



**Universität Augsburg**

Institut für  
Mathematik

---

---

Ali Ünlü

**A Note on the Connection between Knowledge Structures and  
Latent Class Models**

---

Preprint Nr. 13/2009 — 15. Mai 2009

Institut für Mathematik, Universitätsstraße, D-86135 Augsburg

<http://www.math.uni-augsburg.de/>

---

## **Impressum:**

*Herausgeber:*

Institut für Mathematik

Universität Augsburg

86135 Augsburg

<http://www.math.uni-augsburg.de/pages/de/forschung/preprints.shtml>

*ViSdP:*

Ali Ünlü

Institut für Mathematik

Universität Augsburg

86135 Augsburg

*Preprint:* Sämtliche Rechte verbleiben den Autoren © 2009

# A note on the connection between knowledge structures and latent class models

Ali Ünlü  
Institute of Mathematics  
University of Augsburg  
D-86135 Augsburg, Germany  
E-mail: ali.uenlue@math.uni-augsburg.de  
Phone: +49 821 598 2236  
Fax: +49 821 598 2200

May 9, 2009

## Abstract

This is an addendum to the paper by Schrepp [(2005). About the connection between knowledge structures and latent class models. *Methodology*, 1(3), 93–103]. The note adds additional information about the relationship between knowledge space theory and latent class analysis.

**Keywords:** knowledge space theory, latent class analysis

## Introduction

In Schrepp (2005), the connection between knowledge space theory (KST; Albert et al., 2009; Albert and Lukas, 1999; Doignon and Falmagne, 1985, 1999; Falmagne et al., 2006; Falmagne and Doignon, 2009; Falmagne et al., 1990) and classical latent class analysis (LCA; Andersen, 1982; Dayton, 1998; Goodman, 1978; Hagenaars and McCutcheon, 2002; Lazarsfeld and Henry, 1968; McCutcheon, 1987; Vermunt and Magidson, 2004) was discussed. A probabilistic model for knowledge structures was ‘introduced,’ which, in fact, is a well-known and fundamental model in KST. This model was viewed as a constrained latent class model, and based on that fact, a method was proposed for deriving knowledge structures from data. The relationships to other works were addressed, including the relationship to a latent class

scaling model.

In this note, I give a more comprehensive overview of the literature and the probabilistic models that are at the interface of KST and LCA. The KST models are referred to with their common names, including, for instance, the basic local independence model, which, in a restricted form, is the model used in Schrepp (2005). The view that allows interpreting this model as a constrained latent class model is not a new one and goes back to traditional latent class scaling models. I elaborate on that issue more carefully. KST and LCA are also compared concerning the statistical inference methodologies applied in their fields. Albeit without a reference to LCA, parameter estimation and model testing have also been discussed in KST publications. This does not become clear from Schrepp's text. The connection between KST and LCA at the level of inference methodologies is not astonishing though, as both approaches are special cases of classical maximum likelihood methodology for multinomial probability models. I conclude with an outline and some remarks about possible future research arising from a connection of KST to latent variable modeling approaches.

## Classical unrestricted latent class model

LCA is a statistical approach to examining unobserved categorical variables. The population of reference is assumed to be partitioned into  $T$  mutually exclusive and exhaustive subpopulations (classes)  $C_1, \dots, C_T$ , with unknown proportions  $p(C_t) > 0$ . The latent classes can be viewed as the realizations of a random variable  $X$ . Let  $I_1, I_2, \dots, I_L$  be the dichotomous<sup>1</sup> manifest variables or indicators, with realizations  $i_l \in \{0, 1\}$  ( $1 \leq l \leq L$ ). In vector notation,  $I = (I_l)_{1 \leq l \leq L}$  and  $i = (i_l)_{1 \leq l \leq L}$ . For each of the  $T$  latent classes, there is a set of conditional probabilities for every indicator:  $0 \leq r(I_l = i_l | X = C_t) \leq 1$  for any  $1 \leq l \leq L$ ,  $i_l \in \{0, 1\}$ , and  $1 \leq t \leq T$ . Within any of the  $T$  latent classes, the observed indicator scores are assumed to be independent. This is the assumption of local independence which is fundamental in (classical) LCA.

The classical unrestricted ( $T$ -class) latent class model (LCM) is a multinomial

---

<sup>1</sup>For comparison to KST, only dichotomous indicators are considered.

probability model:

$$\begin{aligned}\rho(i) &= \sum_t \left\{ p(C_t) r(I = i | X = C_t) \right\} \\ &= \sum_t \left\{ p(C_t) \prod_l r(I_l = i_l | X = C_t) \right\},\end{aligned}$$

where  $\rho(i)$  is the occurrence probability of the response pattern  $i = (i_l)$ , and the data—the observed absolute counts of the response patterns—are assumed to follow a multinomial distribution. The unrestricted  $T$ -class LCM contains  $(T - 1) + T \cdot L$  independent model parameters to be estimated from the data.<sup>2</sup> Imposing (e.g., equality or fixed value) restrictions on the parameters of the LCM (some examples follow below) gives restricted LCMs.

In particular, the LCM is a finite mixture model (Everitt and Hand, 1981; McLachlan and Basford, 1988; McLachlan and Peel, 2000; Titterington et al., 1985), and LCA is often referred to as finite mixture modeling. There are various extensions of the classical LCM; for instance, latent class factor models (including more than one latent variable; Magidson and Vermunt, 2001) or latent class regression models (parameters of the regression model may differ across unobserved subgroups; Vermunt and Van Dijk, 2001). In the context of KST specifically, Ünlü (2006) discussed a generalization of the classical model and of the basic local independence and latent class scaling models using random effects, which allows to weaken the assumption of local independence and/or to incorporate covariate information.

## KST as unrestricted and restricted LCA

This section integrates KST probabilistic models into the latent class scaling framework and literature. Viewing the knowledge states of a knowledge structure as the latent classes, the fundamental KST basic probabilistic model and basic local independence model (Doignon and Falmagne, 1999, Chapter 7) can be seen as unrestricted and restricted latent class models, respectively. It is important to note here that the latent classes possess an ‘inner structure,’ composed of the indicators, which determines the constraints imposed

---

<sup>2</sup>More generally, for  $L$  indicators  $I_l$  with corresponding numbers of levels  $D_l$ , this model contains  $(T - 1) + T \cdot \sum_{l=1}^L (D_l - 1)$  independent parameters.

on the conditional probabilities. This view is not a new one and goes back to traditional latent class measurement or scaling 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. They originated as probabilistic generalizations of the deterministic, linear Guttman (1944, 1950) model. For a description of these models and their relationships, see Dayton (1998) and Heinen (1996).

Let the indicators  $I_1, I_2, \dots, I_L$  be dichotomous items; that is, let the domain of questions or problems be  $Q = \{I_l : 1 \leq l \leq L\}$ . Scoring 1 or 0 to an item is interpreted as solving or failing to solve that item, respectively. Let  $\mathcal{K}$  be a knowledge structure on  $Q$ , that is, a family of subsets of  $Q$  containing at least the empty set and  $Q$ . (If no response errors are made, these subsets are the only response patterns possible.) The elements of  $\mathcal{K}$  are called knowledge states. Let the latent classes be represented by the knowledge states; the random variable  $X$  now assumes values in the knowledge structure  $\mathcal{K}$ .

The basic probabilistic model (Doignon and Falmagne, 1999, p. 144)

$$\rho(i) = \sum_{K \in \mathcal{K}} \left\{ p(K) r(I = i | X = K) \right\}$$

obviously is an unrestricted latent class (finite mixture) model. It is a very general model and grounds on the basic idea underlying any type of latent class model. The probability of observing a response pattern  $i$ ,  $\rho(i)$ , is a finite weighted average of the  $|\mathcal{K}|$  class-specific conditional probabilities  $r(I = i | X = K)$ .<sup>3</sup> Under the assumption of local independence, the basic probabilistic model specializes to the classical unrestricted LCM

$$\rho(i) = \sum_{K \in \mathcal{K}} \left\{ p(K) \prod_l r(I_l = i_l | X = K) \right\}.$$

This model assumes that the item responses of an examinee are independent given the knowledge state of the examinee, and that the response probabilities  $r(I_l = i_l | X = K)$  are attached to the items (item-specific) *and* knowledge states (state-specific). This model is implicitly referred to in Doignon and Falmagne (1999, p. 146).

---

<sup>3</sup> $|\cdot|$  denotes set-cardinality.

Imposing restrictions on the parameters of the classical unrestricted LCM gives the following restricted latent class (scaling) models, in the order of increasing generality.

1. A single response error rate across all items and knowledge states. That is, for a constant (error rate)  $\tau \in [0, 1)$ ,

$$\begin{aligned} r(I_l = 0|X = K) &= \tau & (I_l \in Q, K \in \mathcal{K}, I_l \in K), \\ r(I_l = 1|X = K) &= \tau & (I_l \in Q, K \in \mathcal{K}, I_l \notin K). \end{aligned}$$

In other words, careless error and lucky guess rates equal  $\tau$ . This restricted LCM is a ‘de-linearized’ Proctor (1970) model.<sup>4</sup>

2. Two response error rates across all items and knowledge states. That is, for constants (error rates)  $\mu \in [0, 1)$  and  $\tau \in [0, 1)$ ,

$$\begin{aligned} r(I_l = 0|X = K) &= \mu & (I_l \in Q, K \in \mathcal{K}, I_l \in K), \\ r(I_l = 1|X = K) &= \tau & (I_l \in Q, K \in \mathcal{K}, I_l \notin K). \end{aligned}$$

In other words, careless error and lucky guess rates equal  $\mu$  and  $\tau$ , respectively. This restricted LCM is a ‘de-linearized’ Dayton and Macready (1976) intrusion–omission model (intrusion rate  $\tau$  and omission rate  $\mu$ ).

3. For each item, two response error rates across all knowledge states (item-specific). That is, for constants (error rates)  $\mu_l \in [0, 1)$  and  $\tau_l \in [0, 1)$  ( $1 \leq l \leq L$ ),

$$\begin{aligned} r(I_l = 0|X = K) &= \mu_l & (I_l \in Q, K \in \mathcal{K}, I_l \in K), \\ r(I_l = 1|X = K) &= \tau_l & (I_l \in Q, K \in \mathcal{K}, I_l \notin K). \end{aligned}$$

In other words, to every item is attached careless error and lucky guess rates  $\mu_l$  and  $\tau_l$  (item-specific), respectively. This restricted LCM is called the basic local independence model (BLIM) in KST (Doignon and Falmagne, 1999, pp. 144–145). The interpretation of the BLIM as a restricted LCM is discussed in Schrepp (2005) and Ünlü (2006). It is a sort of ‘de-linearized’ Lazarsfeld and Henry (1968) latent distance

---

<sup>4</sup>In the following, ‘de-linearized’ means that the knowledge structure  $\mathcal{K}$  is not necessarily totally ordered with respect to set-inclusion, as is the case for the traditional latent class scaling models.

model. At this point, it is important to note that the BLIM is not identifiable in general. This is why in the latent distance model, for two items, say  $I_1$  and  $I_L$ , single response error rates across all states are assumed. That is, for constants  $\tau_1 \in [0, 1)$  and  $\tau_L \in [0, 1)$ ,

$$\begin{aligned} r(I_1 = 0|X = K) &= \tau_1 & (K \in \mathcal{K}, I_1 \in K), \\ r(I_1 = 1|X = K) &= \tau_1 & (K \in \mathcal{K}, I_1 \notin K), \\ r(I_L = 0|X = K) &= \tau_L & (K \in \mathcal{K}, I_L \in K), \\ r(I_L = 1|X = K) &= \tau_L & (K \in \mathcal{K}, I_L \notin K). \end{aligned}$$

For the remaining  $|Q| - 2$  items, the same restrictions as in the BLIM are imposed.

## Inferential statistics

As a statistical method, LCA comes with corresponding inference methodology. Parameter estimation and model testing are typically performed using maximum likelihood. This statistical inference methodology, at least theoretically (in realistic contexts a prohibitively large number of state proportions may have to be estimated), in particular applies to the KST probabilistic models, as they are special latent class scaling models. Albeit without a reference to LCA, this has been already demonstrated by Doignon and Falmagne (1999, Chapters 7 and 8), and the statistical techniques presented there are consistent with latent class modeling methodology. This does not become clear from Schrepp (2005), so I recapitulate that next.

Estimates of the model parameters are obtained maximizing the likelihood function. Let the model parameters of the BLIM be summarized in a parameter vector  $\theta = (p(K), \mu_l, \tau_l)$ , which ranges over a parameter space  $\Theta$ . Let the data be denoted by  $x = (N(i))$ , where  $N(i)$  is the observed absolute count of response pattern  $i$ . The kernel of the likelihood function that is maximized is (multinomial sampling distribution for the data):

$$\mathcal{L}(\theta; x) = \prod_i \left\{ \sum_{K \in \mathcal{K}} \left\{ p(K) \prod_l r(I_l = i_l | X = K) \right\} \right\}^{N(i)}$$

where  $r(I_l = i_l | X = K)$  is  $\mu_l$ ,  $1 - \mu_l$ ,  $\tau_l$ , or  $1 - \tau_l$  depending on whether  $i_l = 0$  and  $I_l \in K$ ,  $i_l = 1$  and  $I_l \in K$ ,  $i_l = 1$  and  $I_l \notin K$ , or  $i_l = 0$  and



$I_l \notin K$ , respectively. Let  $\hat{\theta} \in \Theta$  be the maximum likelihood estimate of the parameter vector  $\theta$ , that is,  $\mathcal{L}(\hat{\theta}; x) = \max_{\theta \in \Theta} \mathcal{L}(\theta; x)$ . The maximum of the kernel of the likelihood function cannot in general be obtained by analytical methods. Numerical optimization methods are required. Popular iterative methods for solving the maximum likelihood estimation problem are expectation–maximization (EM) or Newton–Raphson type algorithms. For instance, the software package Latent GOLD (Vermunt and Magidson, 2000) for latent class and finite mixture modeling implements a hybrid EM–Newton–Raphson algorithm to do maximum likelihood estimation. In KST, the conjugate gradient search algorithm principal axis (PRAXIS) by Brent (1973) has been commonly used, which is a modification of a direction-set method by Powell (1964). For a brief description of Powell’s method and Brent’s modification, see Gegenfurtner (1992), who provides a freely available implementation of the PRAXIS algorithm in the programming language C.

The goodness-of-fit of an estimated BLIM can be tested based on the log-likelihood ratio statistic (deviance)

$$L^2(\theta; x) = 2 \sum_i \left\{ N(i) \ln \left( \frac{N(i)}{N \rho_\theta(i)} \right) \right\}$$

(Doignon and Falmagne, 1999, Section 7.10). Given certain regularity conditions are satisfied, the random variable  $L^2(\hat{\theta}; x)$  is asymptotically chi-squared with degrees of freedom  $df = (2^{|Q|} - 1) - \{(|K| - 1) + 2|Q|\}$ . Instead of  $L^2$ , Pearson’s  $X^2$  (Doignon and Falmagne, 1999, Section 7.5), or more generally, any of the Read–Cressie goodness-of-fit statistics of the power-divergence family (Read and Cressie, 1988) can be used as well; these statistics are asymptotically equivalent. In KST, also general likelihood ratio tests based on the general deviance have been discussed (Doignon and Falmagne, 1999, Section 7.12). Since it is not valid to compare LCMs with different numbers of latent classes by general likelihood ratio tests, model selection information criteria such as the AIC or BIC are often used. The software package Latent GOLD, for instance, implements these procedures.

In other words, KST and LCA are related also at the level of inferential statistics used in both theories. More generally, both approaches are special cases of maximum likelihood methodology for multinomial probability models. For more information about inferential statistics for the class of multinomial probability models, see, for instance, Bishop et al. (1975).

A remark regarding the use of these methodologies in Schrepp (2005) is in order. The data analysis method for deriving knowledge structures proposed in Schrepp (2005), described as an application of exploratory latent class cluster analysis, in fact is fitting as many as possible multinomial probability models to the data and choosing among them using the BIC criterion. In particular, this approach does not carefully address the important issue of identifiability for the models considered, which, as mentioned before, is a crucial problem for BLIM type models. Connected with the identifiability issue is the fact that the BIC criterion, strictly speaking, is based on the number of independent, and at the same time, estimable model parameters.

## Conclusion

Two publications about knowledge structures explicitly point out and build upon the outlined connection between KST and LCA. Schrepp (2005) proposes a method for constructing knowledge structures from the data. Candidate knowledge structures are generated, they are considered as restricted latent class models and fitted to the data, and the BIC criterion is used to choose among them. (Similar to Schrepp (2005), Stefanutti and Robusto (2009) propose a special case of the BLIM where the response error probabilities are constrained, in order to recover a probabilistic knowledge structure.) Ünlü (2006) investigates latent class modeling with random effects for the estimation of response error rates for dichotomous test items in KST. In particular, this approach is extended to give a generalization of the BLIM and of a number of other LCMs. This allows for local dependence among the items given the knowledge state of an examinee and/or for the incorporation of covariates.

An interesting and important direction for future research is investigating models based on stochastic processes. So far, dynamic KST probabilistic models such as the stochastic learning paths systems (Doignon and Falmagne, 1999) have not been compared with statistical approaches to latent variable modeling. For instance, latent Markov (hidden Markov or latent transition) models (e.g., Hagenars, 1990; MacDonald and Zucchini, 1997; Vermunt, 1997) or generalized latent variable modeling (e.g., Skrondal and Rabe-Hesketh, 2004) are interesting topics to study, in further comparisons of KST to statistical analyses of latent variables.

## References

- Albert, D., Doble, C., Eppstein, D., Falmagne, J.-Cl., & Hu, X. (forthcoming, 2009). *Knowledge spaces: Applications to education*. Springer.
- Albert, D., & Lukas, J. (Eds.) (1999). *Knowledge spaces: Theories, empirical research, and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Andersen, E.B. (1982). Latent structure analysis: A survey. *Scandinavian Journal of Statistics*, 9, 1–12.
- Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: MIT Press.
- Brent, R.P. (1973). *Algorithms for minimization without derivatives*. Englewood Cliffs, NJ: Prentice Hall.
- Dayton, C.M. (1998). *Latent class scaling analysis*. Thousand Oaks: Sage Publications.
- 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. (1999). *Knowledge spaces*. Berlin: Springer.
- Everitt, B.S., & Hand, D.J. (1981). *Finite mixture distributions*. London: Chapman and Hall.
- Falmagne, J.-Cl., Cosyn, E., Doignon, J.-P., & Thiéry, N. (2006). The assessment of knowledge, in theory and in practice. In B. Ganter, & L. Kwuida (Eds.), *Formal Concept Analysis, 4th International Conference, ICFCA 2006, Dresden, Germany, Lecture Notes in Artificial Intelligence* (pp. 61–79). Berlin, Heidelberg, and New York: Springer.
- Falmagne, J.-Cl., & Doignon, J.-P. (forthcoming, 2009). *Learning spaces*. Springer.

- 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.
- Gegenfurtner, K.R. (1992). PRAXIS: Brent’s algorithm for function minimization. *Behavior Research Methods, Instruments, and Computers*, 24, 560–564.
- Goodman, L.A. (1978). *Analysing qualitative/categorical variables: Loglinear models and latent structure analysis*. Cambridge: Cambridge University Press.
- 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.
- Hagenaars, J.A. (1990). *Categorical longitudinal data—loglinear analysis of panel, trend and cohort data*. Newbury Park: Sage Publications.
- Hagenaars, J.A., & McCutcheon, A.L. (Eds.) (2002). *Applied latent class analysis*. Cambridge: Cambridge University Press.
- Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks: Sage Publications.
- Lazarsfeld, P.F., & Henry, N.W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- MacDonald, I.L., & Zucchini, W. (1997). *Hidden Markov models and other types of models for discrete-valued time series*. London: Chapman and Hall.
- Magidson, J., & Vermunt, J.K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays. *Sociological Methodology*, 31, 223–264.
- McCutcheon, A.L. (1987). *Latent class analysis*. Newbury Park: Sage Publications.

- McLachlan, G.J., & Basford, K.E. (1988). *Mixture models: Inference and application to clustering*. New York: Marcel Dekker.
- McLachlan, G.J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
- Powell, M.J.D. (1964). An efficient method for finding the minimum of a function in several variables without calculating derivatives. *Computer Journal*, 7, 155–162.
- Proctor, C.H. (1970). A probabilistic formulation and statistical analysis of Guttman scaling. *Psychometrika*, 35, 73–78.
- Read, T.R.C., & Cressie, N.A.C. (1988). *Goodness-of-fit statistics for discrete multivariate data*. New York: Springer.
- Schrepp, M. (2005). About the connection between knowledge structures and latent class models. *Methodology*, 1(3), 93–103.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: Chapman and Hall.
- Stefanutti, L., & Robusto, E. (2009). Recovering a probabilistic knowledge structure by constraining its parameter space. *Psychometrika*, 74, 83–96.
- Titterton, D.M., Smith, A.F.M., & Makov, U.E. (1985). *Statistical analysis of finite mixture distributions*. New York: Wiley.
- 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.
- Vermunt, J.K. (1997). *Log-linear models for event histories*. Thousand Oaks: Sage Publications.
- Vermunt, J.K., & Magidson, J. (2000). *Latent GOLD 2.0 user's guide*. Belmont: Statistical Innovations.
- Vermunt, J.K., & Magidson, J. (2004). Latent class analysis. In M.S. Lewis-Beck, A.E. Bryman, & T.F. Liao (Eds.), *The Sage Encyclopedia of Social Science Research Methods* (pp. 549–553). Thousand Oaks: Sage Publications.

Vermunt, J.K., & Van Dijk, L. (2001). A nonparametric random-coefficients approach: The latent class regression model. *Multilevel Modelling Newsletter*, 13, 6–13.