Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence

Reasoning about Saturated Conditional Independence under Uncertainty: Axioms, Algorithms, and Levesque's Situations to the Rescue*

Sebastian Link
Department of Computer Science, The University of Auckland
Private Bag 92019, Auckland 1142, New Zealand
s.link@auckland.ac.nz

Abstract

The implication problem of probabilistic conditional independencies is investigated in the presence of missing data. Here, graph separation axioms fail to hold for saturated conditional independencies, unlike the known idealized case with no missing data. Several axiomatic, algorithmic, and logical characterizations of the implication problem for saturated conditional independencies are established. In particular, equivalences are shown to the implication problem of a propositional fragment under Levesque's situations, and that of Lien's class of multivalued database dependencies under null values.

1 Introduction

Background. The concept of conditional independence is important for capturing structural aspects of probability distributions, for dealing with knowledge and uncertainty in artificial intelligence, and for learning and reasoning in intelligent systems (Dawid 1979; Pearl 1988; Darwiche 2009). Application areas include natural language processing, speech processing, computer vision, robotics, computational biology, and error-control coding (Halpern 2005; Darwiche 2009). A conditional independence (CI) statement $I(Y, Z \mid X)$ represents the independence of two sets of random variables relative to a third: given three mutually disjoint subsets X, Y, and Z of a set V of random variables, if we have knowledge about the state of X, then knowledge about the state of Y does not provide additional evidence for the state of Z and vice versa. Fundamental to the task of building a Bayesian network is the implication problem of CI statements, which is to decide for an arbitrary set V of random variables, and an arbitrary set $\Sigma \cup \{\varphi\}$ of CI statements over V, whether every probability model that satisfies every CI statement in Σ also satisfies φ. Indeed, if some CI statement φ is not implied by Σ, then adding φ to Σ results in new opportunities to construct complex probability models with polynomially many parameters and to efficiently organize distributed probability computations (Geiger and Pearl 1990). The implication problem for CI statements is not axiomatizable by a finite set of Horn clauses (Studený 1992), and every axiom for CI statements is an axiom for graph separation, but not vice versa (Geiger and Pearl 1993). Recently, the implication problem of stable CI statements (Matúš 1992; de Waal and van der Gaag 2005) has been shown to be finitely axiomatizable (Niepert, Van Gucht, and Gyssens 2008), and *coNP*-complete to decide (Niepert, Van Gucht, and Gyssens 2010). Stability means that the validity of $I(Y, Z \mid X)$ over V implies the validity of every $I(Y, Z \mid X')$ where $X \subseteq X' \subseteq V - YZ$. An important subclass of CI statements are saturated conditional independence (SCI) statements. These are CI statements $I(Y, Z \mid X)$ over V that satisfy $XYZ = V$. Indeed, graph separation and SCI statements enjoy the same axioms (Geiger and Pearl 1993), and their implication problem is decidable in almost linear time (Galil 1982; Link 2012b). These results contribute to the success of Bayesian networks (Darwiche 2009; Geiger and Pearl 1993).

Motivation. Surprisingly, the implication problem of CI statements has not been studied yet in the presence of missing data. Indeed, AI has long recognized the need to reveal missing data, to explain where they come from, and to develop imputation techniques. Significant contributions towards that aim have been made, e.g. (Batista and Monard 2003; Chickering and Heckerman 1997; Dempster, Laird, and Rubin 1977; Fayyad, Piatetsky-Shapiro, and Smyth 1996; Friedman 1997; Lauritzen 1995; Lou and Obradovic 2012; Marlin et al. 2011; Saar-Tsechansky and Provost 2007; Singh 1997; Zhang, Jin, and Zhu 2011; Zhu et al. 2007). However, it is impossible to reveal many missing data because they do not exist, the process is too inaccurate, or resources are insufficient to reveal them. It is thus a natural question to ask what a CI statement in the presence of missing data constitutes, and what can be said about the implication problem. The question is even more important to address than in the classical case, where idealistic assumptions are made that no data are missing or all missing data can be revealed correctly in time. Our findings may start a foundation for reasoning about conditional independence in the presence of missing data, much like the findings of (Dawid 1979; Geiger and Pearl 1990) for complete data. For practice, we recommend to reveal as much missing data as possible, and then to reason about conditional independence under the remaining missing data.

*This research is supported by the Marsden fund council from Government funding, administered by the Royal Society of New Zealand.

Contributions. As a first contribution we assign a suitable semantics to CI statements under uncertainty, that is, in the presence of missing data. Here, it is important to select an appropriate representation of missing data. For this purpose, we choose the most primitive approach in which we use a marker, denoted by μ. An occurrence of μ as a marked "value" of some random variable is interpreted as *no information*, i.e., either a value does not exist, or it exists but is currently unknown. While there is a potential loss in representing knowledge with this interpretation, it is possible to model missing values for which it is not known whether a value exists or is currently unknown. For example, it is difficult to decide whether a missing *maiden name* does not exist at all, or is currently unknown. Our other contributions further justify this interpretation. Due to the infeasibility of the implication problem for CI statements, we focus in the remaining contributions on SCI statements. It is shown that some graph separation axioms fail to hold for SCI statements under uncertainty. Our second contribution is an axiomatic characterization of their associated implication problem. The axiomatization is similar to that of Geiger and Pearl's in the case of certainty (Geiger and Pearl 1990), but one of their rules is no longer sound under uncertainty. We show that completeness under uncertainty is regained by exploiting another sound rule. Our completeness argument is based on special probability models in which two events are assigned probability one half, showing that the implication problem for SCI statements under uncertainty is equivalent to that over special probability models. This insight has remarkable consequences. As a third contribution we establish a logical characterization of the implication problem. In fact, an equivalence is shown to the implication of formulae in a fragment of propositional logic, where implication is defined in terms of Levesque's situations. That is, truth values are assigned to atoms and their negations, but both cannot be false (but they can both be true). As a fourth contribution an equivalence is established to the implication of multivalued database dependencies in the presence of null values, investigated by (Lien 1982). Hence, we establish a counterpart of the known trinity between SCI statements under certainty, classical propositional logic, and multivalued dependencies in purely relational databases (Link 2012b; Malvestuto 1992; Sagiv et al. 1981), for the realistic case of missing data. The equivalence to multivalued dependencies leads to our fifth contribution, an algorithmic characterization showing that the implication of SCI statements under uncertainty is decidable in almost linear time.

Organization. We introduce CI statements under uncertainty in Section 2. In Section 3 we establish an axiomatic characterization of the implication problem for SCI statements under uncertainty. Our completeness argument is used in Section 4 to derive a logical characterization in terms of Levesque's situations. Section 5 establishes a characterization in terms of Lien's multivalued dependencies. An equivalence between instances of the implication problems under uncertainty and sliced instances of the implication problems under certainty allows us in Section 6 to exploit Galil's almost linear time algorithm to decide implication. We conclude and comment on future work in Section 7.

2 Conditional Independence and Uncertainty

We denote by V a finite set $\{v_1,\ldots,v_n\}$ of *random variables*. A *domain mapping* associates a set *dom*(v_i) with each random variable v_i. This set is called the *domain* of v_i and each of its elements is an *event* of v_i. We assume that each domain *dom*(v_i) contains the element μ, which we call the *marker*. Although we use μ like any other event, we think of μ as a marker, denoting that *no information* is currently available about the event of v_i. The interpretation of this marker as no information means that an event does either not exist (known as a structural zero in statistics, and the null value inapplicable in databases), or an event exists but is currently unknown (known as a sampling zero in statistics, and the null value applicable in databases). The disadvantage of using this interpretation is a loss in knowledge when representing values known to not exist or known to exist but currently unknown. One advantage of this interpretation is its simplicity. As another advantage one can represent missing values, even if it is unknown whether they do not exist or exist but are currently unknown. Further advantages will be revealed by the results established in this paper. For $X=\{v_1,\ldots,v_k\}\subseteq V$ we say that $\mathbf{e}$ is an event of X, if $\mathbf{e}\in dom(v_1)\times\cdots\times dom(v_k)$. For an event $\mathbf{e}$ of X we write $\mathbf{e}(y)$ for the projection of $\mathbf{e}$ onto $Y\subseteq X$. We say that $\mathbf{e}=(\mathbf{e}_1,\ldots,\mathbf{e}_k)$ is X-*certain*, if $e_i\neq\mu$ for all $i=1,\ldots,k$.

A *probability model* over a finite set $V=\{v_1,\ldots,v_n\}$ of random variables is a pair (dom,P) where *dom* is a domain mapping that maps each v_i to a finite domain *dom*(v_i), and $P: dom(v_1)\times\cdots\times dom(v_n)\rightarrow[0,1]$ is a probability distribution having the Cartesian product of these domains as its sample space.

The expression $I(Y,Z\mid X)$ where X,Y and Z are disjoint subsets of V is called a *conditional independence* (CI) *statement* over V. The set X is called the *condition* of $I(Y,Z\mid X)$. If $XYZ=V$, we call $I(Y,Z\mid X)$ a *saturated* CI (SCI) statement. Let (dom,P) be a probability model over V. A CI statement $I(Y,Z\mid X)$ is said to *hold for* (dom,P) if for every certain event $\mathbf{x}$ of X, and for every event $\mathbf{y},\mathbf{z}$ of Y and Z, respectively,

$$P(\mathbf{y},\mathbf{z},\mathbf{x})\cdot P(\mathbf{x})=P(\mathbf{y},\mathbf{x})\cdot P(\mathbf{x},\mathbf{z}). \qquad (1)$$

Equivalently, (dom,P) is said to *satisfy* $I(Y,Z\mid X)$.

The satisfaction of $I(Y,Z\mid X)$ requires Equation 1 to hold for *certain* events $\mathbf{x}$ of X only. That is, we exclude events that condition on the missing marker, i.e. for $\mathbf{x}$, but not when the independence is judged for the missing marker, i.e. for $\mathbf{y}$ and $\mathbf{z}$. Indeed, the independence between an event $\mathbf{y}$ and an event $\mathbf{z}$ is conditional on the event $\mathbf{x}$. If there is no information about $\mathbf{x}$, then there should not be any requirement on the independence between $\mathbf{y}$ and $\mathbf{z}$.

Consider a simplified version of the classic burglary example. A *r(obbery)* hopefully sets off an *a(larm)*, hopefully causing *s(heldon)* or *b(atman)* to call security. The independence between *s* and *b*, given *r* and *a*, can be stated as the SCI statement $I(s,b\mid ar)$ over $V=\{b,a,r,s\}$. Assume all domains contain `true`, `false` and μ. Clearly, the independence between *Sheldon* calling in and *Batman* calling in should not be conditional on events for which the alarm went off and it is unknown whether a robbery took place. Events

$$\frac{}{I(V-X,\emptyset \mid X)}\ \text{(saturated trivial independence, } \mathcal{T}) \qquad \frac{I(Y,Z \mid X)}{I(Z,Y \mid X)}\ \text{(symmetry, } \mathcal{S})$$

$$\frac{I(ZW,Y \mid X) \quad I(Z,W \mid XY)}{I(Z,YW \mid X)}\ \text{(weak contraction, } \mathcal{C}) \qquad \frac{I(Y,ZW \mid X)}{I(Y,Z \mid XW)}\ \text{(weak union, } \mathcal{W})$$

Table 1: Axiomatization $\mathfrak{C}$ under Certainty

$$\frac{}{I(V-X,\emptyset \mid X)}\ \text{(saturated trivial independence, } \mathcal{T}) \qquad \frac{I(Y,Z \mid X)}{I(Z,Y \mid X)}\ \text{(symmetry, } \mathcal{S})$$

$$\frac{I(YY',ZZ' \mid X) \quad I(YZ,Y'Z' \mid X)}{I(YY'Z,Z' \mid X)}\ \text{(algebra, } \mathcal{A}) \qquad \frac{I(Y,ZW \mid X)}{I(Y,Z \mid XW)}\ \text{(weak union, } \mathcal{W})$$

Table 2: Axiomatization $\mathfrak{U}$ under Uncertainty

for which no information is available about a robbery are excluded from judging the independence between *Sheldon* and *Batman* calling in.

If we exclude μ from the domains, we recover the standard semantics of CI statements. We will use the phrase *under uncertainty* to indicate that the marker is present in every domain, and the phrase *under certainty* when μ is absent from the domains. The phrases are short for *under (un)certain semantics*.

Let $\mathcal{C}$ denote a class of CI statements, for example, SCI statements under uncertainty. Let $\Sigma \cup \{\varphi\}$ be a set of CI statements over V in $\mathcal{C}$. We say that Σ *implies* φ, denoted by $\Sigma \models \varphi$, if every probability model over V that satisfies every CI statement in Σ also satisfies the CI statement φ. The *implication problem for* $\mathcal{C}$ is to decide whether for any given V and any given set $\Sigma \cup \{\varphi\}$ over V in $\mathcal{C}$, $\Sigma \models \varphi$. For Σ we let $\Sigma^* = \{\varphi \in \mathcal{C} \mid \Sigma \models \varphi\}$ be the *semantic closure* of Σ, that is, the set of all CI statements implied by Σ. We use a syntactic approach to determine implied CI statements by inference rules of the form $\frac{\text{premise}}{\text{conclusion}}$, where inference rules without any premise are called axioms. An inference rule is *sound*, if the set of CI statements in the premise imply the CI statement in the conclusion. We let $\Sigma \vdash_{\mathfrak{R}} \varphi$ denote the *inference* of φ from Σ by the set $\mathfrak{R}$ of inference rules. That is, there is some sequence $\gamma = [\sigma_1, \ldots, \sigma_n]$ of CI statements such that $\sigma_n = \varphi$ and every σ_i is an element of Σ or results from an application of an inference rule in $\mathfrak{R}$ to some elements in $\{\sigma_1, \ldots, \sigma_{i-1}\}$. For Σ, let $\Sigma^+_{\mathfrak{R}} = \{\varphi \mid \Sigma \vdash_{\mathfrak{R}} \varphi\}$ be its *syntactic closure* under inferences by $\mathfrak{R}$. A set $\mathfrak{R}$ of inference rules is said to be *sound* (*complete*) for the implication of CI statements in class $\mathcal{C}$, if for every V and for every set Σ of CI statements over V in $\mathcal{C}$ we have $\Sigma^+_{\mathfrak{R}} \subseteq \Sigma^*$ ($\Sigma^* \subseteq \Sigma^+_{\mathfrak{R}}$). The (finite) set $\mathfrak{R}$ is said to be a (finite) *axiomatization* for the implication of CI statements in $\mathcal{C}$ if $\mathfrak{R}$ is both sound and complete.

For instance, Geiger and Pearl established the set $\mathfrak{C} = \{\mathcal{T}, \mathcal{S}, \mathcal{C}, \mathcal{W}\}$ from Table 1 as a finite axiomatization for the implication of SCI statements under certainty.

Under certainty the SCI statements $I(sb, r \mid a)$ and $I(s, b \mid ar)$ imply the SCI statement $I(s, br \mid a)$, as the soundness of the weak contraction rule $\mathcal{C}$ shows. That situation is quite different under uncertainty.

Lemma 1 *The weak contraction rule* ($\mathcal{C}$) *is not sound for the implication of SCI statements under uncertainty.*

Proof The probability distribution that assigns one half to each of the events $(a = 1, r = \mu, s = 1, b = 1)$ and $(a = 1, r = \mu, s = 0, b = 0)$ satisfies $I(sb, r \mid a)$ and $I(s, b \mid ar)$, but violates $I(s, br \mid a)$. ∎

Intuitively, $I(s, br \mid a)$ is not implied by $I(sb, r \mid a)$ and $I(s, b \mid ar)$ under uncertainty since we cannot find certain values for r in the distribution of the proof above that satisfy both $I(sb, r \mid a)$ and $I(s, b \mid ar)$. For examples, $(a = 1, r = 1, s = 1, b = 1)$ and $(a = 1, r = 1, s = 0, b = 0)$ satisfies $I(sb, r \mid a)$ and violates $I(s, b \mid ar)$; and $(a = 1, r = 0, s = 1, b = 1)$ and $(a = 1, r = 1, s = 0, b = 0)$ violates $I(sb, r \mid a)$ and satisfies $I(s, b \mid ar)$. These certain probability models show that $I(sb, r \mid a)$ does not imply $I(s, br \mid a)$ under certainty, and $I(s, b \mid ar)$ does not imply $I(s, br \mid a)$ under certainty, respectively.

Note that probability distributions, under uncertainty, can feature markers in one event but not in others, for any random variable. The example above is just a special case.

Lemma 1 means that $\mathfrak{C}$ does not axiomatize the implication of SCI statements under uncertainty. In particular, not all axioms for graph separation are also axioms of SCI statements under uncertainty, in contrast to the case of certainty (Geiger and Pearl 1990).

3 Axiomatic Characterization

We show that the set $\mathfrak{U}$ from Table 2 is sound and complete for the implication of SCI statements under uncertainty. The completeness argument uses special probability models where two events have probability one half. This insight will later be used to derive further characterizations.

The rules $\mathcal{T}$, $\mathcal{S}$, $\mathcal{A}$, and $\mathcal{W}$ are all sound for the implication of SCI statements under uncertainty. The key observation is that, for each rule, the condition in the conclusion contains the condition of each of its premises. If there is a probability model that violates the conclusion, then there is an event which is certain on the condition and violates Equation (1). Hence, the same event is certain on the conditions of all premises. The soundness of $\mathfrak{C}$ (Geiger and Pearl 1993) means that one of the premises is also violated. The key observation does not apply to the rule $\mathcal{C}$, in which the condition of the second premise is not contained in the condition of its conclusion.

Independence basis. For some V, some set Σ of SCIs over V, and some $X \subseteq V$ let $IDep_\Sigma(X) := \{Y \subseteq V - X \mid \Sigma \vdash_{\mathfrak{U}} I(Y, Z \mid X)\}$ denote the set of all $Y \subseteq V - X$

such that $I(Y, Z \mid X)$ can be inferred from Σ by $\mathfrak{U}$. The soundness of the algebra rule $\mathcal{A}$ implies that $(\mathit{IDep}_\Sigma(X), \subseteq, \cup, \cap, (\cdot)^\mathcal{C}, \emptyset, V - X)$ forms a finite Boolean algebra where $(\cdot)^\mathcal{C}$ maps a set W to its complement $V - XW$. Recall that an element $a \in P$ of a poset $(P, \sqsubseteq, 0)$ with least element 0 is called an *atom* of $(P, \sqsubseteq, 0)$ precisely when $a \neq 0$ and every element $b \in P$ with $b \sqsubseteq a$ satisfies $b = 0$ or $b = a$. Further, $(P, \sqsubseteq, 0)$ is said to be *atomic* if for every element $b \in P - \{0\}$ there is an atom $a \in P$ with $a \sqsubseteq b$. In particular, every finite Boolean algebra is atomic. Let $\mathit{IDepB}_\Sigma(X)$ denote the set of all atoms of $(\mathit{IDep}_\Sigma(X), \subseteq, \emptyset)$. We call $\mathit{IDepB}_\Sigma(X)$ the *independence basis* of X with respect to Σ.

Theorem 2 *Let Σ be a set of SCI statements. Then $\Sigma \vdash_\mathfrak{U} I(Y, Z \mid X)$ iff $Y = \bigcup \mathcal{Y}$ for some $\mathcal{Y} \subseteq \mathit{IDepB}_\Sigma(X)$.*

Proof Let $Y \in \mathit{IDep}_\Sigma(X)$. Since every element b of a Boolean algebra is the union over those atoms a with $a \subseteq b$ it follows that $Y = \bigcup \mathcal{Y}$ for $\mathcal{Y} = \{W \in \mathit{IDepB}_\Sigma(X) \mid W \subseteq Y\}$. Vice versa, let $Y = \bigcup \mathcal{Y}$ for some $\mathcal{Y} \subseteq \mathit{IDepB}_\Sigma(X)$. Since $I(W, W' \mid X) \in \Sigma^+_\mathfrak{U}$ holds for every $W \in \mathcal{Y}$, applications of the algebra rule result in $I(Y, Z \mid X) \in \Sigma^+_\mathfrak{U}$. ∎

For $\Sigma = \{I(sb, r \mid a), I(s, b \mid ar)\}$ the independence basis is $\mathit{IDepB}_\Sigma(a) = \{sb, r\}$. Hence, $\Sigma \not\models I(s, br \mid a)$ since s is not a union of members of the independence basis.

Completeness. Our completeness proof is based on the assumption that every domain contains at least two different events plus the marker. In some relevant scenarios, e.g. in causal reasoning, this assumption in regard to specific interventional distributions (upon the creation of certain functional constraints) does not hold, but this case lies outside the scope of this work (Tian and Pearl 2000).

Theorem 3 *The set $\mathfrak{U}$ is complete for the implication of SCI statements under uncertainty.*

Proof Let $\Sigma \cup \{I(Y, Z \mid X)\}$ be a set of SCIs over V, and suppose that $I(Y, Z \mid X)$ cannot be inferred from Σ using $\mathfrak{U}$. We will show that $I(Y, Z \mid X)$ is not implied by Σ. For this purpose, we will construct a probability model that satisfies all SCI statements of Σ, but violates $I(Y, Z \mid X)$.

Let $\mathit{IDepB}_\Sigma(X)$ be the disjoint union of $\{\{v\} \mid v \in X\}$ and $\{W_1, \ldots, W_k\}$, in particular $V = XW_1 \cdots W_k$. Since $I(Y, Z \mid X) \notin \Sigma^+_\mathfrak{U}$ we conclude by Theorem 2 that Y is not the union of some elements of $\mathit{IDepB}_\Sigma(X)$. Consequently, there is some $i \in \{1, \ldots, k\}$ such that $Y \cap W_i \neq \emptyset$ and $W_i - Y \neq \emptyset$ hold. For every $v \in V$ we define $\mathit{dom}(v) = \{\mathbf{0}, \mathbf{1}, \mu\}$. We define the following two events $\mathbf{e}_1$ and $\mathbf{e}_2$ of V. We define $\mathbf{e}_1(v) = \mathbf{0}$ iff $v \in XW_i$, and $\mathbf{e}_1(v) = \mu$ otherwise. We further define $\mathbf{e}_2(v) = \mathbf{0}$ iff $v \in X$, $\mathbf{e}_2(v) = \mathbf{1}$ iff $v \in W_i$, and $\mathbf{e}_2(v) = \mu$ otherwise. As probability measure we define $P(\mathbf{e}_1) = P(\mathbf{e}_2) = 0.5$. The construction implies that (dom, P) does not satisfy $I(Y, Z \mid X)$.

It remains to show that (dom, P) satisfies every SCI statement $I(S, T \mid R)$ in Σ. Suppose that for some value $\mathbf{r}$ of R, $P(\mathbf{r}) = 0$. Then Equation (1) will always be satisfied. If $P(\mathbf{r}, \mathbf{s}) = 0$ or $P(\mathbf{r}, \mathbf{t}) = 0$ for some event $\mathbf{r}$ of R, and for some event $\mathbf{s}$ of S or for some event $\mathbf{t}$ of T, then $P(\mathbf{r}, \mathbf{s}, \mathbf{t}) = 0$. Then Equation (1) is also satisfied. Suppose that for some event $\mathbf{r}$ of R, $P(\mathbf{r}) = 0.5$. If for some event $\mathbf{s}$ of S and for some event $\mathbf{t}$ of T, $P(\mathbf{r}, \mathbf{s}) = P(\mathbf{r}, \mathbf{t}) = 0.5$, then $P(\mathbf{r}, \mathbf{s}, \mathbf{t}) = 0.5$, too. It remains to consider the case where $\mathbf{r}$ is a certain event of R such that $P(\mathbf{r}) = 1$. In this case, the construction of the probability model tells us that $R \subseteq X$. Consequently, we can apply the weak union and symmetry rules to $I(S, T \mid R) \in \Sigma$ to infer $I(S - X, T - X \mid X) \in \Sigma^+_\mathfrak{U}$. It follows from Theorem 2 that $S - X$ and $T - X$ are each the union of elements from $\mathit{IDepB}_\Sigma(X)$. Suppose first that $W_i \subseteq S - X$. Then, we are either in the previous case where $P(\mathbf{r}, \mathbf{s}) = 0$ or $P(\mathbf{r}, \mathbf{t}) = 0$, or $P(\mathbf{r}, \mathbf{s}) = 0.5$, $P(\mathbf{r}, \mathbf{t}) = 1$ and $P(\mathbf{r}, \mathbf{s}, \mathbf{t}) = 0.5$. Otherwise, $W_i \subseteq T - X$. Then, we are either in the previous case where $P(\mathbf{r}, \mathbf{s}) = 0$ or $P(\mathbf{r}, \mathbf{t}) = 0$, or $P(\mathbf{r}, \mathbf{s}) = 1$, $P(\mathbf{r}, \mathbf{t}) = 0.5$ and $P(\mathbf{r}, \mathbf{s}, \mathbf{t}) = 0.5$. This concludes the proof. ∎

Recall, for $\Sigma = \{I(sb, r \mid a), I(s, b \mid ar)\}$ we had $\Sigma \not\models I(s, br \mid a)$ as $\mathit{IDepB}_\Sigma(a) = \{sb, r\}$. The probability model, according to the proof of Theorem 3, defined by

r	a	s	b	P
μ	`true`	`true`	`true`	0.5
μ	`true`	`false`	`false`	0.5

satisfies Σ, but violates φ.

Normalizing implication. A probability model (dom, P) over V is *special*, if for all $v \in V$, $\mathit{dom}(v) = \{\mathbf{0}, \mathbf{1}, \mu\}$, and there are two events of V such that $P(\mathbf{e}_1) = 0.5 = P(\mathbf{e}_2)$. Here, we sometimes write $(\mathit{dom}, \{\mathbf{e}_1, \mathbf{e}_2\})$ instead of (dom, P). The proof of Theorem 3 shows that it suffices to consider special probability models.

Corollary 4 *Let $\Sigma \cup \{\varphi\}$ be a set of SCI statements where φ is not implied by Σ. Then there is a special probability model* (dom, P) *that satisfies Σ and violates φ.* ∎

4 Logical Characterizations

We characterize implication by a propositional fragment using the concept of situations from (Levesque 1989). This complements the equivalence between the implication of SCI statements under certainty and the Boolean implication of the same fragment. We then establish an equivalence between instances of implication problems under uncertainty and sliced instances of implication problems under certainty.

Levesque's situations. Let L^* denote the *propositional language* over a finite set L of propositional variables, generated from negation $\neg$, conjunction $\wedge$ and disjunction $\vee$. Elements of L^* are also called formulae of L, and denoted by φ', ψ' or their subscripted versions. Sets of formulae are denoted by Σ'. We omit parentheses if we can.

Let L^ℓ denote the set of all literals over L, i.e., $L^\ell = L \cup \{\neg v' \mid v' \in L\}$. A *situation* of L is a total function $\omega : L^\ell \to \{\mathbb{F}, \mathbb{T}\}$ that does not map both a propositional variable $v' \in L$ and its negation $\neg v'$ to $\mathbb{F}$ (we must not have $\omega(v') = \mathbb{F} = \omega(\neg v')$ for any $v' \in L$).

A situation $\omega : L^\ell \to \{\mathbb{F}, \mathbb{T}\}$ of L can be lifted to a total function $\Omega : L^* \to \{\mathbb{F}, \mathbb{T}\}$. Assuming that φ' is in Negation Normal Form, this lifting is defined as follows:

- $\Omega(\varphi') = \omega(\varphi')$, if $\varphi' \in L^\ell$,
- $\Omega(\varphi' \vee \psi') = \mathbb{T}$ if and only if $\Omega(\varphi') = \mathbb{T}$ or $\Omega(\psi') = \mathbb{T}$,
- $\Omega(\varphi' \wedge \psi') = \mathbb{T}$ if and only if $\Omega(\varphi') = \mathbb{T}$ and $\Omega(\psi') = \mathbb{T}$.

A situation ω is a *model* of a set Σ' of L-formulae if and only if $\Omega(\sigma') = \mathbb{T}$ holds for every $\sigma' \in \Sigma'$. We say that Σ' *implies* an L-formula φ', denoted by $\Sigma' \models_L \varphi'$, if and only if every situation that is a model of Σ' is also a model of φ'.

Equivalences. Let $\phi : V \to L$ denote a bijection between a set V of random variables and the set $L = \{v' \mid v \in V\}$ of propositional variables. We extend ϕ to a mapping Φ from the set of SCI statements over V to the set L^*. For an SCI statement $I(Y, Z \mid X)$ over V, let $\Phi(I(Y, Z \mid X))$ denote

$$\bigvee_{v\in X} \neg v' \vee \left(\bigwedge_{v\in Y} v'\right) \vee \left(\bigwedge_{v\in Z} v'\right).$$

Disjunctions over zero disjuncts are $\mathbb{F}$ and conjunctions over zero conjuncts are $\mathbb{T}$. We will denote $\Phi(\varphi) = \varphi'$ and $\Phi(\Sigma) = \{\Phi(\sigma) \mid \sigma \in \Sigma\} = \Sigma'$.

For $\varphi = I(s, br \mid a)$ we have $\varphi' = \neg a' \vee s' \vee (b' \wedge r')$, and for $\Sigma = \{I(sb, r \mid a), I(s, b \mid ar)\}$ we have $\Sigma' = \{\neg a' \vee (s' \wedge b') \vee r', \neg a' \vee \neg r' \vee s' \vee b'\}$.

We will now show that for any set $\Sigma \cup \{\varphi\}$ of SCI statements over V there is a probability model $\pi = (dom, P)$ over V that satisfies Σ and violates φ if and only if there is a situation ω_π that is a model of Σ' but not a model of φ'. For arbitrary probability models π it is not obvious how to define the situation ω_π. However, Corollary 4 tells us that for deciding the implication problem $\Sigma \models \varphi$ it suffices to examine special probability models. For the special probability model $\pi = (dom, \{\mathbf{e}_1, \mathbf{e}_2\})$ let ω_π denote the following special situation of L:

$$\omega_\pi(v') = \begin{cases} \mathbb{T} & \text{, if } \mathbf{e}_1(v) = \mathbf{e}_2(v) \\ \mathbb{F} & \text{, otherwise} \end{cases} \quad \text{, and}$$

$$\omega_\pi(\neg v') = \begin{cases} \mathbb{T} & \text{, if } \mathbf{e}_1(v) = \mu = \mathbf{e}_2(v) \text{ or} \\ & \quad \mathbf{e}_1(v) \neq \mathbf{e}_2(v) \\ \mathbb{F} & \text{, otherwise} \end{cases} .$$

Recall that $\Sigma = \{I(sb, r \mid a), I(s, b \mid ar)\} \not\models I(s, br \mid a) = \varphi$ as the special probability model π defined by

r	a	s	b	P
μ	**0**	**0**	**0**	0.5
μ	**0**	**1**	**1**	0.5

satisfies Σ, but violates φ. The special situation where $\omega_\pi(r') = \mathbb{T} = \omega_\pi(\neg r')$, $\omega_\pi(\neg a') = \omega_\pi(s') = \omega_\pi(b') = \mathbb{F}$ is a model of $\Sigma' = \{\neg a' \vee (s' \wedge b') \vee r', \neg a' \vee \neg r' \vee s' \vee b'\}$, but not a model of $\varphi' = \neg a' \vee s' \vee (b' \wedge r')$.

The special situation becomes a Boolean interpretation under certainty. The following lemma justifies the definition of the special situation semantically.

Lemma 5 *Let $\pi = (\text{dom}, \{\mathbf{e}_1, \mathbf{e}_2\})$ be a special probability model over V, and let φ denote an SCI statement over V. Then π satisfies φ if and only if ω_π is a model of φ'.*

Proof Let $\varphi = I(Y, Z \mid X)$ and $\varphi' = \bigvee_{v\in X} \neg v' \vee (\bigwedge_{v\in Y} v') \vee (\bigwedge_{v\in Z} v')$. Suppose π satisfies φ. We show that ω_π is a model of φ'. Assume $\omega_\pi(\neg v') = \mathbb{F}$ for all $v \in X$. According to the special situation, $\mathbf{e}_1(v) = \mathbf{e}_2(v) \neq \mu$ for all $v \in X$. That means $P(\mathbf{e}_1(X)) = 1$. Suppose for some $v \in Y$, $\omega_\pi(v') = \mathbb{F}$. Then $\mathbf{e}_1(v) \neq \mathbf{e}_2(v)$ according to the special situation. Then $P(\mathbf{e}_1(XY)) = P(\mathbf{e}_1) = 0.5$. Since π satisfies φ, $P(\mathbf{e}_1(XZ)) = 1$. Hence, for every $v \in Z$, $\mathbf{e}_1(v) = \mathbf{e}_2(v)$. This means that for all $v \in Z$, $\omega_\pi(v') = \mathbb{T}$. Hence, ω_π is a model of φ'.

Suppose ω_π is a model of φ'. We show that π satisfies φ. That is, for every certain event $\mathbf{x}$ of X, and every event $\mathbf{y}$ and $\mathbf{z}$ over Y and Z, respectively, $P(\mathbf{x}) \cdot P(\mathbf{x}, \mathbf{y}, \mathbf{z}) = P(\mathbf{x}, \mathbf{y}) \cdot P(\mathbf{x}, \mathbf{z})$. We distinguish between a few cases.

Case 1. If $P(\mathbf{x}, \mathbf{y}) = 0$ or $P(\mathbf{x}, \mathbf{z}) = 0$, then $P(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 0$. For the remaining cases we can assume $P(\mathbf{x}, \mathbf{y}) > 0$ and $P(\mathbf{x}, \mathbf{z}) > 0$, i.e., also $P(\mathbf{x}) > 0$.

Case 2. Suppose that $P(\mathbf{x}) = 1$. Since $\mathbf{x}$ is certain over X, it follows that $\mathbf{e}_1(v) = \mathbf{e}_2(v) \neq \mu$ for all $v \in X$. Consequently, $\omega_\pi(v') = \mathbb{T}$ for all $v \in X$. Since ω_π is a model of φ' we conclude that $\omega_\pi(v') = \mathbb{T}$ for all $v \in Y$, or $\omega_\pi(v') = \mathbb{T}$ for all $v \in Z$. This, however, would mean that $P(\mathbf{x}, \mathbf{y}) = 1$ or $P(\mathbf{x}, \mathbf{z}) = 1$. Since φ is saturated, it follows that exactly one of $P(\mathbf{x}, \mathbf{y})$ and $P(\mathbf{x}, \mathbf{z})$ is 1, and the other 0.5. Consequently, $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ equals $\mathbf{e}_1$ or $\mathbf{e}_2$. Hence, $P(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 0.5$, too. It follows that π satisfies φ.

Case 3. Suppose that $P(\mathbf{x}) = 0.5$. Then $P(\mathbf{x}, \mathbf{y}) = 0.5 = P(\mathbf{x}, \mathbf{z})$. Then $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ equals $\mathbf{e}_1$ or $\mathbf{e}_2$, as $P(\mathbf{x})$ would have to be 1 otherwise. Hence, $P(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 0.5$. ∎

Corollary 4 and Lemma 5 allow us to establish the anticipated logical characterization under uncertainty.

Theorem 6 *Let $\Sigma \cup \{\varphi\}$ be a set of SCI statements over some set V of random variables, and let $\Sigma' \cup \{\varphi'\}$ denote the set of its corresponding formulae over L. Then $\Sigma \models \varphi$ if and only if $\Sigma' \models_L \varphi'$.*

Proof Suppose $\Sigma \models \varphi$ does not hold. Corollary 4 shows there is some special probability model π over V that satisfies Σ and violates φ. By Lemma 5, ω_π is a model of Σ' but not a model of φ'. Hence, $\Sigma' \models_L \varphi'$ does not hold.

Suppose $\Sigma' \models_L \varphi'$ does not hold. Then some situation ω over L is a model of Σ' but not a model of φ'. Define the following special probability model $\pi = (dom, \{\mathbf{e}_1, \mathbf{e}_2\})$ over V. For $v \in V$, let $dom(v) = \{\mathbf{0}, \mathbf{1}, \mu\}$. We now define $\mathbf{e}_1$ and $\mathbf{e}_2$ as follows. If $\omega(v') = \mathbb{T}$ and $\omega(\neg v') = \mathbb{F}$, then $\mu \neq \mathbf{e}_1(v) = \mathbf{e}_2(v) \neq \mu$. If $\omega(v') = \mathbb{T}$ and $\omega(\neg v') = \mathbb{T}$, then $\mathbf{e}_1(v) = \mu = \mathbf{e}_2(v)$. Finally, if $\omega(v') = \mathbb{F}$ and $\omega(\neg v') = \mathbb{T}$, then $\mu \neq \mathbf{e}_1(v) \neq \mathbf{e}_2(v) \neq \mu$. Since ω is not a model of φ', $\mathbf{e}_1 \neq \mathbf{e}_2$. Hence, $\omega_\pi = \omega$. By Lemma 5, π satisfies Σ but violates φ. Hence, $\Sigma \models \varphi$ does not hold. ∎

Equivalences to Boolean implication. For a set Σ of SCI statements over V, and some $X \subseteq V$ define the X-cut $\Sigma[X]$ of Σ as the set of all SCI statements $I(V, W \mid U) \in \Sigma$ where $U \subseteq X$, i.e., whose condition U is a subset of X.

Theorem 7 *Let $\Sigma \cup \{\varphi\}$ be a set of SCI statements over V, where $\varphi = I(Y, Z \mid X)$. Then $\Sigma \models I(Y, Z \mid X)$ under uncertainty if and only if $\Sigma[X] \models I(Y, Z \mid X)$ under certainty.*

Proof We show the equivalence in the logical setting, i.e., $\Sigma \models_L \varphi'$ if and only if $(\Sigma[X])' \models_{CL} \varphi'$ where $\models_{CL}$ denotes Boolean implication. Suppose $\tau : L \to \{\mathbb{T}, \mathbb{F}\}$ is a model of $(\Sigma[X])'$, but not a model of φ'. Consequently, $\tau(v') = \mathbb{T}$ for all $v \in X$, $\tau(v_0') = \mathbb{F}$ for some $v_0 \in Y$, and

$\tau(v_1') = \mathbb{F}$ for some $v_1 \in Z$. Define a situation $\omega : L^\ell \to \{\mathbb{T}, \mathbb{F}\}$ that is a model of Σ', but not a model of φ'. In fact, set $\omega(\neg v') = \mathbb{F}$ for all $v \in X$, $\omega(v_0') = \mathbb{F}$, $\omega(v_1') = \mathbb{F}$, and $\omega(v') = \mathbb{T} = \omega(\neg v')$ for all $v \in V - (X \cup \{v_0, v_1\})$. Note that every atom that is true under τ is true under ω, and every negated atom, not corresponding to a variable in X, is $\mathbb{T}$ under ω. Certainly, ω is not a model of φ'. For $\sigma \in \Sigma[X]$, ω is a model of σ' since the variables in X that are not in the condition of σ correspond to atoms interpreted as $\mathbb{T}$ under ω. For $\sigma \in \Sigma - \Sigma[X]$ there is some $v \in V - X$ that is part of the condition of σ and whose negated atom $\neg v'$ is interpreted $\mathbb{T}$ under ω.

Suppose there is some $\omega : L^\ell \to \{\mathbb{T}, \mathbb{F}\}$ that is a model of Σ', but not a model of φ'. Hence, $\omega(\neg v') = \mathbb{F}$ for all $v \in X$, $\omega(v_0') = \mathbb{F}$ for some $v_0 \in Y$, and $\omega(v_1') = \mathbb{F}$ for some $v_1 \in Z$. Define a truth assignment $\tau : L \to \{\mathbb{T}, \mathbb{F}\}$ that is a model of $(\Sigma[X])'$, but not a model of φ'. In fact, set $\tau(v_0') = \mathbb{F} = \tau(v_1')$, and $\tau(v') = \mathbb{T}$ for all $v \in V - \{v_0, v_1\}$. Again, every atom true under ω is true under τ. It follows that τ is not a Boolean model of φ'. For $\sigma \in \Sigma[X]$, every variable in X that is not in the condition of σ corresponds to an atom interpreted as $\mathbb{T}$. Therefore, τ is a model of σ'. ∎

For $\Sigma = \{I(sb, r \mid a), I(s, b \mid ar)\}$ we have $\Sigma[a] = \{I(sb, r \mid a)\}$ and the probability model π' defined by

r	a	s	b	P
0	**0**	**0**	**0**	0.5
0	**0**	**1**	**1**	0.5

satisfies $\Sigma[a]$, but violates $\varphi = I(s, br \mid a)$. The Boolean truth assignment where $\tau(r') = \tau(a') = \mathbb{T}$ and $\tau(s') = \mathbb{F} = \omega_\pi(b)$ is a model of $\Sigma[a]' = \{\neg a' \vee (s' \wedge b') \vee r'\}$, but not a model of $\varphi' = \neg a' \vee s' \vee (b' \wedge r')$.

5 Characterization by Data Dependencies

Let $\mathfrak{A} = \{\hat{v}_1, \hat{v}_2, \ldots\}$ be an infinite set of distinct symbols, called attributes. A *relation schema* is a finite non-empty subset R of $\mathfrak{A}$. Each attribute $\hat{v} \in R$ has an infinite domain $dom(\hat{v})$. In order to encompass missing values the domain of each attribute contains the null marker μ. The intention of μ is to mean "no information" (Lien 1982). A *tuple* over R is a function $t : R \to \bigcup_{\hat{v} \in R} dom(\hat{v})$ with $t(\hat{v}) \in dom(\hat{v})$ for all $\hat{v} \in R$. For $X \subseteq R$ let $t(X)$ denote the restriction of t to X. A *relation* r over R is a finite set of tuples over R. For a tuple t over R and a set $X \subseteq R$, t is said to be X-*total*, if for all $\hat{v} \in X$, $t(\hat{v}) \neq \mu$. A relation over R is a *total relation*, if it is R-total. According to (Lien 1982), a *multivalued dependency* (MVD) over R is a statement $X \twoheadrightarrow Y$ where $X, Y \subseteq R$. The MVD $X \twoheadrightarrow Y$ over R is satisfied by a relation r over R if and only if for all $t_1, t_2 \in r$ the following holds: if t_1 and t_2 are X-total and $t_1(X) = t_2(X)$, then there is some $t \in r$ such that $t(XY) = t_1(XY)$ and $t(X(R - Y)) = t_2(X(R - Y))$. Thus, the relation r satisfies $X \twoheadrightarrow Y$ when every X-total value determines the set of values on Y independently of the set of values on $R - Y$.

One may associate an SCI statement $\varphi = I(Y, Z \mid X)$ over V with an MVD $\hat{\varphi} = X \twoheadrightarrow Y$ over $R = \{\hat{v} \mid v \in V\}$; and a set Σ of SCI statements over V with the set $\hat{\Sigma} = \{\hat{\sigma} \mid \sigma \in \Sigma\}$. In the same way we did for SCI statements, one can prove that $\hat{\Sigma} \models \hat{\varphi}$ if and only if $\Sigma' \models_L \varphi'$. In particular, special probability models reduce to two-tuple relations (the same events as before, but without probabilities).

Theorem 8 *Let $\Sigma \cup \{\varphi\}$ be a set of SCI statement over V. Then $\Sigma \models \varphi$ if and only if $\hat{\Sigma} \models \hat{\varphi}$.* ∎

6 Algorithmic Characterization

We exploit Theorems 7 and 8 to derive that, for a set $\Sigma \cup \{\varphi\}$ of SCI statements over V, $\Sigma \models \varphi$ if and only if $\widehat{\Sigma[X]} \models \hat{\varphi}$, that is, every total relation over R that satisfies $\widehat{\Sigma[X]}$ also satisfies $\hat{\varphi}$. The latter problem has a nice algorithmic solution (Galil 1982).

Corollary 9 *Using the algorithm in (Galil 1982), the implication problem $\Sigma \models I(Y, Z \mid X)$ of the set $\Sigma \cup \{I(Y, Z \mid X)\}$ of SCI statements over V can be decided in time $\mathcal{O}(|\hat{\Sigma}| + \min\{k_{\widehat{\Sigma[X]}}, \log \bar{p}_{\Sigma[X]}\} \times |\widehat{\Sigma[X]}|)$. Herein, $|\hat{\Sigma}|$ denotes the total number of attributes occurring in $\hat{\Sigma}$, $k_{\widehat{\Sigma[X]}}$ denotes the cardinality of $\widehat{\Sigma[X]}$, and $\bar{p}_{\Sigma[X]}$ denotes the number of sets in IDepB$_{\Sigma[X]}(X)$ that have non-empty intersection with Y, respectively.* ∎

7 Conclusion and Future Work

We established characterizations of the implication problem for SCI statements in the presence of missing values. These include equivalences in terms of axioms, algorithms, classical and non-classical propositional as well as database logic, which all form counterparts of well-known results in the absence of missing values. Our findings hold under the robust interpretation of a missing value marker as "no information", and are consistent with approaches to sampling and structural zeros in statistics, and applicable and inapplicable null values in databases. We recommend to apply our results after known techniques have been exploited to reveal missing values. Our reasoning tools for SCI statements under uncertainty provide sound lower bounds on the opportunities to factorize joint probability distributions.

While the tools enable us to reason about independence in the presence of missing values, it is not obvious how they can be utilized for learning or inference with Bayesian networks. This should be studied in future work, including the relationship to d-separation (in latent projections) (Pearl 1988; Pearl and Verma 1991). The implication problem of general CI statements under uncertainty should be explored, and other fragments including marginal and stable CI statements. For multivalued dependencies the implication problem has also been studied when the underlying set of attributes is left undetermined, both in the presence (Link 2008; 2006) and absence (Biskup 1980; Hartmann, Link, and Schewe 2004; Biskup and Link 2012; Link 2012a) of missing values. For SCI statements this has only been done in the absence of missing values (Biskup, Hartmann, and Link 2012). Finally, one should consider to specify random variables as certain, i.e., missing values are not permitted to occur for certain random variables. Again, this idea has only been explored in the database framework yet (Hartmann and Link 2012).

References

Batista, G. E. A. P. A., and Monard, M. C. 2003. An analysis of four missing data treatment methods for supervised learning. *Applied Artificial Intelligence* 17(5-6):519–533.

Biskup, J., and Link, S. 2012. Appropriate inferences of data dependencies in relational databases. *Ann. Math. Artif. Intell.* 63(3-4):213–255.

Biskup, J.; Hartmann, S.; and Link, S. 2012. Probabilistic conditional independence under schema certainty and uncertainty. In *SUM*, 365–378.

Biskup, J. 1980. Inferences of multivalued dependencies in fixed and undetermined universes. *Theor. Comput. Sci.* 10(1):93–106.

Chickering, D. M., and Heckerman, D. 1997. Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables. *Machine Learning* 29(2-3):181–212.

Darwiche, A. 2009. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press.

Dawid, A. P. 1979. Conditional independence in statistical theory. *Journal of the Royal Statistical Society. Series B (Methodological)* 41(1):1–31.

de Waal, P. R., and van der Gaag, L. C. 2005. Stable independence in perfect maps. In *UAI*, 161–168.

Dempster, A.; Laird, N. M.; and Rubin, D. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* 39:139.

Fayyad, U. M.; Piatetsky-Shapiro, G.; and Smyth, P. 1996. From data mining to knowledge discovery in databases. *AI Magazine* 17(3):37–54.

Friedman, N. 1997. Learning belief networks in presence of missing values and hidden variables. In *ICML*, 125–133.

Galil, Z. 1982. An almost linear-time algorithm for computing a dependency basis in a relational database. *J. ACM* 29(1):96–102.

Geiger, D., and Pearl, J. 1990. Logical and algorithmic properties of independence and their application to Bayesian networks. *Ann. Math. Artif. Intell.* 2:165–178.

Geiger, D., and Pearl, J. 1993. Logical and algorithmic properties of conditional independence and graphical models. *The Annals of Statistics* 21(4):2001–2021.

Halpern, J. 2005. *Reasoning about uncertainty*. MIT Press.

Hartmann, S., and Link, S. 2012. The implication problem of data dependencies over SQL table definitions. *ACM Trans. Database Syst.* 37(2):Article 13.

Hartmann, S.; Link, S.; and Schewe, K.-D. 2004. Reasoning about functional and multi-valued dependencies in the presence of lists. In *FoIKS*, 134–154.

Lauritzen, S. 1995. The EM algorithm for graphical association models with missing data. *Computational Statistics and Data Analysis* 19:191–201.

Levesque, H. 1989. A knowledge-level account of abduction. In *IJCAI*, 1061–1067.

Lien, E. 1982. On the equivalence of database models. *J. ACM* 29(2):333–362.

Link, S. 2006. On the logical implication of multivalued dependencies with null values. In *CATS*, 113–122.

Link, S. 2008. On the implication of multivalued dependencies in partial database relations. *Int. J. Found. Comput. Sci.* 19(3):691–715.

Link, S. 2012a. Characterizations of multivalued dependency implication over undetermined universes. *J. Comput. Syst. Sci.* 78(4):1026–1044.

Link, S. 2012b. Propositional reasoning about saturated conditional independence. In *WoLLIC*, 257–267.

Lou, Q., and Obradovic, Z. 2012. Margin-based feature selection in incomplete data. In *AAAI*, 1040–1046.

Malvestuto, F. 1992. A unique formal system for binary decompositions of database relations, probability distributions, and graphs. *Inf. Sci.* 59(1-2):21–52.

Marlin, B. M.; Zemel, R. S.; Roweis, S. T.; and Slaney, M. 2011. Recommender systems, missing data and statistical model estimation. In *IJCAI*, 2686–2691.

Matúš, F. 1992. Ascending and descending conditional independence relations. In *Transactions of the 11th Prague Conference on Information Theory*, 189–200.

Niepert, M.; Van Gucht, D.; and Gyssens, M. 2008. On the conditional independence implication problem: A lattice-theoretic approach. In *UAI*, 435–443.

Niepert, M.; Van Gucht, D.; and Gyssens, M. 2010. Logical and algorithmic properties of stable conditional independence. *Int. J. Approx. Reasoning* 51(5):531–543.

Pearl, J., and Verma, T. 1991. A statistical semantics for causation. *Statistics and Computing* 2:91–95.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Franciso, U.S.A.: Morgan Kaufmann.

Saar-Tsechansky, M., and Provost, F. J. 2007. Handling missing values when applying classification models. *Journal of Machine Learning Research* 8:1623–1657.

Sagiv, Y.; Delobel, C.; Parker Jr., D. S.; and Fagin, R. 1981. An equivalence between relational data dependencies and a fragment of propositional logic. *J. ACM* 28(3):435–453.

Singh, M. 1997. Learning Bayesian networks from incomplete data. In *AAAI/IAAI*, 534–539.

Studený, M. 1992. Conditional independence relations have no finite complete characterization. In *Transactions of the 11th Prague Conference on Information Theory*, 377–396.

Tian, J., and Pearl, J. 2000. Probabilities of causation: Bounds and identification. *Ann. Math. Artif. Intell.* 28(1-4):287–313.

Zhang, S.; Jin, Z.; and Zhu, X. 2011. Missing data imputation by utilizing information within incomplete instances. *Journal of Systems and Software* 84(3):452–459.

Zhu, X.; Zhang, S.; Zhang, J.; and Zhang, C. 2007. Cost-sensitive imputing missing values with ordering. In *AAAI*, 1922–1923.