The Implication of Probabilistic Conditional Independence and Embedded Multivalued Dependency

S.K.M. Wong
Department of Computer Science
University of Regina
Regina, Saskatchewan, Canada, S4S 0A2
wong@cs.uregina.ca

C.J. Butz
School of Information Technology & Engineering
University of Ottawa
Ottawa, Ontario, Canada, K1N 6N5
butz@site.uottawa.ca

Abstract

It has been suggested that Bayesian networks and relational databases are *different* because the implication problems for probabilistic conditional independence and embedded multivalued dependency do not always coincide. The present study indicates that the implication problems coincide on *solvable* classes of dependencies and differ on *unsolvable* classes. We therefore maintain that Bayesian networks and relational databases are the *same* in a practical sense, since only *solvable* classes of dependencies are useful in the design and implementation of both knowledge systems.

1 Introduction

The *relational database model* [3] is well established as the basis for developing database systems. The lossless decomposition of a relation into two projections (smaller relations) is based upon the notion of *embedded multivalued dependency* (EMVD). On the other hand, *Bayesian networks* [6] have become a proven framework for uncertainty management. In this approach, *probabilistic conditional independence* is used to factorize a probability distribution into two smaller distributions. It has been suggested [9] that Bayesian networks and relational databases are *different* because the implication problems for probabilistic conditional independence and EMVD do not always coincide. This remark, however, does not take into consideration the important issue of the *solvability* of the implication problem.

Our Bayesian database model [11, 13] provides a *unified* framework for modeling both Bayesian network and relational database applications. We extend the standard relation by adding a column to store the probability values. We introduce *Bayesian embedded multivalued dependency* (BEMVD) as the necessary and sufficient condition for the lossless decomposition of a probabilistic relation into two smaller probabilistic relations. It is important to realize that BEMVD is simply probabilistic conditional independence expressed as a database dependency. Moreover, EMVD is a necessary but not a sufficient condition for BEMVD. Therefore, our Bayesian model offers a convenient tool to present a comprehensive study of the implication of BEMVDs and EMVDs.

The *implication problem* [1, 2, 6, 7, 9, 10, 12, 14] is to determine whether a set $\sum$ of dependencies *logically implies* another dependency σ. We say $\sum$ logically implies σ, written $\sum \models \sigma$, if every relation which satisfies $\sum$ also satisfies σ. That is, there is no counter-example relation such that all of the dependencies in $\sum$ are satisfied but σ is not. We would like to know:

> *Do the implication problems coincide in these two database models?*

That is, we would like to know whether the following proposition holds:

$$\mathbf{C} \models \mathbf{c} \quad \Longleftrightarrow \quad C \models c, \tag{1}$$

where $\mathbf{C}$ and $\mathbf{c}$ are *probabilistic* dependencies in our Bayesian database model, and C and c are the corresponding *data* dependencies in the relational database model. In this paper, we study whether Proposition (1) holds in the following four pairs of dependencies: (BMVD,MVD), (conflict-free BMVD, conflict-free MVD), (conflict-free BEMVD, conflict-free EMVD), and (BEMVD,EMVD), as illustrated in Figure 1. We will show that the only pair for which the implication problems do *not* coincide is (BEMVD,EMVD). This is an important observation since the BEMVD class does *not* have a finite complete axiomatization [9, 14]. Similarly, the EMVD class does *not* have a finite complete axiomatization [5, 7]. On the contrary, each of the classes in the former three pairs *do* indeed have a finite complete axiomatization. These results suggest that the implication problems coincide on the *solvable* classes of dependencies and differ on the *unsolvable* classes, as depicted in Figure 2. We therefore maintain that Bayesian networks and relational databases are the *same* in a practical sense, since only *solvable* classes of dependencies are useful in the design and implementation of both knowledge systems.

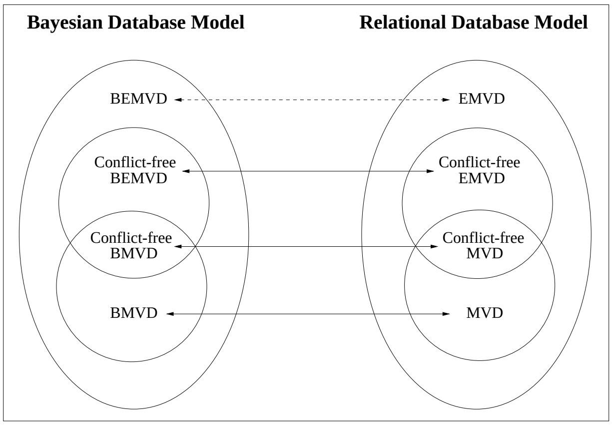

Figure 1: The four classes of *generalized* dependencies in the Bayesian database model and the corresponding class of *data* dependencies in the standard relational database model. A double solid arrow means that Proposition (1) holds, while a double dashed arrow indicates that Proposition (1) does *not* hold.

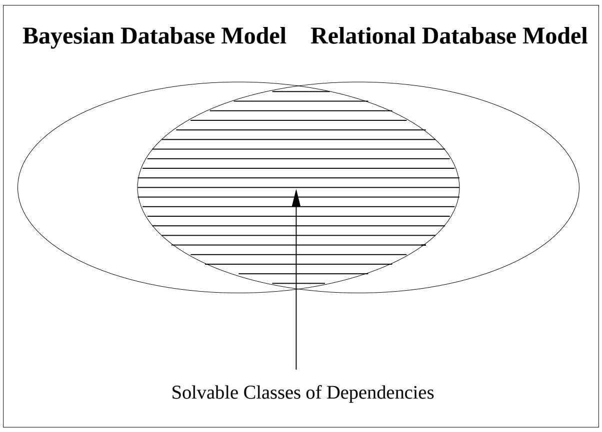

Figure 2: The implication problems coincide on the *solvable* classes of dependencies.

This paper is organized as follows. In Section 2, we review the notions of embedded multivalued dependency (EMVD) and Bayesian embedded multivalued dependency (BEMVD), i.e., probabilistic conditional independence. We define three subclasses of BEMVD in Section 3. In Section 4, we use the implication problem to study the relationship between the relational database model and the Bayesian database model. In Section 5, we examine the role solvability plays in this relationship. The conclusion is presented in Section 6.

2 Embedded Multivalued Dependency and Bayesian Embedded Multivalued Dependency

Before introducing our Bayesian database model [11, 13], we briefly review the notions of *relation* and *embedded multivalued dependency* (EMVD) in the standard relational database model [3].

A *relation scheme* $R = \{A_1, A_2, \ldots, A_m\}$ is a finite set of *attributes*. Corresponding to each attribute A_i is a nonempty finite set D_{A_i}, $1 \leq i \leq m$, called the *domain* of A_i. Let $D = D_{A_1} \cup D_{A_2} \ldots \cup D_{A_m}$. A *relation* r on the relation scheme R, written $r(R)$, is a finite set of mappings $\{t_1, t_2, \ldots, t_s\}$ from R to D with the restriction that for each mapping $t \in r$, $t(A_i)$ must be in D_{A_i}, $1 \leq i \leq m$, where $t(A_i)$ denotes the value obtained by restricting the mapping to A_i. The mappings are called *tuples* and $t(A)$ is called the A-value of t. To simplify the notation, we will simply write a relation r on R as $r(A_1A_2 \ldots A_m)$.

Let X, Y, Z, W be pairwise disjoint subsets of attributes of scheme $R = XYZW$. We say relation $r(XYZW)$ satisfies the *embedded multivalued dependency* (EMVD) $X \rightarrow\rightarrow Y|Z$ in the context XYZ, if the projection $\pi_{XYZ}(r)$ of $r(XYZW)$ satisfies the condition:

$$\pi_{XYZ}(r) \quad = \quad \pi_{XY}(r) \bowtie \pi_{XZ}(r), \qquad (2)$$

where π and $\bowtie$ are the *projection* and *natural join* operators, respectively. For example, relation $r(ABCD)$ on the top of Figure 3 satisfies the EMVD $B \rightarrow\rightarrow A|C$, since $\pi_{ABC}(r) = \pi_{AB}(r) \bowtie \pi_{BC}(r)$.

We *generalize* a traditional relation $r(R)$ by adding an additional column A_p. A *probabilistic* relation is denoted by $\mathbf{r}(RA_p)$, where the column labelled by A_p stores the probability values. Note that $\mathbf{t}(A_p) > 0$, for all $\mathbf{t} \in \mathbf{r}(RA_p)$, namely, tuples with zero probability are *not* stored in relation $\mathbf{r}(R)$. For convenience, we will write $\mathbf{r}(RA_p)$ as $\mathbf{r}(R)$ and say relation $\mathbf{r}$ is on R with the attribute A_p understood by context. That is, relations denoted by boldface represent probabilistic relations.

Let $\mathbf{r}(R)$ be a relation and X be a subset of R. The *marginalization of* $\mathbf{r}$ *onto* X, written $\tau_X(\mathbf{r})$, is:

$$\tau_X(\mathbf{r}) \quad = \quad \{ \ \mathbf{t}(XA_{p(X)}) \ | \ \mathbf{t}(X) \in \pi_X(\mathbf{r}) \text{ and } \mathbf{t}(A_{p(X)}) = \sum_{\mathbf{t}' \in \mathbf{r}, \mathbf{t}'(X) = \mathbf{t}(X)} \mathbf{t}'(A_p) \ \}.$$

$r(ABCD)$ =

A	B	C	D
0	0	0	0
0	0	0	1
0	0	1	1
1	0	0	0
1	0	1	0
1	1	1	1

$\pi_{ABC}(r)$ =

A	B	C
0	0	0
0	0	1
1	0	0
1	0	1
1	1	1

=

A	B
0	0
1	0
1	1

$\bowtie$

B	C
0	0
0	1
1	1

Figure 3: Relation $r(ABCD)$ satisfies the EMVD $B \rightarrow\rightarrow A|C$, since $\pi_{ABC}(r) = \pi_{AB}(r) \bowtie \pi_{BC}(r)$.

In the literature [6], the relation $\tau_X(\mathbf{r})$ is called the *marginal distribution* $p(X)$ of $p(R)$ onto X.

The *product join* of two relations $\mathbf{r}_1(X)$ and $\mathbf{r}_2(Y)$, written $\mathbf{r}_1(X) \times \mathbf{r}_2(Y)$, is defined as

$$\begin{aligned} & \mathbf{r}_1(X) \times \mathbf{r}_2(Y) \\ = \; & \{\mathbf{t}(XYA_{p(X)\cdot p(Y)}) \mid \mathbf{t}(XY) \in \pi_X(\mathbf{r}_1) \bowtie \pi_Y(\mathbf{r}_2) \\ & \text{ and } \mathbf{t}(A_{p(X)\cdot p(Y)}) = \mathbf{t}(A_{p(X)}) \cdot \mathbf{t}(A_{p(Y)})\}. \end{aligned}$$

Thus, $\mathbf{r}_1(X) \times \mathbf{r}_2(Y)$ denotes the product $p(X) \cdot p(Y)$ of the two distributions $p(X)$ and $p(Y)$.

Let X, Y, Z, W be pairwise disjoint subsets of attributes of scheme $R = XYZW$. A probabilistic relation $\mathbf{r}(XYZW)$ satisfies the *Bayesian embedded multivalued dependency* (BEMVD), $X \Rightarrow\Rightarrow Y|Z$, if

$$\tau_{XYZ}(\mathbf{r}) \;=\; \tau_{XY}(\mathbf{r}) \times \tau_{XZ}(\mathbf{r}) \times \tau_X(\mathbf{r})^{-1}, \quad (3)$$

where the inverse relation $\tau_X(\mathbf{r})^{-1}$ is defined using $\tau_X(\mathbf{r})$ as follows:

$$\begin{aligned} \tau_X(\mathbf{r})^{-1} \;=\; & \{\mathbf{t}(XA_{1/p(X)}) \mid \mathbf{t}(X) = \mathbf{t}'(X) \in \tau_X(\mathbf{r}) \\ & \text{ and } \mathbf{t}(A_{1/p(X)}) = 1/\mathbf{t}'(A_{p(X)})\}. \end{aligned}$$

Note that this relation $\tau_X(\mathbf{r})^{-1}$ is well defined because by definition $\tau_X(\mathbf{r})$ does not contain any tuples with zero probability. By introducing a binary operator $\otimes$ called *Markov join*, we say that a relation $\mathbf{r}(XYZ)$ satisfies the BEMVD $X \Rightarrow\Rightarrow Y|Z$, if

$$\begin{aligned} \tau_{XYZ}(\mathbf{r}) \;\equiv\; & \tau_{XY}(\mathbf{r}) \otimes \tau_{XZ}(\mathbf{r}) \\ =\; & \tau_{XY}(\mathbf{r}) \times \tau_{XZ}(\mathbf{r}) \times \tau_X(\mathbf{r})^{-1}. \end{aligned}$$

For example, relation $\mathbf{r}(ABCD)$ on the top of Figure 4 satisfies the BEMVD $B \Rightarrow\Rightarrow A|C$, since the marginal $\tau_{ABC}(\mathbf{r})$ can be written as $\tau_{ABC}(\mathbf{r}) = \tau_{AB}(\mathbf{r}) \otimes \tau_{BC}(\mathbf{r})$.

$\mathbf{r}(ABCD)$ =

A	B	C	D	$A_{p(ABCD)}$
0	0	0	0	0.1
0	0	0	1	0.1
0	0	1	1	0.2
1	0	0	0	0.1
1	0	1	0	0.1
1	1	1	1	0.4

$\tau_{ABC}(\mathbf{r})$ =

A	B	C	$A_{p(ABC)}$
0	0	0	0.2
0	0	1	0.2
1	0	0	0.1
1	0	1	0.1
1	1	1	0.4

=

A	B	$A_{p(AB)}$
0	0	0.4
1	0	0.2
1	1	0.4

$\otimes$

B	C	$A_{p(BC)}$
0	0	0.3
0	1	0.3
1	1	0.4

=

A	B	C	$A_{\frac{p(AB)p(BC)}{p(B)}}$
0	0	0	$(0.4)(0.3)/(0.6) = 0.2$
0	0	1	$(0.4)(0.3)/(0.6) = 0.2$
1	0	0	$(0.2)(0.3)/(0.6) = 0.1$
1	0	1	$(0.2)(0.3)/(0.6) = 0.1$
1	1	1	$(0.4)(0.4)/(0.4) = 0.4$

Figure 4: Relation $\mathbf{r}(ABCD)$ satisfies the BEMVD $B \Rightarrow\Rightarrow A|C$, since $\tau_{ABC}(\mathbf{r}) = \tau_{AB}(\mathbf{r}) \otimes \tau_{BC}(\mathbf{r})$.

Two important remarks need to made. First, EMVD is a *necessary* but not a sufficient condition for BEMVD [11]. It is straightforward to construct a probabilistic relation $\mathbf{r}(R)$ which does *not* satisfy a given BEMVD $X \Rightarrow\Rightarrow Y|Z$, yet the traditional relation $r(R)$ (obtained by striking out the column A_p) does indeed satisfy the corresponding EMVD $X \rightarrow\rightarrow Y|Z$. Second, we say that Y and Z are *conditionally independent* given X in a probability distribution $p(XYZW)$, written $I(Y, X, Z)$, if

$$p(XYZ) \;=\; \frac{p(XY) \cdot p(XZ)}{p(X)}, \quad (4)$$

where X, Y, Z, W are pairwise disjoint. By comparing Equations (3) and (4), it can be easily seen that BEMVD is simply *probabilistic conditional independence* expressed as a database dependency. Let $\mathbf{r}(XYZW)$ be the probabilistic relation representing the probability distribution $p(XYZW)$. Saying that $\mathbf{r}(XYZW)$ satisfies the BEMVD $X \Rightarrow\Rightarrow Y|Z$ is the same as saying that Y and Z are *conditionally independent* given X in $p(XYZW)$, namely,

$$X \Rightarrow\Rightarrow Y|Z \iff I(Y, X, Z). \quad (5)$$

Thus, we use the terms BEMVD and probabilistic conditional independency interchangeably.

3 Subclasses of Bayesian Embedded Multivalued Dependency

Here we define three special subclasses of BEMVD. The corresponding classes of *data* dependencies are defined by replacing the notion of BEMVD with EMVD.

In the special case when the BEMVD $X \Rightarrow\Rightarrow Y|Z$ involves all the attributes in a relation scheme R, i.e., $R = XYZ$, we call $X \Rightarrow\Rightarrow Y|Z$ a *full* BEMVD, or simply *Bayesian multivalued dependency* (BMVD). We write the BMVD $X \Rightarrow\Rightarrow Y|Z$ as $X \Rightarrow\Rightarrow Y$, if the context is understood.

A graphical structure $\mathcal{G}$ is called a *perfect-map* [2, 6] of a set $\sum$ of dependencies, if every dependency logically implied by $\sum$ can be inferred from $\mathcal{G}$, and every dependency inferred from $\mathcal{G}$ is logically implied by $\sum$. We can use the *separation* method [2] to infer BMVDs from an acyclic hypergraph. We say that a BMVD $X \Rightarrow\Rightarrow Y$ is inferred from an acyclic hypergraph $\mathcal{R}$, if and only if Y is the union of some disconnected components of $\mathcal{R}$ with the set X of nodes deleted. (For example, consider the acyclic hypergraph $\mathcal{R} = \{R_1 = AB, R_2 = BCD, R_3 = DE, R_4 = DFG, R_5 = DFH\}$ on the set $R = ABCDEFGH$ of attributes. The disconnected components obtained by deleting node D in $\mathcal{R}$ are $S_1 = ABC$, $S_2 = E$, and $S_3 = FGH$. By definition, the BMVDs $D \Rightarrow\Rightarrow ABC$, $D \Rightarrow\Rightarrow E$, $D \Rightarrow\Rightarrow FGH$, and $D \Rightarrow\Rightarrow ABCE$ can be inferred from $\mathcal{R}$. On the other hand, the BMVD $D \Rightarrow\Rightarrow BC$ is *not* inferred from $\mathcal{R}$ since BC is not equal to the union of some of the sets in $\{S_1, S_2, S_3\}$.)

In general, not every set of BMVDs can be faithfully represented by a single acyclic hypergraph. For example, there is no single acyclic hypergraph that can simultaneously encode the set $\mathbf{C} = \{A_1 \Rightarrow\Rightarrow A_2,\ A_3 \Rightarrow\Rightarrow A_2\}$ of BMVDs on $R = A_1A_2A_3$. The *conflict-free BMVD* [2] class contains every set of BMVDs which has a perfect-map in the form of an acyclic hypergraph.

We now introduce a new subclass within BEMVD called *conflict-free BEMVD*. The *d-separation* [6] method for inferring BEMVDs from a *directed acyclic graph* (DAG) in a manner similar to using the *separation* method for inferring BMVDs from an acyclic hypergraph. Just as there are some sets of BMVDs which cannot be faithfully represented by a single acyclic hypergraph, there are some sets of BEMVDs which cannot be faithfully represented by a single DAG. For example, there is no *single* DAG that can simultaneously encode the set $\mathbf{C} = \{A_2 \Rightarrow\Rightarrow A_1|A_3,\ A_3 \Rightarrow\Rightarrow A_1|A_2,\ A_3 \Rightarrow\Rightarrow A_1A_2|A_4\}$ of BEMVDs on $R = A_1A_2A_3A_4$. The *conflict-free BEMVD* class is defined as those sets of BEMVDs which have a perfect-map in the form of a DAG.

4 Comparing the Bayesian and Relational Database Models

In this section, we show that the implication problems in the Bayesian database model and the relational database model coincide on those classes of dependencies with a finite complete axiomatization. We remind the reader that $\mathbf{C}$ and C denote corresponding sets of BEMVDs and EMVDs, respectively. That is, $C = \{X \to\to Y|Z \mid X \Rightarrow\Rightarrow Y|Z \in \mathbf{C}\}$. Similarly, c denotes the EMVD corresponding to the BEMVD $\mathbf{c}$. We begin our analysis with the pair (BMVD,MVD).

The following two inference axioms (M1) and (M2) are both minimal [4] and *complete* [1] for the MVD class:

(M1) If $Y \subseteq X$, then $X \to\to Y$,

(M2) If $X \to\to Y$ and $Y \to\to Z$, then $X \to\to Z - Y$.

Since the corresponding inference axioms (B1) and (B2) are *sound* [12] for the BMVD class:

(B1) If $Y \subseteq X$, then $X \Rightarrow\Rightarrow Y$,

(B2) If $X \Rightarrow\Rightarrow Y$ and $Y \Rightarrow\Rightarrow Z$, then $X \Rightarrow\Rightarrow Z - Y$,

it immediately follows that the implication problems coincide in the pair (BMVD, MVD).

Theorem 1 $\mathbf{C} \models \mathbf{c} \iff C \models c$ in the pair (BMVD,MVD).

Proof: ($\Rightarrow$) $\mathbf{C} \models \mathbf{c} \implies C \models c$ is a *tautology* in the pair (BMVD,MVD) [11].

($\Leftarrow$) Let $C \models c$. Since the MVD class has a complete axiomatization, $C \models c$ implies that $C \vdash c$. That is, there exists a derivation sequence s of the MVD c by applying the MVD inference axioms (M1) and (M2) to the MVDs in C. This means there exists a derivation sequence $\mathbf{s}$ of the BMVD $\mathbf{c}$ using the BMVDs inference axioms (B1) and (B2) on the BMVDs in $\mathbf{C}$, which parallels the derivation sequence s of the MVD c. That is, $\mathbf{C} \vdash \mathbf{c}$. Since these BMVD axioms are sound, $\mathbf{C} \vdash \mathbf{c}$ implies that $\mathbf{C} \models \mathbf{c}$. $\square$

Obviously the implication problems coincide in the pair (conflict-free BMVD, conflict-free MVD), as this pair is a subclass of the pair (BMVD,MVD).

Theorem 2 $\mathbf{C} \models \mathbf{c} \iff C \models c$ in the pair (conflict-free BMVD, conflict-free MVD).

$r(A_1A_2A_3A_4)$ =

A_1	A_2	A_3	A_4
0	0	0	0
0	0	0	1
0	1	0	0
1	0	0	0
1	1	0	0
1	1	1	0

Figure 5: Relation r satisfies all of the EMVDs in C but does not satisfy the EMVD c, where C and c are defined in Example 1. Therefore, $C \not\models c$.

This concludes our brief discussion of *full* (nonembedded) dependencies.

We now consider *embedded* dependencies. The special classes of conflict-free BEMVD and conflict-free EMVD both have a finite *complete* axiomatization [6].

Theorem 3 [6] $\mathbf{C} \models \mathbf{c} \iff C \models c$ in the pair (conflict-free BEMVD, conflict-free EMVD).

Theorems 1, 2, and 3 are significant since they indicate that testing the implication of *probabilistic* dependencies is the *same* as testing the implication of *data* dependencies. An immediate consequence is that the *chase* [3] algorithm can be directly applied as a *nonaxiomatic* method to test the implication of BMVDs [10, 12].

5 The Role of Solvability

In the last section, it was shown that the implication problems coincide for some classes of dependencies. However, Studeny [9] pointed out that the implication problems for EMVD and BEMVD do *not* always coincide.

Example 1 Consider the set $\mathbf{C} = \{A_3A_4 \Rightarrow\Rightarrow A_1|A_2,\ A_1 \Rightarrow\Rightarrow A_3|A_4,\ A_2 \Rightarrow\Rightarrow A_3|A_4,\ \emptyset \Rightarrow\Rightarrow A_1|A_2\}$ of BEMVDs, and $\mathbf{c}$ the single BEMVD $\emptyset \Rightarrow\Rightarrow A_3|A_4$. In [8], Studeny showed that $\mathbf{C} \models \mathbf{c}$. Now consider the set $C = \{X \rightarrow\rightarrow Y|Z \mid X \Rightarrow\Rightarrow Y|Z \in \mathbf{C}\}$ of EMVDs corresponding to the set $\mathbf{C}$ of BEMVDs, and the single EMVD $\emptyset \rightarrow\rightarrow A_3|A_4$ corresponding to the BEMVD $\mathbf{c}$. Consider the relation $r(A_1A_2A_3A_4)$ in Figure 5. It can be verified that $r(A_1A_2A_3A_4)$ satisfies all of the EMVDs in C but does not satisfy the EMVD c. That is, $C \not\models c$. □

Example (1) indicates that

$$\mathbf{C} \models \mathbf{c} \not\Longrightarrow C \models c. \qquad (6)$$

Example 2 Consider the set $C = \{A_1 \rightarrow\rightarrow A_3|A_4,\ A_2 \rightarrow\rightarrow A_3|A_4,\ A_3A_4 \rightarrow\rightarrow A_1|A_2\}$ of EMVDs, and let c be the single EMVD $A_1A_2 \rightarrow\rightarrow A_3$. It can be shown [12] that $C \models c$. Now consider the corresponding set of BEMVDs $\mathbf{C} = \{A_1 \Rightarrow\Rightarrow A_3|A_4,\ A_2 \Rightarrow\Rightarrow A_3|A_4,\ A_3A_4 \Rightarrow\Rightarrow A_1|A_2\}$ and $\mathbf{c}$ is the BMVD $A_1A_2 \Rightarrow\Rightarrow A_3$. It is easily verified that relation $\mathbf{r}(A_1A_2A_3A_4)$ in Figure 6 satisfies all of the BEMVDs in $\mathbf{C}$ but does not satisfy the BEMVD $\mathbf{c}$. Therefore, $\mathbf{C} \not\models \mathbf{c}$. □

$\mathbf{r}(A_1A_2A_3A_4)$ =

A_1	A_2	A_3	A_4	A_p
0	0	0	0	0.2
0	0	0	1	0.2
0	0	1	0	0.2
0	0	1	1	0.1
0	1	1	1	0.1
1	0	1	1	0.1
1	1	1	1	0.1

Figure 6: Relation $\mathbf{r}$ satisfies all of the BEMVDs in $\mathbf{C}$ but does not the BEMVD $\mathbf{c}$, where $\mathbf{C}$ and $\mathbf{c}$ are defined in Example 2. Therefore, $\mathbf{C} \not\models \mathbf{c}$.

Example 2 indicates that

$$\mathbf{C} \models \mathbf{c} \not\Longleftarrow C \models c. \qquad (7)$$

Based on Equations (6) and (7), Studeny [9] argued that the Bayesian database model and the relational database model are *different*. This remark, however, does not take into consideration one important issue. The question naturally arises as to why the implication problem coincides for some classes of dependencies but not for others. The answer lies in the *solvability* of the implication problem.

There is *no* single DAG which can faithfully represent all of the BEMVDs in the set $\mathbf{C}$ of BEMVDs in Example 1. This means that set $\mathbf{C}$ of BEMVDs belongs to the *general* BEMVD class. Similarly, there is *no* single DAG which can faithfully represent all the EMVDs in the set C of EMVDs in Example 2. This means that the set C of EMVDs belongs to the *general* EMVD class. This means that Studeny's argument that Bayesian networks and standard relational databases are *different* was based on the analysis of the implication problems in the general pair (BEMVD,EMVD). This is an important observation since the general BEMVD class does *not* have a finite complete axiomatization [9, 14], contrary to Pearl's [6] conjecture. Similarly, the general EMVD class does *not* have a finite complete axiomatization [5, 7]. This supports our argument that there is no *real* difference between Bayesian networks and standard relational databases in a practical sense, since only *solvable* classes of dependencies are useful in the design and implementation of both knowledge systems.

6 Conclusion

It has been suggested in [9] that Bayesian networks are *different* from relational databases since the implication problems of probabilistic conditional independence and embedded multivalued dependency do not always coincide. In this paper, we reviewed the fact that our *Bayesian database model* [11, 13] serves as a *unified* model for both Bayesian networks and relational databases. In particular, probabilistic conditional independence can be expressed as *Bayesian embedded multivalued dependency* (BEMVD). We pointed out that Studeny's observation [9] was based on an analysis of the *general* pair (BEMVD,EMVD). This is important since both the general EMVD class [5, 7] and the general BEMVD class [9, 14] do *not* have a finite complete axiomatization. This means that there is no *real* difference between Bayesian networks and standard relational databases in a practical sense, since only *solvable* classes of dependencies are useful in the design and implementation of both knowledge systems.

References

[1] C. Beeri, R. Fagin, and J.H. Howard. A complete axiomatization for functional and multivalued dependencies in database relations. In *Proceedings of ACM-SIGMOD International Conference on Management of Data*, pages 47–61, 1977.

[2] C. Beeri, R. Fagin, D. Maier, and M. Yannakakis. On the desirability of acyclic database schemes. *Journal of the ACM*, 30(3):479–513, July 1983.

[3] D. Maier. *The Theory of Relational Databases.* Principles of Computer Science. Computer Science Press, Rockville, Maryland, 1983.

[4] A. Mendelzon. On axiomatizing multivalued dependencies in relational databases. *Journal of the ACM*, 26(1):37–44, 1979.

[5] D. Parker and K. Parsaye-Ghomi. Inference involving embedded multivalued dependencies and transitive dependencies. In *Proceedings of ACM-SIGMOD International Conference on Management of Data*, pages 52–57, 1980.

[6] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, San Francisco, California, 1988.

[7] Y. Sagiv and F. Walecka. Subset dependencies and a complete result for a subclass of embedded multivalued dependencies. *Journal of the ACM*, 20(1):103–117, 1982.

[8] M. Studeny. Multiinformation and the problem of characterization of conditional-independence relations. *Problems of Control and Information Theory*, 18(1):3–16, 1989.

[9] M. Studeny. Conditional independence relations have no finite complete characterization. In *Eleventh Prague Conference on Information Theory, Statistical Decision Foundation and Random Processes*, pages 377–396. Kluwer, 1990.

[10] S.K.M. Wong. Testing implication of probabilistic dependencies. In *Twelfth Conference on Uncertainty in Artificial Intelligence*, pages 545–553. Morgan Kaufmann Publishers, 1996.

[11] S.K.M. Wong. An extended relational data model for probabilistic reasoning. *Journal of Intelligent Information Systems*, 9:181–202, 1997.

[12] S.K.M. Wong, C.J. Butz, and D. Wu. On the implication problem for probabilistic conditional independency. Technical Report CS-99-03, University of Regina, Regina, Saskatchewan, Canada, S4S 0A2, September 1999.

[13] S.K.M. Wong, C.J. Butz, and Y. Xiang. A method for implementing a probabilistic model as a relational database. In *Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 556–564. Morgan Kaufmann Publishers, 1995.

[14] S.K.M. Wong and Z.W. Wang. On axiomatization of probabilistic conditional independence. In *Tenth Conference on Uncertainty in Artificial Intelligence*, pages 591–597. Morgan Kaufmann Publishers, 1994.