Ünlü, A., & Kickmeier-Rust, M. D. (2006). Prediction of solvability dependencies between dichotomous test items: A local order-theoretic measure of association. In A. Rizzi & M. Vichi (Eds.), COMPSTAT 2006: Proceedings in Computational Statistics (pp. 631–638). Heidelberg, Germany: Physica/Springer-Verlag.

Prediction of Solvability Dependencies between Dichotomous Test Items: A Local Order-Theoretic Measure of Association

Ali Ünlü[1] and Michael D. Kickmeier-Rust[2]

[1] Department of Psychology, University of Graz, Universitätsplatz 2/III, A-8010 Graz, AUSTRIA ali.uenlue@uni-graz.at

[2] Department of Psychology, University of Graz, Universitätsplatz 2/III, A-8010 Graz, AUSTRIA michael.kickmeier@uni-graz.at

Summary. Solvability dependencies between dichotomous test items play an important role in the psychometric theory of knowledge spaces. Knowledge space theory (KST), based on hypothesized solvability dependencies between dichotomous items, has been successfully applied for the computerized, adaptive assessment and training of knowledge. For instance, see the ALEKS system, a fully automated math tutor on the Internet: http://www.aleks.com/. A crucial problem in KST, however, is the empirical detection of such dependencies. In this paper, we present an easy-to-use association measure for the detection of solvability dependencies from multinomial response data. The approach is illustrated with simulated data using a general KST finite mixture latent variable model.

Key words: Dichotomous test item; Solvability dependency; Association measure; Psychometrics; Knowledge space theory; Order theory; Multinomial probability model; Latent variable; Maximum likelihood; Simulation

1 Introduction

Suppose a teacher is probing a student's knowledge of, say, elementary Euclidean geometry. The teacher asks a question, and depending on the student's response, selects the next question to be asked. That way the teacher tries to *efficiently* diagnose the student's state of knowledge capable of explaining all the responses.

This simple example indicates the implicit use of *solvability dependencies* between test items (questions) of the following type: An examinee who is able to solve question J is also able to solve question I (abbreviated $I \leq J$). Such dependencies for a given set Q of items are modeled by a surmise relation (quasi order; see Definition 3) $\leq$ on Q. Surmise relations are mathematical models that belong to the theory of knowledge spaces (see Section 2).

A crucial problem in this combinatorial, qualitative psychometric theory is the empirical detection of such dependencies. Given a set of test items and the responses of a sample of subjects, how can solvability dependencies between the items be detected from the data? Approaches so far have been rather insufficient (for a critical analysis, see, e.g., [UeA04]), or theoretically restrictive and/or technically/computationally intensive (for a comprehensive list of references, see [DF99]).

The aim of this paper is to propose a simple but effective data-analytic method to detect solvability dependencies between dichotomous test items.

2 Knowledge Space Theory

Knowledge space theory (KST) was introduced by Doignon and Falmagne [DF85, DF99]. This section reviews basic deterministic and probabilistic concepts of KST which are relevant for this work.

2.1 Basic Deterministic Concepts

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

Definition 1. *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 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. We also say that $\mathcal{K}$ is a knowledge structure on Q.*

The set Q is supposed to be a set of *dichotomous* items. In this paper, we interpret Q as a set of questions/problems that can either be solved or not be solved. For a set X, let 2^X denote its power-set. The observed responses of a subject are represented by the subset $R \subset Q$ containing exactly the items solved by the subject (*response pattern*). Similarly, the true latent state of knowledge of a subject is represented by the subset $K \subset Q$ containing exactly the items the subject is capable of mastering (*knowledge state*). Given a knowledge structure $\mathcal{K}$, we assume that the only states of knowledge possible are the ones in $\mathcal{K}$.

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

There is also the concept of a *surmise relation*.

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

As mentioned in Section 1, a surmise relation may model a latent hierarchy among the test items based on solvability dependencies.

Birkhoff's theorem (Theorem 1; see [Bir37]) provides a linkage between quasi ordinal knowledge spaces and surmise relations on an item set.

Theorem 1. *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 $\leq$ on Q. Such a correspondence is defined through the two equivalences ($p, q \in Q$, $K \subset Q$):*

$$p \leq q :\Longleftrightarrow [\forall K \in \mathcal{K} : \{q \in K \Longrightarrow p \in K\}] \, ;$$
$$K \in \mathcal{K} :\Longleftrightarrow [\forall (p \leq q) : \{q \in K \Longrightarrow p \in K\}] \, . \quad \square$$

2.2 Basic Probabilistic Concepts

We fix a population of reference and examinees are randomly drawn from the population. We consider a random sample of size $N \in \mathbb{N} := \{1, 2, \ldots\}$. The *response data* are the absolute counts $N(R) \in \mathbb{N}_0$ of response patterns $R \in 2^Q$; that is, $\mathbf{x} := (N(R))_{R \in 2^Q}$. We assume that the data are the realization of a *multinomially* distributed random vector, $\mathbf{X} := (X_R)_{R \in 2^Q}$; that is,

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

where $\rho(R) > 0$ is the true probability of occurrence of a response pattern $R \in 2^Q$.

In Section 4, we simulate data using a *basic local independence model.*

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

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,*

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

with constants $\beta_q, \eta_q \in [0, 1[$ for each $q \in Q$, respectively called careless error and lucky guess probabilities 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 (point 2). For a response pattern $R \in 2^Q$ and a state $K \in \mathcal{K}$, $r(R, K)$ specifies the conditional probability of response pattern R for a subject in state K (point 3). The item responses of a subject are assumed to be independent given the knowledge state of the subject, and the response error probabilities β_q, η_q ($q \in Q$) are attached to the items and do not vary with the knowledge states (point 4).

The BLIM represents a finite mixture latent variable model which can be also viewed as a constrained latent class model (see [Uen06]).

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

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

3 Measure of Association α

In this section, we propose a measure of association, α, for the prediction of solvability dependencies between dichotomous test items. This measure is derived in the framework of an operational *prediction paradigm* based on the unitary method of *proportional reduction in predictive error* (PRPE).

The method of PRPE was originally introduced by Guttman [Gut41], and it was systematically applied in the series of papers by Goodman and Kruskal [GK54, GK59, GK63, GK72].

3.1 Prediction Paradigm and Method of PRPE

Let $Q := \{I_l : 1 \leq l \leq m\}$ be a set of $m \in \mathbb{N}$ dichotomous test items. For any $I_s, I_t \in Q$, let $\rho_{1.}$, $\rho_{.1}$, and ρ_{11} be the marginal totals in $I_s = 1$, $I_t = 1$, and $I_s = I_t = 1$ respectively, taken in the entire m-dimensional cross classification of probabilities $\rho(R)$ ($R \in 2^Q$). In other words, $\rho_{1.} = \sum_{R \in 2^Q, I_s \in R} \rho(R)$, $\rho_{.1} = \sum_{R \in 2^Q, I_t \in R} \rho(R)$, and $\rho_{11} = \sum_{R \in 2^Q, I_s, I_t \in R} \rho(R)$.

The derivation of α rests on the following *prediction paradigm.* Assume a subject is randomly chosen from the population of reference, and we are asked to guess whether this subject solves item I_s, given either

(no info). no further information (than the multinomial distribution on the response patterns); or

(info). further information that this subject has already solved item I_t.

Suppose we follow the *prediction strategy* of guessing item I_s is solved by the subject. What are then the probabilities of a *prediction error* in both cases? In the 'no info' case, the probability of a prediction error is $1-\rho_{1.}$, and in the 'info' case, it is $1-\rho_{11}/\rho_{.1}$ (note, $\rho_{.1} \neq 0$).

The general probability formula of the method of PRPE quantifies the *predictive utility*, PU(info), of given information. Informally,

$$PU(\text{info}) := \frac{\text{Prob. of error (no info)} - \text{Prob. of error (info)}}{\text{Prob. of error (no info)}}.$$

3.2 Definition, Properties, and Maximum Likelihood Estimate

Inserting the afore mentioned prediction error probabilities into the PRPE formula of predictive utility, we obtain the population analogue of α.

Definition 5. *Let $I_s, I_t \in Q$. Let $\rho_{1.} \neq 1$ and $\rho_{.1} \neq 0$. The measure α is defined as*

$$\alpha := \frac{(1-\rho_{1.}) - (1-\rho_{11}/\rho_{.1})}{1-\rho_{1.}} = \frac{\rho_{11}/\rho_{.1} - \rho_{1.}}{1-\rho_{1.}}.$$

The next lemma formulates obvious properties of α.

Lemma 1. *Let $I_s, I_t \in Q$. If α is defined, we have*

1. $-\infty < \alpha \leq 1$;
2. $\alpha = 0 \Longleftrightarrow I_s, I_t$ *statistically independent;*
3. $\alpha = 1 \Longleftrightarrow \rho_{11} = \rho_{.1}$ *(sure predictability);*
4. α *is invariant with respect to permutations of rows and/or columns of the two-by-two table corresponding to the items I_s and I_t.* □

Maximum likelihood estimates (MLEs) for the population probabilities $\rho(R)$ $(R \in 2^Q)$ are given by the relative frequencies $\widehat{\rho(R)} = N(R)/N$.

Let $I_s, I_t \in Q$. Inserting these MLEs, we obtain the MLE for α (if defined):

$$\widehat{\alpha} = \frac{NN_{11} - N_{1.}N_{.1}}{N_{.1}(N - N_{1.})},$$

where $N_{1.}$, $N_{.1}$, and N_{11} are the absolute counts of subjects solving I_s, I_t, and both items, respectively.

4 Simulation Example

To investigate the merit of α, we ran four simulation trials based on a surmise relation to be detected from the simulated data. Simulations were realized using `Microsoft VB.NET`.[3]

[3] Source files for all the computations in this section are freely available from the authors.

4.1 Data Generating Models

We considered the quasi ordinal knowledge space

$$\mathcal{K} := \{\emptyset, \{a\}, \{c\}, \{a, b\}, \{a, c\}, \{a, b, c\}, Q\}$$

defined on the domain $Q := \{a, b, c, d\}$ of four dichotomous test items. The surmise relation derived from this quasi ordinal knowledge space according to Theorem 1 is depicted in Fig. 1.

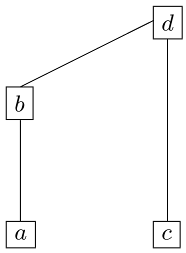

Fig. 1. Hasse diagram of the surmise relation derived from $\mathcal{K}$ according to Birkhoff's theorem (Theorem 1).

In each simulation trial, the aim was to detect this underlying surmise relation (i.e., set of solvability dependencies) from the simulated data. Based on the deterministic model $\mathcal{K}$, we used the following four BLIM specifications BLIM1, BLIM2, BLIM3, and BLIM4 for the respective simulation trials.[4]

BLIM1: Uniform probability distribution on $\mathcal{K}$—that is, $p(K) = 1/7$ for any $K \in \mathcal{K}$—; the same careless error $\beta = 0,01$ and lucky guess $\eta = 0,01$ probabilities for each item.
BLIM2: Uniform probability distribution on $\mathcal{K}$; the same careless error $\beta = 0,02$ and lucky guess $\eta = 0,07$ probabilities for each item.
BLIM3: Uniform probability distribution on $\mathcal{K}$; the same careless error $\beta = 0,30$ and lucky guess $\eta = 0,00$ probabilities for each item.
BLIM4: Uniform probability distribution on $\mathcal{K}$; the same careless error $\beta = 0,10$ and lucky guess $\eta = 0,01$ probabilities for each item.

In each of the four trials, based on the corresponding BLIM specification, we simulated ten binary 100×4 data matrices, each representing the response patterns of $N = 100$ fictitious subjects to $m = 4$ dichotomous test items.

[4] Note that the response error rates were chosen from *psychologically realistic* ranges: Careless error rates ranged from 1 to 30 percent, lucky guess rates from 0 to 7 percent. (From an empirical point of view, guessing effects can be nearly eliminated by appropriate item formulation.)

4.2 Results

In each simulation trial, for any pair of items, we calculated ten α values (given ten data matrices), and subsequently derived the arithmetic mean of these values. In Table 1, the relational matrices for the four simulation trials are presented, containing these average α values.

Table 1. Simulation results for the four trials with respective BLIM specifications BLIM1, BLIM2, BLIM3, and BLIM4.[a]

	BLIM1 $\beta = 0,01,\ \eta = 0,01$					**BLIM2** $\beta = 0,02,\ \eta = 0,07$			
	a	b	c	d		a	b	c	d
a	1,00	0,92	−0,09	0,97	a	1,00	0,66	−0,10	0,73
b	0,21	1,00	0,02	0,86	b	0,20	1,00	0,00	0,68
c	−0,06	0,03	1,00	0,84	c	−0,05	0,04	1,00	0,69
d	0,04	0,13	0,07	1,00	d	0,03	0,12	0,07	1,00
	BLIM3 $\beta = 0,30,\ \eta = 0,00$					**BLIM4** $\beta = 0,10,\ \eta = 0,01$			
	a	b	c	d		a	b	c	d
a	1,00	0,32	−0,03	0,49	a	1,00	0,67	−0,09	0,72
b	0,15	1,00	0,03	0,54	b	0,19	1,00	0,00	0,75
c	0,02	−0,05	1,00	0,51	c	−0,04	−0,02	1,00	0,74
d	0,01	0,08	0,04	1,00	d	0,03	0,08	0,05	1,00

[a] In each specification, the same underlying deterministic knowledge space model $\mathcal{K}$ is used, a uniform probability distribution on the knowledge states, and the same careless error (β) and lucky guess (η) rates for each item. Highlighted cells denote pairs of items with high association-degree of solvability dependency as measured by the coefficient α, as well as pairs of items in the surmise relation derived from $\mathcal{K}$.

The results show that in all trials α allowed for the detection of the surmise relation (i.e., set of solvability dependencies between a, b, c, and d) underlying the simulated data. Even when a comparably high (psychologically unrealistic) careless error probability of 30 percent was applied in third trial specification BLIM3, α could differentiate comparable and noncomparable item pairs. The ability to differentiate was clearer in trials with response error probabilities that may occur in practice; for instance, a careless error rate of 10 percent and a lucky guess rate of 1 percent in fourth trial specification BLIM4.

5 Conclusion

In this paper, we have introduced a new measure α designed to quantify the association-degree of solvability dependency between two dichotomous items. We have seen that α could be successfully applied for the detection of a surmise relation from simulated data using a general KST finite mixture latent variable model (BLIM). The detection of a surmise relation is a crucial and prevalent problem in KST and its applications (see Section 1). This paper thus provides a novel, simple but effective data-analytic method to approach this problem.

Of course, the current simulation studies are a starting point for more in-depth analyses of the measure α. Future research may address the effects of the variation of sample size (especially small sample sizes), the underlying surmise relation model, and the BLIM parameters. In particular, inferential statistics (e.g., confidence intervals) and application to real psychological test data are important and indispensable directions for further analyses.

Acknowledgements. This work was supported by the Austrian Science Fund (FWF) project grant P17071-N04 to Dietrich Albert.

References

[Bir37] Birkhoff, G.: Rings of sets. Duke Mathematical Journal, **3**, 443–454 (1937)

[DF85] Doignon, J.-P., Falmagne, J.-Cl.: Spaces for the assessment of knowledge. International Journal of Man-Machine Studies, **23**, 175–196 (1985)

[DF99] Doignon, J.-P., Falmagne, J.-Cl.: Knowledge Spaces. Springer, Berlin Heidelberg New York (1999)

[GK54] Goodman, L.A., Kruskal, W.H.: Measures of association for cross classifications. Journal of the American Statistical Association, **49**, 732–764 (1954)

[GK59] Goodman, L.A., Kruskal, W.H.: Measures of association for cross classifications, II: Further discussion and references. Journal of the American Statistical Association, **54**, 123–163 (1959)

[GK63] Goodman, L.A., Kruskal, W.H.: Measures of association for cross classifications, III: Approximate sampling theory. Journal of the American Statistical Association, **58**, 310–364 (1963)

[GK72] Goodman, L.A., Kruskal, W.H.: Measures of association for cross classifications, IV: Simplification of asymptotic variances. Journal of the American Statistical Association, **67**, 415–421 (1972)

[Gut41] Guttman, L.: An outline of the statistical theory of prediction. In: Horst, P. (ed) The Prediction of Personal Adjustment. Social Science Research Council, New York (1941)

[Uen06] Ünlü, A.: 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, in press (2006)

[UeA04] Ünlü, A., Albert, D.: The Correlational Agreement Coefficient CA($\leq$,D)—a mathematical analysis of a descriptive goodness-of-fit measure. Mathematical Social Sciences, **48**, 281–314 (2004)