Combining Conditional Log-Linear Structures with pieces of Information on Model Structures

SUNG-HO KIM

KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY

The possibility of representing many statistical models graphically is of special value when large numbers of variables are involved. Restrictions upon experiments and other forms of data collection may result in our being able to estimate only parts of a large graphical model. Confined to graphical log-linear models, Fienberg and Kim (1999) derived a theory to the effect that parts of a large model given in the form of conditional log-linear model can be combined into a larger one adding the conditional variable to the conditional model. In this paper, it is shown how we apply, in the model-combining process, the concept of prediction-refinement and pieces of information on the relationship among a subset of variables involved in a given model. An artificial example is used for a detailed description of the application and an example from real data supports the main result of the paper.

Key words: graphical models; hypermodel; hybrid of tree and log-linear structures ; strong hierarchy principle; model traceability; calibration; refinement; conditional independence.

1 Introduction

During the last two decades, much psychological research has been focused on tasks that better approximate the meaningful learning and problem-solving activities that engage people in everyday life. There has been increasing attention to the fine structure of abilities underlying task performance (see, for example, Haertel and Wiley (1992) and Mislevy (1994)). One approach to characterize the fine structure uses graphical models to represent relationships among abilities and test items, where abilities include problem-solving strategies (Mislevy et al., 1999)

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, 1987, 1990). Skills used in solving a problem may vary according to the strategies selected by test-takers, with some skills used in one strategy not being of use in another. Thus it may be desirable to build models in two steps when there are multiple strategies available. First we build separate models for each problem-solving strategy, then we combine them into a graphical model where the categorical variable for the problem-solving strategies is included as a new variable.

We can find a variety of instances of such a two-step modelling in AI, where the Bayesian network (Pearl, 1986, 1988) is one of the most popular forms of graphical models. The relative efficiency of computational techniques for performing inference over the network makes the graphical model an extremely powerful tool for dealing with uncertainty in AI. Generating a Bayesian network from a knowledge base or a database has been an important issue in the AI research (Cooper and Herskovits, 1991; Poole, 1993; Goldman and Charniak, 1993; Bacchus, 1993; Haddawy, 1994), and there has been growing interest in learning Bayesian networks from statistical data (Chickering and Heckerman, 1999; Friedman and Goldszmidt, 1998; Heckerman, Geiger, and Chickering, 1995; Neil, Wallace, and Korb, 1999). A key feature of the network-learning method is that we assume a prior over the set of network structures and apply a Bayesian scheme in search of the most probable model structure. In particular, Thiesson, Meek, Chickering, and Heckerman (1999) consider a method for Bayesian model selection among graphical models that are given in a form of mixtures of directed acyclic graphical models. In the mixture model, the model of each mixture component is defined conditional on some discrete random variable, which is regarded as latent in Thiesson et al. (1999).

While there has been remarkable improvements in learning Bayesian networks from data, the learning is mainly instrumented by heuristic searching algorithms since the model searching is usually NP-hard except some simple model structures (Chickering, 1996). A good review is given in Cooper(1999) on structural discovery of Bayesian or causal networks from data.

As for model representation, D'Ambrosio (1995), Geiger and Heckerman (1996), and Mahoney and Laskey (1999) consider a problem of representing Bayesian networks, using conditional probability models and conditioning variables, which are capable of explicitly capturing much of lower-level structural details. But they do not address an issue of combining conditional models.

Consider a graphical model of the n random variables, $X_1, \cdots, X_n$. The lower-level model-structures (or local structures) for a subset of the variables are defined conditional on some of the X variables. The local structures may vary across the values of the conditioning variable. This detailed representation helps in model searching and inference making since it reduces the number of parameters of the model and thus prevents the model from being overly large. A set of graphical models in the aforementioned problem-solving example is an instance of the local structures in a model corresponding to a given conditioning variable. Although such a model with lower-level local structures is beneficial to inference-making, it does not let us read the independence/dependence relationships among the X variables right off the model representation. For the structural interpretation of the whole model, Fienberg and Kim (1999) (FK hereafter) propose a theory to the effect that the graphical models can be combined in a consistent manner provided that the models are convertible into log-linear models. We will address in this paper an approach for the structural interpretation of a whole model through structure combination.

Suppose that the log-linear model (LLM) of $\mathbf{X} = (X_1, \cdots, X_n)$ is graphical and that X_1 is multinomial with category levels $1, \cdots, I$ and let $\mathbf{X}_{(i)}$ be the $(n-1)$-vector obtained by deleting X_i from $\mathbf{X}$. Consider a conditional LLM of $\mathbf{X}_{(1)}$ given that $X_1 = x_1$ and denote its log-linear structure by CS_{x_1}. Since X_1 takes on I different values, we can think of I such CS's, whose collection actually makes the LLM of $\mathbf{X}$. In other words, the collection is another form of model representation for $\mathbf{X}$. The conditional log-linear structures (CLLSs) are analogous to the local structures in the preceding paragraphs. We will depict the collection in a tree shape, with a node for X_1 and I arrows from it to the I nodes of CS's. Since this tree shape is a hybrid of tree and log-linear structures, we will call it a *hybrid*, represent it as

$$Hyb(X_1;\ CS_1, \cdots, CS_I), \tag{1}$$

and call it a *one-node hybrid.*

As is well known, a log-linear model of $\mathbf{X}$ is used to represent the logarithm of the joint probability distribution of $\mathbf{X}$ as a linear combination of non-negative terms, where each term is defined on a subset of $X_1, \cdots, X_n$. When the model is hierarchical, the model structure of $\mathbf{X}$ is determined by the terms whose domain subsets are maximal, i.e., each of the maximal domain subsets is not contained in any other domain subset in the model. This is why we can represent the model structure of a hierarchical log-linear model by a set of maximal domain subsets. Adapting our notation from Bishop, Fienberg, and Holland (1975), we will represent a log-linear structure by $\{\theta_1, \cdots, \theta_k\}$, where θ_i's are maximal domain subsets. For notational convenience, we use the indexes of the X variables to represent the log-linear structure.

We will call a log-linear structure corresponding to a one-node hybrid h a *hypermodel corresponding to h.* If all the CLLSs of a one-node hybrid are the same, we will say that the hybrid is *homogeneous*; if the hybrid is not homogeneous but there is a set, θ' say, which is shared by all the CLLSs of the hybrid, we will call the hybrid as *partially homogeneous* with respect to θ'; if there is no such common set at all, we will say that the hybrid is *heterogeneous.*

Suppose that a hybrid is obtained from a graphical log-linear model, m. FK show that m is contained in the set of the hypermodels each of which corresponds to the hybrid (Theorem 5 of FK).

According to FK, the number of hypermodels corresponding to a hybrid depends upon the level of homogeneity of the hybrid and the number of variables involved. When the hybrid is heterogeneous, only one hypermodel corresponds to it. If $n + 1$ variables are involved in a graphical hypermodel and the corresponding one-node hybrid as in (1) is homogeneous, then there are 2^n graphical hypermodels each of which is obtained by either connecting or disconnecting the X_1 node with some of the other n variables. But if we want to find a LLM for $\mathbf{X}$ based on a given hybrid of $\mathbf{X}$, it is desirable to have the size of the set of the hypermodels as small as possible. With this in mind, we will explore methods of reducing the set size of the hypermodels for a given hybrid. In this paper, we will consider hybrids that serve the purposes of prediction and structure interpretation and so the conditional variable of each one-node hybrid is the most informative for the predicted variable among a set of predictor variables. We will apply, for the reduction of the set size of the hypermodels, the notions of calibration and refinement (DeGroot and Fienberg, 1982) in addition to pieces of information on subsets of the variables of a given log-linear model.

This paper consists of 6 sections. Section 2 describes the relationship between a log-linear structure and its conditionals and carries the relationship over to the relation between hypermodel and hybrid, which is then summarized into the hypermodelling process. Section 3 describes the concepts of calibration and refinement and shows how the refinement concept serves in the hypermodelling process. Section 4 then illustrates how the refinement concept and pieces of information on subsets of the variables involved in a hybrid contribute to reducing the set size of the hypermodels. Section 5 is parallel to section 4 in spirit, differing in that we use real data and focus more on the use and interpretation of a hybrid and on how pieces of information are obtained and contribute to the hypermodelling. This paper concludes in section 6 with some remarks on possible applications of the main idea of the paper.

2 Hypermodelling Process

In this section, we briefly review, through three examples, a LLM and its conditionals and carry their relationship over to the relation between hybrid and hypermodel. The first two examples are about how CLLSs are related with a given log-linear structure (LLS for short), and the third example demonstrates searching for the set of the hypermodels that correspond to a given one-node hybrid.

Example 2.1 Consider a LLS for 5 variables, $X_1, \cdots, X_5$, as given by

$$\{\{1,2,3\},\{1,3,4\},\{3,4,5\}\}. \tag{2}$$

If we denote the three component sets in (2), respectively by $\{\theta_1, \theta_2, \theta_3\}$, we can write the corresponding LLM as

$$\log p_{12345}(x_1, \cdots, x_5) = u_{\theta_1} + u_{\theta_2} + u_{\theta_3} + R, \tag{3}$$

where R includes a constant u-term plus the summation of the u_θ-terms, each of whose subscript set is a strict subset of θ_i for some $i \in \{1,2,3\}$.

Let

$$\begin{aligned} w_\theta^{(x_1)} &= u_\theta + u_{\theta\cup\{1\}}, \\ w^{(x_1)} &= u + u_1 - \log p_1(x_1), \end{aligned}$$

Then the logarithm of the conditional probability of $X_2 = x_2, \cdots, X_5 = x_5$ given $X_1 = x_1$ is given by

$$\log p_{2345|1}(x_2, \cdots, x_5; x_1)$$
$$= w^{(x_1)} + w_2^{(x_1)} + w_3^{(x_1)} + w_4^{(x_1)} + u_5 + w_{23}^{(x_1)} + w_{34}^{(x_1)} + u_{35} + u_{45} + u_{345}. \quad (4)$$

Note in (4) that the CLLS is determined by the terms $w_{23}^{(x_1)}$ and u_{345} if neither of them equals zero. Also note that the index sets of these terms, $\{2,3\}$ and $\{3,4,5\}$, are included in the set

$$\{\theta_1 \setminus \{1\}, \theta_2 \setminus \{1\}, \theta_3 \setminus \{1\}\} = \{\{2,3\},\ \{3,4\},\ \{3,4,5\}\}. \qquad \square$$

As connoted in the last equation, we need know when to expect to see the full set of $u-$ or $w-$terms for a successful trip back to a LLM from a specific version of its conditional model (that depends on the value x_1). The lemma below plays an important role in searching for the set of the CLLSs that appear at the same time in a hybrid. Although the proof of the lemma is simple and appears in FK, it is presented here because it gives us an insight into the relation between hybrid and hypermodel.

Lemma 2.1 *Let $\theta \cap \{1\} = \emptyset$. Then, $u_\theta = u_{\{1\}\cup\theta} = 0$, iff $w_\theta^{(x_1)} = 0$ for all $x_1 = 1, \cdots, I$.*

Proof: Suppose that $w_{\theta(\underset{\sim}{x}_\theta)}^{(x_1)} = 0$, for all possible realizations $\underset{\sim}{x}_\theta$, and for all $x_1 = 1, \cdots, I$. Then it follows that, for $x_1 = 1, \cdots, I$,

$$u_{\{1\}\cup\theta(x_1,\ \underset{\sim}{x}_\theta)} = -u_{\theta(\underset{\sim}{x}_\theta)}, \quad \forall\ \underset{\sim}{x}_\theta. \quad (5)$$

Since $\sum_{x_1=1}^{I} u_{\{1\}\cup\theta(x_1,\ \underset{\sim}{x}_\theta)} = 0$, equation (5) implies that $u_\theta = 0$. The proof for the other direction is straightforward. $\square$

According to Lemma 2.1, it is possible that $w_{34}^{(x_1)} = 0$ for some value x_1 of X_1 when $u_{134} \neq 0$. If $w_{34}^{(x_1)}$ in (4) equals 0 for all values of x_1, then the full LLM is not hierarchical since $u_{345} \neq 0$; similarly, it is also possible that $w_3^{(x_1)} = 0$ while $w_{34}^{(x_1)} \neq 0$. These "non-hierarchical" situations are difficult to interpret. We shall assume, throughout the remainder of the paper, that the hierarchy principle holds for both the $CLLM$ and the LLM to avoid such situations. We refer to this as the *strong hierarchy principle* or the SHP for short.

Under the SHP, the CLLS is subject to whether a particular w-term is zero or not. For instance, the $CLLS$ of the model represented by expression (4) is determined by u_{345} and

Table 1: *The possible hybrids for the 5 random variables of Example 2.1.* ○ *indicates that the corresponding pair of $CLLS$s make a hybrid;* × *indicates that the pair does not make it.*

$X_1 = 1$	$X_1 = 2$		
	$\{\{2,3\},\{3,4,5\}\}$	$\{\{2\},\{3,4,5\}\}$	$\{\{3,4,5\}\}$
$\{\{2,3\},\{3,4,5\}\}$	○	○	○
$\{\{2\},\{3,4,5\}\}$	○	×	×
$\{\{3,4,5\}\}$	○	×	×

$w_{23}^{(x_1)}$ when $w_{23}^{(x_1)} \neq 0$, determined by u_{345} and $w_2^{(x_1)}$ when $w_{23}^{(x_1)} = 0$ and $w_2^{(x_1)} \neq 0$, and determined by u_{345} only when both $w_{23}^{(x_1)}$ and $w_2^{(x_1)}$ are non-zero. We refer to such situations where $w^{(\cdot)} = 0$ as *zero-w* phenomena.

Taking the zero-w phenomena into consideration in (4), we come up with the possible $CLLS$s for $X_2, \cdots, X_5$ given X_1 as

$$\{\{2,3\},\{3,4,5\}\}, \quad \{\{2\},\{3,4,5\}\}, \quad \{\{3,4,5\}\}. \tag{6}$$

We can also have the same result directly from the LLS as in (2). Once we condition on X_1 in the model given in (2), $\{1,3,4\}\setminus\{1\} = \{3,4\}$ is a subset of $\{3,4,5\}$ while $\{1,2,3\}\setminus\{1\} = \{2,3\}$ is not. Applying the zero-w phenomenon to $\{2,3\}$ leads us to the list in (6).

By applying Lemma 2.1 to the list in (6), we can obtain hybrids for the 5 random variables in Example 2.1. The possible hybrids are indicated in Table 1 under the assumption that $I = 2$. The table says that the CLLS $\{\{2,3\},\{3,4,5\}\}$ appears in every possible hybrid that corresponds to the LLM in expression (2), but not for the other CLLSs in (6). For example, if the model $Hyb(X_1; \{\{3,4,5\}\}, \{\{2\}, \ \{3,4,5\}\})$ holds true, then by Lemma 2.1 the set $\{1,2,3\}$ can not show up in (2). Note that $\{3,4,5\}$ appears in all the structures in (6), but this is not guaranteed when a set in a LLS contains "1". To make it clearer, we consider a simple LLS which consists of the sets only that contains "1" each.

Example 2.2 Consider a submodel of the model in (2):

$$\{\{1,2,3\}, \ \{1,3,4\}\}. \tag{7}$$

Since "1" is contained in both of the sets in expression (7), we represent the $CLLM$s in terms of w-terms only. Thus, as in Example 2.1, considering all the possible zero-w phenomena concerning both $\{1,2,3\}$ and $\{1,3,4\}$ gives rise to possible $CLLS$s of the form:

$$\{\varphi_1, \varphi_2\}, \qquad \text{where} \qquad \varphi_1 \subseteq \{2,3\},\ \varphi_2 \subseteq \{3,4\}. \tag{8}$$

□

According to Lemma 2.1, the sets $\{2,3\}$ and $\{3,4\}$ must show up in every hybrid that corresponds to the model in (7). It is worthwhile to note in both of the examples that for each set θ in a given LLS, $\theta \setminus \{1\}$ shows up in at least one of the CLLSs unless $\theta \setminus \{1\}$ is a subset of any other in LLS. Thus we may claim that for every set θ that is maximal in $\bigcup_{i=1}^{I} CS_i$, either θ or $\theta \cup \{1\}$ must show up in the corresponding LLS. This relationship between the hybrid and the LLS is a useful tool for finding the set of the LLMs corresponding to a given hybrid model. While we assumed a LLS in each of the preceding examples, we will assume a hybrid, in the next example, which is possible with respect to the model in (2).

We will say that a collection $\mathcal{C}$ consists of maximal sets if there is no non-empty component set in $\mathcal{C}$ that is a proper subset of others in $\mathcal{C}$. We will denote by $\langle \mathcal{C} \rangle$ the collection which is obtained by removing all the sets in $\mathcal{C}$ that are not maximal therein. For example, $\langle \{1,2\}, \{2,3\}, \{2,3,4\} \rangle = \{\{1,2\}, \{2,3,4\}\}$.

Example 2.3 Let $I = 2$ and consider the hybrid $Hyb(X_1;\ CS_1, CS_2)$ of $X_1, \cdots, X_5$ where

$$CS_1 = \{\{2\},\ \{3,4,5\}\} \ \text{ and } \ CS_2 = \{\{2,3\},\ \{3,4,5\}\}. \tag{9}$$

We see both u- and w-terms appear in expression (4). It is important to note that a u_θ-term in a $CLLM$ means that $u_{\{1\}\cup\theta} = 0$ while a w_θ-term in a $CLLM$ implies, by Lemma 2.1, that $u_{\{1\}\cup\theta} \neq 0$.

When a set is contained in all the CLLSs of a hybrid, we need to keep in mind that the corresponding term may be either a u-term or a w-term in the CLLMs. For instance, the set $\{3,4,5\}$ shows up in both of the $CLLS$s. This implies that the LLS H of $X_1, \cdots, X_5$ contains either $\{3,4,5\}$ or $\{1,3,4,5\}$. In the former situation, u_{345} appears in both of the $CLLM$s, while w_{345} appears in both of the $CLLM$s in the latter situation. As for $\{2,3\}$, it is obvious that $u_{123} \neq 0$, since $\{2,3\} \notin CS_1$. That is, $\{1,2,3\} \in H$. Thus, $\{\{1,2,3\}, \{3,4,5\}\}$ and $\{\{1,2,3\}, \{1,3,4,5\}\}$ are candidates for H.

We can obtain the full list of the candidates for H by carefully counting in the sets, $\{1\} \cup \varphi$

for $\emptyset \subseteq \varphi \subset \{3,4,5\}$. If we assume a graphical hypermodel, the full list is as follows:

$$\{\{1,2,3\},\ \{1,3,4,5\}\} \text{ and } \{\langle\{1,2,3\},\ \{3,4,5\},\ \{1\}\cup\varphi\rangle;\ \emptyset \subseteq \varphi \subset \{3,4,5\}\}. \tag{10}$$

According to the property of a graphical LLM (Fienberg, 1980; Darroch, Lauritzen, and Speed, 1980), it is apparent that at most one φ can show up in $\langle\cdot\rangle$ in (10). For example, let $\varphi = \{3\}$ and $\varphi' = \{5\}$. Then, the LLS $\langle\{1,2,3\},\ \{3,4,5\},\ \{1\}\cup\varphi, \{1\}\cup\varphi'\rangle$ is the same as $\langle\{1,2,3\},\ \{3,4,5\},\ \{1,3\},\ \{1,5\}\rangle = \{\{1,2,3\},\ \{1,5\},\ \{3,4,5\}\}$, which is not graphical for the same reason that $\{\{1,3\},\ \{1,5\},\ \{3,5\}\}$ is not.

Note that $Hyb(X_1;\ CS_1, CS_2)$ matches to any model in (10). For example, if the LLS is $\langle\{1,2,3\},\{3,4,5\},\{1,5\}\rangle = \{\{1,2,3\},\{3,4,5\},\{1,5\}\}$, we can easily see that the hybrid with the CS's in (9) is possible. In the CLLS, the terms for $\{2\}$ and $\{2,3\}$ are w-terms and the term for $\{3,4,5\}$ is either a u-term or a w-term. When the first model in (10) is true, we need a w-term for $\{3,4,5\}$ in the CLLS; otherwise, we need a u-term there. □

Although our argument is through examples, we have discussed the relationship between the LLS and the hybrid in the context of a log-linear model. The relationship is recapitulated as follows:

(i) If a set in a LLS does not contain $\{1\}$, the set shows up in every of its CLLSs.

(ii) If a set θ in a LLS contains $\{1\}$ and $\theta\setminus\{1\}$ is maximal in the LLS, then $\theta\setminus\{1\}$ must show up in a corresponding hybrid; but if $\theta\setminus\{1\}$ is a subset of some other set in the LLS, it does not show up in any of its CLLSs.

(iii) If a set θ shows up in every CLLS of a hybrid, then either θ or $\theta\cup\{1\}$ must be in a corresponding LLS.

(iv) If a set θ shows up in some but not all CLLSs of a hybrid, then $\theta\cup\{1\}$ must be in a corresponding LLS.

(v) In (iii), $\{1\}\cup\varphi$ is possible for at most one $\varphi\subset\theta$ in case that θ is in the LLS in stead of $\theta\cup\{1\}$.

We will refrain from describing in detail the relationship in general terms in this section but interested readers are referred to Appendix A. However, we will state more formally, for later use, (iii) and (v) into a theorem and (iv) into another.

Theorem 2.1 *(Theorem 3 of FK) Consider a hybrid $Hyb(X_1;\ CS_1, \cdots, CS_I)$, and suppose that the hybrid is partially homogeneous with respect to a set θ and that the corresponding hypermodel is graphical. Then, under the SHP and the condition that there is a main effect of X_1 in the hypermodel, the graphical hypermodel corresponding to $Hyb(X_1;\ CS_1, \cdots, CS_I)$ contains one of*

$$\{\langle\theta, \{1\} \cup \varphi\rangle,\ \ \emptyset \subseteq \varphi \subseteq \theta\}.$$

Theorem 2.2 *(Theorem 4 of FK) Consider a hybrid $Hyb(X_1;\ CS_1, \cdots, CS_I)$ and suppose that θ is an element set of some CLLS but not common to all CLLSs, and that there is no set in $\cup_{x_1=1}^{I} CS_{x_1} \setminus \theta$ that contains θ. Then, $\{1\} \cup \theta$ is a set in the corresponding hypermodel.*

Example 2.3 illustrates what to consider, under the assumption that the corresponding hypermodel is graphical, to find the set of the hypermodels corresponding to a given hybrid. We will elaborate more on this searching process for a general situation of one-node hybrids.

Assuming the SHP, we have only to concern ourselves with those element sets that satisfy the conditions of either Theorem 2.1 or Theorem 2.2. As a matter of fact, the set of such element sets can be expressed as $\langle\bigcup_{x_1=1}^{I} CS_{x_1}\rangle$. This point is well observed in the process described below.

Suppose that $\langle\bigcup_{x_1=1}^{I} CS_{x_1}\rangle$ consists of K element sets, $\psi_1, \cdots, \psi_K$. If ψ_i satisfies the condition of Theorem 2.1, then we let

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

and if ψ_i satisfies the condition of Theorem 2.2, we 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 C in the collection $\mathcal{C}$, if there is an element set c in C which is a subset of another element set in C, then we remove c from C. Removing all such c's yields $\langle C\rangle$. Note that we do not need such an element set c under the hierarchy principle. This is similar to the general approach to minimal sufficient statistics and generating classes for LLMs (e.g., see Bishop, Fienberg, and Holland (1975) and Whittaker (1990)). When we remove element subsets from all the sets in $\mathcal{C}$, we denote the resulting collection by $\mathcal{C}^*$. We refer to the process of moving from a hybrid to $\mathcal{C}^*$ as the *hypermodelling* *process*. The hypermodelling process leads us to the set of the hypermodels corresponding to a given hybrid.

3 Conditional Variable and Prediction Refinement

In this section, we will briefly review the concepts of calibration and refinement and discuss how we can make use of them in hypermodelling from a one-node hybrid. Consider a set of variables $X_1, \cdots, X_n, Y$, where predictions are to be made on Y, and suppose that we construct a one-node hybrid of the variables where the conditional variable is the most informative for Y among the X variables. Such a hybrid is dualpurpose in that it shows the most informative variable for Y and the interrelationship among the rest $n-1$ X variables and Y. In other words, the hybrid serves dual purposes of prediction and interpretation. In subsequent two sections, we deal with a general form of hybids that are tuned to these purposes. The hypermodelling process is a main tool for interpretation of the interrelationship among the variables in a hybrid and the concept of refinement would help the tool work better. The concept is rooted in the concept of calibration.

The concept of calibration is initially introduced in evaluating probability forecasters (Murphy and Epstein, 1967). Consider a long sequence of weather forecasts (rain/dry), look at those days for which the forecast probability of rain equals x and determine the long run proportion, $\rho(x)$, of such days on which the forecast event (rain) in fact occurred. The plot of $\rho(x)$ against x is the forecaster's empirical *calibration* curve. If the curve is diagonal, i.e. if $\rho(x) = x$, we say that the forecaster is empirically *well-calibrated*. In a more general sense, we may view $\rho(x)$ as our conditional probability of the event (e.g. rain tomorrow) to which the forecaster assigns the value x. Then the forecaster is well-calibrated if

$$\rho(x) = x, \text{ for each value } x \text{ the forecaster used.} \tag{11}$$

In comparing forecasters, calibration alone is not suitable as a criterion. To address the evaluation problem, DeGroot and Fienberg (1982) introduced the concept of *refinement* which is linked to the notion of Blackwell sufficiency in the comparison of experiments and can be used to induce a partial ordering on the class of all well-calibrated forecasters for the same sequence of events.

For two random variables, X and Y, whose support sets are Ω_X and Ω_Y respectively, we define a stochastic transformation $f(y|x)$ on $\Omega_Y \times \Omega_X$ as follows:

$$f(y|x) \geq 0 \text{ for } x \in \Omega_X \text{ and } y \in \Omega_Y,$$

$$\sum_{y \in \Omega_Y} f(y|x) = 1 \text{ for } x \in \Omega_X.$$

Let X_A be a random variable representing the probability predictions by forecastor A, χ_A the support set of X_A, and ν_A the distribution function over χ_A. Similarly, let χ_B be the set of forecasts by B and let ν_B be the distribution function over χ_B. Assuming that forecasters A and B are well-calibrated, we say that A is at least as refined as B if there exists a stochastic transformation f such that the following relations are satisfied:

$$\sum_{x \in \chi_A} f(y|x)\nu_A(x) = \nu_B(y) \text{ for } y \in \chi_B, \tag{12}$$

$$\sum_{x \in \chi_A} f(y|x)x\nu_A(x) = y\nu_B(y) \text{ for } y \in \chi_B. \tag{13}$$

A multivariate version of this definition is the same except that x and y in these expressions are in vector (DeGroot and Fienberg, 1986).

The function f determines the auxiliary randomization that is to be carried out. If A makes a prediction x for a particular case, then we can generate a prediction y by means of an auxiliary randomization in accordance with the conditional probability distribution $f(y|x)$. The relation (12) guarantees that in this way we will make each prediction y with the same frequency $\nu_B(y)$ that B does, and the relation (13) guarantees that our predictions will again be well calibrated. If A is at least as refined as B and ν_A and ν_B are not identically equal (i.e, the predictions by A and B are not the same in distribution), then A is said to be *more refined* than B (DeGroot and Fienberg, 1986, p. 250). Two explicative statements of refinement are given below.

Theorem 3.1 *(Theorem 1 of DeGroot and Eriksson (1985)) Suppose that A and B are well-calibrated predictors. Then the relationship that A is at least as refined as B is equivalent to that there exist disrete random variables X_A and X_B such that*

$$E(X_A|X_B) = X_B.$$

Theorem 3.2 *(Theorem 15 of DeGroot and Fienberg (1986)) Suppose that both forecasters A and B are well-calibrated for the forecasting of events with $k \geq 2$ outcomes. Then, the condition* *that A is at least as refined as B is equivalent to the condition that, for every continuous convex function $g(x)$ defined on the $(k-1)$-dimensional simplex,*

$$E(g(X_A)) \geq E(g(X_B)).$$

A common feature in the definition of refinement and the above two theorems is that more refined predictions $\{x_A\}$ are spread over a larger domain than their counterpart $\{x_B\}$ while centered at the same point of average. We can see from Theorem 3.2 that a more refined probability forecaster gives predictions that are closer to actual outcomes than the predictions made by a less refined forecaster. Note that the actual outcomes correspond to the vertices of the $(k-1)$-dimensional simplex and the probability predictions all the points on the simplex. A formal description of this follows. Let X denote the actual outcomes. Then the mean squared error of X_A is $MSE(X_A) = E(X_A - X)^2$.

$$\begin{aligned} MSE(X_A) - MSE(X_B) &= E(X_A^2 - X_B^2) - 2E((X_A - X_B)X) \\ &= E(X_B^2) - E(X_A^2) \qquad (14) \\ &\leq 0, \end{aligned}$$

where equality (14) follows from that $E(X_A X) = E(X_A^2)$ and $E(X_B X) = E(X_B^2)$ which is possible by the well-calibratedness of the predictors and the inequality is from Theorem 3.2. So, if A is at least as refined as B, then $MSE(X_A) \leq MSE(X_B)$. This result holds whether the predictions are given in scalar or in vector.

Theorem 3.3 is immediate from Theorem 3.1.

Theorem 3.3 *Suppose that Y, X_1, and X_2 are discrete random variables taking on a finite number of values. Then the set of predictions made in terms of $E(Y|X_1, X_2)$ is at least as refined as that made in terms of $E(Y|X_1)$.*

Proof: In Theorem 3.1, we replace X_A and X_B by $E(Y|X_1, X_2)$ and $E(Y|X_1)$. Then

$$\begin{aligned} E(X_A|X_B) &= E(E(Y|X_1, X_2)|E(Y|X_1)) \\ &= E(Y|E(Y|X_1)) \\ &= E(E(Y|X_1)|E(Y|X_1)) \\ &= E(Y|X_1) \\ &= X_B, \end{aligned}$$

where the second and the third equalities are possible by Theorem 6.5.10 of Ash (1972). □

If A makes predictions in terms of $E(Y|X_1, X_2)$ and B in terms of $E(Y|X_1)$, then A is more refined than B when $\nu_A \neq \nu_B$. The corollary below is immediate from Theorem 3.3, and so its proof is omitted.

Corollary 3.4 *If the LLM of X_1, X_2, and Y is given by $\{\{1,2\},\{2,Y\}\}$, and predictions are made in terms of $E(Y|X)$, then the set of predictions by $E(Y|X_2)$ is at least as refined as that by $E(Y|X_1)$.*

In multivariate prediction problems, suppose that probability predictions are made for k events at a time. In the theorem below a "partition" refers to a partition of the k events into some number of non-empty, mutually exclusive, and exhaustive subsets of the k events.

Theorem 3.5 *(Theorem 13 of DeGroot and Fienberg (1986)). If A and B are well-calibrated multivariate predictors, and A is at least as refined as B, then A is also marginally at least as refined as B with respect to all possible partitions of the predicted events.*

The following corollary is a generalized version of Corollary 3.4.

Corollary 3.6 *Suppose that the LLM of $X_1, X_2, \cdots, X_n$, and Y is graphical with the graph where X_1 is connected to X_2 by an edge and X_1 is conditionally independent of $X_3, \cdots, X_n, Y$ given X_2. Then the set of predictions by $E(Y|X_2)$ is at least as refined as that by $E(Y|X_1)$.*

Proof: Regard the conjunction of $X_3, \cdots, X_n$, and Y as a new random vector Y^*. Then, it follows from the condition of the corollary and by Theorem 3.3 that $E(Y^*|X_2)$ is at least as refined as $E(Y^*|X_1)$. So, by Theorem 3.5, we have the desired result. □

If predictions are made on Y of Corollary 3.4, then the prediction by $E(Y|X_2)$ is more refined than the prediction by $E(Y|X_1)$ when $E(Y|X_1)$ is different from $E(Y|X_2)$ in distribution. According to the structure $\{\{1,2\},\{2,Y\}\}$ in Corollary 3.4, X_2 is more informative for Y unless X_1 is a one-to-one function of X_2. Thus, if we build a hybrid by selecting a more informative X variable for Y as a conditional variable, then the resulting hybrid is as in Figure 1.

Fienberg and Kim (1998) proved a theorem that branching at a node of a probability classification tree improves the prediction refinement of the tree. In other words, a probability

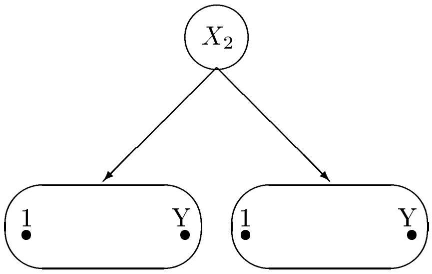


Figure 1: *The hybrid of 3 random variables, X_1, X_2, and Y, whose LLS is as in Corollary 3.4.*

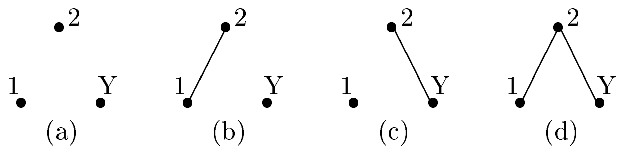


Figure 2: *The hypermodels corresponding to the hybrid in Figure 1*

classification tree that is obtained by branching a tree τ using informative predictor variables gives more refined predictions than τ. Note that the convex function g in Theorem 3.2 has little thing to do with the selection of predictor variables. The x in $g(x)$ is a prediction value, not the value of a predictor variable X. When we build a probability classification tree, we usually use the squared error loss function in such a way that the expected prediction error is minimized. The squared error loss gives both the prediction value (in the form of conditional mean) and the expected error (in the form of mean squared error). Furthermore, it is worthwhile to note in (14) that the difference in the mean squared errors of a pair of predictions is the same as the negative of the difference in the means of the corresponding squared predictions.

Let $Y_{X_i} = E(Y|X_i)$, $i = 1, 2$. Then

$$MSE(Y_{X_2}) < MSE(Y_{X_1}) \tag{15}$$

implies, by (14) and Theorem 3.2, that the prediction by Y_{X_1} is not as refined as the prediction by Y_{X_2}. In particular, when we use the squared error loss in building a probability classification tree, the inequality (15) implies, by the theorem in Fienberg and Kim (1998), that Y_{X_2} is more refined than Y_{X_1}. This point is well observed in the hypermodelling process in the subsequent sections. The hybrid in Figure 1 is homogeneous. So by applying Theorem 2.1, we obtain the corresponding hypermodels as in Figure 2. In this figure, graphs a and b are negligible from the viewpoint of a prediction tree. Note that graph d is the structure considered in Corollary

3.4.

4 An Illustration: Reduction of the Number of the Hypermodels

In this section we will use an artificial example to show explicitly how the set-size of the hypermodels is reduced and how much. Those who are more interested in the use and interpretation of the hybrid and in obtaining information about a subset of variables such as in (16) may skip this section right after Theorem 4.1. Although both this and next sections do illustrate reduction of the set-size, each is different from the other in that this section focuses, by using a relatively simple hybrid, more on the hypermodelling process and a detailed description of the set-size reduction, while the next section focuses more on the use and interpretation of a hybrid, how pieces of information on subsets of variables are obtained, and on how the pieces of information alleviate our working load during the hypermodelling process.

Definition 4.1 *Suppose that a LLS S of $X_1, X_2, \cdots, X_n$ is given by*

$$S = \{\theta_1, \theta_2, \cdots, \theta_k\}.$$

Let φ be a subset of the index set of $X_1, X_2, \cdots, X_n$. Then, the LLS represented by

$$\langle \theta_1 \cap \varphi, \cdots, \theta_k \cap \varphi \rangle$$

is called the **submodel structure of S confined to** *φ and is denoted by S_φ.*

Theorem 4.1 *Consider a hybrid $Hyb(X_1;\ CS_1, CS_2, \cdots, CS_I)$ and let*

$$\mathcal{D} = \langle CS_1, \cdots, CS_I \rangle.$$

Then, if $\mathcal{D}$ is not graphical, neither is any hypermodel corresponding to the hybrid.

Proof: Let H be a hypermodel from the hybrid of the theorem. According to the hypermodelling process, it is obvious that $\mathcal{D} = H_\varphi$, where φ is the index set of the variables that are involved in at least one of the CLLSs of the hybrid of the theorem.

The hypermodel is obtained by connecting the node of X_1 to a subset of the variables that are indexed in $\mathcal{D}$. Hence, the result of the theorem follows. □

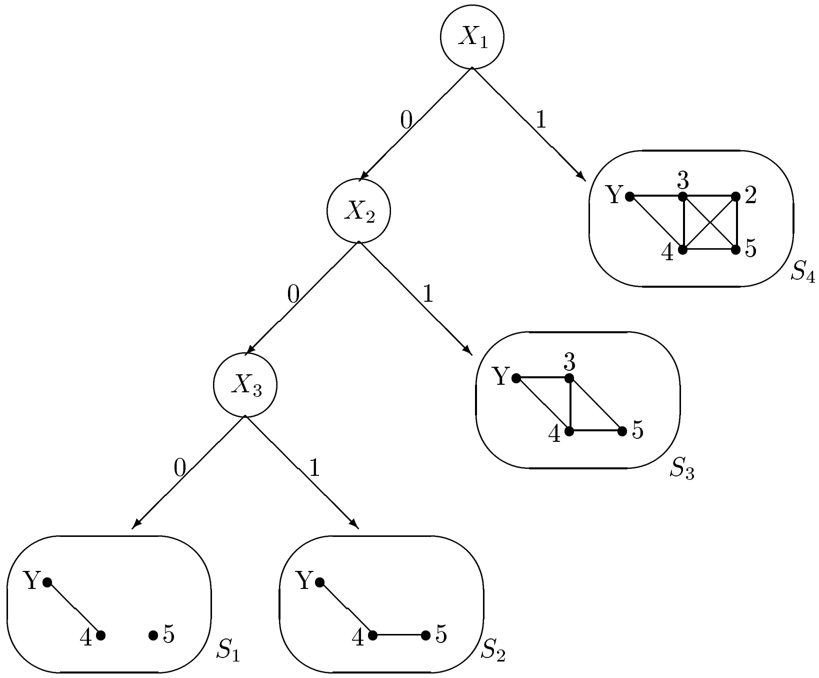
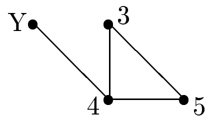


Figure 3: *A hybrid of* $X_1, \cdots, X_5$*, and* Y*. The number on each arrow represents the value of the random variable at the arrow-tail.*

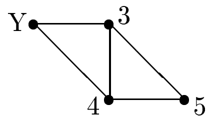
This theorem states that if any submodel structure is not graphical, neither is the LLS. We will call this the *graphicality condition.*

We will assume that all the hypermodels are graphical in the rest of the paper. The graphicality condition, the nature of a probability prediction tree that a more informative variable comes before a less informative one in any path of a tree, and experts' partial knowledge concerning the interrelationship among the variables will be employed together as useful criteria in the set-size reduction process toward the set of the hypermodels.

Suppose that we are given a hybrid as in Figure 3 where 6 variables $X_1, \cdots, X_5, Y$ are involved and that all the X variables are binary. Also suppose that the hybrid is constructed as a statistical decision-aid for making predictions about Y. Observing X_1 first and when $X_1 = 0$, continue observing X_2, otherwise stop observation and look into structure S_4; when $X_2 = 0$, continue observing X_3, otherwise stop observation and look into structure S_3; variable X_3 is the last observation; when $X_3 = 0$, look into S_1, otherwise S_2. The squared error loss function is used in tree construction and the variables are selected from the set of the predictor variables one at each node so that the prediction refinement may be improved the most (Breiman et al., 1984; Kim, 1994). In this regard, we may interpret the hybrid as that X_1 is the most informative for Y among $X_1, \cdots, X_5$, X_2 is the most informative for Y among the X variables given that $X_1 = 0$, and that X_3 is the most informative for Y among the X



(a) $H_{3,1}$



(b) $H_{3,2}$

Figure 4: *The possible hypermodels corresponding to* $Hyb(X_3; S_1, S_2)$.

variables given that $X_1 = 0, X_2 = 0$.

Suppose that we have the following information about the interrelationship among the variables:

$$Y \perp X_5 | (X_3, X_4), \tag{16}$$

that is, Y and X_5 are conditionally independent given the outcome of (X_3, X_4). This kind of information about the relationship structure would often play a crucial role in reducing the set-size of the hypermodels unless the given hybrid consists of heterogeneous one-node hybrids only. Such information can be obtained through consulting relevant experts. Actually, in building a LLM by applying a stepwise procedure (e.g., Ch. 5 of Fienberg (1980)), we often resort to our common sense on the intrinsic relationship among a set of variables. We will see in next section how such information is obtained when working with real data.

Note that $Hyb(X_3; S_1, S_2)$ is partially homogeneous, the same structure for X_4 and Y but not for X_4 and X_5. Thus applying Theorems 2.1 and 2.2 respectively to the pair of X_4 and Y and the pair of X_4 and X_5 yields the hypermodels $\{H_{3,1}, H_{3,2}\}$ which are depicted in Figure 4. But, in $H_{3,1}$, $Y \perp \{3, 5\}|4$, meaning that X_4 is more informative for Y than X_3 and X_5, which is incompatible with the hybrid in Figure 3 by Corollary 3.6. Hence, $H_{3,2}$ is the only hypermodel corresponding to $Hyb(X_3; S_1, S_2)$. We denote by $H_{i,j}$ the jth of the hypermodels that correspond to $Hyb(X_i;\ \cdot, \cdot)$ and by S_l the lth CLLS of a given hybrid.

Proceeding to $Hyb(X_2; H_{3,2}, S_3)$, we have the following collection $\mathcal{C}$ of the hypermodels:

$$\mathcal{C} = \{H_{2,1}, H_{2,2}, \cdots, H_{2,16}\}.$$

The 16 graphs of the set $\mathcal{C}$ are in Figure B.1 in Appendix B.

By Corollary 3.6, we can remove $H_{2,3}$, $H_{2,4}$, and $H_{2,5}$ from the set $\mathcal{C}$ since X_2 is less informative for Y than X_3, X_4, and X_5, respectively in $H_{2,3}$, $H_{2,4}$, and $H_{2,5}$. And in $H_{2,1}$, X_2 is independent of Y, while it is selected as informative for Y in the hybrid in Figure 3. Thus $H_{2,1}$ is also removed.

Table 2: *A summary of the final hypermodelling process with the hybrid in Figure 3. The second column lists the submodel structures, confined to the set* $\{2, 3, 4, 5, Y\}$, *of the hypermodels in the third column.*

Hybrid	Submodel structure	Hypermodels
$Hyb(X_1; H_{2,2}, S_4)$	$\{\{2,3,4,5\},\{3,4,Y\},\{2,Y\}\}$	The submodel structures are not graphical.
$Hyb(X_1; H_{2,6}, S_4)$	$\{\{2,3,4,5\},\{3,4,Y\},\{2,3,Y\}\}$	$H_{1,1} = \{\{1,2,3,4,5\},\{1,3,4,Y\},$ $\{1,2,3,Y\}\}$
$Hyb(X_1; H_{2,7}, S_4)$	$\{\{2,3,4,5\},\{3,4,Y\},\{2,4,Y\}\}$	The submodel structures are not graphical.
$Hyb(X_1; H_{2,9}, S_4)$	$\{\{2,3,4,5\},\{3,4,Y\}\}$	$H_{1,2} = \{\{1,2,3,4,5\},\{3,4,Y\}\}$ $H_{1,3} = \{\{1,2,3,4,5\},\{1,3,4,Y\}\}$
$Hyb(X_1; H_{2,10}, S_4)$	$\{\{2,3,4,5\},\{3,4,Y\}\}$	$H_{1,2}, H_{1,3}$
$Hyb(X_1; H_{2,11}, S_4)$	$\{\{2,3,4,5\},\{3,4,Y\}\}$	$H_{1,2}, H_{1,3}$
$Hyb(X_1; H_{2,12}, S_4)$	$\{\{2,3,4,5\},\{2,3,4,Y\}\}$	$H_{1,4} = \{\{1,2,3,4,5\},\{1,2,3,4,Y\}\}$
$Hyb(X_1; H_{2,15}, S_4)$	$\{\{2,3,4,5\},\{3,4,Y\}\}$	As in the note below.

NOTE: There are 32 corresponding hypermodels which are obtained by connecting X_1 to $X_2, \cdots, X_5$, and Y in 32 different ways.

According to Lemma 2.1, once a graph in $\mathcal{C}$ is taken as a possible hypermodel, its graphical feature is retained in the follow-up hypermodelling process. For example, if $H_{2,6}$ is taken, then in the follow-up hypermodels, $H_{2,6}$ is embedded in the graphical feature among $\{X_2, X_3, X_4, X_5, Y\}$. Hence, by condition (16) and the refinement condition, the collection $\mathcal{C}'$ of the possible hypermodels is given by

$$\mathcal{C}' = \{H_{2,2}, H_{2,6}, H_{2,7}, H_{2,9}, H_{2,10}, H_{2,11}, H_{2,12}, H_{2,15}\}.$$

None of $H_{2,8}, H_{2,13}, H_{2,14}, H_{2,16}$ satisfy condition (16). Condition (16) and the refinement concept have cut down the size of $\mathcal{C}$ into half.

Now from $Hyb(X_1; H', S_4)$ for each $H' \in \mathcal{C}'$, we can get the set of the final possible hypermodels. The last hypermodelling process is summarized in Table 2. In the table, when the submodel structures are not graphical, the corresponding hypermodels are ignored. Hypermodel $H_{1,1}$ is not graphical, and neither hypermodel $H_{1,3}$ nor $H_{1,4}$ does satisfy condition (16). As for the submodel structure in the last row of the table, note that the hybrid $Hyb(X_1;\ H_{2,15}, S_4)$ is homogeneous. And so by applying Theorem 2.1, we obtain 32 ($= 2^5$) graphical hypermodels by connecting the X_1 node to 32 different subsets of $X_2, \cdots, X_5$, and Y including the empty subset. Since hypermodel $H_{1,2}$ is one of the 32 models, we may ignore it, and so there are 32 possible graphical hypermodels. Note in Table 2 that the submodel structure is given by $\{\{2,3,4,5\},\ \{3,4,Y\}\}$, where condition (3) is satisfied and that the corresponding hypermodels are obtained by putting edges between X_1 and the other 5 variables in the submodel structure.

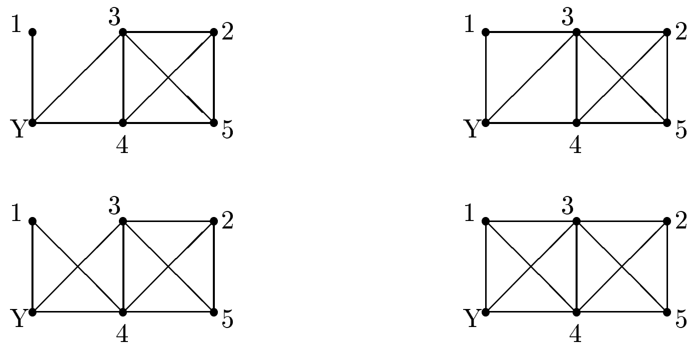


Figure 5: *The hypermodels obtained from the hybrid in Figure 3 under condition (16).*

Therefore, the hypermodels where both X_5 and Y are connected to X_1 must be removed from the 32 models since they violate condition (3). Further removals take place in light of Corollary 3.6. The hypermodels where X_1 is connected only to one of $X_2, \cdots, X_5$ are incompatible with the hybrid in Figure 3 by Corollary 3.6. After these removals, we come up with the four graphs in Figure 5. Note that condition (16) and the refinement concept have served to reduce the number of the possible hypermodels from 34 (the 32 models plus $H_{1,3}$ and $H_{1,4}$) down to only 4 at the last hypermodelling.

If we look back at the whole hypermodelling process for the hybrid in Figure 3, we can see that, as for the X_3-hybrid, condition (16) and the refinement concept have reduced the number of the hypermodels down to half of the number of all the possible hypermodels that would be obtained without considering condition (3) and the refinement concept, the same for the X_2-hybrid, and down to nearly one eighth for the X_1-hybrid. In other words, condition (16) and the refinement concept have reduced the number of the hypermodels corresponding to the hybrid in Figure 3 down to about 1/32 of the number of all the possible hypermodels from the hybrid in Figure 3.

5 Reduction of the Number of the Hypermodels with Real Data

In this section a real data set of size 10,025 that is collected for lecture evaluation at the end of the spring semester of 1998 at a university in South Korea is used. The survey items that are selected for use are 7 out of 25. The selected 7 items are about lecture and teaching assistance(TA)'s performance. We will label the random variables for the 7 items by A, H, F,

Table 3: The 7 labels of items and their meanings

label	key words	item contents
A	Aim of lecture	Has the lecture been carried out with a clear aim of lecture and a profound knowledge of the course?
H	Homework	Was the homework assignments helpful in understanding the lecture and related subjects?
F	Feedback	Did you get a satisfactory feedback from the comments on your homeworks?
R	Recommendation	Would you recommend this course to your friends?
O	Organization	Was the lecture well organized throughout the course?
S	Sincerity	Was the lecture given with full sincerity with regard to lecture preparation, response to students' questions in class, and the likes?
T	Teaching assistance	Are you satisfied with the TAs' performance?

R, O, S, and T, which are explained in Table 3. Each item has three options, *negative, half-and-half, positive*, and so the random variables are all ternary taking on values 1 (for negative), 2 (for half-and-half), or 3 (for positive). The frequency table of the data is of $3^7 = 2187$ cells and is displayed in Appendix C.

The faculty of the university is in general interested in the response to item R and so they want to see what affects R. To see this, I ran a computer algorithm CART (Classification And Regression Trees) (see Breiman et al., 1984) to construct a regession-tree with R as a response variable and the others explanatory. CART is a computer program for a non-parametric regression analysis. It selects explanatory variables one after another in such a way that each selection improves the prediction accuracy mostly among the explanatory variables conditional on the outcomes of the already selected explanatory variables. A final tree is determined by a test-set method or a cross-validation method so that the sum of the prediction accuracy and a penalty of tree-size is minimized.

Figure 6 shows the regression result by CART, which is obtained via a test-set method so that the tree size may be minimized while keeping the prediction error as small as possible. The meanings of arrows in the figure and the numbers on them are the same as those in Figure 3. An arrow connecting a pair of circles represents a sequence of variable selection, from the variable at its tail to the variable at its head, provided that the variable at the tail of the arrow takes on one of the values on the arrow; and an arrow from a circle to a CLLS means that when the variable at the arrow-tail takes on the value on the arrow, no more variable selection is made and a CLLS is tagged for the variables that are not selected yet. Actually, CART yields

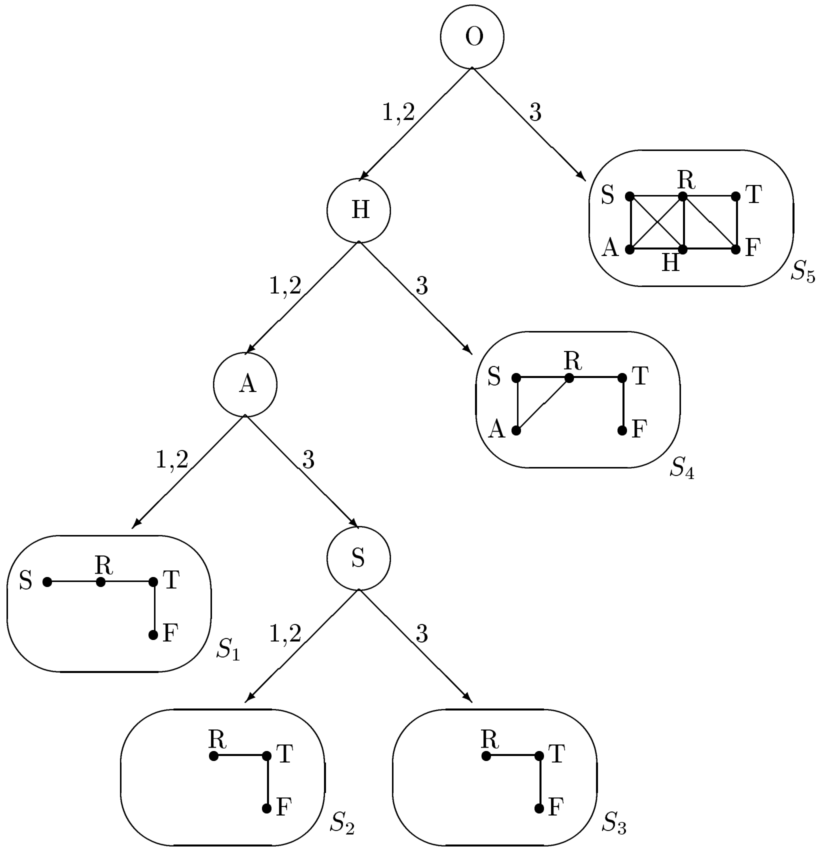


Figure 6: Hybrid of regression tree and CLLSs. The structures at the terminal nodes of the tree are labelled $S_1, \cdots, S_5$ from left to right. The number on each arrow represents the value of the random variable at the tail of the arrow.

prediction probabilities instead of the CLLSs, which were obtained by a separate job using SAS. The prediction probabilities for R are listed in Table 4, which shows that the strength of recommendation increases as we move from left to right of the terminal nodes (actually ovals) of the tree in Figure 6, which are labeled as $S_1, \cdots, S_5$. The goodness-of-fit levels of these 5 structures are given in Table 5. S_5' in the table is a proper submodel of S_5 but is not graphical, and so in the hybrid we use S_5 instead which is the smallest of the graphical models that contain S_5' as a submodel although S_5' looks better than S_5 with regard to goodness-of-fit.

According to the hybrid in Figure 6, variable O is selected first as the most informative for R among the rest 6 variables; when O takes on 1 or 2, variable H is then selected as the most informative for R among the rest 5 variables, A, H, F, S, and T, and when O takes on 3, no further variable-selection improves the prediction accuracy enough and the CLLS among the

Table 4: Probabilities of recommendation (P(R=1), P(R=2), P(R=3)) at the terminal nodes of the hybrid in Figure 6

S_1	S_2	S_3	S_4	S_5
(0.6, 0.3, 0.1)	(0.22, 0.5, 0.28)	(0.12, 0.37, 0.51)	(0.1, 0.33, 0.57)	(0.03, 0.16, 0.81)

Table 5: Goodness-of-fit levels by SAS of the five structures $S_1, \cdots, S_5$ in Figure 6. n stands for sample size and G^2 is the likelihood ratio chi-squared statistic.

Structure	n	G^2	d.f.	p-value
S_1	364	37.11	45	0.792
S_2	1314	19.37	12	0.080
S_3	799	10.46	12	0.575
S_4	2170	189.18	175	0.220
S_5	5378	273.86	239	0.060
S_5'		276.71	250	0.118*
h_{O3}^{**}	10025	1421.21	1015	0.000
h_{O4}	10025	814.34	914	0.992
h_{O8}	10025	643.09	810	1.000

*: For a submodel S_5' of S_5 represented by
$\{\{H, R, S\}, \{A, R, S\}, \{A, H, S\}, \{A, H, R\}, \{F, H, R\}, \{F, R, T\}\}$.
**: See Figure 8.

remaining variables is as in S_5; similar stories for H and A; as for variable S, no subsequent variable-selection is recommended and the corresponding CLLSs, S_2 and S_3, are the same.

It is worthwhile to note in the CLLSs except S_5 that F is not informative for R when T is known. It reflects the fact that TAs of the school grade homework and assist students in experiments or problem solving, and so F may be subordinate to T from the perspectives of students. Another thing to note is that R separates lecture-related variables from TA-related variables in S_4 and S_1 and H and R do the same thing in S_5. Whether O takes on 3 or not we can read in the log-linear structures of the hybrid that lecture-related variables are separated from their counter part by R and H or by R only. Having noticed this, a group of three psychometricians agreed on that

$$(T,\ F) \perp (A,\ S) | (H,\ O,\ R). \tag{17}$$

Item O refers to whether the whole course is well scheduled including a homework list and score weights over examinations and homeworks. The score weights can be a reflection of the types and uses of examinations and homeworks. This may possibly explain why O is the most informative variable for R among the 6 explanatory variables. Considering this, the

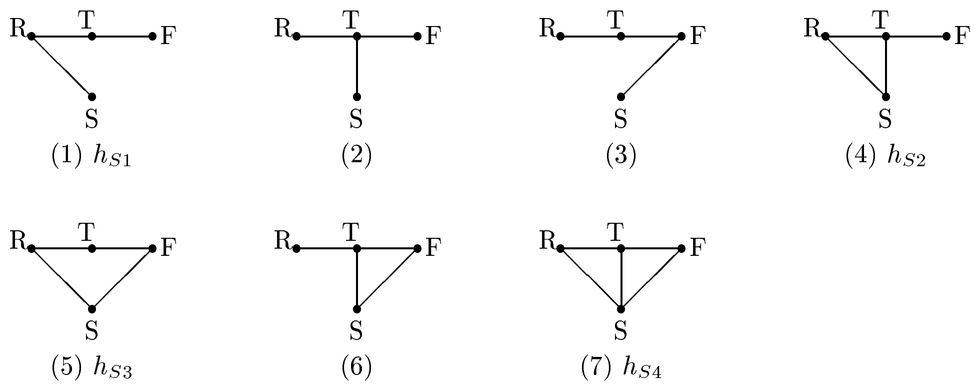


Figure 7: The hypermodels corresponding to $Hyb(S;\ S_2, S_3)$

psychometricians argued that once students have made up their minds about O, the information about what they think about F and T must be contained in H and R only. It is obvious that F has much to do with homework grading by TAs. The psychometricians made another point that in general TAs do their job according to the course schedule which pertains to O and that a poor performance of TAs may affect R negatively. Following this line of reasoning, they concluded that A and S might not be informative for T and F given that the values of H, R, and O are known.

Turning to the hypermodelling using the hybrid in Figure 6, the process begins with the bottom-most one-node hybrid, $Hyb(S;\ S_2, S_3)$. That S is selected for predicting about R implies that S is connected to at least one of R, T, and F, resulting into 7 possible connections as in Figure 7, where graphs 2, 3, and 6 are impossible by Corollary 3.6.

Next, we move up to $Hyb(A;\ h_{Si}, S_1)$, where $h_{Si}, i = 1, \cdots, 4$, are given in Figure 7. The corresponding hypermodels are displayed in Figure B.2 in Appendix B. In this figure, the four graphs, graphs 1, 3, 4, and 10, are impossible by Corollary 3.6. For instance, in graphs 3, 4, and 10, A is not informative for R provided that the value of T is known, but according to the hybrid in Figure 6, S is more informative for R than T is. If graph 1 were true, then A should not show up in the hybrid as an informative variable for R. Graphs 1 through 15 correspond to $Hyb(A;\ h_{S1}, S_1)$; graphs 16 and 17 correspond to $Hyb(A;\ h_{S2}, S_1)$; graphs 18 through 21 correspond to $Hyb(A;\ h_{S3}, S_1)$; and graph 22 corresponds to $Hyb(A;\ h_{S4}, S_1)$.

Note in Figure B.2 that the actual hypermodel corresponding to the hybrid in Figure 6 which satisfies condition (17) must be found in the collection that stems from one of the two one-node hybrids,

$$Hyb(H;\ h_{A1}, S_4) \text{ and } Hyb(H;\ h_{A2}, S_4). \tag{18}$$

This is because A is connected to T or F in graphs h_{A3} through h_{A18} and Lemma 2.1 implies that any edge appearing in a CLLS also appears in the corresponding hypermodel and so that the hypermodel stemming from one of h_{A3} through h_{A18} violates condition (17).

Hypermodelling with respect to the hybrids in (18) ends up with the hypermodels in Figure B.3. The first four graphs therein correspond to both of the hybrids in (18), and the others correspond to $Hyb(H;\ h_{A2}, S_4)$ only. As for $Hyb(H;\ h_{A1}, S_4)$, it is partially homogeneous; the same LLS for the set $\{F, R, T\}$ but not for the set $\{A, R, S\}$ between h_{A1} and S_4. Thus we apply, in hypermodelling, Theorem 2.2 to the set $\{A, S, R\}$ and apply Theorem 2.1 to the set $\{R, T, F\}$. As for $Hyb(H;\ h_{A2}, S_4)$, $h_{A2} = S_4$. So, by Theorem 2.1, $Hyb(H;\ h_{A2}, S_4)$ gives rise to a number of hypermodels, h_{H1} through h_{H31}.

While there are as many as 31 hybrids, $Hyb(O;\ h_{Hi}, S_5)$, $\ i = 1, \cdots, 31$, hypermodelling for these results into only 8 new hypermodels as displayed in Figure 8. None of h_{Hi}, $i = 1, \cdots, 31$, shares the same structure with S_5 as far as the subset of variables, F, H, R, and T are concerned. That is why O must be connected to all of them by Theorem 2.2. As for the subset of variables, A, H, R, and S, the same structure of them appears in h_{H1} through h_{H4} and S_5. So, by Theorem 2.1, O is connected to none or at least one of A and S. Note that O must be connected to H and R as mentioned above. The eight graphs in the figure satisfies condition (17), and the largest of them is h_{O8}.

Table 5 displays the goodness-of-fit levels of 3 structures, h_{O3}, h_{O4}, and h_{O8} in Figure 8. Out of the three graphical models, h_{O4} looks most appropriate in the context of parsimoniousness and goodness-of-fit. Actually it was found that the four-way interaction effect of the variables S, A, O, and H, is apparent in the LLM for the seven variables of Table 3. And so the six graphs, h_{Oi}, $i = 1, 2, 3, 5, 6, 7$, are inappropriate for the data. Much smaller submodels of h_{O4} were found appropriate but none of them were graphical. Hence, if we confine ourselves to graphical models only, we end up with h_{O4} as a most recommended model for the lecture evaluation data.

In addition to (17), we can see from h_{O4} that (i) $T \perp (A, H, S)|(F, O, R)$ and that (ii) S, A, O, R, and H are fully interactive. Statement (i) means that students' opinions on TA's performance are directly associated with F (feedback from homework), O(course organization), and R(course recommendation) only. It is interesting to see that H(homework contents) does

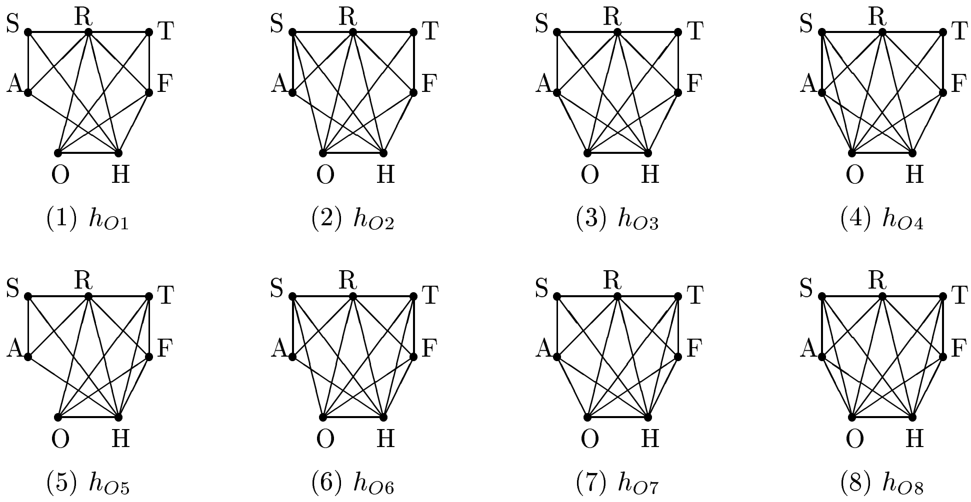


Figure 8: The hypermodels corresponding to $Hyb(O;\ h_{Hi}, S_5)$

not affect T directly. We have already noted that S, A, O, and H are fully interactive. All of them concern the lecturer. And statement (ii) implies that they may represent distinct features of a lecturer and be associated with R in different dimensions.

6 Discussion

Under the assumption that the hypermodel is graphical, we can search for it by focusing on the maximal sets in the union of the CLLSs as illustrated in the description of the hypermodelling process. When hybrid h consists of at least one one-node hybrid which is partially homogeneous, there can be multiple hypermodels corresponding to the hybrid h. To make the set-size of the hypermodels as small as possible, we applied the concept of prediction refinement and pieces of extraneous information (e.g., conditions (16) and (17)), if any, on a subset of the variables of a given model and illustrated how the extraneous information reduces the size of the set of the hypermodels. The refinement concept plays an important role for the set-size reduction when we deal with a hybrid of prediction tree and CLLSs. By applying the concept and the extraneous information, we may see a more reduction in the set-size than when applying only one of them. As long as the hybrid model is concerned, the contribution level for the set-size reduction may vary between the two according to the situations that hybrids are applied to. For instance, in hypermodelling with the hybrid in Figure 6, we could reduce the number of the possible hypermodels corresponding to $Hyb(S;\ S_2, S_3)$ down to 4/7 by applying the refinement concept, while the contribution was smaller at the next one-node hybrid $Hyb(A;\ h_{S_i}, S_1)$ compared with the extraneous information. The number of possible hypermodels may increase exponentially during the hypermodelling. In this respect, it is important to note that a contribution to the set-size reduction at an earlier stage of the hypermodelling process is in general more effective for reduction of the set-size of the hypermodels than a later contribution.

In the hybrid of section 5, all the variables involved are ternary. So, it is possible that the same predictor variable appear more than once in the tree part. To avoid this multiple appearance, it is desirable to split the data at the node of the predictor variable (X say) into as many subsets as the number of the levels of X. Such multiple appearances did not show up in the hybrids of this paper.

The hypermodelling method can be applied sequentially for such hybrids as in Figures 3 and 6. We can construct a tree-structured prediction system (or tree for short) using CART. But the structure of the tree is sensitive to the random variation in data (Breiman et al., 1984; Kim, 1992), and experts may not agree with some sequences of variables embodied in the tree. For example, medical doctors may have their preferences regarding the sequence of symptom examination. Thus they may wish to construct a tree up to a certain tree-size in order to minimize the effect of the noise in data, and then attach CLLSs at the end of the tree. Hybrids of this kind may serve the dual purposes of prediction and interpretation well. Prediction with a LLS can be easily handled by the method described in Lauritzen and Spiegelhalter (1988). The interpretation of the model structure is via a sequential application of the hypermodelling method to each one-node hybrid. Of course, if we have enough data, then we can fit a LLM to data. But when we select a LLM from a set of good looking, candidate LLMs for a data set, it may be desirable to check whether the structure of the selected one is in the set of the possible hypermodels which is obtained from the complex hybrid. If we can not have enough data for all the interested variables but can collect data for subsets of the variables conditional on the outcomes of some other variables, then the hypermodelling approach may be very useful in formulating a LLS for all the interested variables.

In the previous section, we dealt with 7 ternary variables. It may take about a month with an IBM pc with a 200MHz CPU to build a LLM for these variables by a stepwise procedure using SAS. For instance, it took 8 days and 2 hours for fitting h_{O8} and 7 days and 2 hours for fitting h_{O4}. On the other hand, the whole process of obtaining from data the regression tree and the CLLSs as in Figure 6 and then hypermodelling up to the graphs in Figure 8 took less than half a day. In this respect, we may safely recommend that hybrids such as in Figures 3 and 6 be used when searching for a relatively large LLS suitable to data. Another useful feature of such a hybrid is that it serves both the purposes of prediction and interpretation of the structural relationship among variables. As was the case in the example of the preceeding section, the CLLSs of a hybrid themselves may sometimes be a clue to pieces of information about the relationship among variables.

The hybrid such as in Figures 3 and 6 may also be constructed based on experts' opinions when data are sparse or not evenly dispersed across the paths of the tree-part. When data are sparse along a particular path, the corresponding CLLS may have to be obtained from an expert. This implies that some of the CLLSs of a hybrid may be based on real data and the others on experts' opinions. It is possible that more than half of the CLLSs of a hybrid are expert-based and the others data-based. When data are hard to collect and we need to consult experts, it is desirable that we restrict the problem domain as much as possible so that experts could give unambiguous opinions over a certain knowledge domain. For instance, in case that the structures S_2 and S_3 are not provided in Figure 6, experts may feel more comfortable in giving opinions on the corresponding set of variables than when they are consulted for one of S_1, S_4, and S_5. There may be cases that experts are consulted for some paths of the tree part of a hybrid in addition to some CLLSs of a hybrid. We can also do hypermodelling, by applying the theorems and corollaries of sections 2 and 3 and some pieces of information about the interrrelationship among the variables, with such hybrids as are half expert-based and half data-based in search of the hypermodel that reflects both the data and the experts' opinions. The hypermodel can then be updated or modified as data accumulate.

Finally, whether it is data-based, expert-based, or inbetween, a hybrid h gives birth to hypermodels, and the number of the hypermodels is determined subject to the states of homogeneity of the element one-node hybrids of h and pieces of information about the interrelationship among variables. If we have good pieces of such information, the number of the hypermodels will be reduced as we have seen in sections 4 and 5. When building a model based on data, we impose some stochastic or mathematical restrictions on the model or pick some mathematical formula and see how it fits data. The pieces of information on local relationships play the same role as the restrictions on modelling. The quality of such information is up to the experts being consulted.

Appendix A: Relationship between LLS and CLLS

Let $X_1, X_2, \cdots, X_n$ be categorical variables. We consider CLLSs of $X_2, X_3, \cdots, X_n$ conditional on the outcomes of X_1, and assume that X_1 takes on I values $1, \cdots, I$ and that it is effective for the LLM of $X_1, X_2, \cdots, X_n$. Thus we may restrict ourselves to the LLMs that include the u_1 term, i.e.,

$$u_1 \neq 0. \tag{19}$$

Consider a LLS given by

$$\{\theta_1, \theta_2, \cdots, \theta_k\}, \tag{20}$$

where $\theta_1, \theta_2, \cdots, \theta_k$ are distinct subsets of $\{1, 2, \cdots, n\}$. Under the assumption in (19), there must exist at least one set in expression (20) which contains "1". Suppose that there are r of these, $\theta_1, \theta_2, \cdots, \theta_r$. In Example 2.1 we saw that a component set (e.g., $\{3, 4, 5\}$) in the LLS which does not contain "1" appears in every conditional of the LLS but not for the component sets (e.g., $\{1, 2, 3\}$, $\{1, 3, 4\}$) that contain "1".

When a LLM is conditioned by X_1, the "1" disappears into the $w^{(x_1)}$-terms in the CLLM and the terms affect the CLLS under the SHP. If the index set of a w-term (e.g., $w^{(x_1)}_{\{3,4\}}$) is a subset of the index set of some u-term, the w-term does not affect the CLLS; otherwise (e.g., $w^{(x_1)}_{\{2,3\}}$), the w-term affect the CLLS. We call a set such as $\{1, 3, 4\}$ in Example 2.1 as a *disappearing* set, a set such as $\{1, 2, 3\}$ as a *remaining* set, and a set such as $\{3, 4, 5\}$ as a *settled* set. If a set of a LLS is free of "1", it is *settled*; otherwise, it is either *remaining* or *disappearing*.

We may suppose, without loss of generality, that there are d disappearing sets, $\theta_1, \cdots, \theta_d$, $0 \leq d < r$, in expression (20). Thus the sets, $\theta_{d+1}, \cdots, \theta_r$, are the remaining sets in (20). After appropriate rearrangements, we have $d = 1,\ r = 2,\ k = 3$ in Example 2.1, and $d = 0,\ r = k = 2$ in Example 2.2. When all the sets in a LLS contain "1", $r = k$ and $d = 0$.

Many of the $CLLS$s associated with a LLS are the results of zero-w phenomena involving the remaining sets in the LLS. When the LLS is given by (20), the generic form of the $CLLS$s for $X_2, \cdots, X_n$ given $X_1 = x_1$ is given, under the SHP, by

$$\langle \varphi_{d+1}, \varphi_{d+2}, \cdots, \varphi_r, \theta_{r+1}, \cdots, \theta_k \rangle, \ \text{ for } \ \varphi_j \in \mathcal{C}_j, \ \ j = d+1, \cdots, r, \tag{21}$$

where $\mathcal{C}_j$ is a collection of subsets of the remaining set, $\theta_j \setminus \{1\}$ for $d + 1 \leq j \leq r$. The largest of the $CLLS$s is given by $\{\theta_{d+1} \setminus \{1\}, \cdots, \theta_r \setminus \{1\}, \theta_{r+1}, \cdots, \theta_k\}$, and the smallest is given by

$\{\theta_{r+1}, \cdots, \theta_k\}$. The settled sets show up in each of the $CLLS$s.

Appendix B: Some hypermodels referred to in sections 4 and 5

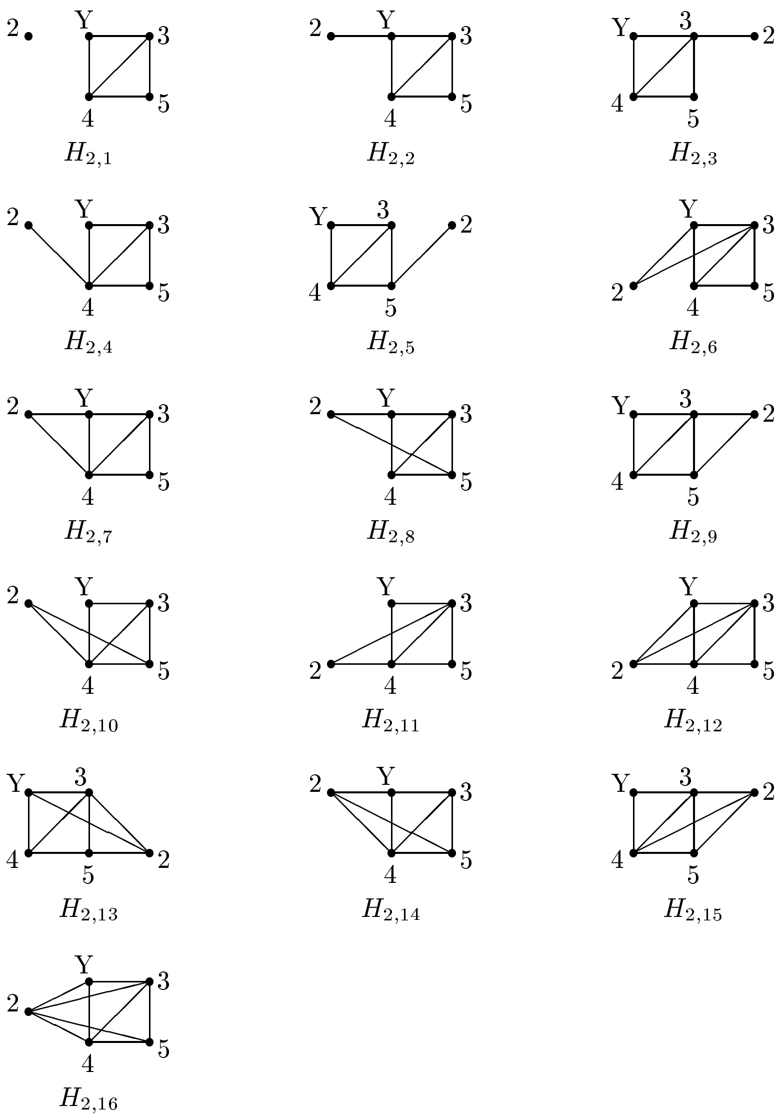


Figure B.1: *Hypermodels from the hybrid* $Hyb(X_2;\ H_{3,2}, S_3)$

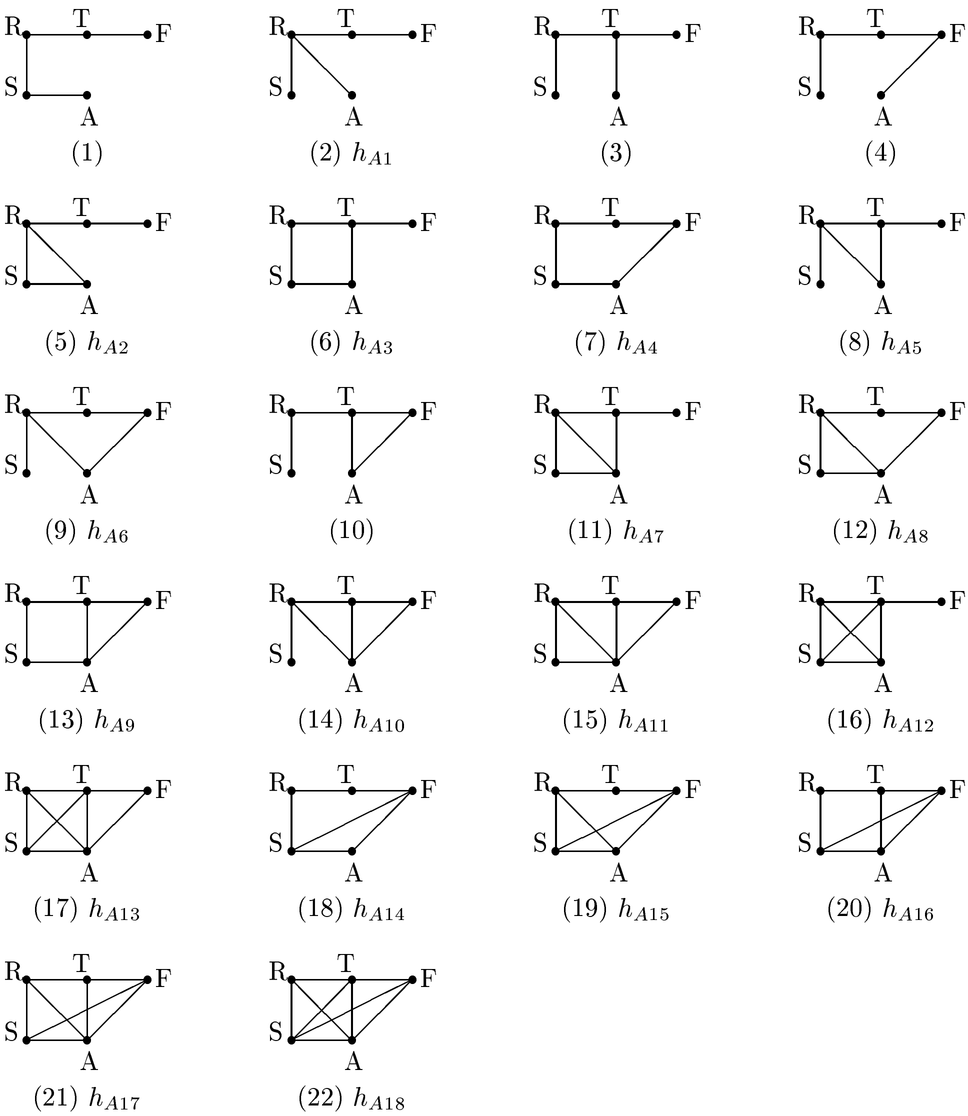
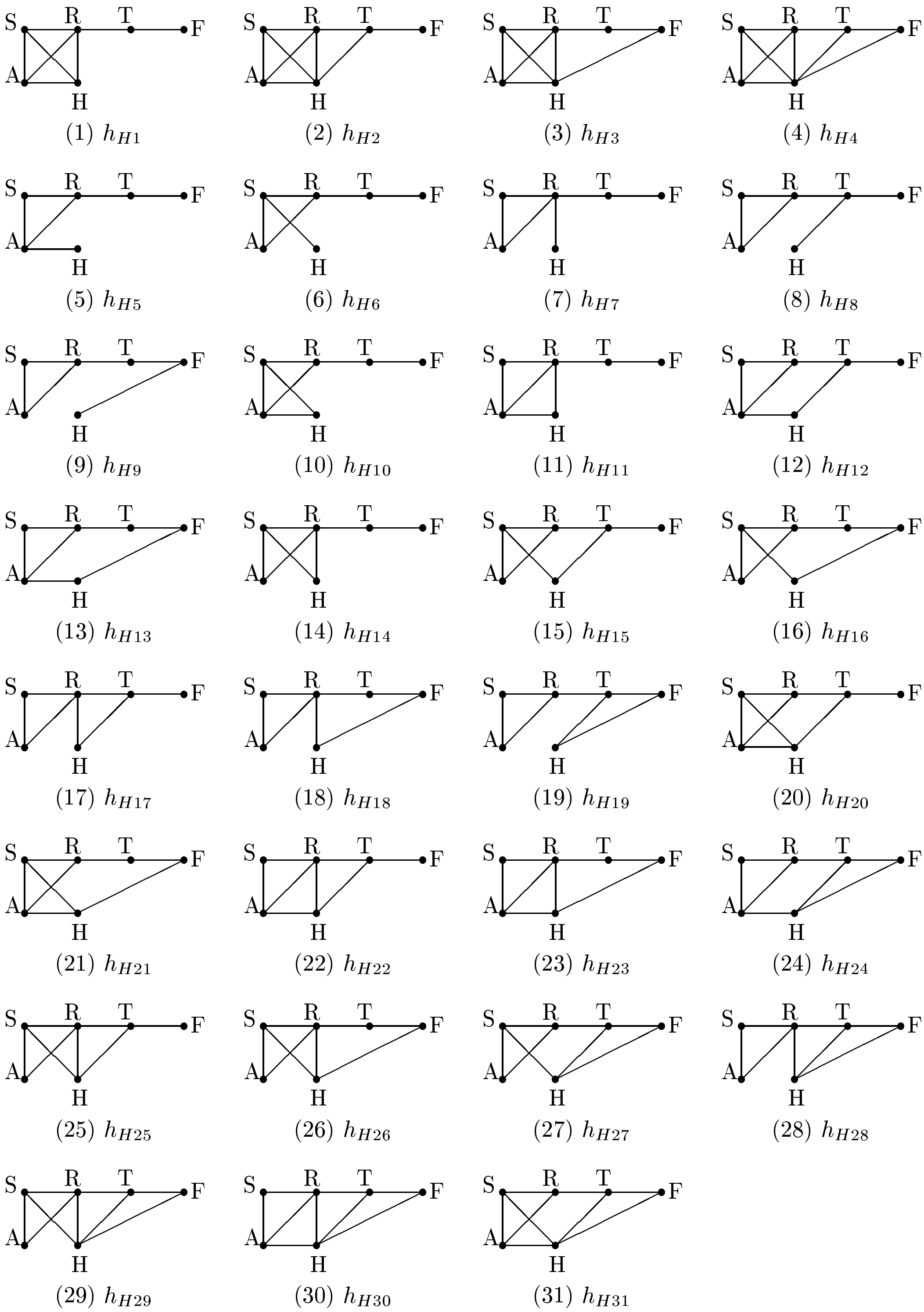


Figure B.2: The hypermodels corresponding to $Hyb(A;\ h_{Si}, S_1)$



(1) h_{H1} (2) h_{H2} (3) h_{H3} (4) h_{H4} (5) h_{H5} (6) h_{H6} (7) h_{H7} (8) h_{H8} (9) h_{H9} (10) h_{H10} (11) h_{H11} (12) h_{H12} (13) h_{H13} (14) h_{H14} (15) h_{H15} (16) h_{H16} (17) h_{H17} (18) h_{H18} (19) h_{H19} (20) h_{H20} (21) h_{H21} (22) h_{H22} (23) h_{H23} (24) h_{H24} (25) h_{H25} (26) h_{H26} (27) h_{H27} (28) h_{H28} (29) h_{H29} (30) h_{H30} (31) h_{H31}

Figure B.3: The hypermodels corresponding to $Hyb(H;\ h_{Ai}, S_4)$

Appendix C: The frequency table of the data used in section 5

The table consists of 20 columns. The 7-digit columns list the configurations of the 7 random variables, S, O, A, H, F, T, R, in the same order. Each random variable has levels, 1, 2, and 3. The 7-digit column is followed by an integer column which lists the cell frequencies at the cells indicated in the 7-digit column. For example, the cell frequency at (1,1,1,1,1,1,1) is 29.

1111111	29	1111112	6	1111113	2	1111121	10	1111122	2	1111123	1	1111131	14	1111132	3	1111133	1	1111211	7
1111212	1	1111213	0	1111221	13	1111222	2	1111223	1	1111231	5	1111232	2	1111233	1	1111311	2	1111312	0
1111313	0	1111321	2	1111322	0	1111323	1	1111331	4	1111332	0	1111333	1	1112111	6	1112112	2	1112113	1
1112121	5	1112122	8	1112123	0	1112131	5	1112132	1	1112133	0	1112211	3	1112212	2	1112213	0	1112221	7
1112222	3	1112223	1	1112231	5	1112232	4	1112233	0	1112311	2	1112312	0	1112313	0	1112321	6	1112322	0
1112323	1	1112331	10	1112332	1	1112333	0	1113111	6	1113112	0	1113113	4	1113121	1	1113122	3	1113123	2
1113131	2	1113132	3	1113133	0	1113211	3	1113212	1	1113213	0	1113221	2	1113222	3	1113223	2	1113231	1
1113232	2	1113233	2	1113311	0	1113312	0	1113313	1	1113321	4	1113322	2	1113323	2	1113331	8	1113332	1
1113333	4	1121111	5	1121112	4	1121113	1	1121121	1	1121122	6	1121123	2	1121131	4	1121132	1	1121133	0
1121211	1	1121212	0	1121213	0	1121221	4	1121222	0	1121223	0	1121231	3	1121232	0	1121233	0	1121311	1
1121312	0	1121313	0	1121321	1	1121322	0	1121323	1	1121331	1	1121332	0	1121333	0	1122111	5	1122112	2
1122113	0	1122121	2	1122122	6	1122123	1	1122131	2	1122132	1	1122133	1	1122211	2	1122212	3	1122213	0
1122221	3	1122222	7	1122223	1	1122231	3	1122232	5	1122233	1	1122311	2	1122312	1	1122313	0	1122321	5
1122322	2	1122323	0	1122331	3	1122332	3	1122333	1	1123111	2	1123112	5	1123113	1	1123121	2	1123122	2
1123123	3	1123131	1	1123132	1	1123133	0	1123211	1	1123212	1	1123213	2	1123221	2	1123222	3	1123223	1
1123231	0	1123232	2	1123233	2	1123311	1	1123312	0	1123313	0	1123321	2	1123322	3	1123323	2	1123331	3
1123332	4	1123333	1	1131111	0	1131112	0	1131113	1	1131121	2	1131122	1	1131123	2	1131131	1	1131132	0
1131133	0	1131211	0	1131212	0	1131213	0	1131221	0	1131222	0	1131223	0	1131231	0	1131232	0	1131233	1
1131311	0	1131312	0	1131313	1	1131321	0	1131322	0	1131323	0	1131331	0	1131332	0	1131333	0	1132111	5
1132112	0	1132113	0	1132121	0	1132122	1	1132123	0	1132131	0	1132132	0	1132133	0	1132211	0	1132212	1
1132213	0	1132221	1	1132222	1	1132223	0	1132231	1	1132232	2	1132233	1	1132311	0	1132312	0	1132313	1
1132321	0	1132322	1	1132323	0	1132331	0	1132332	0	1132333	0	1133111	0	1133112	1	1133113	1	1133121	1
1133122	0	1133123	2	1133131	1	1133132	1	1133133	1	1133211	1	1133212	0	1133213	0	1133221	0	1133222	0
1133223	0	1133231	0	1133232	1	1133233	0	1133311	0	1133312	0	1133313	2	1133321	0	1133322	0	1133323	0
1133331	0	1133332	3	1133333	3	1211111	1	1211112	1	1211113	0	1211121	0	1211122	0	1211123	1	1211131	0
1211132	1	1211133	0	1211211	0	1211212	0	1211213	0	1211221	1	1211222	1	1211223	0	1211231	2	1211232	0
1211233	0	1211311	0	1211312	0	1211313	0	1211321	0	1211322	0	1211323	0	1211331	0	1211332	0	1211333	0
1212111	1	1212112	1	1212113	0	1212121	0	1212122	2	1212123	0	1212131	3	1212132	1	1212133	0	1212211	2
1212212	2	1212213	0	1212221	0	1212222	4	1212223	2	1212231	0	1212232	0	1212233	0	1212311	0	1212312	1
1212313	0	1212321	2	1212322	1	1212323	1	1212331	0	1212332	1	1212333	0	1213111	0	1213112	0	1213113	0
1213121	1	1213122	3	1213123	1	1213131	0	1213132	2	1213133	1	1213211	0	1213212	2	1213213	2	1213221	4
1213222	0	1213223	2	1213231	0	1213232	0	1213233	1	1213311	1	1213312	0	1213313	1	1213321	0	1213322	1
1213323	1	1213331	0	1213332	3	1213333	2	1221111	1	1221112	2	1221113	1	1221121	1	1221122	1	1221123	0
1221131	0	1221132	1	1221133	1	1221211	0	1221212	0	1221213	0	1221221	0	1221222	0	1221223	0	1221231	1
1221232	3	1221233	0	1221311	0	1221312	1	1221313	0	1221321	1	1221322	0	1221323	0	1221331	0	1221332	1
1221333	0	1222111	2	1222112	1	1222113	0	1222121	2	1222122	5	1222123	2	1222131	2	1222132	1	1222133	1
1222211	5	1222212	1	1222213	0	1222221	6	1222222	8	1222223	5	1222231	3	1222232	4	1222233	3	1222311	1
1222312	0	1222313	0	1222321	1	1222322	2	1222323	2	1222331	0	1222332	3	1222333	0	1223111	4	1223112	3
1223113	0	1223121	3	1223122	2	1223123	3	1223131	0	1223132	2	1223133	4	1223211	0	1223212	1	1223213	0
1223221	2	1223222	6	1223223	3	1223231	1	1223232	2	1223233	0	1223311	0	1223312	0	1223313	1	1223321	1
1223322	4	1223323	5	1223331	0	1223332	7	1223333	3	1231111	0	1231112	0	1231113	0	1231121	0	1231122	1
1231123	0	1231131	0	1231132	1	1231133	0	1231211	1	1231212	0	1231213	1	1231221	0	1231222	0	1231223	2
1231231	0	1231232	0	1231233	0	1231311	0	1231312	0	1231313	0	1231321	0	1231322	0	1231323	0	1231331	0
1231332	0	1231333	0	1232111	0	1232112	0	1232113	1	1232121	0	1232122	0	1232123	2	1232131	1	1232132	1
1232133	0	1232211	0	1232212	0	1232213	0	1232221	1	1232222	3	1232223	0	1232231	0	1232232	1	1232233	0
1232311	0	1232312	0	1232313	0	1232321	2	1232322	1	1232323	0	1232331	1	1232332	1	1232333	0	1233111	0
1233112	0	1233113	2	1233121	2	1233122	2	1233123	5	1233131	0	1233132	0	1233133	0	1233211	0	1233212	2
1233213	0	1233221	1	1233222	1	1233223	3	1233231	0	1233232	0	1233233	0	1233311	0	1233312	1	1233313	0
1233321	0	1233322	0	1233323	3	1233331	0	1233332	2	1233333	3	1311111	1	1311112	0	1311113	0	1311121	0
1311122	0	1311123	0	1311131	0	1311132	0	1311133	0	1311211	0	1311212	0	1311213	0	1311221	0	1311222	0
1311223	0	1311231	0	1311232	0	1311233	0	1311311	0	1311312	0	1311313	0	1311321	0	1311322	0	1311323	0
1311331	0	1311332	0	1311333	0	1312111	1	1312112	0	1312113	0	1312121	0	1312122	0	1312123	0	1312131	0
1312132	0	1312133	0	1312211	0	1312212	0	1312213	0	1312221	0	1312222	1	1312223	1	1312231	0	1312232	0
1312233	0	1312311	0	1312312	0	1312313	0	1312321	0	1312322	1	1312323	0	1312331	0	1312332	0	1312333	0
1313111	0	1313112	0	1313113	0	1313121	1	1313122	0	1313123	0	1313131	0	1313132	0	1313133	0	1313211	0
1313212	0	1313213	0	1313221	0	1313222	0	1313223	1	1313231	0	1313232	0	1313233	1	1313311	0	1313312	0
1313313	2	1313321	0	1313322	0	1313323	0	1313331	0	1313332	0	1313333	0	1321111	0	1321112	0	1321113	0
1321121	1	1321122	0	1321123	0	1321131	0	1321132	1	1321133	0	1321211	0	1321212	0	1321213	0	1321221	0
1321222	0	1321223	1	1321231	0	1321232	0	1321233	0	1321311	1	1321312	0	1321313	0	1321321	0	1321322	0
1321323	0	1321331	0	1321332	0	1321333	0	1322111	0	1322112	0	1322113	0	1322121	2	1322122	1	1322123	0
1322131	1	1322132	0	1322133	0	1322211	0	1322212	0	1322213	0	1322221	0	1322222	1	1322223	0	1322231	0
1322232	2	1322233	2	1322311	1	1322312	0	1322313	0	1322321	0	1322322	2	1322323	0	1322331	0	1322332	1
1322333	1	1323111	0	1323112	1	1323113	1	1323121	1	1323122	1	1323123	1	1323131	0	1323132	0	1323133	2
1323211	0	1323212	0	1323213	0	1323221	0	1323222	1	1323223	1	1323231	0	1323232	1	1323233	0	1323311	0
1323312	0	1323313	0	1323321	1	1323322	1	1323323	3	1323331	0	1323332	1	1323333	2	1331111	0	1331112	0
1331113	1	1331121	1	1331122	0	1331123	0	1331131	0	1331132	0	1331133	0	1331211	0	1331212	0	1331213	1
1331221	0	1331222	1	1331223	0	1331231	0	1331232	0	1331233	0	1331311	0	1331312	0	1331313	0	1331321	0
1331322	0	1331323	0	1331331	0	1331332	0	1331333	1	1332111	0	1332112	0	1332113	0	1332121	0	1332122	0
1332123	0	1332131	0	1332132	0	1332133	0	1332211	0	1332212	0	1332213	0	1332221	1	1332222	0	1332223	1
1332231	2	1332232	0	1332233	0	1332311	0	1332312	0	1332313	0	1332321	0	1332322	0	1332323	1	1332331	1
1332332	1	1332333	0	1333111	0	1333112	1	1333113	2	1333121	0	1333122	0	1333123	1	1333131	0	1333132	0

1333133	0	1333211	0	1333212	0	1333213	1	1333221	0	1333222	1	1333223	0	1333231	0	1333232	1	1333233	1
1333311	0	1333312	0	1333313	1	1333321	0	1333322	0	1333323	1	1333331	1	1333332	1	1333333	9	2111111	3
2111112	3	2111113	0	2111121	1	2111122	0	2111123	0	2111131	2	2111132	1	2111133	0	2111211	1	2111212	0
2111213	1	2111221	1	2111222	0	2111223	0	2111231	2	2111232	0	2111233	0	2111311	2	2111312	0	2111313	0
2111321	0	2111322	0	2111323	0	2111331	1	2111332	1	2111333	0	2112111	6	2112112	1	2112113	1	2112121	1
2112122	4	2112123	3	2112131	2	2112132	0	2112133	1	2112211	6	2112212	0	2112213	1	2112221	7	2112222	5
2112223	0	2112231	6	2112232	0	2112233	0	2112311	1	2112312	1	2112313	0	2112321	2	2112322	0	2112323	0
2112331	3	2112332	2	2112333	0	2113111	2	2113112	1	2113113	1	2113121	2	2113122	0	2113123	0	2113131	1
2113132	0	2113133	0	2113211	2	2113212	1	2113213	1	2113221	2	2113222	3	2113223	0	2113231	2	2113232	1
2113233	4	2113311	0	2113312	0	2113313	0	2113321	0	2113322	1	2113323	4	2113331	1	2113332	3	2113333	3
2121111	3	2121112	2	2121113	1	2121121	4	2121122	6	2121123	0	2121131	3	2121132	0	2121133	0	2121211	1
2121212	1	2121213	0	2121221	2	2121222	4	2121223	1	2121231	0	2121232	1	2121233	0	2121311	0	2121312	0
2121313	0	2121321	2	2121322	1	2121323	2	2121331	1	2121332	0	2121333	1	2122111	9	2122112	5	2122113	3
2122121	5	2122122	8	2122123	1	2122131	1	2122132	5	2122133	0	2122211	2	2122212	0	2122213	5	2122221	9
2122222	15	2122223	4	2122231	3	2122232	4	2122233	0	2122311	0	2122312	1	2122313	1	2122321	3	2122322	3
2122323	3	2122331	7	2122332	1	2122333	3	2123111	0	2123112	9	2123113	2	2123121	2	2123122	3	2123123	1
2123131	1	2123132	5	2123133	4	2123211	3	2123212	1	2123213	0	2123221	2	2123222	1	2123223	6	2123231	5
2123232	2	2123233	3	2123311	2	2123312	2	2123313	2	2123321	2	2123322	5	2123323	6	2123331	4	2123332	5
2123333	4	2131111	3	2131112	4	2131113	0	2131121	2	2131122	1	2131123	0	2131131	2	2131132	0	2131133	1
2131211	0	2131212	0	2131213	0	2131221	1	2131222	1	2131223	0	2131231	0	2131232	1	2131233	0	2131311	0
2131312	0	2131313	0	2131321	0	2131322	0	2131323	1	2131331	1	2131332	0	2131333	0	2132111	1	2132112	3
2132113	1	2132121	1	2132122	7	2132123	1	2132131	1	2132132	0	2132133	1	2132211	1	2132212	2	2132213	0
2132221	0	2132222	8	2132223	3	2132231	0	2132232	0	2132233	0	2132311	1	2132312	0	2132313	0	2132321	2
2132322	1	2132323	1	2132331	1	2132332	0	2132333	1	2133111	2	2133112	1	2133113	1	2133121	0	2133122	3
2133123	1	2133131	1	2133132	2	2133133	1	2133211	0	2133212	1	2133213	0	2133221	2	2133222	2	2133223	3
2133231	2	2133232	1	2133233	3	2133311	2	2133312	1	2133313	1	2133321	0	2133322	2	2133323	4	2133331	6
2133332	6	2133333	9	2211111	1	2211112	0	2211113	0	2211121	1	2211122	1	2211123	0	2211131	2	2211132	0
2211133	0	2211211	2	2211212	0	2211213	0	2211221	0	2211222	0	2211223	0	2211231	1	2211232	0	2211233	0
2211311	0	2211312	0	2211313	0	2211321	1	2211322	0	2211323	0	2211331	0	2211332	0	2211333	0	2212111	0
2212112	1	2212113	1	2212121	0	2212122	2	2212123	0	2212131	1	2212132	1	2212133	0	2212211	0	2212212	2
2212213	1	2212221	2	2212222	4	2212223	0	2212231	1	2212232	3	2212233	0	2212311	1	2212312	1	2212313	1
2212321	1	2212322	1	2212323	3	2212331	0	2212332	0	2212333	1	2213111	1	2213112	5	2213113	1	2213121	0
2213122	0	2213123	0	2213131	0	2213132	0	2213133	2	2213211	0	2213212	0	2213213	2	2213221	1	2213222	0
2213223	2	2213231	0	2213232	0	2213233	1	2213311	0	2213312	0	2213313	1	2213321	1	2213322	0	2213323	1
2213331	1	2213332	2	2213333	2	2221111	8	2221112	11	2221113	5	2221121	2	2221122	9	2221123	7	2221131	5
2221132	6	2221133	5	2221211	4	2221212	4	2221213	1	2221221	3	2221222	6	2221223	4	2221231	0	2221232	1
2221233	1	2221311	0	2221312	2	2221313	0	2221321	1	2221322	1	2221323	2	2221331	2	2221332	1	2221333	0
2222111	10	2222112	16	2222113	10	2222121	9	2222122	43	2222123	14	2222131	2	2222132	11	2222133	8	2222211	4
2222212	23	2222213	9	2222221	24	2222222	143	2222223	65	2222231	10	2222232	31	2222233	32	2222311	5	2222312	5
2222313	1	2222321	6	2222322	29	2222323	14	2222331	8	2222332	21	2222333	15	2223111	2	2223112	17	2223113	10
2223121	3	2223122	17	2223123	21	2223131	2	2223132	6	2223133	9	2223211	3	2223212	5	2223213	9	2223221	2
2223222	33	2223223	29	2223231	5	2223232	12	2223233	21	2223311	1	2223312	8	2223313	8	2223321	4	2223322	36
2223323	40	2223331	5	2223332	24	2223333	47	2231111	2	2231112	5	2231113	4	2231121	2	2231122	8	2231123	5
2231131	0	2231132	2	2231133	0	2231211	2	2231212	1	2231213	1	2231221	2	2231222	2	2231223	2	2231231	0
2231232	0	2231233	0	2231311	1	2231312	0	2231313	0	2231321	1	2231322	0	2231323	1	2231331	1	2231332	1
2231333	2	2232111	2	2232112	10	2232113	4	2232121	7	2232122	10	2232123	1	2232131	0	2232132	4	2232133	4
2232211	3	2232212	9	2232213	4	2232221	8	2232222	26	2232223	25	2232231	3	2232232	12	2232233	14	2232311	0
2232312	5	2232313	1	2232321	3	2232322	13	2232323	11	2232331	3	2232332	12	2232333	6	2233111	1	2233112	8
2233113	13	2233121	1	2233122	8	2233123	9	2233131	0	2233132	5	2233133	11	2233211	2	2233212	2	2233213	8
2233221	6	2233222	22	2233223	20	2233231	1	2233232	7	2233233	17	2233311	1	2233312	2	2233313	6	2233321	1
2233322	15	2233323	27	2233331	1	2233332	25	2233333	49	2311111	0	2311112	0	2311113	0	2311121	1	2311122	1
2311123	0	2311131	0	2311132	0	2311133	0	2311211	0	2311212	1	2311213	1	2311221	0	2311222	0	2311223	1
2311231	0	2311232	0	2311233	0	2311311	0	2311312	0	2311313	0	2311321	0	2311322	0	2311323	0	2311331	0
2311332	0	2311333	0	2312111	0	2312112	0	2312113	0	2312121	0	2312122	0	2312123	0	2312131	2	2312132	0
2312133	0	2312211	0	2312212	0	2312213	0	2312221	1	2312222	1	2312223	0	2312231	0	2312232	0	2312233	0
2312311	0	2312312	0	2312313	1	2312321	0	2312322	0	2312323	1	2312331	0	2312332	1	2312333	1	2313111	0
2313112	0	2313113	1	2313121	0	2313122	0	2313123	1	2313131	1	2313132	0	2313133	0	2313211	1	2313212	1
2313213	1	2313221	0	2313222	0	2313223	1	2313231	1	2313232	0	2313233	2	2313311	0	2313312	0	2313313	0
2313321	1	2313322	0	2313323	0	2313331	0	2313332	1	2313333	2	2321111	2	2321112	2	2321113	1	2321121	2
2321122	0	2321123	0	2321131	0	2321132	0	2321133	2	2321211	0	2321212	0	2321213	0	2321221	0	2321222	0
2321223	0	2321231	0	2321232	1	2321233	1	2321311	0	2321312	1	2321313	0	2321321	0	2321322	0	2321323	0
2321331	1	2321332	0	2321333	1	2322111	1	2322112	3	2322113	3	2322121	1	2322122	3	2322123	6	2322131	1
2322132	0	2322133	2	2322211	0	2322212	1	2322213	2	2322221	3	2322222	9	2322223	3	2322231	1	2322232	1
2322233	1	2322311	1	2322312	0	2322313	1	2322321	0	2322322	4	2322323	4	2322331	0	2322332	4	2322333	3
2323111	1	2323112	0	2323113	6	2323121	2	2323122	2	2323123	7	2323131	0	2323132	1	2323133	3	2323211	0
2323212	3	2323213	3	2323221	1	2323222	10	2323223	13	2323231	0	2323232	3	2323233	8	2323311	0	2323312	0
2323313	1	2323321	0	2323322	6	2323323	10	2323331	3	2323332	8	2323333	18	2331111	0	2331112	4	2331113	0
2331121	0	2331122	2	2331123	2	2331131	2	2331132	2	2331133	2	2331211	0	2331212	3	2331213	0	2331221	0
2331222	1	2331223	7	2331231	1	2331232	1	2331233	1	2331311	0	2331312	0	2331313	0	2331321	0	2331322	1
2331323	1	2331331	0	2331332	1	2331333	2	2332111	0	2332112	7	2332113	6	2332121	0	2332122	5	2332123	8
2332131	1	2332132	2	2332133	1	2332211	0	2332212	5	2332213	4	2332221	4	2332222	19	2332223	32	2332231	0
2332232	10	2332233	4	2332311	0	2332312	0	2332313	2	2332321	2	2332322	2	2332323	9	2332331	1	2332332	8
2332333	13	2333111	2	2333112	10	2333113	18	2333121	0	2333122	2	2333123	22	2333131	1	2333132	3	2333133	11
2333211	2	2333212	2	2333213	8	2333221	4	2333222	11	2333223	27	2333231	1	2333232	4	2333233	17	2333311	1
2333312	1	2333313	9	2333321	3	2333322	12	2333323	38	2333331	4	2333332	12	2333333	76	3111111	2	3111112	2
3111113	0	3111121	1	3111122	1	3111123	0	3111131	0	3111132	0	3111133	0	3111211	3	3111212	0	3111213	0
3111221	0	3111222	0	3111223	0	3111231	0	3111232	1	3111233	0	3111311	0	3111312	0	3111313	0	3111321	0
3111322	0	3111323	0	3111331	0	3111332	0	3111333	0	3112111	0	3112112	2	3112113	0	3112121	0	3112122	1
3112123	0	3112131	0	3112132	0	3112133	0	3112211	0	3112212	1	3112213	0	3112221	3	3112222	0	3112223	0
3112231	0	3112232	2	3112233	0	3112311	0	3112312	0	3112313	2	3112321	0	3112322	0	3112323	0	3112331	2
3112332	0	3112333	0	3113111	0	3113112	0	3113113	1	3113121	1	3113122	0	3113123	1	3113131	1	3113132	0
3113133	0	3113211	0	3113212	0	3113213	0	3113221	0	3113222	0	3113223	1	3113231	0	3113232	0	3113233	0
3113311	0	3113312	0	3113313	1	3113321	0	3113322	1	3113323	0	3113331	1	3113332	4	3113333	0	3121111	0
3121112	2	3121113	1	3121121	2	3121122	0	3121123	1	3121131	0	3121132	0	3121133	0	3121211	2	3121212	0
3121213	0	3121221	0	3121222	3	3121223	0	3121231	1	3121232	0	3121233	0	3121311	0	3121312	0	3121313	0
3121321	1	3121322	0	3121323	0	3121331	1	3121332	0	3121333	0	3122111	0	3122112	1	3122113	0	3122121	2
3122122	2	3122123	3	3122131	0	3122132	0	3122133	1	3122211	0	3122212	1	3122213	0	3122221	1	3122222	7

3122223	0	3122231	0	3122232	3	3122233	5	3122311	0	3122312	2	3122313	0	3122321	0	3122322	3	3122323	2
3122331	0	3122332	2	3122333	2	3123111	2	3123112	1	3123113	1	3123121	0	3123122	0	3123123	1	3123131	0
3123132	1	3123133	1	3123211	0	3123212	0	3123213	1	3123221	0	3123222	4	3123223	2	3123231	2	3123232	1
3123233	1	3123311	0	3123312	0	3123313	2	3123321	1	3123322	5	3123323	3	3123331	3	3123332	3	3123333	4
3131111	4	3131112	0	3131113	0	3131121	3	3131122	1	3131123	0	3131131	0	3131132	1	3131133	2	3131211	1
3131212	0	3131213	0	3131221	0	3131222	1	3131223	0	3131231	0	3131232	2	3131233	2	3131311	0	3131312	0
3131313	0	3131321	2	3131322	0	3131323	1	3131331	0	3131332	1	3131333	0	3132111	2	3132112	1	3132113	1
3132121	2	3132122	6	3132123	1	3132131	0	3132132	1	3132133	1	3132211	0	3132212	0	3132213	0	3132221	2
3132222	5	3132223	0	3132231	0	3132232	3	3132233	2	3132311	1	3132312	0	3132313	1	3132321	1	3132322	2
3132323	0	3132331	0	3132332	2	3132333	3	3133111	0	3133112	4	3133113	4	3133121	1	3133122	4	3133123	8
3133131	2	3133132	2	3133133	4	3133211	1	3133212	1	3133213	3	3133221	0	3133222	4	3133223	5	3133231	2
3133232	1	3133233	5	3133311	1	3133312	0	3133313	1	3133321	3	3133322	4	3133323	5	3133331	5	3133332	7
3133333	9	3211111	2	3211112	1	3211113	0	3211121	0	3211122	0	3211123	0	3211131	2	3211132	0	3211133	0
3211211	0	3211212	0	3211213	0	3211221	1	3211222	0	3211223	0	3211231	0	3211232	0	3211233	0	3211311	1
3211312	0	3211313	0	3211321	0	3211322	0	3211323	1	3211331	0	3211332	0	3211333	0	3212111	0	3212112	1
3212113	0	3212121	0	3212122	0	3212123	0	3212131	1	3212132	0	3212133	2	3212211	0	3212212	0	3212213	0
3212221	1	3212222	0	3212223	1	3212231	0	3212232	0	3212233	0	3212311	0	3212312	0	3212313	0	3212321	0
3212322	2	3212323	1	3212331	0	3212332	0	3212333	0	3213111	1	3213112	0	3213113	0	3213121	0	3213122	0
3213123	0	3213131	0	3213132	0	3213133	0	3213211	1	3213212	1	3213213	0	3213221	0	3213222	1	3213223	0
3213231	0	3213232	0	3213233	1	3213311	0	3213312	1	3213313	0	3213321	0	3213322	0	3213323	0	3213331	1
3213332	1	3213333	2	3221111	2	3221112	3	3221113	6	3221121	0	3221122	2	3221123	5	3221131	1	3221132	2
3221133	2	3221211	1	3221212	2	3221213	0	3221221	0	3221222	2	3221223	3	3221231	1	3221232	1	3221233	0
3221311	1	3221312	0	3221313	0	3221321	2	3221322	2	3221323	1	3221331	1	3221332	0	3221333	2	3222111	1
3222112	7	3222113	5	3222121	1	3222122	16	3222123	7	3222131	0	3222132	3	3222133	3	3222211	2	3222212	7
3222213	4	3222221	5	3222222	25	3222223	33	3222231	2	3222232	9	3222233	9	3222311	0	3222312	2	3222313	2
3222321	5	3222322	11	3222323	13	3222331	1	3222332	12	3222333	7	3223111	0	3223112	7	3223113	11	3223121	1
3223122	12	3223123	13	3223131	0	3223132	3	3223133	11	3223211	1	3223212	4	3223213	9	3223221	1	3223222	13
3223223	25	3223231	1	3223232	7	3223233	19	3223311	0	3223312	7	3223313	5	3223321	4	3223322	12	3223323	25
3223331	4	3223332	10	3223333	32	3231111	6	3231112	8	3231113	3	3231121	0	3231122	6	3231123	10	3231131	1
3231132	0	3231133	6	3231211	1	3231212	2	3231213	2	3231221	2	3231222	1	3231223	10	3231231	1	3231232	2
3231233	2	3231311	1	3231312	1	3231313	1	3231321	1	3231322	0	3231323	1	3231331	1	3231332	2	3231333	2
3232111	1	3232112	11	3232113	14	3232121	4	3232122	21	3232123	24	3232131	0	3232132	3	3232133	5	3232211	5
3232212	7	3232213	9	3232221	9	3232222	40	3232223	64	3232231	1	3232232	16	3232233	21	3232311	4	3232312	3
3232313	8	3232321	6	3232322	20	3232323	37	3232331	3	3232332	20	3232333	27	3233111	4	3233112	16	3233113	17
3233121	2	3233122	16	3233123	23	3233131	3	3233132	4	3233133	16	3233211	0	3233212	11	3233213	23	3233221	6
3233222	42	3233223	85	3233231	6	3233232	19	3233233	49	3233311	4	3233312	11	3233313	13	3233321	11	3233322	42
3233323	84	3233331	3	3233332	36	3233333	160	3311111	0	3311112	0	3311113	0	3311121	0	3311122	0	3311123	0
3311131	0	3311132	0	3311133	0	3311211	1	3311212	0	3311213	0	3311221	0	3311222	0	3311223	0	3311231	0
3311232	0	3311233	0	3311311	0	3311312	0	3311313	0	3311321	0	3311322	0	3311323	0	3311331	0	3311332	0
3311333	0	3312111	0	3312112	0	3312113	0	3312121	0	3312122	0	3312123	0	3312131	0	3312132	0	3312133	1
3312211	0	3312212	1	3312213	0	3312221	1	3312222	0	3312223	0	3312231	0	3312232	0	3312233	0	3312311	0
3312312	0	3312313	1	3312321	0	3312322	0	3312323	0	3312331	0	3312332	0	3312333	1	3313111	0	3313112	0
3313113	1	3313121	1	3313122	2	3313123	0	3313131	0	3313132	0	3313133	2	3313211	0	3313212	0	3313213	0
3313221	1	3313222	0	3313223	0	3313231	0	3313232	0	3313233	0	3313311	0	3313312	1	3313313	1	3313321	1
3313322	0	3313323	2	3313331	0	3313332	0	3313333	1	3321111	1	3321112	1	3321113	1	3321121	0	3321122	1
3321123	2	3321131	0	3321132	0	3321133	0	3321211	0	3321212	0	3321213	1	3321221	0	3321222	1	3321223	2
3321231	0	3321232	1	3321233	2	3321311	0	3321312	0	3321313	1	3321321	0	3321322	0	3321323	1	3321331	1
3321332	0	3321333	2	3322111	1	3322112	1	3322113	6	3322121	0	3322122	2	3322123	4	3322131	1	3322132	0
3322133	1	3322211	1	3322212	0	3322213	1	3322221	2	3322222	10	3322223	20	3322231	1	3322232	3	3322233	4
3322311	0	3322312	1	3322313	2	3322321	1	3322322	6	3322323	4	3322331	1	3322332	3	3322333	3	3323111	1
3323112	2	3323113	8	3323121	0	3323122	4	3323123	5	3323131	1	3323132	1	3323133	6	3323211	0	3323212	2
3323213	4	3323221	1	3323222	6	3323223	15	3323231	0	3323232	4	3323233	13	3323311	1	3323312	3	3323313	3
3323321	1	3323322	4	3323323	20	3323331	2	3323332	8	3323333	39	3331111	9	3331112	7	3331113	18	3331121	2
3331122	6	3331123	22	3331131	3	3331132	5	3331133	11	3331211	0	3331212	2	3331213	4	3331221	4	3331222	5
3331223	14	3331231	1	3331232	2	3331233	6	3331311	0	3331312	1	3331313	1	3331321	0	3331322	3	3331323	9
3331331	0	3331332	2	3331333	8	3332111	5	3332112	13	3332113	24	3332121	6	3332122	13	3332123	54	3332131	0
3332132	3	3332133	35	3332211	2	3332212	9	3332213	33	3332221	2	3332222	44	3332223	161	3332231	3	3332232	12
3332233	81	3332311	1	3332312	4	3332313	6	3332321	4	3332322	26	3332323	78	3332331	3	3332332	16	3332333	90
3333111	6	3333112	27	3333113	118	3333121	9	3333122	27	3333123	145	3333131	0	3333132	18	3333133	113	3333211	3
3333212	15	3333213	60	3333221	10	3333222	53	3333223	373	3333231	4	3333232	28	3333233	271	3333311	7	3333312	13
3333313	102	3333321	10	3333322	94	3333323	538	3333331	18	3333332	127	3333333	1280						

References

Ash, R. B. (1972). *Real Analysis and Probability.* New York, N.Y.: Academic Press, Inc.

Bacchus, F. (1993). Using first-order probability logic for the construction of Bayesian networks. In *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence* (pp. 219-226). Los Angeles, CA: Morgan Kaufmann.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975) *Discrete Multivariate Analysis: Theory and Practice.* MIT Press, Cambridge, MA.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees.* Wadsworth, Belmont, CA.

Chickering, D. (1996). Learning Bayesian networks is NP-complete. In D. Fisher and H. Lenz (Eds.), *Learning from Data* (pp. 121-130). Springer-Verlag.

Chickering, D.M. and Heckerman, D. (1999). Fast learning from sparse data. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence* (pp. 109-115). San Francisco, CA: Morgan Kaufmann.

Cooper, G.F. (1999). An overview of the representation and discovery of causal relationships using Bayesian networks. In C. Glymour and G.F. Cooper (Eds.), *Computation, Causation, and Discovery*, (pp. 3-62). The MIT Press.

Cooper, G.F. and Herskovits, E. (1991). A Bayesian method for constructing Bayesian belief networks from databases. In *Proceedings of Seventh Conference on Uncertainty in Artificial Intelligence* (pp. 86-94). Los Angeles, CA: Morgan Kaufmann.

D'Ambrosio, B. (1995). Local expression languages for probabilistic dependence, *International Journal of Approximate Reasoning*, **13**, 61-81.

Darroch, J. N., Lauritzen, S. L. and Speed, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *The Annals of Statistics*,**8**, 3, 522-539.

DeGroot, M. H. and Eriksson, E. A. (1985). Probability forecasting, stochastic dominance, and the Lorenz curve. In J. M. Bernado et al. (Eds.)*Bayesian Statistics* (Vol. 2, pp. 99-118). Elsevier Science Publishers (North-Holland), Amsterdam.

DeGroot, M.H., and Fienberg, S.E. (1982). Assessing probability assessors: Calibration and refinement. In S.S. Gupta and J.O. Berger (Eds.), *Statistical Decision Theory and Related Topics III,* (Vol. 1, 291-314). Academic Press, New York.

DeGroot, M. H. and Fienberg, S. E. (1983). The comparison and evaluation of forecasters. *The Statistician*, **32**, 12-22.

DeGroot, M.H., and Fienberg, S.E. (1986). Comparing probability forecasters: basic binary concepts and multivariate extensions. In P. Goel and A. Zellner (Eds.), *Bayesian Inference and Decision Techniques,* (pp. 247-264). Elsevier Science Publishers B.V.

Friedman, N. and Goldszmidt, M. (1998) Learning Bayesian networks with local structure. In M.I. Jordan, (Ed.), *Learning in Graphical Models* (421-460). Dordrecht, Netherlands: Kluwer.

Fienberg, S. E. (1980). *The Analysis of Cross-Classified Categorical Data.* MIT Press, Cambridge, MA.

Fienberg, S. E. and Kim, S.-H. (1998). Calibration and refinement for classification trees. *Journal of Statistical Planning and Inference*, **70**, 241-254.

Fienberg, S. E. and Kim, S.-H. (1999). Combining conditional log-linear structures. *Journal of the American Statistical Association*, **94**, 445, 229-239.

Geiger, D. and Heckerman, D. (1996). Knowledge representation and inference in similarity networks and Bayesian multinets, *Artificial Intelligence*, **82**, 45-74.

Goldman, R.P. and Charniak, E. (1993). A language for construction of belief networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**, 3, 196-208.

Haddawy, P. (1994). Generating Bayesian networks from probability logic knowledge bases. In *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, (pp. 262-269). Los Angeles, CA: Morgan Kaufmann.

Haertel, E. H. and Wiley, D. E. (1992). Representations of ability structures: Implications for testing. In N. Fredericksen, R. J. Mislevy, and I. I. Bejar (Eds.) *Test theory for a new generation of tests* (pp. 359-384). Hillsdale, NJ.: Erlbaum.

Heckerman, D., Geiger, D., and Chickering, D.M. (1995), Learning Bayesian networks: The combination of knowledge and statistical data, *Machine Learning*, **20**, 197-243.

Kim, S.-H. (1992). Instability in a tree approach to regression. *ETS Research Report* **RR-92-1**. Educational Testing Service, Princeton, NJ.

Kim, S.-H. (1994). A general property among nested, pruned subtrees of a decision-support tree. *Communication in Statistics, Theory and Methods*, **23**, 4, 1227-1238.

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

Mahoney, S.M. and Laskey, K.B. (1999). Representing and combining partially specified CPTs. In K.B. Laskey and H. Prade, (Eds.), *Uncertainty in Artificial Intelligence* (391-400). San Francisco, California: Morgan Kaufmann Publishers.

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika*, **59**, 4, 439-483.

Mislevy, R.J., Almond, R.G., Yan, D., and Steinberg, L.S. (1999), Bayes nets in educational assessment: Where the numbers come from. In K.B. Laskey and H. Prade, (Eds.), *Uncertainty in Artificial Intelligence* (pp.437-446). San Francisco, California: Morgan Kaufmann Publishers.

Murphy, A.H., and Epstein, E.S. (1967). Verification of probabilistic predictions: A brief review. *Journal of Applied Meteorology*, 6, 748–755.

Neil, J.R., Wallace, C.S., and Korb, K.B. (1999). Learning Bayesian networks with restricted causal interactions. In K.B. Laskey and H. Prade, (Eds.), *Uncertainty in Artificial Intelligence* (486-493). San Francisco, California: Morgan Kaufmann Publishers.

Pearl, J. (1986), Fusion, propagation and structuring in belief networks, *Artificial Intelligence*, **29**, 241-288.

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

Poole, D. (1993). Probabilistic horn abduction and Bayesian networks, *Artificial Intelligence*, **64**, 1, 81-129.

Tatsuoka, K. K. (1987). Validation of cognitive sensitivity for item response curves. *Journal of Educational Measurement*, **24**, 233-245.

Tatsuoka, K. K. (1990). Toward an integration of item response theory and cognitive error disgnosis. In N. Frederiksen, R. Glaser, A. Lesgold, and M. G. Shafto, (Eds.), *Diagnostic Monitoring of Skill and Knowledge Acquisition* (pp. 453-488). Hillsdale, Jew Jersey: Erlbaum.

Thiesson, B., Meek, C., Chickering, D.M., and Heckerman, D. (1999). Computationally efficient methods for selecting among mixtures of graphical models (with discussion). In J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, (Eds.), *Bayesian Statistics 6* (pp. 631-656). Oxford: University Press.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics.* Wiley, New York.