukas, J. (1998). *Knowledge Spaces: Theories, Empirical Research, Applications*. Berlin, Heidelberg: Springer-Verlag.

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

Doignon, J.-P., Falmagne, J.-C. (1999). *Knowledge Spaces*. Berlin, Heidelberg: Springer-Verlag.

Dowling, C.E. (1993). Applying the basis of a knowledge space for controlling the questioning of an expert. *Journal of Mathematical Psychology, 37*, 21–48.

Falmagne, J.-C., Doignon, J.-P. (1988). A markovian procedure for assessing the states of a system. *Journal of Mathematical Psychology, 32*, 232–258.

Koppen, M. (1993). Extracting human expertise for constructing knowledge spaces: an algorithm. *Journal of Mathematical Psychology, 37*, 1–20.

Koppen, M., Doignon, J.-P. (1990). How to build a knowledge space by querying an expert. *Journal of Mathematical Psychology, 34*, 311–331.

Stefanutti, L. (2003). Logistic models for probabilistic knowledge structures. Submitted.

Stefanutti, L., Cristante, F. (2001). Knowledge space construction: testing expert agreement by means of log-linear analysis. Paper presented at the *32nd European Mathematical Psychology Group Meeting*, Lisbon, 26–29 September, 2001.

Stefanutti, L., Koppen, M. (2003). A procedure for the incremental construction of a knowledge space. *Journal of Mathematical Psychology*. In press.