DOCUMENT RESUME

ED 384 676 TM 023 998

AUTHOR Kim, Sung-Ho
TITLE Combination of Conditional Log-Linear Structures. Program Statistics Research.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-92-72; ETS-TR-92-27
PUB DATE Nov 92
NOTE 48p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Ability; Data Collection; Equations (Mathematics); *Performance; Simulation; *Test Items
IDENTIFIERS *Graphic Representation; *Log Linear Models

ABSTRACT

There has been increasing attention to the fine structure of abilities underlying task performance (E. H. Haertel and D. E. Wiley). One of the most useful approaches for this is by use of graphical models to represent relationships among abilities and test items. Building a large graphical model is always an issue. Restrictions upon experiments and data collection, among others, may result in parts of the large model. Or it may be convenient for us to build parts of the large models first, and then to try to combine those parts into a larger model. This paper derives a theory, confined to categorical variables only, a theory that may be useful in combining conditional graphical models into a larger one. The main result of the study is that one can see partial information about a true log-linear structure (LLS) from its conditional LLS and use the information in trying to guess the true LLS, assuming that the true LLS is graphical. An application of the result is illustrated using a simulated data set. Six figures and five tables present analysis results. An appendix presents an illustrative chart. (Contains 12 references.) (Author/SLD)

RR-92-72

Combination of Conditional Log-Linear Structures

Sung-Ho Kim
Educational Testing Service

ED 384 676

PROGRAM STATISTICS RESEARCH

TECHNICAL REPORT NO. 92-27

Educational Testing Service
Princeton, New Jersey 08541

Combination of Conditional Log-Linear
Structures

Sung-Ho Kim
Educational Testing Service

Program Statistics Research
Technical Report No. 92-27

Research Report No. 92-72

Educational Testing Service
Princeton, New Jersey 08541

November 1992

The Program Statistics Research Technical Report Series is designed to make the working papers of the Research Statistics Group at Educational Testing Service generally available. The series consists of reports by the members of the Research Statistics Group as well as their external and visiting statistical consultants.

Acknowledgements

Preparation of this report was supported by Educational Testing Service. The author wishes to thank Bob Mislevy and Peter Pashley for their comments on the initial draft of this report. In particular, Section 6 of this report was added according to the suggestion by Bob Mislevy.

Abstract

There has been increasing attention to the fine structure of abilities underlying task performance (Haertel and Wiley (in press)). One of the useful approaches for this is by use of graphical models to represent relationships among abilities and test items. Building a large graphical model is always an issue. Restrictions upon experiments and data collection, among others, may result in parts of the large model. Or it may be convenient for us to build parts of the large models first, and then try to combine those parts into a larger model. This paper derives, confined to categorical variables only, a theory which may be useful in combining conditional graphical models into a larger one. The main result of the paper is that we can see partial information about a true log-linear structure (LLS) from its conditional LLSs and use the information in trying to guess the true LLS, assuming that the true LLS is graphical. An application of the result is illustrated using a simulated data set.

Key words: log-linear model, strong hierarchy assumption, hypermodel, traceability, influence diagram.

1 Introduction and Problem

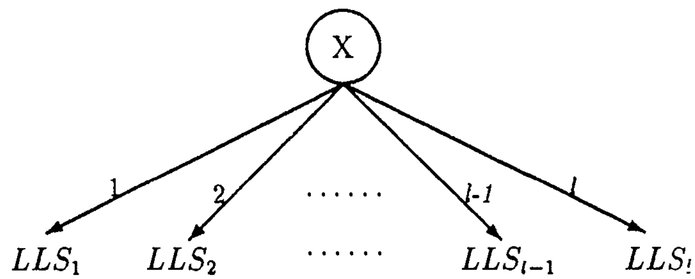
During the last two decades, psychological research has been focused on tasks that better approximate the meaningful learning and problem-solving activities that engage people in real life. There has been increasing attention to the fine structure of abilities underlying task performance (Haertel and Wiley (in press)). As Haertel and Wiley (in press) note, current test theory may not be sufficient to characterize the fine structure of the ability pattern and hence to build upon this structure to more clearly represent the acquisition and the structure of aggregate abilities. One of the appropriate approaches in dealing with this kind of issues is by use of graphical models by which we can represent relationships among abilities and test items.

In this paper, "graphical models" includes graphical log-linear models, influence diagrams and Bayes Networks. One of the problems that we face when we try to build graphical models of task abilities and test items is described below. Consider the test items which can be solved in several different ways. Problem-solving strategies may vary across a group of individuals. For example, in dealing with mixtures of whole numbers and fractions, a student may prefer dealing with those numbers in a fraction form, while another in a mixed form (Tatsuoka (1990)). Skills used in solving a problem may vary according to the strategies selected by test-takers. Some skills used in one strategy may not be used in another. This is one of the reasons why it is desirable to build models in two steps when there are more than one strategy available — first, we build graphical models for each problem-solving strategy; second, we combine them into a graphical model where a categorical variable for the problem-solving strategies is included as a new variable.

The main goal of this paper is to derive a rule which is useful when combining graphical models, where the same set of variables are involved in each of the graphical models. The above-mentioned three types of graphical models share a common ground (see Lauritzen and Spiegelhalter (1988), Pearl (1988), Smith (1989)). That is, Bayes networks and influence diagrams are interpretable in terms of graphical log-linear model under a positivity condition that every configuration of the variables involved in the model has positive probability. Hence in our derivation of the combining rule our discussion will be in the context of graphical log-linear model.

We will define by *log-linear structure* (LLS) the generating class of a log-linear model. If the log-linear model is graphical, then its LLS can be represented in the form of a graph. Suppose that a categorical variable X has l levels, and that we have l conditional LLSs for a set of variables, corresponding to the l possible outcomes of X. In these circumstances it seems desirable to

Figure 1.1: Hybrid tree/log-linear structure representation



develop hybrid tree/log-linear structure representations. Such a hybrid representation is given in Figure 1.1. We will simply call such a representation a *hybrid.* A hybrid can be described in terms of log-linear structures only, instead of a graph. For example, Figure 1.1 can be expressed by

$$(LLS_1, LLS_2, \cdots, LLS_l),$$

where the subscripts indicate the values of X.

There are only finitely many possible log-linear structures corresponding to a hybrid, if the hybrid involves only a finite number of random variables. We will call the log-linear structure of a log-linear model which gives rise to the hybrid h, say, by a *hypermodel corresponding to (abbreviated to "c.t.") h*. There may be many hypermodels corresponding to a hybrid.

This paper consists of 7 sections. In sections 2 through 5, we derive a rule for obtaining hypermodels from a given hybrid. Section 2 deals in detail with a conditional log-linear model. The possible conditional log-linear structures of a given log-linear structure are discussed in section 3. In section 4, we derive basic rules of obtaining the set of the possible graphical hypermodels corresponding to a given hybrid. Section 5 shows that we can find the true graphical hypermodel with some uncertainty by applying the rules derived in section 4. Section 6 illustrates combining two graphical models by applying the main results of this paper. Section 7 concludes the paper.

In the rest of this paper, we will use abbreviations (in the parenthesis) as follows if confusion is not likely: log-linear model (LLM), log-linear structure (LLS), conditional log-linear model (CLLM), conditional log-linear structure (CLLS).

2 A General Expression for the CLLM and the Strong Hierarchy Assumption

Suppose there are n categorical variables $X_1, X_2, \cdots, X_n$, for which we will consider LLMs under the hierarchy assumption. We will borrow most of the notations from Bishop, Fienberg, and Holland (1980). In this paper, we will consider conditional probabilities of a set of variables given an outcome of X_1 variable, and assume that X_1 takes on I_1 values $x_{11}, \cdots, x_{1I_1}$. Let $\theta_1, \theta_2, \cdots, \theta_k$

be distinct subsets of $\{1, 2, \cdots, n\}$. Consider a LLM whose generating class is given by $\{\theta_1, \theta_2, \cdots, \theta_k\}$, namely,

$$\begin{aligned}
&\log p_{12\cdots n}(x_1, x_2, \cdots, x_n) \\
&= u_{\theta_1(\underset{\sim}{x}_{\theta_1})} + u_{\theta_2(\underset{\sim}{x}_{\theta_2})} + \cdots + u_{\theta_k(\underset{\sim}{x}_{\theta_k})} + R(x_1, x_2, \cdots, x_k), \qquad (2.1)
\end{aligned}$$

where $R(x_1, x_2, \cdots, x_k)$ is a constant u-term plus the summation of the u_θ-terms each of whose subscript sets, θ's, is a strict subset of some θ_i.

We denote by $p_{23\cdots k|1}(x_2, x_3, \cdots, x_k|x_1)$ the conditional probability that $X_2 = x_2, X_3 = x_3, \cdots, X_k = x_k$, given $X_1 = x_1$. Then we have that

$$\begin{aligned}
&\log p_{23\cdots k|1}(x_2, x_3, \cdots, x_k|x_1) \\
&= \sum_{i=1}^{k} u_{\theta_i(\underset{\sim}{x}_{\theta_i})} + R(x_1, x_2, \cdots, x_k) - \log p_1(x_1). \qquad (2.2)
\end{aligned}$$

For notational convenience, we will omit the argument of the subscript set θ when confusion is not likely. For θ such that $\theta \cap \{1\} = \emptyset$(the empty set), we let

$$u^{(x_1)}_{\theta(\underset{\sim}{x}_\theta)} = u_{\theta(\underset{\sim}{x}_\theta)} + u_{\{1\}\cup\theta(x_1, \underset{\sim}{x}_\theta)}. \qquad (2.3)$$

Then, we can re-express equation (2.2) in terms of the usual u-terms and the $u^{(\cdot)}$-terms. At this point, we need to know $p_1(x_1)$. Assume that

$$u_1 \neq 0 \quad \text{in equation (2.1).} \qquad (2.4)$$

Let

$$c(x_1) = u + u_{1(x_1)} - \log p_1(x_1). \tag{2.5}$$

In (2.2), we may rearrange, without loss of generality, θ's so that $\theta_1, \cdots, \theta_r$, for $1 \le r \le k$, contain the singleton $\{1\}$ as a subset, while the other θ's do not. Under this rearrangement, let $\theta_i' = \theta_i \setminus \{1\}$, for $i = 1, 2, \cdots, r$. Let $u_\emptyset = 0$, while u symbolizes a constant term in a LLM. Then we have, from (2.2) and (2.5), that

$$\begin{aligned} \log p_{2\cdots n|1}(x_2, \cdots, x_n|x_1) &= \sum_{i=1}^{r} u^{(x_1)}_{\theta_i'(x_{\theta_i'})} + \sum_{i=r+1}^{k} u_{\theta_i(x_{\theta_i})} + \\ & R^{(x_1)}(x_2, \cdots, x_n) - u - u_{1(x_1)} + c(x_1). \end{aligned} \tag{2.6}$$

where

$$R^{(x_1)}(x_2, \cdots, x_n) = R(x_1, x_2, \cdots, x_n) - \sum_{i=1}^{r} u_{\theta_i'(x_{\theta_i'})}. \tag{2.7}$$

Note that $\sum_y u^{(x_1)}_{\theta_i'(x_{\theta_i'})} = 0$, where the summation goes over all the possible value y's of X_j for each $j \in \theta_i'$.

A hierarchical LLM for a joint probability of a set of variables does not necessarily imply a hierarchical LLM for any of its conditional probabilities. For example, consider a LLM whose generating class is given by $\{\{1, 2\}, \{1, 3, 4\}\}$. Then from equation (2.6) we have

$$\log p_{234|1}(x_2, x_3, x_4|x_1) =$$

$$u^{(x_1)}_{\{3,4\}(x_3, x_4)} + u^{(x_1)}_{\{2\}(x_2)} + u^{(x_1)}_{\{3\}(x_3)} + u^{(x_1)}_{\{4\}(x_4)} + (u_{1(x_1)} - \log p_1(x_1)).$$

In this expression, $u^{(x_1)}_{\{3\}}$ or $u^{(x_1)}_{\{4\}}$ may be equal to 0, while $u^{(x_1)}_{\{3,4\}}$ may not.

However, we will confine ourselves, in this paper, to the cases where the hierarchy principle holds in the LLMs for the joint probability and for the conditional probabilities. We will refer to this restriction the **strong hierarchy assumption** (SHA).

We denote a CLLM of a set of variables given $X_1 = x_{1l}$ by CM_l, $l = 1, \cdots, I_1$, and the LLS of CM_l by CS_l. In expression (2.6), there may be some θ'_i such that $\emptyset \subset \theta'_i \subset \theta_j$, for some j, $r+1 \leq j \leq k$. Under the SHA, we don't care about such $u^{(x_{1l})}_{\theta'_i}$-terms in expressing CS_l, and throw such $u^{(x_{1l})}_{\theta'_i}$-terms into $R^{(x_{1l})}$-term. In this sense, we may refer to those terms as the *disappearing* $u^{(\cdot)}$-terms, the rest of them, as the *remaining* $u^{(\cdot)}$-terms. Suppose there are s *disappearing* $u^{(\cdot)}$-terms, and $r-s$ *remaining* $u^{(\cdot)}$-terms, for $0 \leq s \leq r$, in the r.h.s. of equation (2.6). Without loss of generality, we may then rearrange the $u^{(x_{1l})}_{\theta'_i}$-terms, so that $u^{(x_{1l})}_{\theta'_i}$, $i = 1, 2, \cdots, s$, disappear into $R^{(x_{1l})}$-term. We now have

$$\log p_{2\cdots n|1}(x_2, \cdots, x_n | x_{1l}) =$$
$$\sum_{i=s+1}^{r} u^{(x_{1l})}_{\theta'_i(x_{\theta'_i})} + \sum_{i=r+1}^{k} u_{\theta_i(x_{\theta_i})} + R^{(x_{1l})}_{+}(x_2, \cdots, x_n) + c(x_{1l}),$$

where

$$R_+^{(x_{1l})}(x_2,\cdots,x_n) = R(x_{1l},x_2,\cdots,x_n) - u - u_{1(x_{1l})} + \sum_{i=1}^{s} u_{\theta_i(x_{1l},\, \underset{\sim}{x}_{\theta_i'})} - \sum_{i=s+1}^{r} u_{\theta_i'(\underset{\sim}{x}_{\theta_i'})}. \tag{2.8}$$

From (2.7) and (2.8), we can see that $R_+^{(x_{1l})}(x_2,\cdots,x_n)$ is a linear combination of the u-terms each of whose subscript sets is a strict subset of some set in $\{\theta'_{s+1},\cdots,\theta'_r,\theta_{r+1},\cdots,\theta_k\}$. Actually $R_+^{(x_{1l})}$ is composed of the following 3 types of u or $u^{(\cdot)}$-terms:

(a) the $u_{\theta''}^{(x_{1l})}$-terms, for $i = 1,2,\cdots,s$, with $\emptyset \subset \theta'' \subseteq \theta_i'$,

(b) the $u_{\theta''}^{(x_{1l})}$-terms, with $\emptyset \subset \theta'' \subset \theta_i'$, for $i = s+1,\cdots,r$, excluding the terms in (a),

(c) the u-terms of $R(x_{1l},\cdots,x_n) - u - u_{1(x_{1l})}$ except those terms in (a) and (b).

Rewriting $\log p_{2\cdots n|1}(x_2,\cdots,x_n|x_{1l})$ gives

$$\begin{aligned}
&\log p_{2\cdots n|1}(x_2,\cdots,x_n|x_{1l}) \\
&= \sum_{i=s+1}^{r} u_{\theta_i'(\underset{\sim}{x}_{\theta_i'})}^{(x_{1l})} + \sum_{i=r+1}^{k} u_{\theta_i(\underset{\sim}{x}_{\theta_i})} + \text{sum of } \mathbf{(a)} \text{ terms} \\
&\quad + \text{sum of } \mathbf{(b)} \text{ terms} + \text{sum of } \mathbf{(c)} \text{ terms} + c(x_{1l}).
\end{aligned} \tag{2.9}$$

This result is true whether $\{1\}$ is one of $\{\theta_i\}_{i=1}^{k}$ in (2.1) or not, i.e., whether

$R(x_1, x_2, \cdots, x_\pi)$ has u_1-term in its summation or not. Note that $c(x_{1l})$ is the constant term in expression (2.9).

Before concluding this section, we need to note that the u-terms in the second summation in expression (2.9) are all non-zero since they remain the same as they are in the original model (2.1). On the other hand, some $u^{(\cdot)}$-term in the first summation can be zero, in which case some (b) term may affect the CLLS.

In this section, we examined CLLM's and derived a generalized expression (2.9). Under the SHA, the CLLS CS_l, $l = 1, \cdots, I_1$, is determined by u and $u^{(\cdot)}$-terms in

(i) the first two summations in (2.9), if none of the terms is zero;

(ii) the first, second and fourth summations in (2.9), if some $u^{(\cdot)}$-term in the first summation is zero.

3 The CLLSs in a Hybrid Given a Hypermodel

In Section 2, we have seen that the two CLLSs can be different. From the CLLSs, we can guess the possible hypermodels. The relationship between the

CLLSs and the hypermodel is examined in this section. The following lemma plays an important role in searching for the possible CLLSs.

Lemma 3.1 *Let $\theta \cap \{1\} = \emptyset$. Then, $u_\theta = u_{\{1\}\cup\theta} = 0$, iff $u_\theta^{(x_{1l})} = 0$, for all $l = 1, \cdots, I_1$.*

Proof: Suppose that $u_{\theta(\underset{\sim}{x}_\theta)}^{(x_{1l})} = 0, \ \forall \underset{\sim}{x}_\theta$, for all $l = 1, \cdots, I_1$. Then, it follows that, for $l = 1, \cdots, I_1$,

$$u_{\{1\}\cup\theta(x_{1l}, \underset{\sim}{x}_\theta)} = -u_{\theta(\underset{\sim}{x}_\theta)}, \ \forall \underset{\sim}{x}_\theta. \tag{3.1}$$

Since $\sum_{l=1}^{I_1} u_{\{1\}\cup\theta(x_{1l}, \underset{\sim}{x}_\theta)} = 0$, we have, by equation (3.1), $u_\theta = 0$. Hence, the "if part" is proved.

The proof of the other direction is straightforward. □

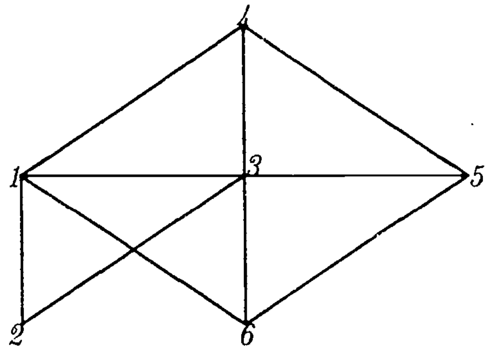
The following example illustrates what CLLSs are possible by applying Lemma 3.1.

Example 3.2 *Suppose that $I_1 = 2$, and that the LLS S for a 6-dimensional contingency table is given by*

$$H = \{\{1,2,3\}, \{1,3,4\}, \{1,3,6\}, \{3,4,5\}, \{3,5,6\}\}; \tag{3.2}$$

it is given in graphical form as in Figure 3.1.

Figure 3.1:



Following the same argument as in Section 2, we have, for $l = 1, 2$.

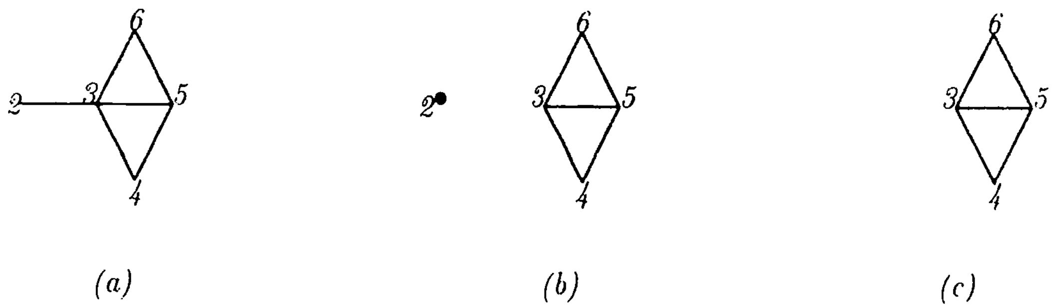
$$
\begin{aligned}
&\log p_{2\cdots 6|1}(x_2 \cdots x_6 | x_{1l}) \\
&= c(x_{1l}) + u_2^{(x_{1l})} + u_3^{(x_{1l})} + u_4^{(x_{1l})} + u_5 + u_6^{(x_{1l})} + u_{23}^{(x_{1l})} + u_{34}^{(x_{1l})} \\
&\quad + u_{35} + u_{36}^{(x_{1l})} + u_{45} + u_{56} + u_{345} + u_{356}, \qquad (3.3)
\end{aligned}
$$

where some $u^{(\cdot)}$-term may be zero.

Denote the element sets in (3.2) by $\theta_1, \cdots, \theta_5$, in that order. Under the SHA, our interest is on the $u_{23}^{(x_{1l})}$-term in (3.3). If $u_{23}^{(x_{1l})} \neq 0$, then the CLLS CS_l is given by $S_1 = \{\theta_1', \theta_4, \theta_5\}$. Otherwise, if $u_2^{(x_{1l})} \neq 0$, then CS_l is given by $S_2 = \{\{2\}, \theta_4, \theta_5\}$; if $u_2^{(x_{1l})} = 0$, then CS_l is given by $S_3 = \{\theta_4, \theta_5\}$.

The LLSs S_1, S_2 and S_3 are depicted by the graphs (a), (b) and (c) in Figure 3.2, respectively.

Figure 3.2:



The point here is that both $u_{23}^{(x_{11})}$ and $u_{23}^{(x_{12})}$ cannot be zero. *If both were equal to zero, then by Lemma 3.1,*

$$u_{123} = u_{23} = 0,$$

which contradicts the LLS S. The same argument applies to $u_2^{(\cdot)}$. *Table 3.1 shows the set of all the possible pairs of CLLSs.*

□

We note that the union of the element sets appearing in CS_1 or CS_2 is $\{2, 3, 4, 5, 6\}$, which is equal to $\cup_{\theta \in S}\theta \setminus \{1\}$. Generalization of this result follows. For a LLM M, denote by $\mathcal{G}(\mathcal{M})$ the generating class of M.

Theorem 3.3 *Let Hybrid(*$CS_1, \cdots, CS_{I_1}$*) be a hybrid from a LLM M. Then, under condition (2.4) and the SHA, we have*

$$\bigcup_{\theta \in \mathcal{G}(M)} \theta = \{1\} \cup \left(\bigcup_{\theta \in CS_1 \cup \cdots \cup CS_{I_1}} \theta \right).$$

Table 3.1: The possible pairs of the CLLSs

$X_1 = x_{11}$

$\longrightarrow$

$X_1 = x_{12} \downarrow$

	S_1	S_2	S_3
S_1	○*	○	○
S_2	○	×*	×
S_3	○	×	×

**Note:* ○ *means that the corresponding pair of CLLSs are possible;* × *impossible.*

Proof: Let $Z = \cup_{i=1}^{I_1} CS_i$. For $\theta \in \mathcal{G}(M)$, $\theta \backslash \{1\} \in Z$ by Lemma 3.1, if $\{1\} \subset \theta$; if $\{1\} \cap \theta = \emptyset$, then the u-terms in the second summation of (2.9) implies $\theta \in Z$. Thus

$$\bigcup_{\theta \in \mathcal{G}(M)} \theta \subseteq \{1\} \cup (\bigcup_{\theta \in Z} \theta).$$

On the other hand, for $\theta \in Z$, we can see, from expression (2.9) and by Lemma 3.1, that the true hypermodel involves at least the X-variables with subscripts in θ. Thus, by assumption (2.4), we have

$$\bigcup_{\theta \in \mathcal{G}(M)} \theta \supseteq \{1\} \cup (\bigcup_{\theta \in Z} \theta).$$

Q.E.D. □

Now we have seen that we don't have to worry about the categorical variables not involved in a hybrid, when we try to guess the true hypermodel corresponding to the hybrid.

By comparing (2.1) with (2.9) of Section 2, and by applying Lemma 3.1, we can summarize the possible CLLSs from a given LLS as in the theorem below, whose proof is briefly outlined.

Theorem 3.4 *Consider a LLS $S = \{\theta_1, \cdots, \theta_k\}$. Suppose, for $1 \leq s \leq r \leq k$, the following three conditions hold:*

(i) $1 \in \theta_j$, *for* $1 \leq j \leq r$; $1 \notin \theta_j$ *for* $r < j \leq k$.

(ii) *For* $1 \leq j \leq s$, *there exists* l, $r < l \leq k$ *such that* $\theta_j \setminus \{1\} \subset \theta_l$.

(iii) *For* $s < j \leq r$, $\theta_j \setminus \{1\} \not\subset \theta_l$ *for every* l, $r < l \leq k$.

Then, under the SIIA, we have the following results:

(i') *For* $1 \leq j \leq s$, $\theta_j \setminus \{1\}$ *does not show up in any CLLS of* S.

(ii') *For* $s < j \leq r$, $\theta_j \setminus \{1\}$ *shows up in at least one CLLS of* S.

(iii') *For* $r < l \leq k$, θ_l *shows up in all the CLLSs of* S.

Proof: Results (i') and (iii') are immediate from (2.1) and (2.9). Result (ii') follows from (2.1), (2.9), and Lemma 3.1. Q.E.D. □

In this section, we have seen the possible CLLSs of a LLS. It is also shown that only the categorical variables involved in the CLLSs are involved in the corresponding LLS. In Theorem 3 4, we have seen that some components in a LLS are kept or shrunk, or disappear in a corresponding hybrid. A careful look at Theorem 3.4 may help us in guessing the unknown LLS (or true hypermodel) from the given hybrid.

4 Basic Results for Hypermodels corresponding to a Hybrid

In the expression of the CLLMs, we must be careful in using u-terms and $u^{(\cdot)}$-terms. We will use $v_\theta^{(\cdot)}$ in the CLLMs to denote either u_θ or $u_\theta^{(\cdot)}$, for $\theta \neq \emptyset$. $v_\theta^{(\cdot)} = u_\theta$ when $u_\theta \neq 0$ and $u_{\{1\}\cup\theta} = 0$; $v_\theta^{(\cdot)} = u_\theta^{(\cdot)}$ when $u_\theta \neq 0$ and $u_{\{1\}\cup\theta} \neq 0$. In this section, we use u-terms only for the log-linear expression of the hypermodel. In other words, we will use u-terms for a LLM and $v^{(\cdot)}$-terms for a CLLM.

Theorem 4.1 *Consider the hybrid Hybrid($CS_1, \cdots, CS_{I_1}$), and suppose that a set θ is common in all the CSs and that the true hypermodel is graphical. Then, under condition (2.4) and the SHA, the true graphical hypermodel c.t*

Hybrid$(CS_1, \cdots, CS_{I_1})$ contains one of the followings:

$$\theta \cup \{1\}, \quad \{\theta, \{1\} \cup \varphi\}, \quad \textit{for } \emptyset \subseteq \varphi \subset \theta. \tag{4.1}$$

Proof: Let

$$Z = \bigcup_{i=1}^{I_1} CS_i,$$

and

$$D = \bigcup_{\rho \in Z} \rho.$$

By Theorem 3.3, it is necessary and sufficient to use only the variables whose subindices are in the set D.

For $X_1 = x_{1l}$, $l = 1, \cdots, I_1$, we express the LLM for CS_l by

$$\log p_{D|1}(\underset{\sim}{x}_D | x_{1l}) = \sum_{\rho \subseteq D} v_\rho^{(x_{1l})} + c(x_{1l}) - u_{1(x_{1l})}. \tag{4.2}$$

Since all of the CLLSs contain θ,

$$v_\theta^{(x_1)} \neq 0, \quad \text{for } x_1 = x_{11}, \cdots, x_{1I_1}. \tag{4.3}$$

From expression (4.2), we have

$$\begin{aligned} \log p_{D \cup \{1\}}(x_{1l}, \underset{\sim}{x}_D) &= \sum_{\rho \subseteq D} v_\rho^{(x_{1l})} + c(x_{1l}) - u_{1(x_{1l})} + \log p_1(x_{1l}) \\ &= \sum_{\rho \subseteq D} v_\rho^{(x_{1l})} + u. \end{aligned} \tag{4.4}$$

The candidacy for the true hypermodel is determined by whether $v_\rho^{(\cdot)}$ in (4.4) is zero or not. From the condition of the theorem, we know that $\theta \cap \{1\} = \emptyset$. For each $l \in \{1, \cdots, I_1\}$, $v_\theta^{(x_{1l})} \neq 0$ implies, under the SHA, one of the followings:

(i) $u_{\{1\}\cup\theta(x_{1l},\, \underset{\sim}{x}_\theta)} = 0$ and $u_{\theta(\, \underset{\sim}{x}_\theta)} \neq 0$.

(ii) $u_{\{1\}\cup\theta(x_{1l},\, \underset{\sim}{x}_\theta)} \neq 0$ and $u^{(x_{1l})}_{\theta(\, \underset{\sim}{x}_\theta)} \neq 0$.

Hence, there are 2^{I_1} possible combinations of cases (i) and (ii) for $(v_\theta^{(x_{1l})}, \cdots, v_\theta^{(x_{1I_1})})$. However, if case (ii) holds for at least one l in $\{1, 2, \cdots, I_1\}$, then

$$\{1\} \cup \theta \text{ is contained in the true hypermodel, } H\text{, say.} \tag{4.5}$$

By Lemma 3.1, any set that contains $\{1\} \cup \theta$ as a subset can not be contained in H, since such a set doesn't show up in any of the CLLSs.

On the other hand, if case (i) holds for all $l = 1, 2, \cdots, I_1$, then we can see that θ is in H but not $\{1\} \cup \theta$. In this case, we must not forget about the possibility of the *disappearing* $u^{(\cdot)}$-terms in some CLLS. Under the premise that the true hypermodel is graphical, the consideration of the *disappearing* $u^{(\cdot)}$-terms gives rise to one of the followings as a subset of H:

$$\{\theta, \varphi \cup \{1\}\}, \text{ for } \emptyset \subseteq \varphi \subset \theta. \tag{4.6}$$

(4.5) and (4.6) are combined into (4.1). Q.E.D. □

An alternative situation of the condition of Theorem 4.1 is considered in the theorem below.

Theorem 4.2 *For the hybrid* Hybrid$(CS_1, \cdots, CS_{I_1})$, *suppose that* θ *is an element set of some CS, and not common in all the CSs, and that there is no* *set in* $\cup_{i=1}^{I_1} LLS_i \setminus \theta$ *that contains* θ. *Then,* $\{1\} \cup \theta$ *is in the true hypermodel.*

Proof: The condition of the theorem says that θ is contained in at least one CS. Without loss of generality, let one of such CSs be CS_1. It is easy to see that there are three disjoint and exhaustive sets A and B such that

$A = \{j: \ CS_j \text{ contains } \theta\}$,

$B = \{1, 2, \cdots, I_1\} \setminus A$.

For CS_1, $v_\theta^{(x_{11})} = u_\theta^{(x_{11})}$ or u_θ. That $v_\theta^{(x_{11})} = u_\theta$ means, for $i = 1$,

$$u_{\{1\}\cup\theta(x_{1i}, \underset{\sim}{x}_\theta)} = 0 \text{ and } u_{\theta(\underset{\sim}{x}_\theta)} \neq 0, \ \forall \underset{\sim}{x}_\theta. \tag{4.7}$$

And that $v_\theta^{(x_{11})} = u_\theta^{(x_{11})}$ means, for $i = 1$,

$$u_{\{1\}\cup\theta(x_{1i}, \underset{\sim}{x}_\theta)} \neq 0 \text{ and } u_{\theta(\underset{\sim}{x}_\theta)}^{(x_{1i})} \neq 0. \tag{4.8}$$

Recall that all the CLLMs for a LLM are given by the same format as in (2.9) and that the variation among the CLLSs is due to whether $v^{(\cdot)}$ is zero or not.

For $j \in B$, either

$$u_{\{1\}\cup\theta(x_{1j}, \underset{\sim}{x}_\theta)} = u_{\theta(\underset{\sim}{x}_\theta)} = 0, \ \forall \ \underset{\sim}{x}_\theta \tag{4.9}$$

or

$$u_{\{1\}\cup\theta(x_{1j}, \underset{\sim}{x}_\theta)} \neq 0 \text{ and } u_{\{1\}\cup\theta(x_{1j}, \underset{\sim}{x}_\theta)} + u_{\theta(\underset{\sim}{x}_\theta)} = 0, \ \forall \ \underset{\sim}{x}_\theta. \tag{4.10}$$

If $A = \{1\}$, either (4.7) or (4.8) is possible. If $\{1, 2, \cdots, I_1\} \supset A \supset \{1\}$ (note that $B \neq \emptyset$ by the condition of the theorem), one or both of (4.7) and (4.8) are possible. It is worth noting that one and only one of (4.9) and (4.10) holds for all $j \in B$, since $u_\theta = 0$ in (4.9) while $u_\theta \neq 0$ in (4.10). Thus, one and only one of (4.9) and (4.10) is possible for all $j \in B$. Considering all of these situations, we have the following 10 cases we can think of from the condition of the theorem:

(Case-1) $A = \{1\}$; (4.7) for CS_1, and (4.9) for CS_j, $j \in B$.

(Case-2) $A = \{1\}$; (4.7) for CS_1, and (4.10) for CS_j, $j \in B$.

(Case-3) $A = \{1\}$; (4.8) for CS_1, and (4.9) for CS_j, $j \in B$.

(Case-4) $A = \{1\}$; (4.8) for CS_1, and (4.10) for CS_j, $j \in B$.

(Case-5) $A \supset \{1\}$; (4.7) for CS_i, $i \in A$, and (4.9) for CS_j, $j \in B$.

(Case-6) $A \supset \{1\}$; (4.7) for CS_i, $i \in A$, and (4.10) for CS_j, $j \in B$.

(Case-7) $A \supset \{1\}$; (4.8) for CS_i, $i \in A$, and (4.9) for CS_j, $j \in B$.

(Case-8) $A \supset \{1\}$; (4.8) for CS_i, $i \in A$, and (4.10) for CS_j, $j \in B$.

(Case-9) $A \supset \{1\}$; (4.7) and (4.8) each at least once for CS_i, $i \in A$, and (4.9) for CS_j, $j \in B$.

(Case-10) $A \supset \{1\}$; (4.7) and (4.8) each at least once for CS_i, $i \in A$, and (4.10) for CS_j, $j \in B$

Only 3 of these 10 cases are possible. They are cases 4, 8 and 10. Now we will investigate why the others are impossible.

For Cases 1, 5 and 9: (4.7) and (4.9) are not compatible (see the u_θ term).

For Case-2: From (4.10), it follows that

$$u_{\{1\}\cup\theta(x_{1_j}, \underset{\sim}{x}_\theta)} = -u_{\theta(\underset{\sim}{x}_\theta)}, \quad \forall \ \underset{\sim}{x}_\theta.$$

If (4.7) and (4.10) hold together and if $A = \{1\}$, then

$$\sum_{x_1} u_{\{1\}\cup\theta(x_1, \underset{\sim}{x}_\theta)} = -(I_1 - 1)u_{\theta(\underset{\sim}{x}_\theta)} \neq 0, \quad \forall \ \underset{\sim}{x}_\theta. \tag{4.11}$$

which is a violation of the constraint

$$\sum_{x_1} u_{\{1\}\cup\theta(x_1, \underset{\sim}{x}_\theta)} = 0. \tag{4.12}$$

For Cases 3 and 7: (4.8) and (4.9) are incompatible (note that by the hierarchy assumption (4.8) implies $u_\theta \neq 0$).

For Case-6: Denote the number of the elements in A by $\#(A)$. Expression (4.11) now becomes

$$\sum_{x_1} u_{\{1\}\cup\theta(x_1, \underset{\sim}{x}_\theta)} = -(I_1 - \#(A))u_{\theta(\underset{\sim}{x}_\theta)} \neq 0, \quad \forall \ \underset{\sim}{x}_\theta,$$

which is again a violation of the constraint (4.12). (Note that $I_1 - \#(A) > 0$ by the condition of the theorem.)

We now turn to the 3 possible cases.

For Cases 4 and 8: It is possible that (4.8) and (4.10) hold under the constraint (4.12).

For Case 10: For $I_1 \geq 3$, it is possible that such $\{u_{\{1\}\cup\theta}\}$ exist that satisfy the conditions of this case under the constraint (4.12).

In each of the above three possible cases, it is obvious that $\{1\} \cup \theta$ is contained in the true hypermodel. By the condition of the theorem, at least one of the three possible cases must occur. Therefore, we have the desired result. $\Box$.

The conditions of Theorems 4.1 and 4.2 do not seem to cover all the possible situations for each element set in $\bigcup_{j=1}^{I_1} CS_j$. The only situation not considered yet is that a set θ is in some CLLS and not common in all the CLLSs and that there is a set in some CLLS which contains θ as a strict subset. But we don't have to worry about the situation. If there is a set φ in some CLLS which is not a strict subset of any other set in the union of all the CLLSs, and if $\theta \subset \varphi$, then under the SHA we can see, in the light of expression (2.9), that the θ corresponds to a (**b**)-type term and φ corresponds to a $u^{(r_1)}$-term in the first summation of (2.9). Thus we don't have to worry about the θ in search of the true hypermodel (since the θ does not affect a LLS under the hierarchy principle). Instead we have to concern ourselves with φ and apply Theorem

4.2. Therefore, we have essentially considered all the necessary situations of the relationship between the sets in $\bigcup_{i=1}^{I_1} CS_i$ in search of the true hypermodel.

Now the question is if the rules derived in Theorems 4.1 and 4.2 lead us towards the true hypermodel from a given hybrid.

5 Traceability of True Hypermodel

Consider a hybrid of $\{CS_i\}_1^{I_1}$. In the searching process for the true hypermodel, we basically have to pay attention to all the element sets of $\bigcup_{i=1}^{I_1} CS_i$. However, as indicated at the end of the preceeding section, we have only to concern ourselves, under the SHA, with those element sets each of which satisfies either the condition of Theorems 4.1 or the condition of 4.2. This point is well noted in the process described below.

Suppose there are L element sets, $\psi_1, \cdots, \psi_L$, in $\bigcup_{i=1}^{I_1} CS_i$. If ψ_i does not satisfy the condition of Theorem 4.1 nor the condition of Theorem 4.2, let $T_i = \emptyset$; otherwise, apply either Theorem 4.1 or Theorem 4.2 to ψ_i accordingly. If Theorem 4.1 is applied, let

$$T_i = \{\psi_i \cup \{1\}, \{\psi_i, \{1\} \cup \varphi\}, \text{ for } \emptyset \subseteq \varphi \subset \psi_i\};$$

if Theorem 4.2 is applied, let

$$T_i = \{\{1\} \cup \psi_i\}.$$

Then we can obtain a collection $\mathcal{C}$ of the sets, where each set is composed of the element sets from the T_i's, one element set from each T_i. For each set $\mathbf{C}$ in the collection $\mathcal{C}$, if there is an element set $\mathbf{c}$ in $\mathbf{C}$ which is a subset of another element set in $\mathbf{C}$, then we remove the element set $\mathbf{c}$ from $\mathbf{C}$. Note that we don't need such an element set $\mathbf{c}$ under the hierarchy principle. We will call this removing process by *Subset-Removing* process. When the subsets are removed from all the sets in $\mathcal{C}$, we denote the new, subset-removed collection by $\mathcal{C}^*$.

For example, we consider a hybrid of $\{\{2\},\{3,4\}\}_1$ and $\{\{2,3\},\{3,4\}\}_2$.

$$\bigcup_{i=1}^{2} CS_i = \{\{2\},\{2,3\},\{3,4\}\}.$$

Let $\psi_1 = \{2\}$, $\psi_2 = \{2,3\}$ and $\psi_3 = \{3,4\}$. Then, $T_1 = \emptyset$, $T_2 = \{1,2,3\}$, and

$$T_3 = \{\{1,3,4\},\{\{3,4\},\{1\}\},\{\{3,4\},\{1,3\}\},\{\{3,4\},\{1,4\}\}\}.$$

By selecting one element set from each T_i, we have

$$\begin{aligned}\mathcal{C} \;=\; & \{\{\{1,2,3\},\{3,4\},\{1\}\},\{\{1,2,3\},\{3,4\},\{1,3\}\},\\ & \{\{1,2,3\},\{3,4\},\{1,4\}\},\{\{1,2,3\},\{1,3,4\}\}\}.\end{aligned}$$

After subset-removing, we have

$$\begin{aligned} C^* &= \{\{\{1,2,3\},\{3,4\}\},\{\{1,2,3\},\{3,4\},\{1,4\}\}, \\ &\quad \{\{1,2,3\},\{1,3,4\}\}\}. \end{aligned}$$

We will call the process from hybrid to C^* the *hypermodelling process*. Now our question is whether we can find the true hypermodel in C^*. The theorem below addresses this issue.

Theorem 5.1 (**Traceability Theorem**) *Suppose that the true hypermodel is graphical. Then under condition (2.4) and the SHA, the subset-removed collection C^* which is obtained through the hypermodelling process contains the true graphical hypermodel.*

Proof: Denote the true hypermodel by H. Without loss of generality, we may take

$$H = \{\theta_1, \cdots, \theta_k\}, \text{ for some subscript sets } \theta_i, \ i = 1, \cdots, k.$$

Suppose that we regroup the element sets of H according to whether an element set contains "1" and whether, for $\theta \in H$, $\theta \setminus \{1\} \subset \theta'$ for some $\theta' \in H$ which does not contain "1". (Note that, by condition (2.4) and the hierarchy principle, there exists at least one element set which contains "1".) Then H

is divided into 3 exhaustive and mutually exclusive subgroups where some subgroups may be empty. If $\{\theta_i\}_1^k$ satisfy the conditions (i), (ii) and (iii) of Theorem 3.4, then the three subgroups are

$$\{\theta_1, \cdots, \theta_s\}, \ \{\theta_{s+1}, \cdots, \theta_r\} \text{ and } \{\theta_{r+1}, \cdots, \theta_k\}. \tag{5.1}$$

(Note that s can be equal to 0 or r and that r can be equal to 1 or k.)

θ_j, $r+1 \leq j \leq k$, appear in all the CLLSs (by (iii') of Theorem 3.4). By Lemma 3.1, each θ_j, $s+1 \leq j \leq r$, appears in at least one of the CLLSs. None of $\{\theta_j\}_1^s$ appears in any CLLS, whose corresponding $u^{(x_1)}$-terms have been dubbed "disappearing" terms in section 2. Thus $\bigcup_1^{I_1} CS_i$ is composed of the sets, θ_j, $j = s+1, \cdots, k$, and some subsets of some $\theta_{j'}$, $j' = s+1, \cdots, r$.

We will use T_i for the same meaning as previously described. As indicated in the above paragraph, $\theta_j \setminus \{1\}$, $j = s+1, \cdots, k$, are in $\bigcup_1^{I_1} CS_i$ (note that for $j = r+1, \cdots, k$, $\theta_j \setminus \{1\} = \theta_j$.) For any set in

$$\left(\bigcup_1^{I_1} CS_i\right) \setminus \left(\bigcup_{s+1}^{k} (\theta_j \setminus \{1\})\right), \tag{5.2}$$

there exists,if $s \geq 1$, $j' \in \{s+1, \cdots, r\}$ such that the set is a subset of $\theta_{j'}$. As shown at the end of the previous section, the sets in (5.2) do not affect the LLS of a hypermodel under the hierarchy principle. Thus the sets may be ignored in the hypermodelling process. In other words, we may concern ourselves only with the sets $\theta_j \setminus \{1\}$, $j = s+1, \cdots, k$, in $\bigcup_1^{l_1} CS_i$ in the hypermodelling process. For $i = s+1, \cdots, k$, let T_i correspond to $\theta_i \setminus \{1\}$ in $\bigcup_1^{l_1} CS_i$.

Assuming H as the true hypermodel, with its element sets regrouped as in expression (5.1), we will prove that $H \in \mathcal{C}^*$.

For θ_j, $j \in \{r+1, \cdots, k\}$, we apply Theorem 4.1 and get

$$T_j = \{\theta_j \cup \{1\}, \{\theta_j, \{1\} \cup \varphi\}, \text{ for } \emptyset \subseteq \varphi \subset \theta_j\}. \tag{5.3}$$

For θ_j, $s+1 \leq j \leq r$, we apply either Theorem 4.1 or Theorem 4.2 according as whether θ_j is common in all the CLLSs or not. If θ_j is common in all the CLLSs, T_j is given by (5.3). Otherwise,

$$T_j = \{\theta_j\}. \tag{5.4}$$

For each θ_j, $j = r+1, \cdots, k$, there is at most one j' in $\{1, \cdots, s\}$ such that

$$(\theta_{j'} \setminus \{1\}) \subset \theta_j,$$

since the true hypermodel is graphical. Without loss of generality, we may suppose that

$$(\theta_i \setminus \{1\}) \subset \theta_{r+i}, \text{ for } i = 1, \cdots, s. \tag{5.5}$$

(Note that under the graphicality assumption $s \leq k - r$.)

Therefore, we have:

(i) For θ_j, $s+1 \leq j \leq r$, $\theta_j \in T_j$. (See (5.3) and (5.4).)

(ii) For $j \in \{1, \cdots, s\}$, $\{\theta_j, \theta_{j+r}\} \in T_{j+r}$. (See (5.3).)

(iii) For θ_j, $r+s < j \leq k$, $\{\theta_j, \{1\}\} \in T_j$. (Note that (2.4) is assumed and that if $r \geq 1$, then $\{1\}$ will not show up in the expression of a LLS under the hierarchy principle.) (See (5.3).)

The results, (i), (ii), and (iii), mean that H is in C^*. So far we have considered the cases where $1 \leq s \leq r \leq k$. But the result also holds for $s = 0$ by (i) and (iii). This completes the proof. □

We will illustrate the Traceability Theorem for the hypermodel given in Example 3.2. Before the illustration, we introduce a subset-removing operator $\langle\ \rangle$ on a set. For example,

$$\langle\{1,2\}, \{2,3\}, \{2,3,4\}\rangle = \{\{1,2\}, \{2,3,4\}\}.$$

Example 5.2 *For the hypermodel S in expression (3.2), the possible pairs of CLLSs are shown in Table 3.1. We will use the same notation as in Example 3(1) For the hybrid $Hybrid(S_1, S_1)$;*

$$\begin{aligned} C^* &= \{\langle\{1\} \cup \varphi_1, \theta_1', \{1\} \cup \varphi_4, \theta_4, \{1\} \cup \varphi_5, \theta_5\rangle; \\ &\quad \textit{for } \emptyset \subseteq \varphi_1 \subset \theta_1', \emptyset \subseteq \varphi_4 \subset \theta_4, \emptyset \subseteq \varphi_5 \subset \theta_5\}. \end{aligned} \tag{5.6}$$

(2) *For the hybrids Hybrid(S_1, S_2) and Hybrid(S_1, S_3):*

$$\begin{aligned} C^* &= \{\langle\{1\} \cup \theta_1', \{1\} \cup \varphi_4, \theta_4, \{1\} \cup \varphi_5, \theta_5\rangle; \\ &\qquad \text{for } \emptyset \subseteq \varphi_4 \subset \theta_4, \emptyset \subseteq \varphi_5 \subset \theta_5\}. \end{aligned} \tag{5.7}$$

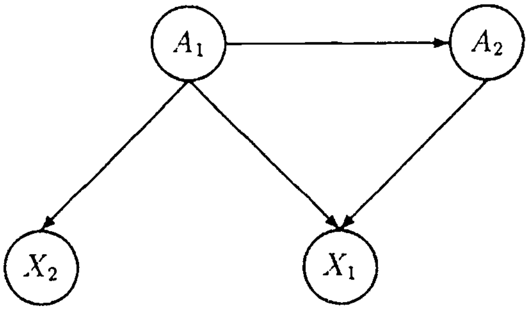
We can see that the true hypermodel S is included in each of (5.6) and (5.7).

□

6 An Illustration

Consider an educational test which consists of items that can be solved by one of two problem-solving strategies. In this section, we will illustrate an application of the main result of this paper supposing that we are given two influence diagrams (Oliver and Smith (1990)) of item responses and abilities for two problem-solving strategies each. Both influence diagrams (IDs) have the same structure as in Figure 6.1. In the figure, "A" stands for ability and "X" stands for item score, where X_i is the item score for item i. The arrow from A_1 to A_2 stands for that *possession of ability* A_2 requires *possession of ability* A_1 as a prerequisite. The arrows to the node of an item score from a set of the nodes of abilities mean that the abilities are tapped by the corresponding item. For instance, item 1 taps abilities denoted by A_1 and A_2; and item 2 taps the ability denoted by A_1. The marginal or conditional probabilities for

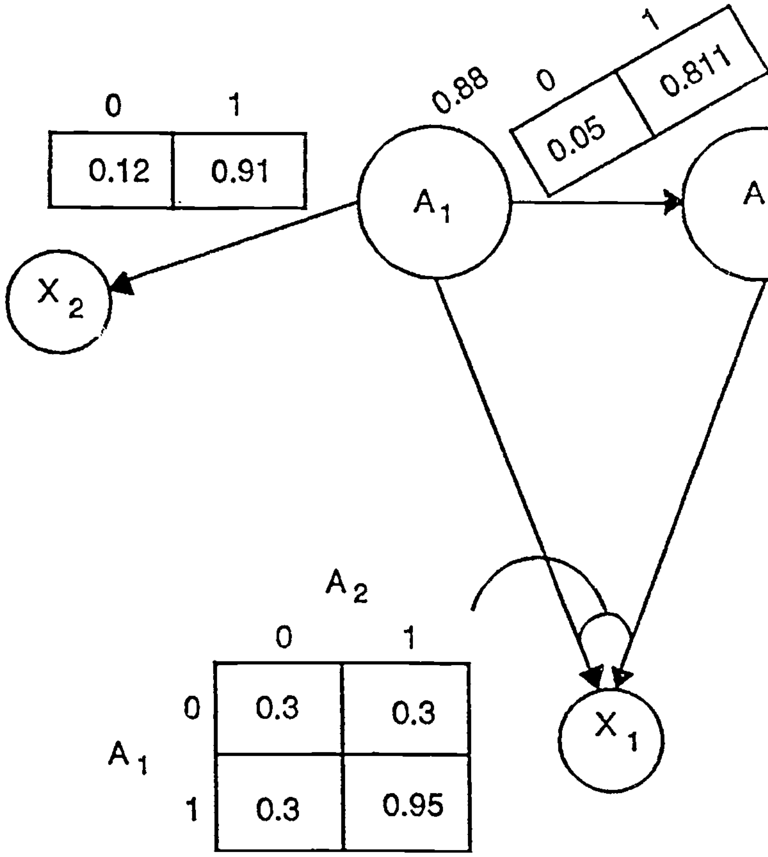
Figure 6.1: A conditional ID



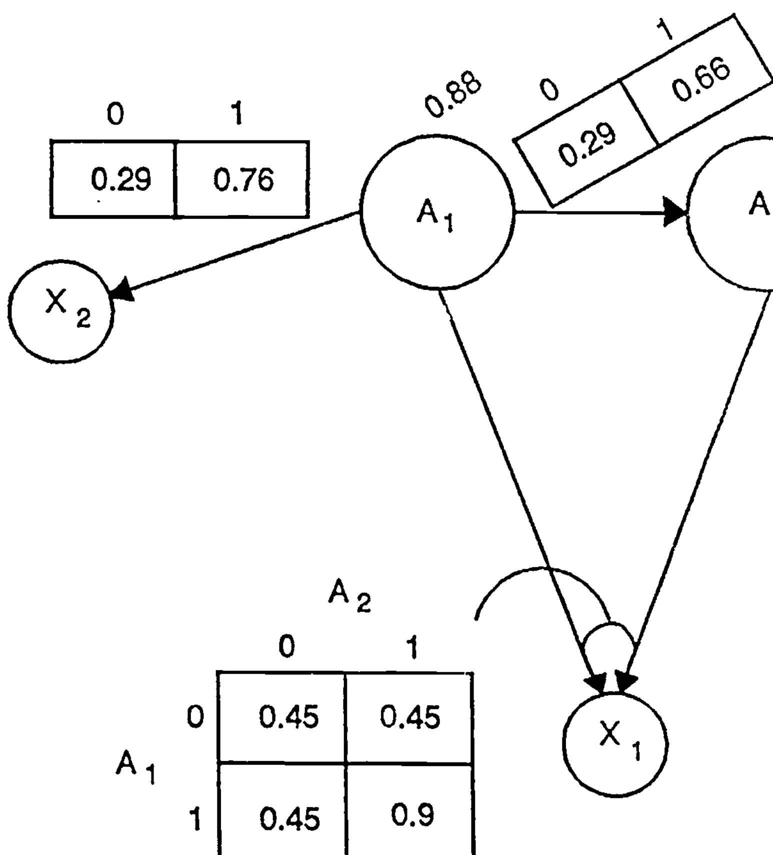
the IDs are given in Figure 6.2, where panel (a) shows the conditional ID given the problem-solving strategy ST-1, as does panel (b) given ST-2. Suppose the probabilities in Figure 6.2 are obtained from two distinct data sets, for each of the two strategies separately. From the conditional IDs in Figure 6.2, assuming that $P(ST = \text{"ST-1"}) = 0.4$, we generated two data sets, one (of size 768) from the model in panel (a) and the other (of size 1232) from panel (b). As anticipated, the log-linear analysis of the two data sets ends up with the LLS as in (6.1) for both data sets. Table 6.1 shows the result of the analysis.

We may ignore the directions in the graphs (see Lauritzen and Spiegelhalter (1988)), and regard each graph as a two-clique graph for the four categorical variables, A_1, A_2, X_1, and X_2. For both of the strategies, the CLLSs of the four categorical variables are

$$\{\{A_1, X_2\}, \{A_1, A_2, X_1\}\}. \tag{6.1}$$


0 1
0.12 0.91
0.88
0 1
0.05 0.811
A1
A2
X2
A2
0 1
A1
0 0.3 0.3
1 0.3 0.95
X1

(a)


0 1
0.29 0.76
0.88
0 1
0.29 0.66
A1
A2
X2
A2
0 1
A1
0 0.45 0.45
1 0.45 0.9
X1

(b)

Figure 6.2

Table 6.1: Log-linear analysis of the data from the panels (a) and (b) of Figure 6.2

(a) For *ST-1* (data size = 768)

Conditional LLS	d.f.	Likelihood-ratio		Pearson	
		χ^2	Prob.	χ^2	Prob.
$\{\{A_1, X_2\}, \{A_1, A_2, X_1\}\}$	6	2.80	0.83	2.82	0.83
$\{\{X_2\}, \{A_1, A_2, X_1\}\}^*$	7	288.11	0.00	357.56	0.00
$\{\{A_1, X_2\}, \{A_1, A_2\}, \{A_2, X_1\}, \{A_1, X_1\}\}^*$	7	36.50	0.00	50.09	0.00

(b) For *ST-2* (data size = 1232)

Conditional LLS	d.f.	Likelihood-ratio		Pearson	
		χ^2	Prob.	χ^2	Prob.
$\{\{A_1, X_2\}, \{A_1, A_2, X_1\}\}$	6	9.00	0.17	9.06	0.17
$\{\{X_2\}, \{A_1, A_2, X_1\}\}^*$	7	144.77	0.00	157.27	0.00
$\{\{A_1, X_2\}, \{A_1, A_2\}, \{A_2, X_1\}, \{A_1, X_1\}\}^*$	7	43.07	0.00	47.89	0.00

*: As appeared in the BMDP output, using a backward-deletion approach.

As shown in the proof of Traceability Theorem, we may confine ourselves to the sets, $\{A_1, X_2\}$ and $\{A_1, A_2, X_1\}$, in the hypermodelling process. Since the two sets are common to both CLLSs, by applying Theorem 4.1 to each of the two sets, we can see that the largest possible hypermodel is given by

$$\{\{ST, A_1, X_2\}, \{ST, A_1, A_2, X_1\}\}, \tag{6.2}$$

where "ST" is for the categorical variable of the strategy. By the collapsibility theorem (Theorem 2.5-1 of Bishop et al. (1975)), we may investigate the interaction within the individual sets marginally. The standardized u-terms (see section 4.4.2 of Bishop et al. (1975)) of the three-factor and the four-factor effects for the first and the second sets respectively in (6.2) can be estimated and used in searching for the true LLS. This may be a useful approach when data are not available for all the five variables.

However, we will take advantage of the simulation set-up where data are available for the five variables. The Traceability Theorem says that the true LLS is either the model in (6.2) or a submodel of it.

Actually, the model in (6.2) fits well to the simulated data set from the hybrid of the conditional IDs. The observed chi-square was 11.80 with 12 d.f., whose upper-tail probability is 0.46. Table 6.2 shows a BMDP output of model-fitting by backward-deletion that starts from the LLS in (6.2).

Table 6.2: Model-fitting result by BMDP

LLS	d.f.	Likelihood-ratio		Pearson	
		χ^2	Prob.	χ^2	Prob.
$\{\{ST,A_1,X_2\},\{ST,A_1,A_2,X_1\}\}$	12	11.80	0.46	11.88	0.46
$\{\{ST,A_1,X_2\},\{A_1,A_2,X_1\},\{ST,A_2,X_1\},$ $\{ST,A_1,X_1\},\{ST,A_1,A_2\}\}$	13	17.72	0.17	16.98	0.20
$\{\{ST,A_1,X_2\},\{A_1,A_2,X_1\},$ $\{ST,A_2,X_1\},\{ST,A_1,A_2\}\}$	14	20.24	0.12	19.27	0.15

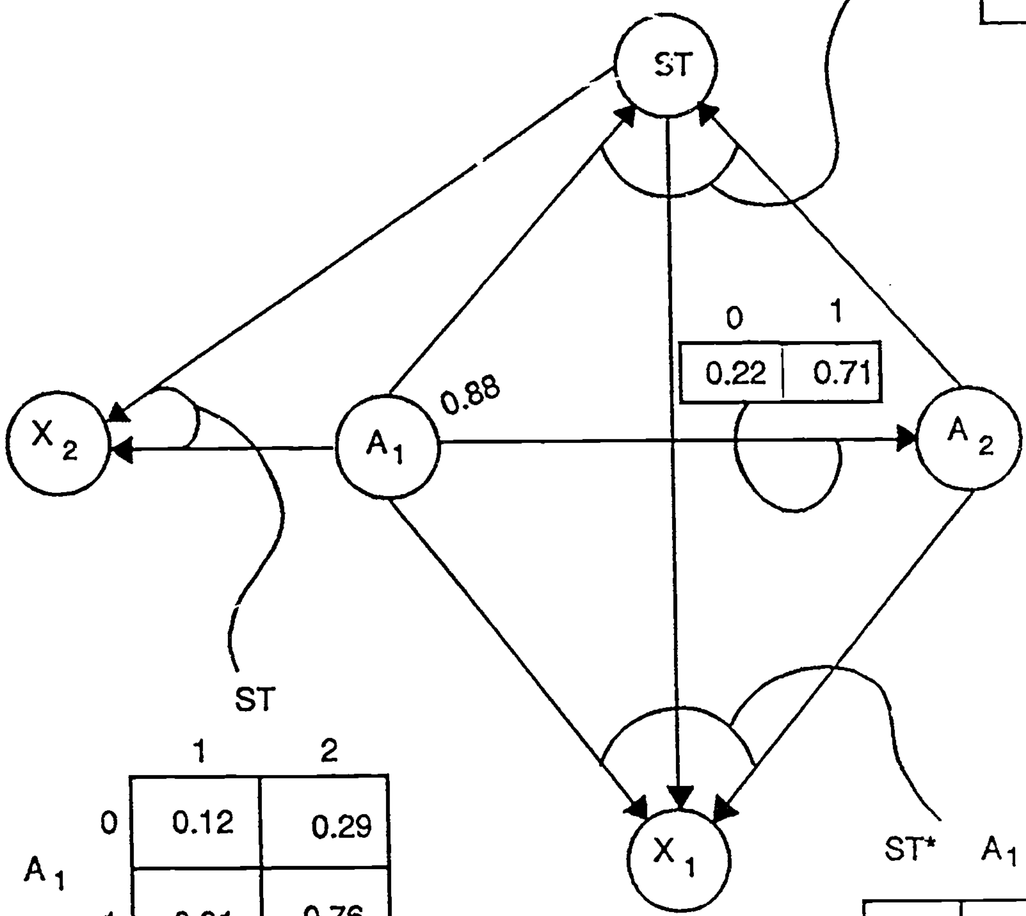
Table 6.2 suggests two other LLSs, $\{\{ST,A_1,X_2\},\{A_1,A_2,X_1\},\{ST,A_2,X_1\},\{ST,A_1,X_1\},\{ST,A_1,A_2\}\}$ and $\{\{ST,A_1,X_2\},\{A_1,A_2,X_1\},\{ST,A_2,X_1\},\{ST,A_1,A_2\}\}$, as possible candidate LLSs for the data. But neither of them is graphical.

We can represent the relationship among these five variables by an ID, still preserving the relationship among them in the language of LLS. We can construct an ID of the five variables in the following order[1] of the variables: A_1, A_2, ST, X_2, X_1. (See Pearl, Geiger, and Verma (1990).) We have an ID of the five variables in Figure 6.3, which is obtained as follows[2]. A_1 and A_2 do not look independent ($\chi^2_{(1)}$=232.18), and by the prerequisite relation between

[1]Any other ordering will do. But this ordering reflects the inherent relationship among the variables.

[2]See the table in the appendix for the simulated frequency table of the five variables.

		A_2	
		0	1
A_1	0	0.52	0.88
	1	0.73	0.56



		ST	
		1	2
A_1	0	0.12	0.29
	1	0.91	0.76

ST*	A_1	A_2	$P(X_1=1 \mid st, a_1, a_2)$
1	0	0	0.3
		1	0.3
	1	0	0.3
		1	0.95
2	0	0	0.45
		1	0.45
	1	0	0.45
		1	0.9

*"ST=i" means ST-i.

Figure 6.3

them, we put an arrow from node A_1 to node A_2. The three variables, A_1, A_2, and ST, look second-order associated (No submodel of the saturated LLS of the three variables fits well). Choice of a problem-solving method is subject to a test-taker's knowledge state and experience (Greeno and Simon (1988)). In other words, the test-taker's ability level may influence his or her choice of a problem-solving method, which is depicted by two converging arrows from the nodes of A_1 and A_2 to the node of ST. The three variables, A_1, ST, and X_2, make a clique in the LLS of the five variables, and the item score X_2 is influenced by the state of A_1 and the value of ST (i.e., the strategy selected). This can be expressed by

$$A_2 \perp X_2 \mid \{A_1, ST\} \tag{6.3}$$

in the ID construction process (see Corollary 4.1 of Pearl, Geiger, and Verma (1990)). The relationship in (6.3) is supported by the simulated data. Marginally for the 4 variables, ST, A_1, A_2 and X_2, the LLS, $\{\{ST, A_1, A_2\}, \{ST, A_1, X_2\}\}$, has the likelihood-ratio chi-square value 5.73 with 4 d.f. (its upper tail prob.=0.22). Finally, the four variables, A_1, A_2, ST, and X_1, make a clique, and, by the same reasoning as above, we drew arrows from A_1, A_2, and ST to X_1. This completes an ID for the five variables, A_1, A_2, ST, X_1, and X_2. The maximal or conditional probabilities for the ID are given in Figure 6.3. These probabilities are frequency probabilities based on Table A.1 in the appendix.

The covariance structures of the five variables from the combined ID (Figure 6.3) and from the hybrid of the conditional IDs (Figure 6.2) were very close to each other. The test statistic for the null hypothesis, that the covariance structures are the same, asymptotically follows the chi-square distribution with 15 d.f. (Section 10.6.1 of Anderson (1984)). For simulated data of size 3,000 (we regenerated them) from each of the two ID models, the combined ID and the hybrid of the conditional IDs, the observed value of the test statistic was 17.246 (p-value=0.304).

This section has illustrated an application of the main result of the paper to the problems of searching for a LLS using its conditional LLSs. In reality, data for "ST" may be missing. In this situation, however, we can apply the main result of the paper by assigning a prior distribution on the unobserved variable.

7 Summary Remarks

A main result in this paper is that we can see partial information about a true LLS from its CLLSs and use the information in trying to guess the true LLS. But we would face uncertainty, as indicated in Theorem 4.1, in the hypermodelling process, when all the CLLSs contain at least one common subscript set. Under the assumption that the true LLS is graphical, we can search for the true LLS by focusing on the "largest cliques" that are obtained from the CLLSs, as illustrated in the previous section.

The hypermodels are confined to be graphical in this paper. However, we can easily extend to the case where the true hypermodel is hierarchical. In such a case, the set of the possible hierarchical hypermodels will become much larger. If all the CLLSs of a hybrid are the same and given by $\{2, 3, \cdots, n+1\}$ for $1 \leq n < \infty$, then the number of the possible graphical hypermodels is given by $g_n = 2^n$; while that of the possible hierarchical hypermodels is given by

$$h_n = 1 + \sum_{i=1}^{n} \binom{n}{i} \sum_{j=1}^{i} f_i(j),$$

where $f_i(j)$ is the number of j non-empty subgroups of a set of i distinct elements, and its values, for $1 \leq j \leq i \leq 6$, are given in Table 7.1. The numbers g_n and h_n of up to $n = 6$ are given in Table 7.2.

Besides the numerical advantage, the graphical log-linear structure would be a natural means when we deal with experts' opinions which are given via causal networks or influence diagrams (see Lauritzen and Spiegelhaulter (1989), Shachter (1986) and Smith (1989)). As illustrated in section 6, the hy-

Table 7.1: Values of $f_i(j)$ for $1 \leq j \leq i \leq 6$

	1	2	3	4	5	6
1	1					
2	1	1				
3	1	3	1			
4	1	10	6	1		
5	1	15	25	10	1	
6	1	31	90	65	15	1

Table 7.2: Values of g_n and h_n for $1 \leq n \leq 6$

n	g_n	h_n
1	2	2
2	4	5
3	8	15
4	16	55
5	32	218
6	64	922

permodelling process may be useful when we want to build a larger graphical model from a set of conditional graphical models. Although a simple example was used in this paper, the approach can be applied to any sized graphical models in principle.

In this paper, each of the CLLSs in a hybrid involves the same set of variables. But it may not be unusual that the CLLSs involve different sets of variables. The difference could be due to nonresponse or variable-omission, the result of this paper is not of immediate use unless the missingness is ignorable. Research in this direction deserves our attention, and the author of this paper is currently exploring this issue.

Appendix

<table>
<tr><th rowspan="2">X_2</th><th rowspan="2">X_1</th><th rowspan="2">A_2</th><th rowspan="2">A_1</th><th colspan="2">ST</th></tr>
<tr><th>ST-1</th><th>ST-2</th></tr>
<tr><td rowspan="8">0</td><td rowspan="4">0</td><td rowspan="2">0</td><td>0</td><td>65</td><td>40</td></tr>
<tr><td>1</td><td>7</td><td>61</td></tr>
<tr><td rowspan="2">1</td><td>0</td><td>8</td><td>19</td></tr>
<tr><td>1</td><td>4</td><td>16</td></tr>
<tr><td rowspan="4">1</td><td rowspan="2">0</td><td>0</td><td>19</td><td>38</td></tr>
<tr><td>1</td><td>5</td><td>47</td></tr>
<tr><td rowspan="2">1</td><td>0</td><td>1</td><td>12</td></tr>
<tr><td>1</td><td>42</td><td>138</td></tr>
<tr><td rowspan="8">1</td><td rowspan="4">0</td><td rowspan="2">0</td><td>0</td><td>9</td><td>20</td></tr>
<tr><td>1</td><td>101</td><td>144</td></tr>
<tr><td rowspan="2">1</td><td>0</td><td>1</td><td>6</td></tr>
<tr><td>1</td><td>24</td><td>64</td></tr>
<tr><td rowspan="4">1</td><td rowspan="2">0</td><td>0</td><td>4</td><td>9</td></tr>
<tr><td>1</td><td>38</td><td>129</td></tr>
<tr><td rowspan="2">1</td><td>0</td><td>0</td><td>5</td></tr>
<tr><td>1</td><td>440</td><td>484</td></tr>
</table>

References

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis.* 2nd Ed. John Wiley & Sons.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1980) *Discrete Multivariate Analysis: Theory and Practice*, The MIT Press.

Darroch, J. N., Lauritzen, S. L., and Speed, T. P. (1980) Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.*, **8**, 522-539.

Greeno, J. G. and Simon, H. A. (1988). Problem solving and reasoning. *Technical Report AIP-85.* Dept. of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213.

Haertel, E. H. and Wiley, D. E. (in press). Representations of ability structures: Implications for testing. In *Test theory for a new generation of tests* (Eds: N. Fredericksen, R. J. Mislevy, and I. I. Bejar). NJ: Erlbaum.

Lauritzen, S. L. and Spiegelhalter, D. J. (1988) Local computations with probabilities on graphical structures and their application to expert systems, *J. R. S. S. Series B*, **50**, 157-224.

Oliver, R. M. and Smith, J. Q. (1990). *Influence Diagrams, Belief Nets and Decision Analysis* (edited). John Wiley & Sons.

Pearl, J. (1988). *Probabilistic Reasoning In Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc., San Mateo, CA.

Pearl, J., Geiger, D., and Verma, T. (1990). The Logic of Influence Diagrams. In *Influence Diagrams, Belief Nets and Decision Analysis* (Eds: R. M. Oliver and J. Q. Smith). John Wiley & Sons.

Shachter, R. D. (1986) Intelligent probabilistic inference. In *Uncertainty in Artificial Intelligence* (Eds: L. N. Kanal and J. F. Lemmer). Elsevier Science Publishers B.V., 371-382.

Smith, J. Q. (1989). Influence Diagrams for statistical modelling. *Ann. Statist.*, **17**, 654-672.

Tatsuoka, K. K. (1990). Toward an integration of item response theory and cognitive error diagnosis. In *Diagnostic monitoring of skill and knowledge acquisition* (Eds: N. Fredericksen, R. Glaser, A. Lesgold, and M. G. Shafto). NJ: Erlbaum.