Probabilistic Logic with Conditional Independence Formulae[1]

Magdalena Ivanovska and **Martin Giese**[2]

Abstract. We investigate probabilistic propositional logic as a way of expressing and reasoning about uncertainty. In contrast to Bayesian networks, a logical approach can easily cope with incomplete information like probabilities that are missing or only known to lie in some interval. However, probabilistic propositional logic as described e.g. by Halpern [9], has no way of expressing conditional independence, which is important for compact specification in many cases. We define a logic with conditional independence formulae. We give an axiomatization which we show to be complete for the kind of inferences allowed by Bayesian networks, while still being suitable for reasoning under incomplete information.

1 Introduction

We report work carried out within the CODIO project on COllaborative Decision support for Integrated Operations. The goal of this project is to develop a decision support system for petroleum drilling operations. Decisions in drilling operations are characterized by a high degree of uncertainty (geology, reservoir size, possible equipment failure, etc.) and certain temporal aspects, like parts of the information about downhole happenings only being available after drilling fluid has been circulated up. In modern days, this is complemented by overwhelming amounts of sensor data, and the collaboration of several people on- and off-shore in decision making processes.

As one part of the CODIO project, we have designed a system based on a Bayesian network [11] model to provide assistance in operational decisions based on real-time sensor readings for a specific kind of typical situation [7]. While this effort was largely successful when tested on case data from a major oil company, the use of Bayesian networks as a modelling tool also had some shortcomings. It seems as if a logic-based approach to decision support would make it possible to resolve some of the issues:

- It was hard to elicit complete information about all probabilities required in the Bayesian network. Some of the values had to be guessed without feedback on the impact of those guesses. By contrast, it is typical for logic-based approaches that information irrelevant to a certain conclusion does not need to be given in a specification. If the specification is not precise enough to arrive at some conclusion, then no conclusion is inferred. For instance, knowing only that some value or probability lies within a certain range, instead of having a concrete number, is not a problem in a logical setting.
- The temporal aspects of the problem could be included in the description of the uncertainties by using temporal logic operators
- The fact that different people with different information are involved in the decision making could be taken into account by using a multi-modal logic of knowledge or belief. One could for instance reason about the necessary knowledge transfer needed to make all involved parties come to the same conclusion.

Though attempts have been made to cope with time, imprecise information, etc. in Bayesian networks, we believe that a logical approach can be a natural framework to merge many different aspects.

We are currently in the process of investigating logics to express decision problems, similarly to Influence Diagrams. The topic of this paper is a small but important part of this endeavor, namely the combination of quantitative reasoning about uncertainty with the qualitative reasoning about conditional independence which is central to approaches based on Bayesian networks, but which has received comparatively little attention from the logical side. More specifically, in contrast to the related work mentioned in Sect. 8 for instance, we consider a purely logical approach, where the whole specification takes the shape of logical formulae, and inference is done with an axiomatic system. This is important if the formalism is to be combined with temporal, epistemic, or other kinds of reasoning.

We begin in Sect. 2 by reviewing some results on probabilistic logic without independence statements. In Sect. 3, we introduce syntax and semantics of a logic with formulae expressing conditional independence. Sect. 4 briefly reviews the definition of Bayesian networks and shows how to transform them into sets of axioms in our logic. We then show two completeness results for this transformation, a semantical one in Sect. 5, and a syntactical one in Sect. 6. In Sect. 7, we discuss reasoning with conditional independence under incomplete information. An account of various related work is given in Sect. 8. We conclude with an outlook on future work in Sect. 9.

2 Probabilistic Propositional Logic

Let a set $\mathbf{P} = \{X_1, X_2, \ldots\}$ of propositional letters be given. Following Fagin, Halpern, and Megiddo ([4], see also [9]), we consider a probabilistic propositional logic obtained by augmenting the propositional logic over the alphabet $\mathbf{P}$ with *linear likelihood formulae*

$$a_1\ell(\varphi_1) + \cdots + a_k\ell(\varphi_k) \geq a \quad ,$$

where $a_1, \ldots, a_k, a$ are real numbers and $\varphi_1, \ldots, \varphi_k$ are *pure propositional formulae*, i.e. formulae which do not themselves contain likelihood formulae. The intention is that $\ell(\varphi)$ expresses the probability of φ being true, and the language allows expressing arbitrary linear relationships between such probabilities.

[1] A short version of this paper is published as an ECAI 2010 short paper.

[2] University of Oslo, Norway, email: magdalei/martingi@ifi.uio.no

This logic is interpreted over (simple, measurable)[3] *probability structures* $M = (W, \mu, \pi)$, where W is a set of possible worlds, μ is a probability measure that assigns a value in $[0,1]$ to any subset of W, and π is an interpretation function. To each element $w \in W$, the function π assigns a truth-value function $\pi_w : \mathbf{P} \to \{0,1\}$. The interpretation π_w is extended to arbitrary formulae in the usual way, where the interpretation of linear likelihood formulae is defined as follows:

> $\pi_w(a_1\ell(\varphi_1) + \cdots + a_k\ell(\varphi_k) \geq a) = 1$ iff $a_1\mu(\varphi_1^M) + \cdots + a_k\mu(\varphi_k^M) \geq a$, where $\varphi^M := \{w | \pi_w(\varphi) = 1\}$ for any formula φ.

Conditional likelihood formulae can be introduced as abbreviations as follows: $\ell(\varphi/\psi) \geq c$ is defined as $\ell(\varphi \wedge \psi) - c\ell(\psi) \geq 0$ and $\ell(\varphi/\psi) \leq c$ is defined as $\ell(\varphi \wedge \psi) - c\ell(\psi) \leq 0$. We also define $\ell(\varphi/\psi) = c$ to be an abbreviation for the conjunction of the previous two formulae. Note that linear combinations of conditional likelihood terms are not allowed.

Fagin et al. [4] give a sound and complete axiomatization consisting of the following axioms:

Prop All the substitution instances of tautologies in propositional logic;
QU1 $\ell(\varphi) \geq 0$;
QU2 $\ell(\top) = 1$;
QU3 $\ell(\varphi) = \ell(\varphi \wedge \psi) + \ell(\varphi \wedge \neg\psi)$, where φ and ψ are pure propositional formulae;
Ineq All substitution instances of valid linear inequality formulae.

and two inference rules:

MP From f and $f \Rightarrow g$ infer g.
QUGen From $\varphi \Leftrightarrow \psi$ infer $\ell(\varphi) = \ell(\psi)$.

From this axiomatization, an NP decision procedure for the logic can be derived.

What is *not* expressible in this logic is stochastic independence of formulae. In fact, it is not hard to see that any statement about independence leads to non-linear statements about probabilities. Fagin et al. discuss an extension of their formalism which includes *polynomial likelihood formulae*, and in which independence can easily be expressed. Unfortunately, this increases the complexity of the satisfiability problem to PSPACE.

To give a complete specification of a large probabilistic system, it is vital to be able to express conditional independence of variables. This is one of the main assets of Bayesian networks. This is our motivation for looking at other ways to include statements about independence in a probabilistic logic.

3 Probabilistic Propositional Logic with Independence Formulae

To the probabilistic propositional logic defined in the previous section, we add *conditional independence formulae* (CI-formulae)

$$I(\mathbf{X}_1, \mathbf{X}_2/\mathbf{X}_3) \quad ,$$

where $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$ are sets of propositional letters.

[3] Halpern and others define variants where not all events need to be measurable, and also variants where several separate subjective probability measures are dealt with. We consider neither of these here.

If any of the sets $\mathbf{X}_i$ is a singleton set, we will omit the braces around it, for example, we will write $I(A, \{B,C\}/D)$ instead of $I(\{A\}, \{B,C\}/\{D\})$. We denote the set of all formulae by $\mathbf{F}$.

Before assigning a semantics to CI-formulae, we recall the definition of conditional independence from probability theory:

Definition 1 *Given a probability space* (W, μ)*, we say that events A and B are* independent conditional on *an event C,* $I_\mu(A, B/C)$*, iff*

$$\mu(B \cap C) \neq 0 \quad \textit{implies} \quad \mu(A/C) = \mu(A/B \cap C)$$

or, equivalently,

$$\mu(A \cap C) \neq 0 \quad \textit{implies} \quad \mu(B/C) = \mu(B/A \cap C) \quad .$$

We extend this to sets of events $\mathbf{A}_i$ *by defining that* $I_\mu(\mathbf{A}_1, \mathbf{A}_2/\mathbf{A}_3)$ *iff* $I_\mu(B_1, B_2/B_3)$ *for all intersections* $B_i = \bigcap_{A \in \mathbf{A}_i} A^{(C)}$ *of possibly complemented events from* $\mathbf{A}_i$.

We then define the interpretation of CI-formulae in a structure $M = (W, \mu, \pi)$ in the following way:

$$\pi_w(I(\mathbf{X}_1, \mathbf{X}_2/\mathbf{X}_3)) = 1 \text{ iff } I_\mu(\mathbf{X}_1^M, \mathbf{X}_2^M/\mathbf{X}_3^M),$$

where $\mathbf{X}_i^M = \{X^M \mid X \in \mathbf{X}_i\}$.

We denote the logic given by this syntax and semantics by $\mathbf{L}$.

We will use the following terminology and notations:

A *literal* is a propositional letter (*positive literal*) or a negation of a propositional letter (*negative literal*).

Given a propositional letter X, an X*-literal* is either X or $\neg X$. An S*-atom* for some set $S \subseteq \mathbf{P}$ is a conjunction of literals containing one X-literal for each $X \in S$.

A formula f from the language $\mathbf{L}$ is *true* (*satisfied*) in the structure M, written $M \models f$, if $\pi_w(f) = 1$ for every $w \in W$. If S is a set formulae *satisfied* in the structure M, we denote that by $M \models S$.

We say that the set of formulae S *semantically entails* the formula f from the language $\mathbf{L}$, if $M \models f$ for every model with $M \models S$. We denote that by $S \models f$.

Note that according to the previous definitions, the conditional likelihood formula $\ell(\varphi/\psi) \geq c$, i.e. $\ell(\varphi \wedge \psi) - c\ell(\psi) \geq 0$, is true in M if and only if $\mu(\varphi^M \cap \psi^M) - c\mu(\psi^M) \geq 0$, which is equivalent to $\mu(\varphi^M/\psi^M) \geq c$ only in the case when $\mu(\psi^M) \neq 0$. If $\mu(\psi^M) = 0$, then $\mu(\varphi^M \cap \psi^M) = 0$ as well, and $\mu(\varphi^M \cap \psi^M) - c\mu(\psi^M) \geq 0$ is true for any real number c. Hence, if $\mu(\psi^M) = 0$, then for arbitrary c, $\ell(\varphi/\psi) \geq c$ is true in the structure, but it does not really mean that $\mu(\varphi^M/\psi^M) \geq c$, since the mentioned conditional likelihood is not even defined in that case. The same observation holds for the other two types of conditional likelihood formulae. The attentive reader will see in the following that this fits nicely with the role of conditional likelihoods for zero-probability conditions in Bayesian networks.

4 Expressing Bayesian Networks

In $\mathbf{L}$, it is possible to express the content of a Bayesian network using a similar amount of space. We will make this precise in the following definitions. To keep the presentation as simple as possible, we restrict ourselves to Bayesian networks where each variable can only have two states. Instead of talking about variables, we talk about events, identified by propositional letters. Our results could be generalized to variables with many states by using one propositional letter per state and adding some more axioms.

Definition 2 *A* binary Bayesian network *BN is a pair* (G, f) *where* $G = (V,E)$ *is a DAG (directed acyclic graph) whose nodes* $V \subseteq \mathbf{P}$ *are propositional letters. We denote by* $Pa(X)$ *resp.* $ND(X)$ *the sets of parents resp. non-descendents of* $X \in V$ *in G. f is a function that associates with each node X in G a cpt (conditional probability table) which contains an entry* $f(X)(\delta) \in [0,1]$ *for each* $Pa(X)$*-atom* δ*.*

A cpt entry of a node X gives the probability of (the degree of belief in) the proposition X, under the assumption that the $Pa(X)$-atom corresponding to that entry takes a true value.

Given a binary Bayesian network BN with nodes V, let W_{BN} be the set of all V-atoms, and define $\pi_w(X) = 1$ iff X occurs positively in w. In other words, W_{BN} has exactly one world for each combination of the propositions in BN being true or false. We define a probability measure μ_{BN} on W_{BN} by defining for any elementary outcome $w \in W_{BN}$:

$$\mu_{BN}(\{w\}) = \prod_{X \text{ pos. in } w} f(X)(w_{Pa(X)}) \cdot \prod_{X \text{ neg. in } w} (1 - f(X)(w_{Pa(X)})) \quad (1)$$

where $w_{Pa(X)}$ is the unique $Pa(X)$-atom consisting of a subset of the conjuncts of w. It is easy to see that this really defines a probability measure on W_{BN}. Let $M_{BN} = (W_{BN}, \mu_{BN}, \pi_{BN})$.

Definition 3 *A DAG* $G = (V,E)$ qualitatively represents *or* is qualitatively compatible *with a probability structure* $M = (W, \mu, \pi)$ *if the* Markov condition $I_\mu(X^M, ND(X)^M / Pa(X)^M)$ *is satisfied for every node* $X \in V$*.*

For any network $BN = (G, f)$, it can be shown that μ_{BN} as defined in (1) satisfies the Markov condition, i.e. that G qualitatively represents M_{BN}.

Definition 4 *The Bayesian network BN* quantitatively represents *or* is quantitatively compatible *with a probability structure* $M = (W, \mu, \pi)$*, if BN qualitatively represents M and the cpts agree with* μ*: for each node* $X \in V$ *and* $Pa(X)$*-atom* δ *with* $\mu(\delta^M) \neq 0$*,* $\mu(X^M / \delta^M) = f(X)(\delta)$*. (It does not matter what* $f(X)(\delta)$ *is when* $\mu(\delta^M) = 0$*.)*

Due to definition (1), M_{BN} is also *quantitatively represented* by $BN = (G, f)$ in this sense. The usual formulation of Bayesian networks with random variables leads to the result that a Bayesian network quantitatively represents a *unique* probability distribution (See e.g. [9]). A corresponding result for our formulation with events is the following:

Lemma 1 *Any models* $M = (W, \mu, \pi)$ *and* $M' = (W', \mu', \pi')$ *quantitatively represented by a Bayesian network BN agree in the probabilities assigned to any combination of events described by letters from V, i.e.* $\mu(\delta^M) = \mu'(\delta^{M'})$ *for all V-atoms* δ*.*

This means that in a given network $BN = (G, f)$, the graph G captures qualitative information about a probability distribution: existence of independencies between events (absence of arrows) and the cpts present the quantitative information about the strength of the probabilistic dependencies.

Both qualitative and quantitative information embedded in the network can be appropriately represented in our logical language $\mathbf{L}$.

For example, if a network consists of the nodes A, B, and C, and edges (B,A) and (C,A) and the cpt contains the following information: $f(B) = b_1$, $f(C) = c_1$ $f(A)(B \wedge C) = a_1$, $f(A)(B \wedge \neg C) = a_2$, $f(A)(\neg B \wedge C) = a_3$ and $f(A)(\neg B \wedge \neg C) = a_4$, then this information can be given in our language in the following way: $\ell(A/B \wedge C) = a_1$, $\ell(A/B \wedge \neg C) = a_2$, $\ell(A/\neg B \wedge C) = a_3$ and $\ell(A/\neg B \wedge \neg C) = a_4$.

The qualitative information from the graph structure of the network is represented by conditional independence formulae. In the case of the previous example we will have: $I(A, \{B,C\}/\{B,C\})$, $I(B, C/\emptyset)$ and $I(C, B/\emptyset)$.

We generalize the whole idea with the following definition:

Definition 5 *Let* $\mathbf{BN}$ *be the class of all binary Bayesian networks and* $\mathbf{F}$ *be the set of all formulae in* $\mathbf{L}$*. The* specific axioms function*,* $Ax : \mathbf{BN} \to 2^{\mathbf{F}}$ *is a function that to each Bayesian network* $BN = (G, f)$ *with* $G = (V,E)$ *assigns the set of formulae containing*

- $\ell(X/\delta) = c$ *for every node* $X \in V$ *and every* $Pa(X)$*-atom* δ *such that* $f(X)(\delta) = c$*, and*
- $I(X, ND(X)/Pa(X))$ *for every node* $X \in V$*.*

5 Theorem for Semantic Entailment

Bayesian networks capture only the probabilities of combinations of certain events identified by the nodes, and not a probability measure on some underlying set of worlds or elementary outcomes. Also the formulae in our logic only describe certain properties of models. It is easy to show however that the inferences possible from a Bayesian network are also logical consequences of the network's set of specific axioms.

By inferences in BNs, we mean the computation of probabilities of arbitrary conjunctions of literals, conditional on other conjunctions of literals. Due to Lemma 1, such a conditional probability follows for *one* model quantitatively represented by the Bayesian network iff it follows for *all* such models.

Theorem 1 *Let BN be a Bayesian network with nodes V and M a probability structure quantitatively represented by BN. If the formula* $\ell(\varphi/\psi) = b$*, where* φ *and* ψ *are conjunctions of literals of letters from V, is satisfied in M, then it is entailed by the specific axioms of BN,* $Ax(BN) \models \ell(\varphi/\psi) = b$*.*

Proof: Let $M = (W, \mu, \pi)$ be quantitatively represented by BN, and $M \models \ell(\varphi/\psi) = b$. It is sufficient to show that $M' \models \ell(\varphi/\psi) = b$ for an arbitrary but fixed structure $M' = (W', \mu', \pi')$ with $M' \models Ax(BN)$.

The axioms in $Ax(BN)$ are a direct encoding of the conditions for qualitative and quantitative representation, so it is easy to see that BN also quantitatively represents M'. Due to Lemma 1, we have $\mu(\delta^M) = \mu'(\delta^{M'})$ for all V-atoms δ.

For an arbitrary conjunction ξ of literals of letters from V and any structure M'',

$$\xi^{M''} = \bigcup_{\substack{\delta \text{ a } V\text{-atom} \\ \xi \text{ a sub-conjunction of } \delta}} \delta^{M''}$$

and since the sets $\delta^{M''}$ are disjoint, additivity of the probability measure gives us:

$$\mu''(\xi^{M''}) = \sum_{\substack{\delta \text{ a } V\text{-atom} \\ \xi \text{ a sub-conjunction of } \delta}} \mu''(\delta^{M''})$$

for any probability measure μ''. Together with Lemma 1 it follows in particular that $\mu(\varphi^M \cap \psi^M) = \mu'(\varphi^{M'} \cap \psi^{M'})$ and $\mu(\psi^M) = \mu'(\psi^{M'})$.

Now $M \models \ell(\varphi/\psi) = b$, i.e. $\mu(\varphi^M \cap \psi^M) - c\mu(\psi^M) = 0$ and therefore also $\mu'(\varphi^{M'} \cap \psi^{M'}) - c\mu'(\psi^{M'}) = 0$, i.e. $M' \models \ell(\varphi/\psi) = b$. **Q.E.D.**

While Theorem 1 guarantees the semantic connection between a Bayesian network BN and its axiomatization $Ax(BN)$ in our logic $\mathbf{L}$, it says nothing about the derivability of conditional likelihood formulae. This will be covered in the following section.

6 Axiomatic System and Theorem for Syntactic Entailment

The axiomatic system given in Sect. 2 is complete for reasoning in a logic without CI-formulae. To accommodate those, we need to add some extra axioms. The axiomatization of conditional independence has been the subject of a certain amount of research, see e.g. [6] for a survey. In particular, it has been shown (see [12] and [10]) that there is no finite complete axiomatization of conditional independence, if the language contains nothing but CI statements. On the other hand, Fagin et al. [5] show that complete reasoning about conditional independence is possible if one is willing to pay the price of reasoning about polynomials.

In this work, we are less interested in deriving new CI formulae. We want to mimic the kind of reasoning possible with a Bayesian network, so we want a system that allows to derive arbitrary statements about conditional likelihood of propositional formulae from the specific axioms $Ax(BN)$ of a Bayesian network.

We will use an axiomatic system that consists of four parts, each dealing with a different type of reasoning: propositional reasoning, reasoning about probability, reasoning about linear inequalities, and reasoning about conditional independence. For the first three parts, we use the axioms from the system AX_{MEAS} in [5], as given in Sect. 2. For conditional independence reasoning, we add the following inference rules:

SYM From $I(\mathbf{X}_1,\mathbf{X}_2/\mathbf{X}_3)$ infer $I(\mathbf{X}_2,\mathbf{X}_1/\mathbf{X}_3)$
DEC From $I(\mathbf{X}_1,\mathbf{X}_2\cup\mathbf{X}_3/\mathbf{X}_4)$ infer $I(\mathbf{X}_1,\mathbf{X}_2/\mathbf{X}_4)$
IND From $I(\mathbf{X}_1,\mathbf{X}_2/\mathbf{X}_3)$ and $\ell(\varphi_1/\varphi_3)\le(\ge)a$ infer $\ell(\varphi_1/\varphi_2\wedge\varphi_3)\le(\ge)a$, where φ_i is an arbitrary $\mathbf{X}_i$-atom, for $i\in\{1,2,3\}$.

We say that a set of formulae S *syntactically entails* a formula f in this axiomatic system (including the axioms and rules from Sect. 2) if f can be derived from S by using the given axioms and inference rules. We denote this by $S\vdash f$.

It can be checked that this system is sound, in the sense that only semantically valid entailments can be inferred: If $S\vdash f$, then $S\models f$.

Before we state our restricted completeness theorem, we show how to use the calculus by proving the following proposition:

Proposition 1 *Let $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ be sets of propositional letters. Then we can derive the following syntactic entailments:*

a) $\{I(\mathbf{X}_1,\mathbf{X}_2/\mathbf{X}_3),\ \ell(\varphi_1/\varphi_3)=a,\ \ell(\varphi_2/\varphi_3)=b\}\vdash\ell(\varphi_1\wedge\varphi_2/\varphi_3)=ab$
b) $\{I(\mathbf{X}_1,\mathbf{X}_2/\mathbf{X}_3),\ \ell(\varphi_1/\varphi_3)=a,\ \ell(\varphi_3/\varphi_2)=b\}\vdash\ell(\varphi_1\wedge\varphi_3/\varphi_2)=ab$

where φ_i are arbitrary $\mathbf{X}_i$-atoms for $i\in\{1,2,3\}$.

Proof: a) We have the following derivation:

1. $I(\mathbf{X}_1,\mathbf{X}_2/\mathbf{X}_3),\ell(\varphi_1/\varphi_3)=a$ .... (premises)
2. $\ell(\varphi_1/\varphi_2\wedge\varphi_3)=a$ .... (1 and IND)
3. $\ell(\varphi_1\wedge\varphi_2\wedge\varphi_3)=a\ell(\varphi_2\wedge\varphi_3)$ ...... (2, def. of cond. likel., Ineq)
4. $\ell(\varphi_2/\varphi_3)=b$ .... (premise)
5. $\ell(\varphi_2\wedge\varphi_3)=b\ell(\varphi_3)$ .... (4, def. of cond. likelihood, Ineq)
6. $\ell(\varphi_1\wedge\varphi_2\wedge\varphi_3)=ab\ell(\varphi_3)$ .... (3, 5, Ineq)
7. $\ell(\varphi_1\wedge\varphi_2/\varphi_3)=ab$ .... (6, def. of conditional likelihood)

b) The derivation is as in a) until step 3, and then

4. $\ell(\varphi_3/\varphi_2)=b$ .... (premise)
5. $\ell(\varphi_3\wedge\varphi_2)=b\ell(\varphi_2)$ .... (4, def. of cond. likelihood, Ineq)
6. $\ell(\varphi_1\wedge\varphi_2\wedge\varphi_3)=ab\ell(\varphi_2)$ .... (3, 5, Ineq, Prop)
7. $\ell(\varphi_1\wedge\varphi_3/\varphi_2)=ab$ .... (6, def. of conditional likelihood)

Q.E.D.

The following lemma is required for the proof of the restricted completeness theorem:

Lemma 2 *Let φ, ψ and ν be pure propositional formulae. Then the following syntactic entailment holds: $\{\ell(\varphi\wedge\nu/\psi)=a_1,\ell(\varphi\wedge\neg\nu/\psi)=a_2\}\vdash\ell(\varphi/\psi)=a_1+a_2$.*

Proof: We have the following derivation steps:

1. $\ell(\varphi\wedge\nu/\psi)=a_1$ .... (premise)
2. $\ell(\varphi\wedge\nu\wedge\psi)=a_1\ell(\psi)$ .... (1, def. of cond. likelihood)
3. $\ell(\varphi\wedge\neg\nu/\psi)=a_2$ .... (premise)
4. $\ell(\varphi\wedge\neg\nu\wedge\psi)=a_2\ell(\psi)$ .... (3, def. of cond. likelihood)
5. $\ell(\varphi\wedge\psi)=\ell(\varphi\wedge\nu\wedge\psi)+\ell(\varphi\wedge\neg\nu\wedge\psi)$ .... (QU3)
6. $\ell(\varphi\wedge\psi)=a_1\ell(\psi)+a_2\ell(\psi)$ .... (3, 4, 5, Ineq)
7. $\ell(\varphi\wedge\psi)=(a_1+a_2)\ell(\psi)$ .... (6, Ineq)
8. $\ell(\varphi/\psi)=a_1+a_2$ .... (7, def. of cond. likelihood)

Q.E.D.

We are now ready to state our limited completeness theorem for the derivation of conditional likelihoods from $Ax(BN)$.

Theorem 2 *Let $BN=(G,f)$ be a Bayesian Network. If the formula $\ell(\varphi/\psi)=b$, where φ and ψ are conjunctions of literals of letters from V, is satisfied in models quantitatively represented by BN, then $Ax(BN)\vdash\ell(\varphi/\psi)=b$.*

Proof: We prove this by deriving $Ax(BN)\vdash\ell(\varphi\wedge\psi)=m$, $\ell(\psi)=n$, where $\ell(\varphi\wedge\psi)=m$ and $\ell(\psi)=n$ are formulae that are satisfied in every model of $Ax(BN)$. If $n=0$, then also $m=0$, and therefore $\ell(\varphi\wedge\psi)-b\ell(\psi)=0$ is an instance of Ineq independently of b. Otherwise, $b=m/n$, and the conclusion follows from the following derivation steps:

1. $\ell(\varphi\wedge\psi)=m$, $\ell(\psi)=n$
2. $\ell(\varphi\wedge\psi)-(m/n)\ell(\psi)=0$ .... (1 and Ineq)
3. $\ell(\varphi/\psi)=m/n$ .... (2, def. of cond. likelihood)

We will describe the procedure of deriving $\ell(\psi)=n$. A similar one can be used for the derivation of $\ell(\varphi\wedge\psi)=m$.

Let $\mathbf{Y}=\{Y_1,Y_2,\dots,Y_s\}$ be ancestors in G of all the propositional letters occurring in ψ. Applying QU3 and Ineq, we perform the following derivation steps:

1. $\ell(\psi)=\ell(\psi\wedge Y_1)+\ell(\psi\wedge\neg Y_1)$
2. $\ell(\psi\wedge Y_1)=\ell(\psi\wedge Y_1\wedge Y_2)+\ell(\psi\wedge Y_1\wedge\neg Y_2)$
3. $\ell(\psi\wedge\neg Y_1)=\ell(\psi\wedge\neg Y_1\wedge Y_2)+\ell(\psi\wedge\neg Y_1\wedge\neg Y_2)$
4. $\ell(\psi)=\ell(\psi\wedge Y_1\wedge Y_2)+\ell(\psi\wedge Y_1\wedge\neg Y_2)+\ell(\psi\wedge\neg Y_1\wedge Y_2)+\ell(\psi\wedge\neg Y_1\wedge\neg Y_2)$

...

j. $\ell(\psi)=\sum_{\delta\text{ is a }\mathbf{Y}\text{-atom}}\ell(\psi\wedge\delta)$

Let $\psi\wedge\delta=Z_1'\wedge\dots\wedge Z_k'$ be one of the conjuncts in the last step, where $Z_i'=Z_i$ or $Z_i'=\neg Z_i$, for some propositional letters $Z_1,\dots,Z_k$.

We put the Z_is into a topological order with respect to G, i.e. if Z_i is a descendent in G of Z_j, then $i > j$.

If we can derive a formula $\ell(Z'_1 \wedge \cdots \wedge Z'_k) = c$ such that $\ell(Z'_1 \wedge \cdots \wedge Z'_k) = c$ is true in every model of $Ax(BN)$, for all conjuncts, then we have derived $\ell(\psi) = n$ as desired. To do this, we start with Z_k. All of $Z_1, \ldots, Z_{k-1}$ are non-descendants of Z_k, hence, by application of DEC to $I(Z_k, ND(Z_k)/Pa(Z_k))$, we obtain that $I(Z_k, \{Z_1, \ldots, Z_{k-1}\}/Pa(Z_k))$ holds. Since $Z'_1 \wedge \cdots \wedge Z'_k$ is a full conjunction (together with a propositional letter, it contains all of its ancestors), all the parents of Z_k are in $\{Z_1, \ldots, Z_{k-1}\}$. Let us denote by $\varphi_{Pa(Z_k)}$ the $Pa(Z_k)$-atom which is a sub-conjunction of $Z'_1 \wedge \cdots \wedge Z'_k$. Then we proceed with the following derivation:

1. $I(Z_k, \{Z_1, \ldots, Z_{k-1}\}/Pa(Z_k)),\ \ell(Z'_k/\varphi_{Pa(Z_k)})) = z_k$
   (If $Z'_k = Z_k$, then $z_k = a_k$, where $\ell(Z_k/\varphi_{Pa(Z_k)}) = a_k$ is a premise. If $Z'_k = \neg Z_k$, then, by using Lemma 2 and Ineq, we obtain $\ell(Z'_k/\varphi_{Pa(Z_k)}) = 1 - a_k$, from the premise $\ell(Z_k/\varphi_{Pa(Z_k)}) = a_k$)
2. $\ell(Z'_k/\varphi_{Pa(Z_k)} \wedge Z'_1 \wedge \cdots \wedge Z'_{k-1}) = z_k$ .... (1 and IND)
3. $\ell(Z'_k/Z'_1 \wedge \cdots \wedge Z'_{k-1}) = z_k$ .... (2, Prop, QUGen)
4. $\ell(Z'_1 \wedge \cdots \wedge Z'_{k-1} \wedge Z'_k) = z_k \ell(Z'_1 \wedge \cdots \wedge Z'_{k-1})$ .... (3 and the definition of conditional likelihood)

This process can be inductively repeated for the letters $Z_1, \ldots, Z_{k-1}$ to obtain:

5. $\ell(Z'_1 \wedge \cdots \wedge Z'_{k-1}) = z_{k-1} \ell(Z'_1 \wedge \ldots \wedge Z'_{k-2})$
6. $\ell(Z'_1 \wedge \cdots \wedge Z'_{k-1} \wedge Z'_k) = z_k z_{k-1} \ell(Z'_1 \wedge \cdots \wedge Z'_{k-2})$ ... (4, 5, Ineq)

for a number $z_{k-1} = \ell(Z'_{k-1}/\varphi_{Pa(Z_{k-1})})$, obtained similarly like z_k, and so on inductively, until we obtain

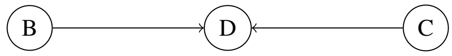
$$\ell(Z'_1 \wedge \cdots \wedge Z'_{k-1} \wedge Z'_k) = z_k z_{k-1} \ldots z_1 \quad ,$$

where the $z_i = \ell(Z'_i/\varphi_{Pa(Z_i)})$ are directly determined by the axioms in $Ax(BN)$. This is easily seen to be in direct correspondence to the product given in (1), so we derived $\ell(Z'_1 \wedge \cdots \wedge Z'_k) = c$ for the same c that can be computed from the Bayesian network. **Q.E.D.**

7 Beyond Bayesian Networks

So far, we have proved that using our language we can completely imitate the reasoning in Bayesian networks, and derive the same probabilistic statements as from a BN. To do this, we worked only with formulae expressing conditional probabilities of conjunctions of literals, and we did not make use of the possibility to express inequalities in our language. In our logic, it is easy to use complex formulae in axioms and conclusions to be derived, which can lead to more natural formulations in some cases. One can show that our axiomatization is complete also for conditional likelihoods of arbitrary pure propositional formulae, since these can be computed from the conditional likelihoods of conjunctions of literals.

Expressing inequalities for conditional probabilities can also be very useful. For instance, in many contexts in modeling using Bayesian networks, it is very hard to elicit the numbers in the cpts, which determine the probability distribution expressed by the network. In some cases, it may be possible to give a range of the cpt entry (or, not to give information about it at all, in which case the range will be [0,1]), instead of a precise value. Reasoning with such a range is then preferable to using an imprecise guess. In Bayesian network research, *credal networks* [2] have been developed for this purpose. Our logic **L** in its current form can be used to express imprecise information about probabilities by using conditional likelihood inequality formulae. The following example shows that our axioms can be used to infer bounds on conditional probabilities, but not necessarily tightest bounds. In other words, the axiomatization is not complete for the combination of inequality reasoning and conditional independence. In future work, we hope to be able to extend our axiom system to be complete also for the inference of tightest bounds on conditional inequalities.



Example: Let us consider BN_3 with nodes B, C and D with "head to head" placed arrows at D and such that the cpts of this network provide the following information:

$$f(B) \in [b_1, b_2] \qquad f(C) \in [c_1, c_2]$$
$$f(D)(B \wedge C) = d_1 \qquad f(D)(B \wedge \neg C) = d_2$$
$$f(D)(\neg B \wedge C) = d_3 \qquad f(D)(\neg B \wedge \neg C) = d_4$$

We translate the given input information about BN_3 into the logical language in the following way:

$$\ell(B) \geq b_1, \quad \ell(B) \leq b_2, \quad \ell(C) \geq c_1, \quad \ell(C) \leq c_2$$
$$\ell(D/B \wedge C) = d_1 \qquad \ell(D/B \wedge \neg C) = d_2$$
$$\ell(D/\neg B \wedge C) = d_3 \qquad \ell(D/\neg B \wedge \neg C) = d_4$$
$$I(B, C/\emptyset), \qquad I(C, B/\emptyset), \qquad I(D, \{B, C\}/\{B, C\})$$

To derive a probability statement of type $\ell(D) \geq d$, for some d, we follow a similar procedure as described in the theorem 2, using axioms and inference rules for linear likelihood inequality formulae:

1. $\ell(D) = \ell(D \wedge B \wedge C) + \ell(D \wedge B \wedge \neg C) + \ell(D \wedge \neg B \wedge C) + \ell(D \wedge \neg B \wedge \neg C)$
2. $\ell(D \wedge B \wedge C) = d_1 \ell(B \wedge C)$
3. $\ell(D \wedge B \wedge \neg C) = d_2 \ell(B \wedge \neg C)$
4. $\ell(D \wedge \neg B \wedge C) = d_3 \ell(\neg B \wedge C)$
5. $\ell(D \wedge \neg B \wedge \neg C) = d_4 \ell(\neg B \wedge \neg C)$
6. $\ell(D) = d_1 \ell(B \wedge C) + d_2 \ell(B \wedge \neg C) + d_3 \ell(\neg B \wedge C) + d_4 \ell(\neg B \wedge \neg C)$

For each $\ell((\neg)B \wedge (\neg)C)$ we can derive two bounds that can be expressed as *linear* likelihood formulae:

7. $\ell(B \wedge C) \geq \ell(B) c_1$
8. $\ell(B \wedge C) \geq b_1 \ell(C)$
9. $\ell(B \wedge \neg C) \geq \ell(B)(1 - c_2)$
10. $\ell(B \wedge \neg C) \geq b_1 \ell(\neg C)$
11. $\ell(\neg B \wedge C) \geq \ell(\neg B) c_1$
12. $\ell(\neg B \wedge C) \geq (1 - b_2) \ell(C)$
13. $\ell(\neg B \wedge \neg C) \geq \ell(\neg B)(1 - c_2)$
14. $\ell(\neg B \wedge \neg C) \geq (1 - b_2) \ell(\neg C)$

Depending on $d_1, \ldots, d_4$, these can be used to derive different lower bounds for $\ell(D)$.

On the other hand, for any structure $M = (W, \mu, \pi)$, such that $M \models Ax(BN)$, we have:

$$\mu(D^M) = d_1 \mu(B^M) \mu(C^M) + d_2 \mu(B^M) \mu((\neg C)^M) + d_3 \mu((\neg B)^M) \mu(C^M) + d_4 \mu((\neg B)^M) \mu((\neg C)^M)$$

Assuming for instance that $d_1 < d_2, d_3 < d_4$, it is not hard to see that this expression has a greatest lower bound of

$$\mu(D^M) \geq d_1 b_2 c_2 + d_2 b_2 (1 - c_2) + d_3 (1 - b_2) c_2 + d_4 (1 - b_2)(1 - c_2)$$

which is larger than any bound that can be obtained as a sum of the individual bounds in formulae 7 to 14.

8 Related work

We are certainly not the first who attempt to combine the expressiveness and the inference mechanisms of probabilistic logics and probabilistic networks. There are several other lines of work with the same or similar goals. Haenni et al. [8] use inference networks (Bayesian networks as well as credal networks [2] for the case with incomplete information) as a calculus for probabilistic logic in general – they provide a network model for a set of (linear inequality) probabilistic statements that describe the problem domain and then make inference using Bayesian and credal network inference algorithms. Cozman et al. [3] also propose inference with imprecise probabilities, but they represent the independence statements by introducing a graph-theoretic model called probabilistic propositional logic network consisting of a DAG and linear inequality probabilistic statements corresponding to its nodes; for inferencing, they construct a credal counterpart of the DAG. In both mentioned approaches the independence structure is captured by a qualitative BN, which means that independence statements are implied by missing edges, according to the Markov condition. We prefer explicitly stating independence statements in the logic because that requires conscious thought while building a specification. In a graph representation, a missing arrow that expresses a conditional independence might be missing by accident. Another advantage of this representation could be the possibility of expressing conditional independencies of a probability distribution that can not be entailed by any graph.

The work on 'Bayesian logic' by Andersen and Hooker [1] is perhaps that which is most closely related to ours. The independence statements there are encoded in a standard probabilistic logic by using polynomial likelihood formulae that come from the definition of independence and Bayes' rule; these formulae together with formulae that come from the cpts and some general probability constraints are used as a basis for a linear programming inference to infer the tightest range for a probability in question. They also go beyond Bayesian networks, considering probabilistic inequality statements. Note that the representation of independence in polynomial likelihood formulae has also been suggested by Fagin and Halpern [4].

What these approaches have in common, and what separates them from our work is that we do not recur to existing Bayesian or credal network algorithms, or linear or non-linear programming for inference. These surely give efficient means of reasoning within their respective limits, but we were interested in a purely logical characterization, i.e. not only are all aspects of a specification expressed as formulae with a standard model semantics, but also inference is done using an axiomatic system. While this surely doesn't give the most efficient reasoning, we consider the sufficiency of a relatively small axiomatic system to be interesting in its own right. And the axiomatic approach should make it easier to combine the logic with epistemic, temporal, and other logics.

9 Conclusion and Future Work

A logical approach to reasoning about uncertainty has a number of advantages over approaches based on Bayesian networks, including the ease of accommodating incomplete knowledge, and the possibility of combining uncertainty reasoning with logics for reasoning about time, knowledge of multiple agents, etc.

Conditional independence is a crucial element for the compact and efficient representation of a system of uncertainties. This is naturally expressed in Bayesian networks, but not usually a part of logics dealing with probabilities. We have introduced a probabilistic logic with conditional independence formulae that allows representing the information in a Bayesian network in a comparable amount of space. We have discussed the difficulties of complete quantitative reasoning in the presences of conditional independence, but we have given an axiom system and shown it to be complete for a small class of problems, including at least the conditional probability statements derivable from a Bayesian network.

Topics for future work include the extension of the calculus to obtain completeness also for ranges of probabilities. It would also be interesting to investigate efficient, maybe approximate, reasoning algorithms for Bayesian networks and try to recast them into an optimized axiom system or other logical calculus, allowing more efficient inference in practice.

Introducing the conditional independence statements directly in the language, and the choice of CI axioms we have made also allows for other choices of semantics, for example, model structures with a possibility instead of a probability measure, or with a more general belief measure, so we consider that as another possibility for extending the scope of our work.

We are currently in the process of extending the given logic with a mechanism for expressing decision scenarios, including options and utilities, which will make it possible to infer optimal strategies from a set of observations.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers of STAIRS and ECAI 2010 for many useful suggestions on previous versions of this paper. The CODIO project is partially funded by grants from the Research Council of Norway, Petromaks Project 175899/S30, 2007–2010.

REFERENCES

[1] K. A. Andersen and J. N. Hooker, 'Bayesian logic', *Decis. Support Syst.*, **11**(2), 191–210, (1994).
[2] F. G. Cozman, 'Credal networks', *Artif. Intell.*, **120**(2), 199–233, (2000).
[3] F. G. Cozman, C. P. de Campos, and J. C. Ferreira da Rocha, 'Probabilistic logic with independence', *Int. J. Approx. Reasoning*, **49**(1), 3–17, (2008).
[4] R. Fagin and J. Y. Halpern, 'Reasoning about knowledge and probability', in *Proc. TARK '88*, ed., Moshe Y. Vardi, pp. 277–293, (1988).
[5] R. Fagin, J. Y. Halpern, and N. Megiddo, 'A logic for reasoning about probabilities', *Information and Computation*, **87**, 78–128, (1990).
[6] D. Geiger and J. Pearl, 'Logical and algorithmic properties of conditional independence and graphical models', *Ann. Statist.*, **21**(4), 2001–2021, (1993).
[7] M. Giese and R. B. Bratvold, 'Probabilistic modelling for decision support in drilling operations, SPE 127761', in *Proc. SPE Intelligent Energy Conference*, (2010).
[8] R. Haenni, J. W. Romeijn, G. Wheeler, and J. Williamson. Probabilistic logic and probabilistic networks, December 2008. Draft.
[9] Joseph Y. Halpern, *Reasoning about Uncertainty*, MIT Press, 2003.
[10] Sanjiang Li, 'Causal models have no complete axiomatic characterization', *CoRR*, **abs/0804.2401**, (2008).
[11] R. E. Neapolitan, *Learning Bayesian Networks*, Prentice Hall, 2003.
[12] Milan Studený, 'Conditional independence relations have no finite complete characterization', in *11th Prague Conf. on Information Theory, Statistical Decision Functions and Random Processes*, volume B, pp. 377–396. Kluwer, (1992).