ELSEVIER

Artificial Intelligence 117 (2000) 297–308

Artificial Intelligence

www.elsevier.com/locate/artint

Research Note

# On the complexity of inference about probabilistic relational models

Manfred Jaeger [1]

*Max-Planck-Institut für Informatik, Im Stadtwald, 66123 Saarbrücken, Germany*

Received 1 July 1999; received in revised form 16 November 1999

**Abstract**

We investigate the complexity of probabilistic inference from knowledge bases that encode probability distributions on finite domain relational structures. Our interest here lies in the complexity in terms of the domain under consideration in a specific application instance. We obtain the result that assuming NETIME $\neq$ ETIME this problem is not polynomial for reasonably expressive representation systems. The main consequence of this result is that it is unlikely to find inference techniques with a better worst-case behavior than the commonly employed strategy of constructing standard Bayesian networks over ground atoms (knowledge based model construction). © 2000 Elsevier Science B.V. All rights reserved.

*Keywords:* Knowledge based model construction; Bayesian networks; First-order logic

## 1. Introduction

A recent development in probabilistic reasoning in AI is the emergence of various systems for the specification of probability distributions on relational structures, or, in the terminology of Friedman et al. [5], the construction of *probabilistic relational models* (Ngo and Haddawy [17], Jaeger [9], Koller and Pfeffer [13]). These systems have evolved out of earlier frameworks that were developed as specification languages for structurally uniform classes of Bayesian networks (Poole [19], Breese [2], Saffiotti and Umkehrer [20]). Given a particular probabilistic query, a specification in such a language would serve as the blueprint for the automatic generation of a Bayesian network in which the probability of the query then is computed. This method has been called *knowledge based model construction* (Wellman, Breese and Goldman [21]).

[1] Email: jaeger@mpi-sb.mpg.de.


PII: S0004-3702(99)00109-5

Initially, only representation languages were considered that are based on some form of probabilistic Horn clauses. Ignoring many particular features present in the representation languages proposed by various authors, these probabilistic Horn clauses are essentially of the form

$$\begin{aligned} &\mathsf{p}(u,v) \xleftarrow{0.3} \mathsf{q}(u,v), \\ &\mathsf{p}(u,v) \xleftarrow{0.5} \mathsf{r}(u), \mathsf{s}(v), \end{aligned} \tag{1}$$

where $\mathsf{p}, \mathsf{q}, \mathsf{r}, \mathsf{s}$ are relation symbols, and $u, v$ logical variables. The intuitive meaning of, e.g., the first clause is: for all $u, v$, the conditional probability of $\mathsf{p}(u, v)$ given that $\mathsf{q}(u, v)$ holds is 0.3. Given constants $a, b$ for which we have evidence $\mathsf{q}(a, b)$ the rule allows us to compute a posterior probability of 0.3 for $\mathsf{p}(a, b)$ [2] (if (1) expresses statistical knowledge, this computation would be an instance of *direct inference*, cf. Bacchus [1]). This coincides with the interpretation of similar rules in certain probabilistic logics (Ng and Subrahmanian [16], Lakshmanan and Sadri [14]). The difference between knowledge based model construction and its outgrowths on the one hand, and probabilistic logics on the other, emerges when we consider conditional probabilities that are not fully determined by the rules. The conditional probability

$$\rho := P\big(\mathsf{p}(a,b) | \mathsf{q}(a,b), \mathsf{r}(a), \mathsf{s}(b)\big),$$

for instance, is not defined by either of the rules in (1). Moreover, only the trivial bounds $[0, 1]$ are strictly implied for $\rho$ by instantiations of the rules (1) with $a, b$. In most probabilistic logics, therefore, one will be unable to derive from (1) any nontrivial bounds for $\rho$.

In purely propositional settings, Bayesian networks have proven to be more useful in practice than propositional probabilistic logics (Nilsson [18], Frisch and Haddawy [6]) because they define a unique probability distribution on the set of propositional models (i.e., truth assignments), and therefore (at the cost of a greater specification effort) allow us to derive a unique probability value for every query. It is natural to extend this approach to certain forms of first-order probabilistic information, and to develop tools for defining probability distributions on models for first-order logic. In knowledge based model construction this is done by interpreting the probabilistic rules (1) as rules for the construction of standard Bayesian networks over ground atoms. Given a ground query $P(\mathsf{p}(a,b) | \mathsf{q}(a,b), \mathsf{r}(a), \mathsf{s}(b)) =?$ the model construction will yield a Bayesian network containing nodes for the atoms $\mathsf{p}(a,b), \mathsf{q}(a,b), \mathsf{r}(a), \mathsf{s}(b)$ (and possibly a large number of additional nodes), and thereby determine a unique value for $P(\mathsf{p}(a,b) | \mathsf{q}(a,b), \mathsf{r}(a), \mathsf{s}(b))$. As noted above, the intended semantics of the rules (1) alone will not uniquely determine the desired probabilities, so that it is clear that at some point additional information or assumptions—not directly expressed by (1)—must enter the construction process. Essentially, these additional assumptions have to determine how

[2] As long as no constant symbols appear in the rules, the same will be true for constants $c, d$ for which we have the same evidence $\mathsf{q}(c, d)$. Most concrete representation systems provide for constants in the rules, so that the probabilities entailed by the rules are not necessarily invariant under substituting different constants. For the purpose of the present paper we may focus on rules without constants, because our main result is a lower complexity bound, which, obviously, also is applicable to richer systems admitting constant symbols.

the conditional probabilities in clauses with the same head are to be combined to obtain the conditional probability of the head given the conjunction of the bodies of the various clauses.

In early approaches (Breese [2]) this information was supplied implicitly by certain implementation details of the construction algorithm, and consequently the primary representation language did not possess a declarative semantics independent from the network construction process. Haddawy [7] and Ngo and Haddawy [17] have argued that this is unsatisfactory, and have proposed representation systems with additional syntactic constructs that in the knowledge base declare how several applicable clauses are to be combined.

Relational Bayesian networks (Jaeger [9,10]) can be understood as a representation formalism that goes one step further by compiling sets of clauses (1), and the necessary additional conventions for their combination, into a single functional expression $F$, so that the knowledge base now consists of exactly one declaration of the form

$$\mathsf{r}(\boldsymbol{v}) := F(\mathsf{s}_1, \ldots, \mathsf{s}_k, \boldsymbol{v}), \tag{2}$$

for each relation symbol $\mathsf{r}$ (we use boldface letters $\boldsymbol{v}, \boldsymbol{a}, \ldots$ as abbreviations for tuples $(v_1, \ldots, v_k), (a_1, \ldots, a_l), \ldots$ of variables or constants). Formal semantics for this set of declarations then can be defined in a straightforward manner. Another related framework that uses a representation language different from probabilistic Horn clauses are the probabilistic frame-based systems of Koller and Pfeffer [13].

Once one has taken the step to supply the primary representation formalism with descriptive semantics independent from any construction algorithm for standard Bayesian networks, the question arises whether standard Bayesian networks are still needed at all. Their role now has changed from being the subject of our primary representation to being merely a tool of inference: if there were more efficient ways to compute the answer to a probabilistic query than by constructing a Bayesian network over ground atoms, we would be happy to dispense with Bayesian networks altogether. To emphasize this shift of perspective, we refer as *auxiliary network construction* to the process of constructing standard Bayesian networks as an inference technique for representation languages with independent semantics.

It does not seem to be unreasonable to expect more efficient inference techniques than auxiliary network construction to exist, because this approach amounts to a complete "propositionalization" of first-order information. For logic inference problems from (deterministic) Horn-clauses we know that we can avoid this, and, for example, by unification and resolution deduce from

$$\mathsf{p}(v) \leftarrow \mathsf{q}(v, w)$$
$$\mathsf{q}(a, u) \leftarrow$$

that $\mathsf{p}(a)$ holds, without first constructing all the ground atoms $\mathsf{p}(c), \mathsf{q}(c, c'), \ldots$ for all constants $c, c', \ldots$ in the language.

It is natural to look for corresponding techniques for probabilistic inference from first-order probabilistic rules like (1) or (2)—techniques that compute probabilities by manipulating more abstract logical expressions than ground atoms. In this paper we show that it is very unlikely that with such algorithms we can obtain inference techniques that are more efficient than auxiliary network construction.

## 2. Model representation systems

It is our aim to derive our complexity results in as general terms as possible, showing their applicability to a great variety of different representation systems. In order to do this, we have to abstract from the concrete syntactical constructs used in various systems, and analyze these systems in terms of their semantic expressiveness.

To achieve this goal we develop in this section the general concept of a *probabilistic model representation system*, which (very loosely) can be seen as specialized counterpart of the general concept of a logic. Just as different logics can be compared, and their complexity be analyzed, by considering the classes of models they can define, we derive results for model representation systems in terms of the class of models they can describe, where models now are probability distributions.

First, we have to describe the structure of the models that are defined by the representation systems we deal with. To motivate the following definition, consider again the case of probabilistic Horn clauses as the representation language. It is clear that, e.g., the semantics of knowledge base (1) will be used to assign probability values to sentences such as $\mathsf{p}(a,b) \wedge \mathsf{s}(b)$. However, it is not enough to say that the semantics of a knowledge base is given by a probability distribution over sentences: to see why, consider the two rules

$$\mathsf{p}(v) \xleftarrow{0.5} \mathsf{q}(v,u)$$
$$\mathsf{q}(v,u) \xleftarrow{0.8}$$

where the second clause means that the marginal probability of $\mathsf{q}(v,u)$ is 0.8. Also assume that the semantical conventions adopted (perhaps via some additional declarations in the knowledge base) make the conditional probability of $\mathsf{p}(a)$ increase in the number of valid instantiations for $u$ in $\mathsf{q}(a,u)$. In particular, we would have

$$P\big(\mathsf{p}(a)|\mathsf{q}(a,b)\big) < P\big(\mathsf{p}(a)|\mathsf{q}(a,b),\mathsf{q}(a,c)\big).$$

But more than that, since each possible instantiation of $u$ in $\mathsf{q}(a,u)$ has a positive probability of 0.8 of being valid, the probability of $\mathsf{p}(a)$ should also increase in the number of possible instantiations, whether or not they appear in the evidence. Thus, the probability of $\mathsf{p}(a)$ as defined by the given rules, can only be determined with respect to a certain (finite) domain $D$ of elements that we can substitute for $v$ and $u$.

Consequently, the semantics of a knowledge base does not consist of a single probability distribution over sentences, but of one distribution for each (finite) domain $D$. Formally, a probability distribution on sentences containing relation symbols from a vocabulary $S$ and constants from $D$ is most conveniently represented by a distribution on the set of all structures (or models) that interpret the symbols in $S$ over $D$. We denote the set of these structures by $\mathrm{Mod}_D(S)$. As the particular names of the elements of $D$ should be irrelevant, we may restrict attention to the case where $D = n = \{0,1,\ldots,n-1\}$ for some $n \in \mathbb{N}$ (we here avail ourselves of the set-theoretic convention to identify the number $n \in \mathbb{N}$ with the set $\{0,\ldots,n-1\}$).

**Definition 1.** A *finite domain probabilistic relational model representation system* **M** consists of

- A syntax that defines for every relational vocabulary $S = \{\mathtt{r}_1, \ldots, \mathtt{r}_k\}$ a set $M(S)$ of well-formed *model representations.*
- A semantics that assigns to every $\Phi \in M(S)$ and every $n \in \mathbb{N}$ a probability measure $P_n^{\Phi}$ on $\mathrm{Mod}_n(S)$.

Note that a probabilistic model representation system differs from a *probabilistic logic* in that it is required that every model representation $\Phi$ defines for every $n$ a unique measure on $\mathrm{Mod}_n(S)$, whereas a theory $\Phi$ in a probabilistic logic will usually define a (possibly empty) set of such measures. It should also be noted that most existing systems are somewhat more general than described in Definition 1 in that they allow for a set $R$ of predefined, deterministic relations on the domain, so that the semantics maps $R$-structures over $n$ to probability measures over $S$-expansions of the $R$-structure.

In the sequel we simply write "model representation system" and "vocabulary" for "finite domain probabilistic relational model representation system" and "relational vocabulary", respectively.

Next we describe minimal requirements for the expressiveness of model representation systems. Our complexity results will hold for those systems that satisfy these requirements.

The first requirement is very simple: we should be able to represent the uniform distribution on $\mathrm{Mod}_n(S)$. The second requirement is to have the ability to condition the probability that $\boldsymbol{v}$ belongs to some relation $\mathtt{r}$ on certain logical properties of $\boldsymbol{v}$ with respect to other relations $\mathtt{s}_1, \mathtt{s}_2, \ldots$. In a clause based representation language, for instance, this requirement will demand the availability of rules of the form

$$\mathtt{r}(\boldsymbol{v}) \xleftarrow{p} \alpha(\boldsymbol{v}, \mathtt{s}_1, \ldots, \mathtt{s}_k), \tag{3}$$

where $\alpha(\boldsymbol{v}, \mathtt{s}_1, \ldots, \mathtt{s}_k)$ is some logical expression in the variables $\boldsymbol{v}$ and the relation symbols $\mathtt{s}_1, \ldots, \mathtt{s}_k$. Our minimal requirement will be that rules of this form are available for $\alpha$ being an equality constraint $v_i = v_j$, a conjunction $\mathtt{s}_1(\boldsymbol{v}') \wedge \mathtt{s}_2(\boldsymbol{v}'')$, a negated atom $\neg\mathtt{s}(\boldsymbol{v})$, and an existentially quantified atom $\exists w \mathtt{s}(\boldsymbol{v}, w)$. Of these types of rules only the case of $\alpha$ being a conjunction is readily recognized as being provided by existing systems for knowledge based model construction. Existential quantification, on the other hand, might look like a rather strong assumption about a system's expressiveness. It should be noted, however, that a rule like

$$\mathtt{r}(v) \xleftarrow{1} \mathtt{s}(v, w)$$

together with the common convention that multiple instantiations of the right hand side of a rule are to be combined by *noisy-or*, just amounts to existential quantification.

The following definition formulates the availability of rules like (3) in syntax-independent, general semantic terms. In this definition, and in the remainder of the paper, we need some notation for restrictions of structures to sub-vocabularies, and restrictions of measures to sub-algebras: when $\mathcal{M}' \in \mathrm{Mod}_n(S')$ and $S \subseteq S'$, then $\mathcal{M}' \upharpoonright S$ denotes the $S$-structure over domain $n$ that has the same interpretations of the symbols in $S$ as $\mathcal{M}'$. Conversely, a structure $\mathcal{M} \in \mathrm{Mod}_n(S)$ can be identified with the subset $\{\mathcal{M}' | \mathcal{M}' \upharpoonright S = \mathcal{M}\} \subseteq \mathrm{Mod}_n(S')$. When $P'$ is a probability measure on $\mathrm{Mod}_n(S')$, then $P' \upharpoonright \mathrm{Mod}_n(S)$

denotes the probability measure $P$ on $\mathrm{Mod}_n(S)$ defined by $P(\mathcal{M}) = P'(\{\mathcal{M}' | \mathcal{M}' \upharpoonright S = \mathcal{M}\})$, and $P'(\cdot|\mathcal{M})$ denotes the conditional distribution on $\mathrm{Mod}_n(S')$ given $\{\mathcal{M}' | \mathcal{M}' \upharpoonright S = \mathcal{M}\}$. We use the notation $\boldsymbol{v}' \subseteq \boldsymbol{v}$ to express that all variables in the tuple $\boldsymbol{v}'$ are variables that also appear in $\boldsymbol{v}$.

**Definition 2.** A model representation system **M** *allows first-order conditioning* if

- For every vocabulary $S$ there exists $\Phi \in M(S)$ such that $P_n^{\Phi}$ is the uniform distribution on $\mathrm{Mod}_n(S)$ for all $n \in \mathbb{N}$.
- For every vocabulary $S$, every $\Phi \in M(S)$, every $k$-ary $\mathtt{r} \notin S$, for $\boldsymbol{v} := (v_1, \ldots, v_k)$, and for every expression $\alpha(\boldsymbol{v})$ of one of the four forms
  - $v_i = v_j$ $(1 \leqslant i, j \leqslant k)$,
  - $\mathtt{s}_1(\boldsymbol{v}') \wedge \mathtt{s}_2(\boldsymbol{v}'')$ $(\mathtt{s}_1, \mathtt{s}_2 \in S;\ \boldsymbol{v}', \boldsymbol{v}'' \subseteq \boldsymbol{v})$,
  - $\neg \mathtt{s}(\boldsymbol{v}')$ $(\mathtt{s} \in S,\ \boldsymbol{v}' \subseteq \boldsymbol{v})$,
  - $\exists w \mathtt{s}(\boldsymbol{v}', w)$ $(\mathtt{s} \in S,\ \boldsymbol{v}' \subseteq \boldsymbol{v})$,

  there exists $\Phi^{\alpha} \in M(S \cup \{\mathtt{r}\})$, such that $\forall n \in \mathbb{N}$:

$$P_n^{\Phi^{\alpha}} \upharpoonright \mathrm{Mod}_n(S) = P_n^{\Phi}, \tag{4}$$

  and for all $n \in \mathbb{N}$, all $\mathcal{M} \in \mathrm{Mod}_n(S)$, and all $\boldsymbol{m} \in n^k$:

$$P_n^{\Phi^{\alpha}}\big(\mathtt{r}(\boldsymbol{m}) | \mathcal{M}\big) = \begin{cases} 1 & \text{if } \mathcal{M} \models \alpha(\boldsymbol{m}), \\ 0 & \text{if } \mathcal{M} \not\models \alpha(\boldsymbol{m}). \end{cases} \tag{5}$$

The conditions of Definition 2 demand that the probability of $\mathtt{r}(\boldsymbol{v})$ can be conditioned on very simple logical properties of $\boldsymbol{v}$. For relational Bayesian networks it is straightforward to show that, in fact, $\mathtt{r}(\boldsymbol{v})$ can be conditioned on arbitrary first-order expressible properties of $\boldsymbol{v}$ [9]. For systems based on probabilistic Horn rules, on the other hand, it is not so obvious that with rules (3) for simple formulas $\alpha$ we can also encode more complicated conditions like

$$\mathtt{r}(\boldsymbol{v}) \xleftarrow{1} \neg \exists w\big(\mathtt{s}_1(\boldsymbol{v}', w)\big) \vee \mathtt{s}_2(\boldsymbol{v}''). \tag{6}$$

The following lemma, which is instrumental to the proof of Theorem 4 in the next section, shows that the elementary requirements of Definition 2 are sufficient to guarantee that rules like (6) can be encoded.

**Lemma 3.** *Let* **M** *be a model representation system that allows first-order conditioning. Let $\Phi$ be a model representation for a vocabulary $S$, $\mathtt{r} \notin S$ a $k$-ary relation symbol, and $\phi(\boldsymbol{v})$ a first-order $S$-formula whose free variables are among $\boldsymbol{v} = (v_1, \ldots, v_k)$. Then there exists a model representation $\Phi^{\phi}$ for a vocabulary $S^{\phi} \supseteq S \cup \{\mathtt{r}\}$, such that for all $n$*

$$P_n^{\Phi^{\phi}} \upharpoonright \mathit{Mod}_n(S) = P_n^{\Phi}, \tag{7}$$

*and for all $n \in \mathbb{N}$, all $\mathcal{M} \in \mathit{Mod}_n(S)$, and all $\boldsymbol{m} \in n^k$:*

$$P_n^{\Phi^{\phi}}\big(\mathtt{r}(\boldsymbol{m}) | \mathcal{M}\big) = \begin{cases} 1 & \text{if } \mathcal{M} \models \phi(\boldsymbol{m}), \\ 0 & \text{if } \mathcal{M} \not\models \phi(\boldsymbol{m}). \end{cases} \tag{8}$$

*In particular, for all $\boldsymbol{m} \in n^k$:*

$$P_n^{\Phi^{\phi}}\big(\mathtt{r}(\boldsymbol{m})\big) = P_n^{\Phi}\big(\phi(\boldsymbol{m})\big). \tag{9}$$

**Proof.** First note that (7) and (8) directly imply (9). We prove the existence of $\Phi^{\phi}$ with (7) and (8) by induction on the structure of $\phi$.

First, assume that $\phi$ is of the form $\mathtt{s}(\boldsymbol{v}')$ for some $\mathtt{s} \in S$, $\boldsymbol{v}' \subseteq \boldsymbol{v}$. Then the lemma follows from the case $\alpha(\boldsymbol{v}) \equiv \mathtt{s}(\boldsymbol{v}') \wedge \mathtt{s}(\boldsymbol{v}')$ in Definition 2. The case for $\phi$ of the form $v_i = v_j$ is similar.

Now consider $\phi(\boldsymbol{v})$ of the form $\psi(\boldsymbol{v}) \wedge \chi(\boldsymbol{v})$. According to the induction hypothesis, the lemma holds for $\psi$ and $\chi$. Applying the induction hypothesis first to $\psi$, let $\mathtt{r}_{\psi}$ be a new $k$-ary relation symbol, and let $\Phi^{\psi}$ be a model representation for a vocabulary $S^{\psi} \supseteq S \cup \{\mathtt{r}_{\psi}\}$, such that

$$P_n^{\Phi^{\psi}} \restriction \mathrm{Mod}_n(S) = P_n^{\Phi},$$

and for all $\mathcal{M} \in \mathrm{Mod}_n(S)$, $\boldsymbol{m} \in n^k$:

$$P_n^{\Phi^{\psi}}\big(\mathtt{r}_{\psi}(\boldsymbol{m})|\mathcal{M}\big) = \begin{cases} 1 & \text{if } \mathcal{M} \models \psi(\boldsymbol{m}), \\ 0 & \text{else.} \end{cases}$$

Now we apply the induction hypothesis to $\chi$ and the already constructed $\Phi^{\psi}$. This gives us a model representation $\Phi^{\chi}$ for a vocabulary $S^{\chi} \supseteq S^{\psi}$ containing another new $k$-ary relation symbol $\mathtt{r}_{\chi}$, such that

$$\begin{aligned} P_n^{\Phi^{\chi}} \restriction \mathrm{Mod}_n(S) &= \big(P_n^{\Phi^{\chi}} \restriction \mathrm{Mod}_n(S^{\psi})\big) \restriction \mathrm{Mod}_n(S) \\ &= P_n^{\Phi^{\psi}} \restriction \mathrm{Mod}_n(S) \\ &= P_n^{\Phi}, \end{aligned}$$

and for all $\mathcal{M} \in \mathrm{Mod}_n(S^{\psi})$, $\boldsymbol{m} \in n^k$:

$$\begin{aligned} P_n^{\Phi^{\chi}}\big(\mathtt{r}_{\chi}(\boldsymbol{m})|\mathcal{M}\big) &= \begin{cases} 1 & \text{if } \mathcal{M} \models \chi(\boldsymbol{m}), \\ 0 & \text{else.} \end{cases} \\ &= \begin{cases} 1 & \text{if } \mathcal{M} \restriction S \models \chi(\boldsymbol{m}), \\ 0 & \text{else.} \end{cases} \end{aligned}$$

Given $\Phi^{\chi}$ we now can use the case $\alpha \equiv \mathtt{r}_{\psi}(\boldsymbol{v}) \wedge \mathtt{r}_{\chi}(\boldsymbol{v})$ of Definition 2 to find a final model representation $\Phi^{\phi}$ for the vocabulary $S^{\phi} = S^{\chi} \cup \{\mathtt{r}\}$, such that

$$\begin{aligned} P_n^{\Phi^{\phi}} \restriction \mathrm{Mod}_n(S) &= \big(P_n^{\Phi^{\phi}} \restriction \mathrm{Mod}_n(S^{\chi})\big) \restriction \mathrm{Mod}_n(S) \\ &= P_n^{\Phi^{\chi}} \restriction \mathrm{Mod}_n(S) \\ &= P_n^{\Phi}, \end{aligned}$$

and for all $\mathcal{M} \in \mathrm{Mod}_n(S^{\chi})$, $\boldsymbol{m} \in n^k$:

$$\begin{aligned} P_n^{\Phi^{\phi}}\big(\mathtt{r}(\boldsymbol{m})|\mathcal{M}\big) &= \begin{cases} 1 & \text{if } \mathcal{M} \models \mathtt{r}_{\psi}(\boldsymbol{m}) \wedge \mathtt{r}_{\chi}(\boldsymbol{m}), \\ 0 & \text{else.} \end{cases} \\ &= \begin{cases} 1 & \text{if } \mathcal{M} \restriction S \models \psi(\boldsymbol{m}) \wedge \chi(\boldsymbol{m}), \\ 0 & \text{else.} \end{cases} \end{aligned}$$

The last identity establishes (8) for $\mathcal{M} \in \mathrm{Mod}_n(S)$.

The case for $\phi(\boldsymbol{v})$ of the form $\neg\psi(\boldsymbol{v})$ is dealt with in a similar manner.

Finally, consider $\phi(\boldsymbol{v})$ of the form $\exists w \psi(\boldsymbol{v}, w)$. We apply the induction hypothesis to a relation symbol $\mathsf{r}_\psi$ of arity $k+1$ and the formula $\psi(\boldsymbol{v}, w)$ to obtain a model representation $\Phi^\psi$. We then obtain $\Phi^\phi$ by applying Definition 2 for the case $\alpha(\boldsymbol{v}) \equiv \exists w \mathsf{r}_\psi(\boldsymbol{v}, w)$ to $\mathsf{r}$ and $\Phi^\psi$. □

## 3. Complexity: Deterministic, exact inference

Given a model representation system **M** we now are interested in the complexity of answering probabilistic queries, i.e., of computing $P_n^\Phi(\phi(\boldsymbol{m}))$ for a model representation $\Phi$, a domainsize $n$, and a proposition $\phi(\boldsymbol{m})$. Obviously, with the assumptions we have made we cannot derive exact bounds for the complexity of this computation, because these would depend on many specific features of the system **M** that we have left unspecified. Our aim here, therefore, only is to investigate one specific aspect of the overall computational complexity, namely its dependence on the domainsize $n$. The dependency on this parameter is of particular interest, because it is with regard to this parameter that we would expect to obtain a gain in efficiency by replacing auxiliary network construction with more sophisticated inference techniques: when $\Phi$ and $\phi(\boldsymbol{m})$ are fixed, then the number of nodes in an auxiliary network constructed to compute $P_n^\Phi(\phi(\boldsymbol{m}))$ will usually be polynomial in $n$, and the complexity of inference exponential in $n$ (because, in general, we will also have in the auxiliary network a polynomial growth of the maximal number of parents of single nodes). It is not obvious that this exponential blowup in $n$ is inherent in the problem, and cannot be avoided by other inference techniques. Note, in particular, that the well-known complexity results for inference in Bayesian networks [3] are not applicable here, because we cannot represent a suitable class of Bayesian networks that shows that inference is NP-hard in the network size as the set of auxiliary networks constructed for a set of queries $P_n^\Phi(\phi(\boldsymbol{m}))$ ($n \in \mathbb{N}$; $\Phi, \phi$ fixed).

Thus, we here will be concerned with the complexity of computing $P_n^\Phi(\phi(\boldsymbol{m}))$ as a function of $n$ with $\Phi$ and $\phi(\boldsymbol{m})$ being fixed. Moreover, following a common strategy, we will first concentrate on the simpler problem of deciding whether $P_n^\Phi(\phi(\boldsymbol{m})) > 0$. Formally, our problem then becomes that of deciding predicates of the form

$$\mathrm{NONZERO}\big(\Phi, \phi(\boldsymbol{m})\big) := \big\{n \in \mathbb{N} \mid P_n^\Phi\big(\phi(\boldsymbol{m})\big) > 0\big\}$$

defined by model representations $\Phi$ (in some representation system **M**), and a formula $\phi(\boldsymbol{m})$ containing constants $m_1, \ldots, m_k \in \mathbb{N}$ (use the convention that $P_n^\Phi(\phi(\boldsymbol{m})) = 0$ when $m_i > n$ for some $i \leqslant k$, and therefore $\phi(\boldsymbol{m})$ cannot not be interpreted over the domain $n$). For arbitrary subsets $A \subseteq \mathbb{N}$ we use $A^{\mathrm{un}}$ and $A^{\mathrm{bin}}$ to denote the sets of unary and binary encodings, respectively, of the members of $A$. Since we are interested in the complexity in terms of $n$ of deciding $\mathrm{NONZERO}(\Phi, \phi(\boldsymbol{m}))$, not in terms of $\log(n)$, we really are talking about the complexity of deciding $\mathrm{NONZERO}(\Phi, \phi(\boldsymbol{m}))^{\mathrm{un}}$, when complexity is measured in input size.

Adopting the notation of Johnson [11], we denote by (N)ETIME the class of subsets of $\{0,1\}^*$ that can be decided in (nondeterministic) time $\mathrm{O}(2^{cn})$ for some $c > 0$. Note that this class is distinct from (N)EXPTIME, which is characterized by time bounds of the form $\mathrm{O}(2^{n^c})$ ($c > 0$). We can now formulate our main theorem.

**Theorem 4.** *Let* **M** *be a model construction system that allows first-order conditioning. If* NETIME $\neq$ ETIME*, then there exist an* $S$*-model representation* $\Phi$ *in* **M***, and a ground* $S$*-atom* $\mathtt{r}(\boldsymbol{m})$ *such that* $\mathrm{NONZERO}(\Phi, \mathtt{r}(\boldsymbol{m}))^{\mathrm{un}} \notin \mathrm{P}$.

The proof of Theorem 4 is quite straightforward using Lemma 3 and established results due to Jones and Selman [12] on the connection between the class NETIME and spectra of first-order sentences. We briefly review the relevant definitions and results here.

The *spectrum* of a first-order sentence $\phi$ in the vocabulary $S$ is the set $\{n \in \mathbb{N} \mid \exists \mathcal{M} \in \mathrm{Mod}_n(S)\colon \mathcal{M} \models \phi\}$, i.e., the set of all finite cardinalities for which $\phi$ has a model. A subset of $\mathbb{N}$ is called a spectrum if it is the spectrum of some first-order sentence $\phi$ (over an arbitrary vocabulary—but note that without loss of generality we can assume a relational vocabulary). The result of Jones and Selman [12] that we shall use is: a subset $A$ of $\mathbb{N}$ is a spectrum iff $A^{\mathrm{bin}} \in \mathrm{NETIME}$.

**Proof of Theorem 4.** Let **M** be as stated in the theorem, and assume that $A^{\mathrm{bin}} \in \mathrm{NETIME} \setminus \mathrm{ETIME}$. By Jones and Selman's [12] theorem there exists a first-order sentence $\psi$ in a relational vocabulary $S$, such that $A$ is the spectrum of $\psi$. Let $\Phi$ be a model representation for $S$ such that $P_n^{\Phi}$ is the uniform distribution on $\mathrm{Mod}_n(S)$ for all $n \in \mathbb{N}$. By Lemma 3 there exists a model representation $\Phi^{\phi}$ for $S^{\phi} \supseteq S$, such that $S^{\phi}$ contains a unary relation symbol $\mathtt{r}$, and (7) and (9) hold. Since the right hand side of (9) is nonzero iff $n$ is in the spectrum of $\phi$, we obtain

$$\mathrm{NONZERO}\big(\Phi^{\phi}, \mathtt{r}(m)\big) = A.$$

By the assumption $A^{\mathrm{bin}} \notin \mathrm{ETIME}$ it follows that

$$A^{\mathrm{un}} = \mathrm{NONZERO}\big(\Phi^{\phi}, \mathtt{r}(m)\big)^{\mathrm{un}} \notin \mathrm{P}. \qquad \square$$

Our proof of Theorem 4 relies crucially on the requirement of Definition 2 that we can condition on equality constraints. If we deleted the case $\alpha \equiv v_i = v_j$ from Definition 2 then our arguments would only show that we can encode as sets of the form $\mathrm{NONZERO}(\Phi, \mathtt{r}(\boldsymbol{m}))$ spectra of first-order sentences without equality. These, however, are simply sets of the form $\mathbb{N} \setminus \{1, 2, \ldots, k-1, k\}$, which can be decided in constant time.

Theorem 4 gives us a lower complexity bound for some $\mathrm{NONZERO}(\Phi, \mathtt{r}(\boldsymbol{m}))$. For most concrete model construction systems proposed so far, on the other hand, we have the upper bound $\mathrm{NONZERO}(\Phi, \mathtt{r}(\boldsymbol{m})) \in \mathrm{NP}$ for all $\Phi, \mathtt{r}(\boldsymbol{m})$. This suggests to check whether $\mathrm{NONZERO}(\Phi, \mathtt{r}(\boldsymbol{m}))$ might be an NP-complete problem for some representation system **M**, and suitable $\Phi, \mathtt{r}(\boldsymbol{m})$. A general result in complexity theory, however, says that this is unlikely to be the case.

**Theorem 5.** *If* P $\neq$ NP *then* $\mathrm{NONZERO}(\Phi, \mathtt{r}(\boldsymbol{m}))^{\mathrm{un}}$ *is not* NP*-complete for any representation system* **M***, model representation* $\Phi$*, and query* $\mathtt{r}(\boldsymbol{m})$.

**Proof.** This follows immediately from results of Mahaney [15] that so-called *sparse* sets cannot be NP-complete if P $\neq$ NP. All sets in unary encoding (also called tally languages) are examples of sparse sets. $\square$

While it is customary to simplify questions about the complexity of computing a function to a simpler decision problem, it is of course the complexity of computing the value of $P_n^{\Phi}(\mathtt{r}(\boldsymbol{m}))$ that we are ultimately interested in. In practice, one will usually not need to compute the precise probability value, but only an approximation with a certain given precision. This leads us to the subject of approximate inference, which we deal with in the following section.

## 4. Approximate inference

Following Dagum and Luby [4], we may distinguish between four principal variants of approximate inference: the approximation may be within a specified *absolute* or *relative error*, and the approximation algorithm may be either *deterministic* or *randomized*.

When lower complexity bounds for the computation of exact probabilities $P(\cdot)$ are derived by a reduction to a decision problem of the form $P(\cdot) > 0$? (as we did here, and as did Cooper [3] for standard Bayesian network inference), then we cannot gain much by turning from exact inference to approximate inference with a bounded relative error, because an approximation of $P(\cdot)$ with a bounded relative error will still show whether $P(\cdot) > 0$ or $P(\cdot) = 0$. For this reason we here concentrate on computing approximations $z$ for $P_n^{\Phi}(\mathtt{r}(\boldsymbol{m}))$ with a bound $\varepsilon$ on the absolute error, i.e., a number $z$ that satisfies

$$z \in \left[P_n^{\Phi}\left(\mathtt{r}(\boldsymbol{m})\right) - \varepsilon,\, P_n^{\Phi}\left(\mathtt{r}(\boldsymbol{m})\right) + \varepsilon\right].$$

We first turn to deterministic approximations. It turns out that now the complexity of computing $P_n^{\Phi}(\mathtt{r}(\boldsymbol{m}))$ can be as well-behaved as one might hope—at least in theory.

**Theorem 6.** *There exist model construction systems that allow first-order conditioning such that for every model representation $\Phi$, every ground atom $\mathtt{r}(\boldsymbol{m})$, and every $\varepsilon > 0$ the complexity (in $n$) of computing an approximation of $P_n^{\Phi}(\mathtt{r}(\boldsymbol{m}))$ with absolute error at most $\varepsilon$ is* $\mathrm{O}(1)$.

**Proof.** In Jaeger [10] it is shown that for a certain subclass of relational Bayesian networks the probabilities $P_n^{\Phi}(\mathtt{r}(\boldsymbol{m}))$ converge to some limit as $n \to \infty$ for every network $\Phi$ in that subclass, and every query $\mathtt{r}(\boldsymbol{m})$. The subclass identified in Jaeger [10] is rich enough to allow first-order conditioning.

Now assume that $\Phi$ and $\mathtt{r}(\boldsymbol{m})$ are such that $P_n^{\Phi}(\mathtt{r}(\boldsymbol{m})) \to p \in [0,1]$ as $n \to \infty$, and let $\varepsilon > 0$ be given. Then there exists $n_0 \in \mathbb{N}$, such that $P_n^{\Phi}(\mathtt{r}(\boldsymbol{m})) \in [p - \varepsilon,\, p + \varepsilon]$ for all $n \geqslant n_0$. Thus, we obtain an algorithm for computing an $\varepsilon$-approximation of $P_n^{\Phi}(\mathtt{r}(\boldsymbol{m}))$ by exact computation of $P_n^{\Phi}(\mathtt{r}(\boldsymbol{m}))$ (using any available algorithm) when $n < n_0$, and by simply outputting $p$ when $n \geqslant n_0$. The time requirement of this procedure is asymptotically constant in $n$. □

Clearly this result is of theoretical rather than practical interest, because neither does it tell us how to compute the number $n_0$, nor does it provide any bound on the constant characterizing the time requirement. Furthermore, the theorem is not applicable for representation systems in which $P_n^{\Phi}(\mathtt{r}(\boldsymbol{m}))$ need not converge.

In practice, randomized approximation algorithms can be particularly well-suited for computing probabilities $P_n^{\Phi}(\mathtt{r}(\boldsymbol{m}))$. To see why, consider an algorithm that produces random samples $\mathcal{M}_i \in \mathrm{Mod}_n(S)$ according to the distribution $P_n^{\Phi}$. As in *logic sampling* for standard Bayesian networks [8] we could use the fraction of structures $\mathcal{M}_i$ with $\mathcal{M}_i \models \mathtt{r}(\boldsymbol{m})$ in a random sample $\mathcal{M}_1, \ldots, \mathcal{M}_n$ as an estimate for $P_n^{\Phi}(\mathtt{r}(\boldsymbol{m}))$. This is usually not the best use we can make of the sample $\mathcal{M}_1, \ldots, \mathcal{M}_n$, though: when the distribution $P_n^{\Phi}$ is invariant under renaming (as we have always assumed), then we have

$$P_n^{\Phi}\big(\mathtt{r}(\boldsymbol{m})\big) = E_n^{\Phi}\big(\|\mathtt{r}(\boldsymbol{v})\|\big), \tag{10}$$

where by $\|\mathtt{r}(\boldsymbol{v})\|$ we denote the fraction of tuples $\boldsymbol{m}$ that satisfy $\mathtt{r}(\boldsymbol{m})$ in a structure $\mathcal{M} \in \mathrm{Mod}_n(S)$, and by $E_n^{\Phi}$ the expected value under the distribution $P_n^{\Phi}$. Therefore, we also gain an estimate for $P_n^{\Phi}(\mathtt{r}(\boldsymbol{m}))$ by averaging over the structures $\mathcal{M}_1, \ldots, \mathcal{M}_n$ the values of $\|\mathtt{r}(\boldsymbol{v})\|$. The variance of the random variable $\|\mathtt{r}(\boldsymbol{v})\|$ is at most as large as that of the indicator variable for $\mathtt{r}(\boldsymbol{m})$, and usually decreasing in $n$. Depending on how fast the variance of $\|\mathtt{r}(\boldsymbol{v})\|$ decreases the reduction of the size of a random sample needed to estimate $P_n^{\Phi}(\mathtt{r}(\boldsymbol{m}))$ with given error and confidence bounds can offset the increased complexity of sampling a single structure $\mathcal{M}_i$. In the best case we will really obtain a time requirement that is constant in $n$.

## 5. Conclusion

The purpose of this paper was twofold: first, we wanted to develop a conceptual framework that permits us to treat in a unified way a number of systems that have been proposed in the literature for integrating some first-order reasoning capabilities into Bayesian networks. We here have obtained this unified view by characterizing such systems entirely in terms of their semantics, without imposing any restrictions on the specific syntax used.

Our main objective then was to derive within this general setting results on the complexity of probabilistic inference as a function of the size of the specific domain to which the generic knowledge is applied. This is a new complexity problem that is distinctive of the emergent class of probabilistic relational model representation systems. It does not appear in probabilistic logics, because there entailment always is with regard to all models of a knowledge base, not the models over a specific domain. It also is distinct from complexity questions about inference in standard Bayesian networks, because there a change of the domain (by way of a different set of nodes in the network) always is accompanied by a new model representation (i.e., a new network).

Our complexity problem is of particular interest, because polynomial bounds here would have shown that there are more efficient ways for probabilistic inference than auxiliary network construction—an inference technique that a priori looks rather wasteful, because it involves a complete propositionalization of originally first-order knowledge. However, Theorem 4 shows that for reasonably expressive representation systems we are unlikely to find inference techniques that have a better worst-case behavior than auxiliary network construction. The proof of the theorem points to reasoning about equality as the main cause for the complexity bounds we obtained. Investigations of weaker systems with potentially lower complexity should therefore be directed towards systems without equality reasoning.

## References

[1] F. Bacchus, Representing and Reasoning with Probabilistic Knowledge, MIT Press, Cambridge, MA, 1990.
[2] J.S. Breese, Construction of belief and decision networks, Computational Intelligence 8 (4) (1992) 624–647.
[3] G.F. Cooper, The computational complexity of probabilistic inference using Bayesian belief networks, Artificial Intelligence 42 (1990) 393–405.
[4] P. Dagum, M. Luby, Approximating probabilistic inference in Bayesian belief networks is NP-hard, Artificial Intelligence 60 (1993) 141–153.
[5] N. Friedman, L. Getoor, D. Koller, A. Pfeffer, Learning probabilistic relational models, in: Proc. IJCAI-99, Stockholm, Sweden, 1999.
[6] A. Frisch, P. Haddawy, Anytime deduction for probabilistic logic, Artificial Intelligence 69 (1994) 93–122.
[7] P. Haddawy, Generating Bayesian networks from probability logic knowledge bases, in: Proc. 10th Conference on Uncertainty in Artificial Intelligence (UAI-94), Seattle, WA, 1994, pp. 262–269.
[8] M. Henrion, Propagating uncertainty in Bayesian networks by probabilistic logic sampling, in: J.F. Lemmer, L.N. Kanal (Eds.), Uncertainty in Artificial Intelligence, Vol. 2, Elsevier Science, Amsterdam, 1988, pp. 149–163.
[9] M. Jaeger, Relational Bayesian networks, in: Proc. 13th Conference on Uncertainty in Artificial Intelligence (UAI-97), Providence, RI, 1997.
[10] M. Jaeger, Convergence results for relational Bayesian networks, in: Proc. 13th Annual IEEE Symposium on Logic in Computer Science (LICS-98), Indianapolis, IN, 1998.
[11] D.S. Johnson, A catalog of complexity classes, in: J. van Leeuwen (Ed.), Handbook of Theoretical Computer Science, Vol. 1, Elsevier, Amsterdam, 1990, pp. 67–161.
[12] N.D. Jones, A.L. Selman, Turing machines and the spectra of first-order formulas with equality, in: Proc. 4th ACM Symposium on Theory of Computing, Denver, CO, 1972, pp. 157–167.
[13] D. Koller, A. Pfeffer, Probabilistic frame-based systems, in: Proc. AAAI-98, Madison, WI, 1998, pp. 580–587.
[14] V. Lakshmanan, F. Sadri, Probabilistic deductive databases, in: M. Bruynooghe (Ed.), Proc. International Symposium on Logic Programming (ILPS-94), MIT Press, Cambridge, MA, 1994.
[15] S.R. Mahaney, Sparse sets and reducibilities, in: R.V. Book (Ed.), Studies in Complexity Theory, Pitman, London, 1986, pp. 63–118.
[16] R. Ng, V. Subrahmanian, Probabilistic logic programming, Inform. and Comput. 101 (1992) 150–201.
[17] L. Ngo, P. Haddawy, Answering queries from context-sensitive probabilistic knowledge bases, Theoret. Comput. Sci. 171 (1997) 147–177.
[18] N. Nilsson, Probabilistic logic, Artificial Intelligence 28 (1986) 71–88.
[19] D. Poole, Probabilistic Horn abduction and Bayesian networks, Artificial Intelligence 64 (1993) 81–129.
[20] A. Saffiotti, E. Umkehrer, Inference-driven construction of valuation systems from first-order clauses, IEEE Trans. Systems Man Cybernet. 24 (11) (1994) 1611–1624.
[21] M.P. Wellman, J.S. Breese, R.P. Goldman, From knowledge bases to decision models, The Knowledge Engineering Review 7 (1) (1992) 35–53.