

Lower complexity bounds for lifted inference

MANFRED JAEGER

Department of Computer Science, Aalborg University, 9220 Aalborg, Denmark
(e-mail: jaeger@cs.aau.dk)

submitted 14 June 2012; revised 2 January 2013; accepted 7 January 2013

Abstract

One of the big challenges in the development of probabilistic relational (or probabilistic logical) modeling and learning frameworks is the design of inference techniques that operate on the level of the abstract model representation language, rather than on the level of ground, propositional instances of the model. Numerous approaches for such “lifted inference” techniques have been proposed. While it has been demonstrated that these techniques will lead to significantly more efficient inference on some specific models, there are only very recent and still quite restricted results that show the feasibility of lifted inference on certain syntactically defined classes of models. Lower complexity bounds that imply some limitations for the feasibility of lifted inference on more expressive model classes were established earlier in Jaeger (2000; JAEGER, M. 2000. On the complexity of inference about probabilistic relational models. *Artificial Intelligence* 117, 297–308). However, it is not immediate that these results also apply to the type of modeling languages that currently receive the most attention, i.e., weighted, quantifier-free formulas. In this paper we extend these earlier results, and show that under the assumption that $\text{NETIME} \neq \text{ETIME}$, there is no polynomial lifted inference algorithm for knowledge bases of weighted, quantifier-, and function-free formulas. Further strengthening earlier results, this is also shown to hold for approximate inference and for knowledge bases not containing the equality predicate.

KEYWORDS: Probabilistic-logic models, lifted inference

1 Introduction

Probabilistic logic models (a.k.a. probabilistic or statistic relational models) provide high-level representation languages for probabilistic models of structured data (Breese 1992; Poole 1993; Ngo *et al.* 1995; Sato 1995; Jaeger 1997; Friedman *et al.* 1999; Kersting and De Raedt 2001; Taskar *et al.* 2002; Milch *et al.* 2005; Richardson and Domingos 2006; Vennekens *et al.* 2006). While supporting model specifications at an abstract, first-order logic level, inference is typically performed at the level of concrete ground instances of the models, i.e., at the propositional level. This mismatch between model specification and inference methods has been noted early on (Jaeger 1997), and has given rise to numerous proposals for inference techniques that operate at the high level of the underlying model specifications (Poole 2003; de Salvo Braz *et al.* 2005; Milch *et al.* 2008; Kisiński and Poole 2009; Jha

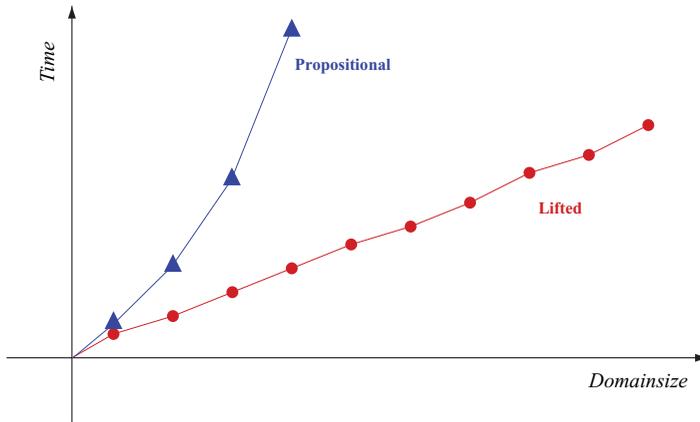


Fig. 1. A typical performance evaluation.

et al. 2010; Fierens *et al.* 2011; Gogate and Domingos 2011; Van den Broeck 2011; Van den Broeck *et al.* 2011). Inference methods of this nature have collectively become known as “lifted” inference techniques.

The concept of lifted inference is mostly introduced on an informal level: “...*lifted*, that is, deals with groups of random variables at a first-order level” (de Salvo Braz *et al.* 2005); “*The act of exploiting the high level structure in relational models is called lifted inference*” (Apsel and Brafman 2011); “*The idea behind lifted inference is to carry out as much inference as possible without propositionalizing* (Kiszyński and Poole 2009); “*lifted inference, which deals with groups of indistinguishable variables, rather than individual ground atoms* (Singla *et al.* 2010). While, thus, the term lifted inference emerges as a quite coherent algorithmic metaphor, it is not immediately obvious what its exact technical meaning should be. Since quite a variety of different algorithmic approaches are collected under the label “lifted”, and since most of them can degenerate for certain models to ground, or propositional inference, it is difficult to precisely define the class of lifted inference techniques in terms of specific algorithmic techniques employed.

A more fruitful approach is to make more precise the concept of lifted inference in terms of its objectives. Here one observes that lifted inference techniques very consistently are evaluated on, and compared against each other, by how well inference complexity scales as a function of the domain (or population) for which the general model is instantiated. Thus, empirical evaluations of lifted inference techniques are usually presented in the form of domainsize versus inference time plots, as shown in Figure 1.

Van den Broeck (2011), therefore, has proposed a formal definition of *domain lifted inference* in terms of polynomial time complexity in the domainsize parameter. Experimental and theoretical analyses of existing lifted inference techniques then show that they provide domain lifted inference in some cases where basic propositional inference techniques would exhibit exponential complexity (as illustrated in Figure 1). However, until recently, these positive results were mostly limited to examples of individual models, and little was known about the feasibility of

lifted inference for certain well-defined classes of models. First results that show the feasibility of lifted inference for whole classes of models are given by Van den Broeck (2011) and Domingos and Webb (2012).

On the other hand, Jaeger (2000) has shown that under certain assumptions on the expressivity of the modeling language, probabilistic inference is not polynomial in the domainsize, thereby demonstrating some inherent limitations in terms of worst-case complexity for the goals of lifted inference. However, the results of Jaeger (2000) are based on types of probabilistic logic models that are somewhat different from the models that presently receive the most attention: first, they essentially assume a *directed* modeling framework, in which the model represents a generative stochastic process for sampling relational structures. The model is defined by specifying marginal and conditional probability distributions for random variables corresponding to ground atoms. Ground instances of the model, then, can be represented by directed graphical models, i.e., Bayesian networks. While the majority of existing model classes fall into the category of directed models (Breese 1992; Poole 1993; Ngo *et al.* 1995; Sato 1995; Jaeger 1997; Friedman *et al.* 1999; Kersting and De Raedt 2001; Milch *et al.* 2005; Vennekens *et al.* 2006), there is currently a lot of interest in *undirected* models that are given by a set of soft constraints on relational structures, specified in the form of potential functions, and in the ground case giving rise to undirected graphical models, i.e., Markov networks. Second, the results of Jaeger (2000) require quite strong assumptions on the expressivity of the probabilistic-logic modeling language, which is required to allow that conditional distributions of atoms can be specified dependent on unrestricted first-order properties. Much current work, in contrast, is concerned with languages that only incorporate certain weak fragments of first-order logic.

In this paper the general approach of (Jaeger 2000) is extended to obtain lower complexity bounds for inference in probabilistic-logic model classes that have emerged as the focus of interest for lifted inference techniques, i.e., undirected models based on quantifier- and function-free fragments of first-order logic.

In a sharp contrast with Jaeger (2000), where a “trivial” constant-time approximate inference method was described, we show that our lower complexity bounds also hold for approximate inference. Further sharpening earlier results, we finally establish that the lower complexity bounds also hold for models not using the equality predicate, which in Jaeger (2000) was conjectured to be the key source of inherent complexity.

A preliminary version of this paper has been published as Jaeger (2012). Its main results were also already included in the survey paper (Jaeger and Van den Broeck 2012), which contains a systematic overview of known results and open problems related to the complexity of lifted inference.

In the following section we introduce a general framework in which classes of undirected probabilistic-logic models and classes of associated inference problems can be defined. Section 3 reviews classic results relating first-order logic models to the complexity class NETIME. Section 4 contains our main results and Section 5 discusses some notable differences that emerge between the results for directed and for undirected models.

2 Weighted feature models

Similarly as Richardson and Domingos (2006), Van den Broeck *et al.* (2011), and Gogate and Domingos (2011) we assume the following framework: a model or knowledge base is given by a set of weighted formulas:

$$KB : \begin{array}{ll} \phi_1(\mathbf{v}_1) & : w_1 \\ \phi_2(\mathbf{v}_2) & : w_2 \\ \dots & \dots \\ \phi_n(\mathbf{v}_N) & : w_N \end{array} \tag{1}$$

where ϕ_i are formulas in first-order predicate logic, $w_i \in \mathbb{R}$ are non-negative *weights*, and $\mathbf{v}_i = (v_{i,1}, \dots, v_{i,k_i})$ are the free variables of ϕ_i . The case $k_i = 0$, i.e., ϕ_i is a sentence without free variables, is also permitted. The ϕ_i use a given signature S of relation-, function-, and constant symbols.

An interpretation (or possible world) (D, I) for S consists of a domain D and an interpretation function I that maps the symbols in S to functions, relations, and elements on D . For a tuple $\mathbf{d} \in D^{k_i}$ then the truth value of $\phi_i(\mathbf{v}_i/\mathbf{d})$ is defined, and we write $(D, I) \models \phi_i(\mathbf{d})$, or simpler $I \models \phi_i(\mathbf{d})$, if $\phi_i(\mathbf{v}_i/\mathbf{d})$ is true in (D, I) . We use $\mathcal{I}(D, S)$ to denote the set of all interpretations for the signature S over the domain D .

In this paper we are only concerned with finite domains and assume without loss of generality that $D = D_n := \{1, \dots, n\}$ for some $n \in \mathbb{N}$.

For $I \in \mathcal{I}(D_n, S)$ let $\#(i, I)$ denote the number of elements \mathbf{d} in D^{k_i} for which $I \models \phi_i(\mathbf{d})$. The *weight* of I then is

$$W_{KB,n}(I) := \prod_{i=1}^N w_i^{\#(i,I)}, \tag{2}$$

where $0^0 = 1$. The probability of I is

$$P_{KB,n}(I) = W_{KB,n}(I)/Z,$$

where Z is the normalizing constant (partition function)

$$Z = \sum_{I \in \mathcal{I}(D_n, S)} W_{KB,n}(I). \tag{3}$$

For a first-order sentence ϕ and $n \in \mathbb{N}$ then

$$P_{KB,n}(\phi) := P_{KB,n}(\{I \in \mathcal{I}(D_n, S) \mid I \models \phi\}) \tag{4}$$

is the probability of ϕ in $\mathcal{I}(D_n, S)$.

We call knowledge base (1) together with the semantics given by (2) and (4) a *weighted feature model*, since it associates weights w_i with model features ϕ_i . Weighted feature models in our sense can be seen as a slight generalization as *weighted model counting* frameworks (Fierens *et al.* 2011; Gogate and Domingos 2011) in which non-zero weights are only associated with literals. Knowledge bases of the form (1) can be translated into weighted model counting frameworks via an introduction of new relation symbols R_1, \dots, R_n , hard constraints $\phi_i(\mathbf{v}_i) \leftrightarrow R_i(\mathbf{v}_i)$, and weighted formulas $R_i(\mathbf{v}_i) : w_i$ (Gogate and Domingos 2011; Van den Broeck *et al.* 2011). Up

to an expansion of the signature, thus, weighted feature models and weighted model counting are equally expressive. *Markov Logic Networks* (Richardson and Domingos 2006) also are based on knowledge bases of the form (1), allowing arbitrary formulas ϕ_i . However, the semantics of the model there depends on a transformation of the formulas into conjunctive normal form, and therefore does not exactly correspond to (2) and (4) unless ϕ_i are clauses.

All types of models here discussed, thus, are very similar in nature, and only differ with respect to certain restrictions on what types of logically defined features can be associated with a weight. The general definition of weighted feature models gives us the flexibility of considering a variety of classes of such restrictions.

A *probabilistic inference problem* $PI(KB, n, \chi, \eta)$ for a weighted feature model is given by a knowledge base KB , a domainsize $n \in \mathbb{N}$, and two first-order sentences χ, η . The solution to the inference problem is the conditional probability $P_{KB, n}(\chi \mid \eta)$.

A *class* of inference problems is defined by allowing arguments KB , χ , and η only from some restricted classes \mathcal{KB} , \mathcal{Q} (the query class), and \mathcal{E} (the evidence class), respectively. We use the notation

$$PI(\mathcal{KB}, \mathcal{Q}, \mathcal{E}) := \{PI(KB, n, \phi, \psi) \mid KB \in \mathcal{KB}, n \in \mathbb{N}, \chi \in \mathcal{Q}, \eta \in \mathcal{E}\}$$

for classes of inference problems.

The results of this paper will be given for the case where \mathcal{Q} consists of all ground atoms, denoted as \mathcal{AT} , and \mathcal{E} is empty. Thus, as far as \mathcal{Q} and \mathcal{E} are concerned, we are considering the most restrictive class of inference problems. Since we are deriving lower complexity bounds, this leads to the strongest possible results, which directly apply also to more general classes \mathcal{Q} and \mathcal{E} .

Classes \mathcal{KB} are defined by various syntactic restrictions on the formulas ϕ_i in the knowledge base. In this paper, we consider the following fragments of first-order logic (FOL): relational FOL (RFOL), i.e., FOL without function and constant symbols; 0-RFOL, which is the quantifier-free fragment of RFOL; and 0-RFOL[≠], which is 0-RFOL without the equality relation.

An algorithm solves a class $PI(\mathcal{KB}, \mathcal{Q}, \mathcal{E})$ if it solves all instances of $PI(KB, n, \chi, \eta)$ in the class. An algorithm ϵ -approximately solves $PI(\mathcal{KB}, \mathcal{Q}, \mathcal{E})$, if for any $PI(KB, n, \chi, \eta)$ in the class it returns a number $p \in [P_{KB, n}(\chi \mid \eta) - \epsilon, P_{KB, n}(\chi \mid \eta) + \epsilon]$. An algorithm that solves $PI(\mathcal{KB}, \mathcal{Q}, \mathcal{E})$ is *polynomial in the domainsize*, if for fixed KB, χ, η the computation of $PI(KB, n, \chi, \eta)$ is polynomial in n .

3 Spectra and complexity

The following definition introduces the central concept for our analysis.

Definition 3.1

Let ϕ be a sentence in first-order logic. The *spectrum* of ϕ is the set of integers $n \in \mathbb{N}$ for which ϕ is satisfiable by an interpretation of size n .

Example 3.2

Let $\phi = \psi_1 \wedge \psi_2 \wedge \psi_3$, where

$$\begin{aligned} \psi_1 &\equiv \forall x, y \ u(x, y) \Leftrightarrow u(y, x) \\ \psi_2 &\equiv \forall x \ \exists y \ y \neq x \wedge u(x, y) \\ \psi_3 &\equiv \forall x, y, y' \ (u(x, y) \wedge u(x, y') \Rightarrow y = y'). \end{aligned}$$

ϕ expresses that the binary relation u defines an undirected graph (ψ_1) in which every node is connected to exactly one other node (ψ_2, ψ_3). Thus, ϕ describes a pairing relation that is satisfiable exactly over domains of even size: $spec(\phi) = \{n \mid n \text{ even}\}$.

The complexity class ETIME consists of problems solvable in time $O(2^{cn})$, for some constant c . The corresponding nondeterministic class is NETIME. Note that these classes are distinct from the more commonly studied classes (N)EXPTIME, which are characterized by complexity bounds $O(2^{n^c})$ (Johnson 1990). For $n \in \mathbb{N}$ let $bin(n) \in \{0, 1\}^*$ denote the binary coding of n , and $un(n) \in \{1\}^*$ the unary coding (i.e., n is represented as a sequence of n 1s). A set $S \subseteq \mathbb{N}$ is in (N)ETIME, iff $\{bin(n) \mid n \in S\}$ is in (N)ETIME, which also is equivalent to $\{un(n) \mid n \in S\}$ being in (N)PTIME.

Like Jaeger (2000), we use the following connection between spectra and NETIME as the key tool for our complexity analysis.

Theorem 3.3

(Jones and Selman 1972) A set $A \subseteq \mathbb{N}$ is in NETIME, iff A is the spectrum of a sentence $\phi \in \text{RFOL}$.

Corollary 3.4

If $\text{NETIME} \neq \text{ETIME}$, then there exists a first-order sentence ϕ , such that $\{un(n) \mid n \in spec(\phi)\}$ is not recognized in deterministic polynomial time.

Thus, by reducing instances $n \in spec(\phi)$ of the spectrum recognition problem to probabilistic inference problems $PI(KB, n, \chi, \eta)$, where $KB \in \mathcal{KB}, \chi \in \mathcal{L}, \eta \in \mathcal{E}$ are fixed for the given ϕ , one establishes that the $PI(\mathcal{KB}, \mathcal{L}, \mathcal{E})$ is not polynomial in the domainsize (under the assumption $\text{ETIME} \neq \text{NETIME}$).

4 Complexity results

This section contains our complexity results. We begin with a result for knowledge bases using full RFOL. This is rather straightforward and (for exact inference) already implied by the results of Jaeger (2000). We then proceed to extend this base result to 0-RFOL and 0-RFOL $^\neq$.

4.1 Base result: the RFOL case

Theorem 4.1

If $\text{NETIME} \neq \text{ETIME}$, then there does not exist an algorithm that 0.25-approximately solves $PI(\text{RFOL}, \mathcal{AF}, \emptyset)$ in time polynomial in the domainsize.

The proof of this theorem provides the general pattern also for subsequent proofs. It is therefore here given in full.

Proof: Let ϕ be a sentence with a non-polynomial spectrum as given by Corollary 3.4. Let S be the relational signature of ϕ . Let $a()$ be a new relation symbol of arity zero (i.e., $a()$ represents a propositional variable). The first weighted formula in our knowledge base then is

$$\neg(\phi \leftrightarrow a()) \quad : \quad 0 \tag{5}$$

We now already have that $P_{KB,n}(a()) > 0$ iff there exists $I \in \mathcal{I}(D_n, S)$ with $I \models \phi$, i.e., iff $n \in spec(\phi)$. This already reduces the decision problem for $spec(\phi)$ to solving $PI(KB, n, a(), \emptyset)$ exactly. However, from the 0-1 laws of first-order logic (Fagin 1976), it follows that for our current $KB: P_{KB,n}(a()) \rightarrow_{n \rightarrow \infty} 0$. Thus, for every $\epsilon > 0$ we could define an ϵ -approximate constant-time inference algorithm by returning 0 for all sufficiently large n .

In order to obtain our result for approximate inference, we will now ensure that for all $n \in spec(\phi)$ the probability $P_{KB,n}(a())$ is greater than 0.5, while it remains zero for $n \notin spec(\phi)$. We do this essentially by calibrating the normalization constant Z in (3). For this we introduce another new relation $b()$ and add to KB :

$$\neg \left(\left(\bigwedge_{R \in S} \forall x \neg R(x) \right) \leftrightarrow b() \right) \quad : \quad 0 \tag{6}$$

Thus, for every n there is exactly one interpretation $I \in \mathcal{I}(D_n, S)$ with non-zero weight in which $b()$ is true (the one in which all relations have empty interpretations). Finally, we give zero weight to all interpretations except those in which $a()$ or $b()$ is true:

$$\neg(a() \vee b()) \quad : \quad 0 \tag{7}$$

Let KB consist of (5), (6), and (7). Every $I \in \mathcal{I}(D_n, S)$ then has weight 0 if it satisfies one of the three formulas, and weight 1 otherwise. Consider the case $n \notin spec(\phi)$. Then, by (5) $W_{KB,n}(a()) = 0$. By (7) this then means that in all interpretations of non-zero weight $b()$ must be true. By (6) there is exactly one such interpretation. Thus, Z in (3) is 1 and $P_{KB,n}(a()) = 0/1 = 0$.

If $n \in spec(\phi)$, then $W_{KB,n}(a()) \geq 1$, and $Z = W_{KB,n}(a())$ (if the interpretation in which all R are empty also is a model of ϕ), or $Z = W_{KB,n}(a()) + 1$ (otherwise). Thus, $P_{KB,n}(a()) \geq 1/2$. A 0.25-approximate inference algorithm for $PI(KB, n, a(), \emptyset)$, thus, would decide $spec(\phi)$. □

4.2 The 0-RFOL case

We now proceed towards our main result, which is going from RFOL to 0-RFOL. If we wanted to allow function and constant symbols in our knowledge base, then one could go to a quantifier-free fragment in a quite straightforward manner using Skolemization. Since satisfiability over a given domain is the same for a formula ϕ and its quantifier-free Skolemized version ϕ^{Skol} , the arguments of the proof of Theorem 4.1 would go through with little change. In order to accomplish the same

using only the relational fragment 0-RFOL, we define the *relational Skolemization* of a formula. The idea is to replace function and constant symbols in the Skolemized version of a formula with relational representations. For example, the Skolemized version of ψ_2 from Example 3.2 is

$$\psi_2^{Skol} \equiv \forall x \ f(x) \neq x \wedge u(x, f(x))$$

with a new function symbol $f()$. Introducing a relational encoding of $f()$ leads to

$$\psi_2^{R-Skol} \equiv \forall x, y \ R^f(x, y) \rightarrow (y \neq x \wedge u(x, y))$$

with R^f , a new binary relation symbol encoding $f()$. This translation must be accompanied by axioms that confine the possible interpretations of R^f to relations that encode functions.

Such relational encodings of functions are well established. However, there does not seem to be a standard account of this technique that serves our purpose. The following proposition, therefore, provides the relevant result in a form tailored for our needs.

Proposition 4.2

Let $\phi(\mathbf{x}) \in 0\text{-FOL}(S \cup S^F)$, where S is a set of relation symbols and S^F a set of function and constant symbols. Let S^+ be a set of new relation symbols that for every k -ary $f \in S^F$ contains a $k + 1$ -ary R^f (constant symbols are treated as 0-ary function symbols). Let *Func* be a set of sentences such that for every $f \in S^F$ it contains

$$\forall \mathbf{x} \ y \ y' \ (R^f(\mathbf{x}, y) \wedge R^f(\mathbf{x}, y') \rightarrow y = y') \tag{8}$$

$$\forall \mathbf{x} \exists y \ R^f(\mathbf{x}, y). \tag{9}$$

Then there exists a formula $\phi^+(\mathbf{x}, \mathbf{z}) \in 0\text{-RFOL}(S \cup S^+)$, such that the following are equivalent for all n :

- (i) there exists $I \in \mathcal{I}(D_n, S \cup S^F)$ with $I \models \forall \mathbf{x} \phi(\mathbf{x})$
- (ii) there exists $I^+ \in \mathcal{I}(D_n, S \cup S^+)$ with $I^+ \models \text{Func} \wedge \forall \mathbf{x} \mathbf{z} \ \phi^+(\mathbf{x}, \mathbf{z})$

If ϕ^{Skol} is the Skolemization of a formula $\phi \in \text{RFOL}$, we then call ϕ^{Skol+} the *relational Skolemization* of ϕ , written as ϕ^{R-Skol} .

Our plan, now, is to prove the analogon of Theorem 4.1 for 0-RFOL by replacing ϕ in (5) with ϕ^{R-Skol} . However, this is not enough, since we also need to constrain the models of our knowledge base (more precisely, those models in which $a()$ is true) to satisfy the axioms (8) and (9). This poses a problem because (9) contains an existential quantifier and so we cannot add this axiom directly as a constraint to a knowledge base restricted to 0-RFOL. Indeed, we almost seem to have gone a full circle, since we are back at knowledge bases in a relational vocabulary with existential quantification! However, we now have reduced arbitrary occurrences of existential quantifiers to occurrences only within in the special formulas (9).

Our strategy, now, is to approximate formulas (9) with weighted formulas of the form

$$a() \wedge R^f(\mathbf{x}, y) \quad : \quad w \tag{10}$$

that reward models of $a()$ in which the existential quantifier of (9) is satisfied for many (all) x . We will no longer be able to ensure that $W_{KB,n}(a()) = 0$ when $n \notin \text{spec}(\phi)$. However, by a suitable choice of w and by a careful calibration of the weight of models of the alternative proposition $b()$, we still can ensure that $W_{KB,n}(b()) \gg W_{KB,n}(a())$ when $n \notin \text{spec}(\phi)$, and $W_{KB,n}(b()) \approx W_{KB,n}(a())$ when $n \in \text{spec}(\phi)$. However, the right calibration of the weights of models of $a()$ and $b()$ within $\mathcal{S}(D_n, S)$ will now require that one sets w to a value $w(n)$ depending on n .

This means that we no longer can reduce the decision problem $n \in \text{spec}(\phi)$ to the probabilistic inference problem $PI(KB, n, a(), \emptyset)$ for a fixed knowledge base KB . We only achieve a reduction to the inference problem $PI(KB(w(n)), n, a(), \emptyset)$, where the logical structure of KB is fixed, but a weight parameter $w(n)$ depends on n . Generally, for a knowledge base KB containing N weighted formulas, we denote with $KB(w_1, \dots, w_N)$ the knowledge base that contains the same formulas as KB , but with the weights set to values w_1, \dots, w_N .

To translate the lower complexity bounds of the original spectrum recognition problem into lower complexity bounds for the resulting inference problem, one now has to be precise about the representation of the inference problem. To this end, we assume that weights w are rational numbers and represented by pairs (u, v) of integers, so that $w = u/v$. We then define the representation size $l(w)$ as $\log(|u| + 1) + \log(|v| + 1)$. The total representation size of the weight parameters $\mathbf{w} = (w_1, \dots, w_N)$ in a knowledge base is $l(\mathbf{w}) := \sum_{i=1}^N l(w_i)$. An inference algorithm for probabilistic inference problems in $PI(\mathcal{KB}, \mathcal{Q}, \mathcal{E})$ is *polynomial in the domainsize and the representation size of the weight parameters*, if for any $KB \in \mathcal{KB}$, $\chi \in \mathcal{Q}$, $\eta \in \mathcal{E}$ the class of inference problems $PI(KB(\mathbf{w}), n, \chi, \eta)$ can be solved in time that is bounded by a polynomial $\sum_{i,j=0}^d \alpha_{i,j} l(\mathbf{w})^i n^j$ ($\alpha_{i,j} \in \mathbb{R}$, $d \in \mathbb{N}$). We can now state the following theorem:

Theorem 4.3

If $\text{NETIME} \neq \text{ETIME}$, then there does not exist an algorithm that 0.2-approximately solves $PI(0\text{-RFOL}, \mathcal{AT}, \emptyset)$ in time polynomial in the domainsize and the representation size of the weight parameters.

The full proof of the theorem is given in the appendix. It consists of a polynomial-time reduction of the $n \in \text{spec}(\phi)$ decision problem to a probabilistic inference problem $PI(KB(\mathbf{w}(n)), n, a(), \emptyset)$, where $l(\mathbf{w}(n))$ is polynomial in n . An inference algorithm that can solve $PI(KB(\mathbf{w}(n)), n, a(), \emptyset)$ in time polynomial in the domainsize and $l(\mathbf{w}(n))$, thus, would yield a polynomial decision procedure for $\text{spec}(\phi)$.

4.3 Polynomiality in $l(\mathbf{w})$

One may wonder how strong or surprising Theorem 4.3 really is in light of its extra runtime polynomial in $l(\mathbf{w})$ condition. It has previously been emphasized that lifted inference procedures should only be expected to be polynomial in the domain size, but not in other parameters that characterize the complexity of KB (Jaeger 2000; Van den Broeck 2011). These remarks, however, have mostly been motivated by considerations of the logical complexity of KB , e.g., in terms of the number and

complexity of its weighted formulas, or the size of the signature. The complexity in terms of numerical parameters, on the other hand, has not received much attention.

To better understand the nature of the condition of being polynomial in the domainsize and $l(\mathbf{w})$, we have to look a little closer at how the parameters affect the complexity of the computation. We consider algorithms that can be described as follows: to compute $PI(KB(\mathbf{w}), n, \chi, \eta)$ the algorithm performs a number of steps, $i = 1, \dots, L$, where step i consists either of executing a constant time operation that does not depend on the numerical model parameters (e.g., a logical operation on formulas), or of a basic operation on numerical parameters.

We consider the executions the algorithm performs on inputs with fixed logical structure KB and fixed χ, η , but varying weight parameters \mathbf{w} and domainsizes n . Let $V_{\mathbf{w},n}(i)$ denote the set of all numerical variables stored by the algorithm before performing step i , when it is run on inputs (\mathbf{w}, n) . Thus, $V_{\mathbf{w},n}(i)$ comprises the original weight parameters of the model, as well as computed intermediate results, etc. We now make two basic assumptions on the algorithm:

Ass.1 The weight parameters \mathbf{w} only influence the numerical values of the variables stored in $V_{\mathbf{w},n}(i)$, but not the sequence of execution steps performed by the algorithm. In particular, the number of execution steps performed by the algorithm only depends on n : $L = L(n)$.

Ass.2 The basic operations performed on numerical variables are polynomial time in the size of their arguments and they produce an output whose size is linear in the size of the inputs. This is the case for the basic arithmetic operations, addition and multiplication, for example.

The total representation size of $V_{\mathbf{w},n}(i)$ then is bounded by $c_n(i)l(\mathbf{w})$, where $c_n(i)$ is a coefficient not depending on \mathbf{w} . Also, let $q(\cdot)$ be a polynomial that provides a common complexity bound for the basic numerical operations that can be performed at one step. The total execution time of the algorithm on input (\mathbf{w}, n) then is bounded by

$$\sum_{i=1}^{L(n)} q(c_n(i)l(\mathbf{w})). \quad (11)$$

If, now, for fixed weight vectors \mathbf{w} the algorithm is polynomial in n (equivalently, the algorithm is polynomial in n under a computation model where basic numeric operations are constant time), then $L(n)$ and $\max_{i=1, \dots, L(n)} c_n(i)$ must be polynomially bounded in n . The combined complexity (11) then, in fact, is polynomial both in n and $l(\mathbf{w})$.

In summary, this shows that an algorithm, that for fixed \mathbf{w} is polynomial in n and satisfies assumptions Ass.1 and Ass.2, actually is polynomial in n and $l(\mathbf{w})$. Thus, for this type of algorithm, the additional restriction of Theorem 4.3 compared to Theorem 4.1 is insignificant.

The remaining question, then, is how restrictive or realistic assumptions Ass.1 and Ass.2 actually are. For exact inference algorithms it appears that Ass.1 and Ass.2 are satisfied by all existing approaches, with a small qualification: algorithms might give special treatment to special weight parameters, such as $w = 0$ or $w = \infty$, which then can lead to a violation of Ass.1 in the strict sense. However, our analysis could

also be performed based on a weakened form of Ass.1 that allows certain special weights to influence the computation differently from proper numerical weights $0 < w < \infty$. A slightly more elaborate argument would then arrive at essentially the same conclusions.

The situation is less clear for approximate inference algorithms. Here the numerical values stored in $V_{w,n}(i)$ may influence the algorithm in multiple ways: for example, they can be used to test a termination condition, or to decide which computations to perform next in order to improve approximation bounds derived so far. In all such cases, the model weights w can have an impact on the sequence and the total number of execution steps, and Ass.1 is not satisfied. Thus, even though the theorem also applies to approximate inference, its implications for the construction of approximate inference algorithms may be less severe, since there might be reasonable ways to build approximate inference algorithms that are polynomial in n , without also being polynomial in $l(w)$.

4.4 The 0-RFOL[≠] case

In a final strengthening of our results, we now move on to the fragment 0-RFOL[≠]. The availability of the equality predicate for the formulas of KB , so far, has been an important prerequisite for our arguments, because Theorem 3.3 crucially depends on equality: spectra for formulas $\phi \in \text{RFOL}^{\neq}$ are always of the form $\mathbb{N} \setminus \{1, \dots, k\}$ for some k , and thus, decidable in constant time. For this reason it was suggested in Jaeger (2000) that one should focus on logical fragments without equality when looking for model classes for which lifted inference scales polynomially in the domainsize. As our final result shows, however, elimination of equality may not have such a large impact on complexity, after all.

Theorem 4.4

If $\text{NETIME} \neq \text{ETIME}$, then there does not exist an algorithm that 0.2-approximately solves $PI(0\text{-RFOL}^{\neq}, \mathcal{A}, \emptyset)$ in time polynomial both in the domainsize and the representation size of the weight parameters.

This theorem is a generalization of Theorem 4.3, and strictly speaking, makes 4.3 redundant. It is only for expository purposes and greater transparency in the proof arguments that we here develop these results in two steps.

The proof of Theorem 4.4 is a refinement of the proof of Theorem 4.3. In addition to approximating Skolem functions f with relations R^f , we now also approximate the equality predicate $=$ with a binary relation $E(\cdot, \cdot)$. Similarly, as we could not impose in 0-RFOL hard constraints that ensure that R^f encodes a function, we also cannot constrain models to always interpret E as the equality relation. However, just as with (8) and (10) we rewarded interpretations with functional R^f , we can penalize interpretations in which E is not true equality by means of the two weighted formulas:

$$a() \wedge \neg E(x, x) \quad : \quad 0 \tag{12}$$

$$a() \wedge E(x, y) \quad : \quad 1/w, \tag{13}$$

where w is a large weight.

5 Approximate inference, convergence, and evidence

There are some notable differences with respect to approximate inference between the results we here obtained for weighted model counting and the results of Jaeger (2000). In Jaeger (2000) it was shown that due to convergence of query probabilities $P_n(a())$ as $n \rightarrow \infty$, in theory a trivial constant time approximation algorithm exists: perform exact inference for all input domains up to a size n^* and output the limit probability for all domains of size $> n^*$. This “algorithm”, however, has no practical use, since for a desired accuracy value ϵ one first would have to determine a sufficiently high threshold value $n^* \in \mathbb{N}$ to make the output indeed be an ϵ -approximation.

Nevertheless, the difference between the existence of an impractical approximation algorithm on the one hand, and the non-existence of any approximation algorithm on the other hand, is just one consequence of a more fundamental difference: while in the models considered in Jaeger (2000) query probabilities $P_n(a())$ converge to a limit, this is not necessarily the case for knowledge bases of weighted formulas – at least when full RFOL is allowed: in the proof of Theorem 4.1 we have constructed knowledge bases KB , such that $P_{KB,n}(a())$ oscillates between zero and values $> 1/2$ as n oscillates between $spec(\phi)$ and its complement. The construction of knowledge bases with this behavior does not require formulas ϕ with a non-polynomial spectrum as in Corollary 3.4, and is not contingent on $NETIME \neq ETIME$. Already a knowledge base as constructed in the proof of Theorem 4.1 with ϕ replaced by ψ of Example 3.2 will show this behavior.

The reason behind these different convergence properties lies in a somewhat different role that conditioning on evidence plays in directed and undirected models: in the former, a conditional probability $P_{M,n}(a() \mid b())$ defined by a model M can, in general, not be defined as an unconditional probability $P_{M',n}(a())$ in a modified model M' . As a result, the convergence guarantees and – theoretical – approximability for certain classes of unconditional queries $P_{M,n}(a())$ do not carry over to conditional queries $P_{M,n}(a() \mid b())$.

For weighted feature knowledge bases KB , on the other hand, there is no fundamental difference between unconditional and conditional queries $P_{KB',n}(a())$ and $P_{KB,n}(a() \mid b())$, respectively. To reduce the conditional to unconditional queries, one can just add to KB the hard constraint $\neg b() : 0$ to obtain KB' with $PP_{KB',n} = P_{KB,n} \mid b()$. This means that as long as \mathcal{E} is not more expressive than \mathcal{KB} , the problem classes $PI(\mathcal{KB}, \mathcal{Q}, \mathcal{E})$ and $PI(\mathcal{KB}, \mathcal{Q}, \emptyset)$ have the same characteristics in terms of complexity as a function of the domainsize. Note, though, that this is only true when we consider complexity of $PI(KB, n, \chi, \eta)$ strictly as a function of n for fixed KB, χ, η . If the evidence is allowed to change with the domainsize, i.e., $\eta = \eta(n)$, then even in cases where restrictions on \mathcal{KB} make $PI(\mathcal{KB}, \mathcal{Q}, \mathcal{E})$ polynomial in n , one can define sequences of inference problems $PI(KB, n, \chi, \eta(n))$ with $KB \in \mathcal{KB}$, $\eta(n) \in \mathcal{E}$ that are no longer polynomial in n (Van den Broeck and Davis 2012).

6 Conclusion

We have shown that for currently quite popular probabilistic-logic models consisting of collections of weighted, quantifier-, and function-free formulas there is likely to be no general polynomial lifted inference method (contingent on $NETIME \neq ETIME$).

Somewhat surprisingly, this even holds for approximate inference. Between this negative result and the positive result of (Van den Broeck 2011), there still could be a lot of room for identifying tractable fragments by restricting 0-RFOL further via limits on the number of variables or the richness of signature S .

References

- APSEL, U. AND BRAFMAN, R. I. 2011. Extended lifted inference with joint formulas. In *Proc. of the 27th Conference on Uncertainty in Artificial Intelligence (UAI-11)*. AUAI Press, 11–18.
- BREESE, J. S. 1992. Construction of belief and decision networks. *Computational Intelligence* 8, 624–647.
- DE SALVO BRAZ, R., AMIR, E. AND ROTH, D. 2005. Lifted first-order probabilistic inference. In *Proc. of 19th International Joint Conference on Artificial Intelligence (IJCAI-05)*. Professional Book Center, 1319–1325.
- DOMINGOS, P. AND WEBB, W. A. 2012. A tractable first-order probabilistic logic. In *Proc. of 26th AAAI Conference on Artificial Intelligence (AAAI-12)*. AAAI Press, 1902–1909.
- FAGIN, R. 1976. Probabilities on finite models. *Journal of Symbolic Logic* 41, 50–58.
- FIERENS, D., DEN BROECK, G. V., THON, I., GUTMANN, B. AND DE RAEDT, L. 2011. Inference in probabilistic logic programs using weighted cnf's. In *Proc. of 27th Conference on Uncertainty in Artificial Intelligence (UAI-11)*. AUAI Press, 211–220.
- FRIEDMAN, N., GETOOR, L., KOLLER, D. AND PFEFFER, A. 1999. Learning probabilistic relational models. In *Proc. of 16th International Joint Conference on Artificial Intelligence (IJCAI-99)*. Morgan Kaufmann, 1300–1309.
- GOGATE, V. AND DOMINGOS, P. 2011. Probabilistic theorem proving. In *Proceedings of the 27th Conference of Uncertainty in Artificial Intelligence (UAI-11)*. AUAI Press, 256–265.
- JAEGER, M. 1997. Relational bayesian networks. In *Proc. of 13th Conference of Uncertainty in Artificial Intelligence (UAI-13)*. Morgan Kaufmann, 266–273.
- JAEGER, M. 2000. On the complexity of inference about probabilistic relational models. *Artificial Intelligence* 117, 297–308.
- JAEGER, M. 2012. Lower complexity bounds for lifted inference. Accessed 15 April 2012. URL: <http://arxiv.org/abs/1204.3255>.
- JAEGER, M. AND VAN DEN BROECK, G. 2012. Liftability of probabilistic inference: Upper and lower bounds. In *Proc. of the 2nd International Workshop on Statistical Relational AI*.
- JHA, A., GOGATE, V., MELIOU, A. AND SUCIU, D. 2010. Lifted inference seen from the other side: The tractable features. In *Proc. of 24th Annual Conference on Neural Information Processing Systems (NIPS-10)*. Curran Associates, Inc., 973–981.
- JOHNSON, D. S. 1990. A catalog of complexity classes. In *Handbook of Theoretical Computer Science*, J. van Leeuwen, Ed, vol. 1. Elsevier, Amsterdam, 67–161.
- JONES, N. D. AND SELMAN, A. L. 1972. Turing machines and the spectra of first-order formulas with equality. In *Proc. of 4th ACM Symposium on Theory of Computing*. ACM, 157–167.
- KERSTING, K. AND DE RAEDT, L. 2001. Towards combining inductive logic programming with Bayesian networks. In *Proc. of 11th International Conference on Inductive Logic Programming (ILP-01)*. LNAI, vol. 2157. Springer, 118–131.
- KISYŃSKI, J. AND POOLE, D. 2009. Lifted aggregation in directed first-order probabilistic models. In *Proc. of 21st International Joint Conference on Artificial Intelligence (IJCAI-09)*. IJCAI Organization, 1922–1929.
- MILCH, B., MARTHI, B., RUSSELL, S., SONTAG, D., ONG, D. AND KOLOBOV, A. 2005. Blog: Probabilistic logic with unknown objects. In *Proc. of the 19th International Joint Conference on Artificial Intelligence (IJCAI-05)*. Professional Book Center, 1352–1359.

- MILCH, B., ZETTMAYER, L. S., KERSTING, K., HAIMES, M. AND KAEHLING, L. P. 2008. Lifted probabilistic inference with counting formulas. In *Proc. of 23rd AAAI Conference on Artificial Intelligence (AAAI-08)*. AAAI Press, 1062–1068.
- NGO, L., HADDAWY, P. AND HELWIG, J. 1995. A theoretical framework for context-sensitive temporal probability model construction with application to plan projection. In *Proc. of 11th Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann, 419–426.
- POOLE, D. 1993. Probabilistic horn abduction and Bayesian networks. *Artificial Intelligence* 64, 81–129.
- POOLE, D. 2003. First-order probabilistic inference. In *Proc. of 18th International Joint Conference on Artificial Intelligence (IJCAI-03)*. Morgan Kaufmann, 985–991.
- RICHARDSON, M. AND DOMINGOS, P. 2006. Markov logic networks. *Machine Learning* 62, 107–136.
- SATO, T. 1995. A statistical learning method for logic programs with distribution semantics. In *Proc. of 12th International Conference on Logic Programming (ICLP'95)*. MIT Press, 715–729.
- SINGLA, P., NATH, A. AND DOMINGOS, P. 2010. Approximate lifted belief propagation. In *Proc. of AAAI-10 Workshop on Statistical Relational AI*.
- TASKAR, B., ABBEEL, P. AND KOLLER, D. 2002. Discriminative probabilistic models for relational data. In *Proc. of 18th Conference in Uncertainty in Artificial Intelligence (UAI-02)*. Morgan Kaufmann, 485–492.
- VAN DEN BROECK, G. 2011. On the completeness of first-order knowledge compilation for lifted probabilistic inference. In *Proc. of 25th Annual Conference on Neural Information Processing Systems (NIPS)*. 1386–1394.
- VAN DEN BROECK, G. AND DAVIS, J. 2012. Conditioning in first-order knowledge compilation and lifted probabilistic inference. In *Proc. of 26th AAAI Conference on Artificial Intelligence*. AAAI Press, 1961–1967.
- VAN DEN BROECK, G., TAGHIPOUR, N., MEERT, W., DAVIS, J. AND DE RAEDT, L. 2011. Lifted probabilistic inference by first-order knowledge compilation. In *Proc. of 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*. IJCAI/AAAI, 2178–2185.
- VENNEKENS, J., DENECKER, M. AND BRUYNNOGHE, M. 2006. Representing causal information about a probabilistic process. In *Logics in Artificial Intelligence, 10th European Conference, JELIA 2006, Proceedings*. Lecture Notes in Computer Science, vol. 4160. Springer-Verlag, Berlin, 452–464.

Appendix A: Proofs

Proof of Proposition 4.2: We begin by defining the *term-depth* of a term t in signature S^F as the maximal nesting depth of function symbols in t . Precisely, we define inductively: if $t \equiv x$, then t has term depth 0. If $t \equiv f()$ (a constant), or $t = f(x_1, \dots, x_k)$ (a function term with only variables as arguments), then t has term depth 1. If $t = f(t_1, \dots, t_k)$, then the term depth of t is one plus the maximal term depth of the t_i .

The term depth of a formula $\phi(\mathbf{x})$ is the maximal term depth of the terms it contains.

We now show that every formula $\phi(\mathbf{x})$ of term depth l can be transformed into a formula $\phi^{l-1}(\mathbf{x}, \mathbf{z})$ of term depth $l-1$ in 0-FOL($S \cup S^F \cup S^+$), such that the statement for ϕ^+ of the proposition holds for ϕ^{l-1} (but with $S \cup S^F \cup S^+$ instead of $S \cup S^+$ in **(ii)**). The proposition then follows by defining ϕ^+ as the result of iteratively applying

l such transformations to ϕ . Since the term depth of the resulting ϕ^+ is zero, then actually $\phi^+(\mathbf{x}, \mathbf{z}) \in 0\text{-RFOL}(S \cup S^+)$.

Let $\{f_i(\mathbf{x}_i) \mid i = 1, \dots, r\}$ be the set of all distinct terms (including sub-terms) of depth 1 appearing in $\phi(\mathbf{x})$. Let z_1, \dots, z_r be new variables. Then $\phi^{l-1}(\mathbf{x}, \mathbf{z})$ would be defined as

$$\bigwedge_{i=1}^r R^{f_i}(\mathbf{x}_i, z_i) \rightarrow \phi(\mathbf{x})[z_1/f_1(\mathbf{x}_1), \dots, z_r/f_r(\mathbf{x}_r)].$$

To now show **(i)** \Rightarrow **(ii)**, let $I \in \mathcal{I}(n, S \cup S^F)$ with $I \models \forall \mathbf{x} \phi(\mathbf{x})$. Define $I^+ \in \mathcal{I}(n, S \cup S^F \cup S^+)$ as the expansion of I in which each $R^f \in S^+$ is interpreted as the relational representation of f , i.e., $I^+ \models R^f(\mathbf{d}, e)$ iff $I \models f(\mathbf{d}) = e$. Clearly, $I^+ \models \text{Func}$. Furthermore, the following are equivalent:

$$\begin{aligned} I &\models \forall \mathbf{x} \phi(\mathbf{x}) \\ I &\models \forall \mathbf{xz} \bigwedge_{i=1}^r f_i(\mathbf{x}_i) = z_i \\ &\quad \rightarrow \phi(\mathbf{x})[z_1/f_1(\mathbf{x}_1), \dots, z_r/f_r(\mathbf{x}_r)] \\ I^+ &\models \forall \mathbf{xz} \bigwedge_{i=1}^r R^{f_i}(\mathbf{x}_i, z_i) \\ &\quad \rightarrow \phi(\mathbf{x})[z_1/f_1(\mathbf{x}_1), \dots, z_r/f_r(\mathbf{x}_r)] \end{aligned}$$

For **(ii)** \Rightarrow **(i)** let I^+ as in **(ii)** be given. Since $I^+ \models \text{Func}$, we can turn I^+ into an interpretation for $S \cup S^F$ by defining $f(\mathbf{d})$ as the unique e for which $R^f(\mathbf{d}, e)$ holds in I^+ . Then, by the same equivalences as above, $I^+ \models \forall \mathbf{xz} \phi^+(\mathbf{x}, \mathbf{z})$ implies $I \models \forall \mathbf{x} \phi(\mathbf{x})$. □

Proof of Theorem 4.3: Let $\phi \in \text{RFOL}$ as given by Corollary 3.4 and $\forall \mathbf{x} \phi^{R\text{-Skol}}(\mathbf{x})$ its relational Skolemization. Let S be the original signature of ϕ and S^+ the relation symbols introduced in the relational Skolemization. Furthermore, for each k -ary $R^+ \in S^+$ we introduce a new $(k - 1)$ -ary relation, R^{++} . These new symbols will be used to calibrate the weight of models for the reference proposition $b()$. Note that the arity of symbols in S^+ is at least 1, and R^{++} , thus, is well-defined, but may contain relations of arity 0. We denote with S^{++} the collection of all the introduced R^{++} symbols. We now reduce the spectrum recognition problem for ϕ to probabilistic inference from a knowledge base in the signature $S \cup S^+ \cup S^{++} \cup \{a(), b()\}$.

The first formula in our knowledge base is

$$a() \wedge \neg \phi^{R\text{-Skol}}(\mathbf{x}) \quad : \quad 0. \tag{A1}$$

We now approximately axiomatize the functional nature of the symbols $R^+ \in S^+$. The sentence (8) can be directly encoded as a weighted formula:

$$R^+(\mathbf{x}, y) \wedge R^+(\mathbf{x}, y') \wedge y \neq y' \quad : \quad 0 \tag{A2}$$

Next, we would like to enforce (9) by means of a weighted formula. However, (9) encodes the essence of the existential quantifiers we are about to eliminate, and thus, it is not surprising that this is not possible to enforce strictly. However, we can reward models in which the existential quantification of (9) is satisfied via the weighted formulas

$$a() \wedge R^+(\mathbf{x}, y) \quad : \quad w \quad (R^+ \in S^+), \tag{A3}$$

where $w > 1$ is a weight whose exact value is to be defined later.

We now proceed with constraining models of the reference proposition $b()$. First, all symbols in $S \cup S^+$ shall have an empty interpretation in models of $b()$:

$$b() \wedge R(x) : 0 \quad (R \in S), \tag{A4}$$

$$b() \wedge R^+(x, y) : 0 \quad (R^+ \in S^+). \tag{A5}$$

In order to allow $b()$ -models to gain some weight, we use the extra symbols in S^{++} :

$$b() \wedge R^{++}(x) : w \quad (R^{++} \in S^{++}), \tag{A6}$$

where w is the same weight as in (A3). To further limit the possible interpretations of $b()$ -models, we also stipulate

$$b() \wedge \neg R^{++}(x) : 0 \quad (R^{++} \in S^{++}). \tag{A7}$$

The extra symbols R^{++} must have empty interpretations in $a()$ -models:

$$a() \wedge R^{++}(x) : 0 \quad (R^{++} \in S^{++}). \tag{A8}$$

Finally, we add

$$\neg(a() \vee b()) : 0. \tag{A9}$$

We now determine (approximately) $W_{KB,n}(a())$ and $W_{KB,n}(b())$ for the cases $n \in spec(\phi)$ and $n \notin spec(\phi)$.

First, consider $b()$: for any n , there exists exactly one interpretation $I_{b()} \in \mathcal{I}(D_n, S \cup S^+ \cup S^{++} \cup \{a(), b()\})$ with nonzero weight in which $b()$ is true. This is the interpretation in which all relations in $S \cup S^+$ are empty ((A4), (A5)), all relations in S^{++} are maximal (A7), and, in consequence of the latter, because of (A8), $a()$ is false.

Assume that $S^+ = \{R_1^+, \dots, R_m^+\}$, where R_i^+ has arity $k_i + 1$. Then $R_i^{++} \in S^{++}$ contributes via (A6), a factor of $w^{n^{k_i}}$, to $W_{KB,n}(I_{b()})$, and the total weight is

$$W_{KB,n}(I_{b()}) = W_{KB,n}(b()) = w^{n^{k_1} + \dots + n^{k_m}} = w^{K(n)}, \tag{A10}$$

using for abbreviation $K(n) := n^{k_1} + \dots + n^{k_m}$.

We next turn to $W_{KB,n}(a())$ in the case $n \in spec(\phi)$. Then there exists at least one interpretation $I \in \mathcal{I}(D_n, S \cup S^+)$, in which $\forall x \phi^{R-Skol}(x)$ is true and the relations from S^+ have a functional interpretation. We can expand this interpretation to an interpretation in $\mathcal{I}(n, S \cup S^+ \cup S^{++} \cup \{a(), b()\})$ by giving all relations in S^{++} an empty interpretation and setting $a()$ to true and $b()$ to false. Then I does not violate any hard constraint in KB and collects from (A3) a total weight of $w^{K(n)}$. Thus

$$W_{KB,n}(a()) \geq w^{K(n)},$$

and therefore, when $n \in spec(\phi)$,

$$P_{KB,n}(a()) \geq W_{KB,n}(a()) / (W_{KB,n}(a()) + W_{KB,n}(b())) \geq 1/2. \tag{A11}$$

Finally, we have to consider $W_{KB,n}(a())$ in the case $n \notin spec(\phi)$. For any I with non-zero weight in which $a()$ is true, because of (A1), also $\forall x \phi^{R-Skol}(x)$ must be true. This, now, only is possible when some $R^+ \in S^+$ is not a functional relation, which, because of (A2) can only mean that for some x there exists no y with $R^+(x, y)$. The

total weight of I accrued from (A3) then is at most $w^{K(n)-1}$. Because of (A8), I cannot obtain any additional weight from (A6), so that

$$W_{KB,n}(I) \leq w^{K(n)-1}. \tag{A12}$$

The total number of interpretations in $\mathcal{I}(D_n, S \cup S^+ \cup S^{++} \cup \{a(), b()\})$ is $2^{L(n)}$ for a polynomial $L(n)$. Thus

$$W_{KB,n}(a()) \leq 2^{L(n)} w^{K(n)-1}. \tag{A13}$$

We now obtain for the case $n \notin \text{spec}(\phi)$

$$P_{KB,n}(a()) \leq W_{KB,n}(a()) / W_{KB,n}(b()) \leq 2^{L(n)} w^{K(n)-1} / w^{K(n)} = 2^{L(n)} / w. \tag{A14}$$

Setting $w = 10 \cdot 2^{L(n)}$, we thus have $P_{KB,n}(a()) \leq 1/10$ if $n \notin \text{spec}(\phi)$. The representation size of w is polynomial in n . Thus, an algorithm that computes $P_{KB,n}(a())$ up to an accuracy of $0.2 = (0.5 - 0.1)/2$ in time polynomial in n and the representation size of w would give a polynomial time decision procedure for $\text{spec}(\phi)$. \square

Proof of Theorem 4.4: The proof is an extension of the proof of Theorem 4.3 and we here just give the necessary modifications.

Let E be a new binary relation symbol. We replace equalities $x = y$ in (A1) and (A2) with $E(x, y)$. To (approximately) axiomatize E as the identity relation in models of $a()$, we add to the knowledge base consisting of (A1)–(A9) the weighted formulas

$$a() \wedge \neg E(x, x) \quad 0, \tag{A15}$$

$$a() \wedge E(x, y) \quad 1/w, \tag{A16}$$

where $w > 1$ is the same weight as in (A3) and (A6), and whose exact value is to be determined later. To calibrate the weight of $b()$ -models, we introduce in analogy to the R^{++} relations a unary relation E^{++} , and in analogy to (A6)–(A8) add to the knowledge base

$$b() \wedge E^{++}(x) \quad 1/w, \tag{A17}$$

$$b() \wedge \neg E^{++}(x) \quad 0, \tag{A18}$$

$$a() \wedge E^{++}(x) \quad 0. \tag{A19}$$

We now obtain for all n ,

$$W_{KB,n}(b()) = w^{K(n)} (1/w)^n = w^{K(n)-n}. \tag{A20}$$

If $n \in \text{spec}(\phi)$, then there exists an interpretation in which $a()$ is true, R^+ have a functional interpretation, and the interpretation of E is the identity relation. We can thus lower-bound the weight of $a()$ by the weight of that interpretation:

$$W_{KB,n}(a()) \geq w^{K(n)} (1/w)^n = w^{K(n)-n}. \tag{A21}$$

As in (A11), one then obtains $P_{KB,n}(a()) \geq 1/2$.

We now turn to the case $n \notin \text{spec}(\phi)$. Consider any I in which $a()$ is true and that has nonzero weight. This now is only possible when in I there is an $R^+ \in S^+$ which is not a functional relation, or when E is not the identity relation in I (or both).

In all cases, the weight of I coming from (A3) and (A16) is at most $w^{K(n)-n-1}$. The total number of interpretations in $\mathcal{I}(D_n, S \cup S^+ \cup S^{++} \cup \{a(), b(), E\})$ is $2^{M(n)}$ for a polynomial $M(n)$. Thus

$$W_{KB,n}(a()) \leq 2^{M(n)} w^{K(n)-n-1}, \quad (\text{A22})$$

from which, as in (A14), then $P_{KB,n}(a()) \leq 2^{M(n)}/w$. Now setting $w = 10 \cdot 2^{M(n)}$ again yields the bound $P_{KB,n}(a()) \leq 1/10$.

□