Reasoning About Infinite Random Structures with Relational Bayesian Networks

Manfred Jaeger
Max-Planck-Institut für Informatik
Im Stadtwald, 66123 Saarbrücken, Germany
jaeger@mpi-sb.mpg.de

Abstract

Relational Bayesian networks extend standard Bayesian networks by integrating some of the expressive power of first-order logic into the Bayesian network paradigm. As in the case of the related technique of knowledge based model construction, so far, decidable semantics only have been provided for finite stochastic domains. In this paper we extend the semantics of relational Bayesian networks, so that they also define probability distributions over countably infinite structures. Using a technique remeniscent of quantifier elimination methods in model theory, we show that probabilistic queries about these distributions are decidable.

1 INTRODUCTION

Bayesian networks (Pearl 1988) currently are the most popular and successful framework for representing and reasoning with probabilistic information. In their basic form, Bayesian networks define a probability distribution over the set of possible values of a finite set of random variables $X_1, \ldots, X_n$, each variable having a finite range of possible values.

Semantically, Bayesian networks describe *attributes* of individual random events or random objects, e.g. the symptoms and diseases of a random patient, or state variables describing a robot and its environment. For many applications the restriction to a fixed number of finite range random variables is a severe limitation. One of the approaches to go beyond these limits is the technique of *knowledge based model construction* (Wellman, Breese & Goldman 1992, Breese 1992, Haddawy 1994, Ngo & Haddawy 1995), where the number of variables is adjusted on a case-by-case basis. The basic idea here is to model the probabilistic domain not by a Bayesian network directly, but to use a knowledge base containing probabilistic rules that are used as a blueprint for the construction of Bayesian networks tailored to each specific inference task.

A simple example of a rule that such a knowledge base might contain is

$$\textit{fever}(v) \stackrel{0.8}{\longleftarrow} \textit{flu}(v), \tag{1}$$

with the intended meaning: if v suffers from the flu, then v will have fever with probability 0.8. A knowledge base containing such rules, and ground atoms *flu(thomas)*, *fever(sylvia)*, ..., gives rise to a standard Bayesian network over binary random variables representing all ground atoms relevant for the processing of a specific query.

When all the rules in the knowledge base only contain a single variable (as in (1)), then this rule-based approach only amounts to a notational variant of standard Bayesian networks, because information about one object mentioned in the knowledge base will not influence our inferences for another. Probabilistic knowledge bases gain their edge in expressive power over Bayesian networks by the ability to also define rules involving two or more variables, and n- ary relation symbols:

$$\textit{infects}(u, v) \stackrel{0.6}{\longleftarrow} \textit{sick}(u) \wedge \textit{contact}(u, v). \tag{2}$$

A rule like (2) no longer merely describes the attributes of single random objects or events, but specifies *relations* between multiple random objects/events. A knowledge base with rules of this form, and ground atoms *infects(thomas,sylvia)*, *sick(thomas)*,... again can be used to generate a Bayesian network over ground atoms. Size and structure of this network now will depend on the number of constants appearing in the knowledge base. No single standard Bayesian network with a fixed number of finite-range random variables can be defined that supports the same inferences as can be drawn from a rule base with multi-variable rules.

In (Jaeger 1997) it was argued that the frameworks for knowledge based model construction proposed so far suffered from two deficiencies: first, they lack in expressiveness, second, the semantic clarity of the Bayesian network paradigm is lost, because the declarative character of the rules makes it hard to distill their meaning into a single probability distribution.

The issue of expressiveness was discussed at some length in (Jaeger 1997). The issue of semantic transparency it may be worthwhile to here elaborate on somewhat more. One of the main advantages of the Bayesian network paradigm is that it provides the user with a firm framework for how to describe a probability distribution: he or she is asked to specify the underlying directed acyclic graph, and the conditional probability tables for all nodes. Providing the required information then guarantees the user that one, and only one, probability distribution is being defined. Moreover, following this procedure gives the user a fairly well-understood control over the distribution he or she is defining. Probabilistic rules, like (1) or (2), on the other hand, only impose certain constraints on the probability distribution described. These constraints, hopefully, have a fairly intuitive meaning for the user, but will require a substantial overhead of semantic definitions and conventions in order to be given an exact interpretation. Breese (1992), for instance, gives the semantics of his knowledge bases indirectly by defining a procedure that, given a knowledge base and a specific query, constructs a Bayesian network in which the probability of the query is determined. Ngo and Haddawy (1995) provide declarative semantics for their representation formalism. However, a knowledge base in their language only defines a unique distribution if certain consistency and completeness conditions are satisfied. It is not clear that it is decidable whether these conditions are fulfilled, so that, in general, it may be impossible for the user to tell whether a knowledge base has semantics at all.

Relational Bayesian networks (RBNs) were proposed in (Jaeger 1997) as an alternative approach for the specification of probability distributions on relations between several random objects. Apart from providing additional expressiveness, RBNs recapture the semantic transparency afforded by the Bayesian network paradigm of probabilistic model construction (but see section 5 for a qualification).

In the present paper we are going to explore another advantage of RBNs: their semantics can be extended to define probability distributions for infinite domains of random objects, and queries for these distributions are still decidable. The framework of Ngo and Haddawy (1995) also is defined for infinite domains. Specifically, their distributions are defined for Herbrand universes as domains, which are infinite if the underlying first-order vocabulary contains at least one constant and one function symbol. Herbrand universes have a richer internal structure than the infinite domains for which the central results of the present paper are obtained. A result given in section 5 indicates why this structure on the domain might lead to inherent undecidability of Ngo and Haddawy's approach.

2 FINITE DOMAINS

2.1 THE BASIC FRAMEWORK

In this section we review the basic definitions introduced in (Jaeger 1997), extending them, where necessary, to deal with the case of infinite domains treated in the subsequent sections.

The purpose of an RBN is to define a probability distribution that models random attributes and random relations in a domain of objects or events. For the time being, assume that this domain $D = \{d_1, \ldots, d_n\}$ is finite. A probabilistic model for a set of relations $S = \{r_1, \ldots, r_k\}$ on this domain then consists of a probability distribution on the set of S-structures (a.k.a. models) with domain D, denoted $\mathrm{Mod}_D(S)$. Relations in S can be of any arity. The arity of r_i is denoted $|\,r_i\,|$. For a single relation r_i, the set $\mathrm{Mod}_D(r_i)$ of possible r_i-structures comprises just the set of possible interpretations $I(r_i) \subseteq D^{|\,r_i\,|}$ of r_i in D. Hence, $\mathrm{Mod}_D(r_i)$ can be identified with the powerset of $D^{|\,r_i\,|}$. Moreover, an S-structure $\mathscr{M} \in \mathrm{Mod}_D(S)$ is given by a tuple $(I(r_1), \ldots, I(r_k))$ of interpretations, so that $\mathrm{Mod}_D(S)$ can be identified with the Cartesian product

$$\mathrm{Mod}_D(S) = \mathrm{Mod}_D(r_1) \times \ldots \times \mathrm{Mod}_D(r_k). \qquad (3)$$

Thus, a probability distribution on $\mathrm{Mod}_D(S)$ can be defined in the form of a joint distribution for the individual $\mathrm{Mod}_D(r_i)$. Viewing each $\mathrm{Mod}_D(r_i)$ as a random variable, such a joint distribution can be defined following the Bayesian network paradigm: specify a directed acyclic graph with a node for each r_i, and at r_i define the conditional probability of each possible interpretation of r_i, given the instantiation of the parent nodes of r_i.

Assume that a directed acyclic graph has been given. For the node r_i denote by $\mathit{Pa}(r_i) = \{r_{j_1}, \ldots, r_{j_m}\}$ the set of parent nodes of r_i in the graph. Also, denote by $\mathscr{M}_i = (I(r_{j_1}), \ldots, I(r_{j_m}))$ some given $\mathit{Pa}(r_i)$-structure, which, in Bayesian network terminology, is just an instantiation of the parent nodes of r_i. For nodes without parents we interpret $\mathscr{M}_i$ simply as the given domain D. Following our program for defining a distribution on $\mathrm{Mod}_D(S)$, we then have to define the conditional probability

$$P(I(r_i) \mid \mathscr{M}_i) \qquad (4)$$

for every $I(r_i) \in \mathrm{Mod}_D(r_i)$ and $\mathscr{M}_i \in \mathrm{Mod}_D(\mathit{Pa}(r_i))$. Given some finite D, both $\mathrm{Mod}_D(r_i)$ and $\mathrm{Mod}_D(\mathit{Pa}(r_i))$

are finite, so that, in principle, all probabilities (4) could be explicitly listed in a huge conditional probability table. This, of course, is infeasible due to the size of such a table. More importantly, however, we aim for a generic definition that does not refer to a specific domain. To obtain such a generic definition, we define a schematic specification of conditional probabilities (4) that work "tuple by tuple", i.e. that determines for each $\boldsymbol{d} \in D^{|r_i|}$ the conditional probability

$$P(\boldsymbol{d} \in I(r_i) \mid \mathscr{M}_i). \tag{5}$$

Subsequently, we put

$$P(I(r_i) \mid \mathscr{M}_i) := \prod_{\boldsymbol{d} \in I(r_i)} P(\boldsymbol{d} \in I(r_i) \mid \mathscr{M}_i) \prod_{\boldsymbol{d} \notin I(r_i)} (1 - P(\boldsymbol{d} \in I(r_i) \mid \mathscr{M}_i)). \tag{6}$$

Conditional probabilities (5) we define by *probability formulas*, which are the key components of RBNs. To motivate the following definitions, consider a very simple example. Suppose we have the relation symbols *contact*(u, v) and *sick*(u), and want to model random structures in which the predicate *sick* depends on *contact*. Specifically, assume that for each person a, and for each $b \neq a$, *contact*(a, b) is understood to be a possible cause for *sick*(a) – each instance *contact*(a, b) causing *sick*(a) with probability 0.1, independently for different b. Then the overall probability of *sick*(a), given complete information about *contact*, would be computed by combining probability values 0.1 by noisy-or for all instances *contact*(a, b):

$$P(\textit{sick}(a) \mid I(\textit{contact})) = \textit{n-o}\{0.1\textit{contact}(a, b) \mid b \neq a\}.$$

The expression on the right hand side of this equation is an (informal) example of a probability formula. Key ingredients of probability formulas are *combination functions* (such as noisy-or) that are applied to multisets.

Definition 2.1 A *countable multiset* over $[0, 1]$ is a mapping $A : [0, 1] \to \{0, 1, 2, \ldots\} \cup \{\omega\}$, where $A(q) > 0$ for at most countably many $q \in [0, 1]$. For countable multisets A, B we say that A is a subset of B ($A \subseteq B$), iff $A(q) \leq B(q)$ for all q. The supremum of a chain $(A_i)_{i \in \omega}$ ($A_i \subseteq A_j$ for $i \leq j$) of multisets is the multiset with $A(q) = \textit{sup}\{A_i(q) \mid i \in \omega\}$. The multiset A is called finite, iff $A(q) \neq \omega$ for all q, and $A(q) > 0$ for only finitely many q.

For the present section only finite multisets are needed. We use the following notations to describe specific multisets: if $q_i \in [0, 1]$ for i from some countable index set I, then $\{\!|q_i \mid i \in I|\!\}$ denotes the multiset A with $A(q) = |\{i \mid q = q_i\}|$. Still more concretely, $\{\!|q_1 : \lambda_1, \ldots, q_m : \lambda_m|\!\}$ denotes the multiset A with $A(q_i) = \lambda_i$, $A(q) = 0$ for $q \notin \{q_1, \ldots, q_m\}$.

Definition 2.2 A *combination function* is any function that maps finite multiset over [0,1] into [0,1]. A combination function *comb* is *defined for countable multisets*, iff for every countable multiset A, and all chains $(A_i)_{i \in \omega}, (A'_i)_{i \in \omega}$ of finite multisets with $A = \textit{sup}\{A_i \mid i \in \omega\} = \textit{sup}\{A'_i \mid i \in \omega\}$, we have $\textit{sup}\{\textit{comb}(A_i) \mid i \in \omega\} = \textit{sup}\{\textit{comb}(A'_i) \mid i \in \omega\} =: \textit{comb}(A)$.

Interesting examples of combination functions are

$$\begin{array}{lll} \textit{noisy-or}: & \textit{n-o}\{\!|a_i \mid i \in I|\!\} & := 1 - \prod_{i \in I}(1 - a_i) \\ \textit{maximum}: & \textit{max}\{\!|a_i \mid i \in I|\!\} & := \textit{max}\{a_i \mid i \in I\} \\ \textit{mean}: & \textit{mean}\{\!|a_i \mid i \in I|\!\} & := \frac{1}{|I|}\sum_{i \in I} a_i \end{array}$$

It is easy to see that whenever *comb* is monotonically increasing (decreasing), i.e. $A \subseteq B$ implies $\textit{comb}(A) \leq (\geq)\textit{comb}(B)$, then *comb* is defined for countable multisets. Thus, noisy-or and max are defined for countable multisets. The mean, on the other hand, is not.

We need to introduce some notational conventions: logical variables (as opposed to random variables) are denoted throughout by letters u, v, w, z. Tuples $(v_1, \ldots, v_l)$ of variables are represented by a single letter $\boldsymbol{v}$ in bold face. The length of the tuple $\boldsymbol{v}$ is denoted $|\boldsymbol{v}|$. We also interpret $\boldsymbol{v}$ loosely as the set of variables it contains, so that expressions like $u \in \boldsymbol{v}$, or $\boldsymbol{w} \subseteq \boldsymbol{v}$ make sense. An *equality constraint* for $\boldsymbol{v}$ is a logical formula $c(\boldsymbol{v})$ that is a boolean combination of equality expressions $v_i = v_j$.

Definition 2.3 The class of *probability formulas* over the relational vocabulary S is inductively defined as follows.

(i) (Constants) Each rational number $q \in [0, 1]$ is a probability formula.

(ii) (Indicator functions) For every n-ary symbol $r \in S$, and every n-tuple $\boldsymbol{v}$ of variables, $r(\boldsymbol{v})$ is a probability formula.

(iii) (Convex combinations) When F_1, F_2, F_3 are probability formulas, then so is $F_1F_2 + (1 - F_1)F_3$.

(iv) (Combination functions) When $F_1, \ldots, F_k$ are probability formulas, *comb* is any combination function, $\boldsymbol{v}$,$\boldsymbol{w}$ are tuples of variables, and $c(\boldsymbol{v}, \boldsymbol{w})$ is an equality constraint, then $\textit{comb}\{\!|F_1, \ldots, F_k \mid \boldsymbol{w}; c(\boldsymbol{v}, \boldsymbol{w})|\!\}$ is a probability formula.

A probability formula $F(\boldsymbol{v})$ over S maps $|\boldsymbol{v}|$- tuples $\boldsymbol{d}$ from the domain of a finite S-structure $\mathscr{M}$ into [0,1], according to the following definition.

Definition 2.4 Let $F(\boldsymbol{v})$ be a probability formula over S, D a finite domain, $\mathscr{M} \in \mathrm{Mod}_D(S)$, and $\boldsymbol{d} \in D^{|\boldsymbol{v}|}$. We inductively define the value $F(\boldsymbol{d})[\mathscr{M}]$.

(i) if $F = q$ then $F(\boldsymbol{d})[\mathscr{M}] = q$.

(ii) if $F = r(\boldsymbol{v})$ then $F(\boldsymbol{d})[\mathscr{M}] = 1$ if $\mathscr{M} \models r(\boldsymbol{d})$, and $F(\boldsymbol{d})[\mathscr{M}] = 0$ else.

(iii) if $F = F_1F_2 + (1 - F_1)F_3$ then $F(\boldsymbol{d})[\mathscr{M}] = F_1(\boldsymbol{d})[\mathscr{M}]F_2(\boldsymbol{d})[\mathscr{M}] + (1 - F_1(\boldsymbol{d})[\mathscr{M}])F_3(\boldsymbol{d})[\mathscr{M}]$.

(iv) if $F = \textit{comb}\{\!|F_1, \ldots, F_k \mid \boldsymbol{w}; c(\boldsymbol{v}, \boldsymbol{w})|\!\}$ then $F(\boldsymbol{d})[\mathscr{M}] = \textit{comb}A$, where A is the multiset that for $q \in [0, 1]$ has

$$A(q) = \ |\{(i, \boldsymbol{d}') \mid i \in \{1, \ldots, k\}, \boldsymbol{d}' \in D^{|\boldsymbol{w}|}; \mathscr{M} \models c(\boldsymbol{d}, \boldsymbol{d}'), q = F_i(\boldsymbol{d}, \boldsymbol{d}')[\mathscr{M}]\}| \,.$$

If F only contains combination functions that are defined for countable multisets, then definition 2.4 also extends to countably infinite D. The following example illustrates case (iv) in the above definition.

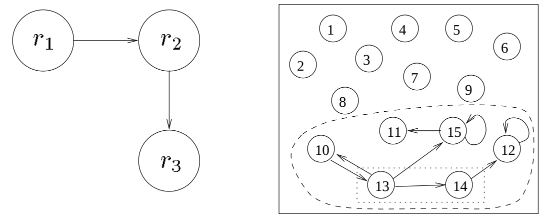
Example 2.5 Let

$$F(v) = \textit{comb}\{\!|0.3r(v), 0.7s(v, w) \mid w; w \neq v|\!\}. \qquad (7)$$

Let $\mathscr{M}$ be an $\{r, s\}$-structure with domain $D = \{d_1, \ldots, d_8\}$; let the interpretation of r in $\mathscr{M}$ be $\{d_1, \ldots, d_4\}$, and let $\{(d_1, d_7), (d_1, d_8)\}$ be that part of the interpretation of s in $\mathscr{M}$ that has d_1 in the first component. To evaluate $F(d_1)[\mathscr{M}]$ we proceed as follows. First, we generate a list of all elements w of the domain that satisfy the constraint $w \neq d_1$. The result is $d_2, \ldots, d_8$. For each tuple (d_1, d'), $d' \in \{d_2, \ldots, d_8\}$, we compute $0.3r(v)[d_1, d'][\mathscr{M}]$ and $0.7s(v, w)[d_1, d'][\mathscr{M}]$. The notation $r(v)[d_1, d'], s(v, w)[d_1, d']$, rather than $r(d_1), s(d_1, d')$, here is used to emphasize that according to definition 2.4 we count each substitution of a tuple $(\boldsymbol{d}, \boldsymbol{d}')$ for the variables $(\boldsymbol{v}, \boldsymbol{w})$ in the F_i separately, no matter whether F_i actually contains the variables for which different values are substituted. The results of these recursive evaluations are 0.3 for the first formula, and 0 for the second, when $d' \in \{d_2, \ldots, d_6\}$, and 0.3, respectively 0.7, for $d' \in \{d_7, d_8\}$. Here, the multiset A in (iv) thus is $\{\!|0.3 : 7, 0 : 5, 0.7 : 2|\!\}$. Applying *comb* to A then yields the result $F(d_1)[\mathscr{M}]$.

The following lemma contains a very useful result on the expressiveness of probability formulas. The simple proof can be found in (Jaeger 1997).

Lemma 2.6 Let $\phi(\boldsymbol{v})$ be a first-order formula in the relational vocabulary S. Then there exists a probability formula $F_\phi(\boldsymbol{v})$ in S, that uses *max* as the only combination function, s.t. for every finite S-structure $\mathscr{M}$, and every $\boldsymbol{d} \in D^{|\boldsymbol{v}|}$: $F_\phi(\boldsymbol{d}) = 1$ iff $\mathscr{M} \models \phi(\boldsymbol{d})$, and $F_\phi(\boldsymbol{d}) = 0$ else.



(a) (b)

Figure 1: A relational Bayesian network and a typical structure

Definition 2.7 A *relational Bayesian network* for the (relational) vocabulary S is given by a directed acyclic graph containing one node for every $r \in S$. The node for an n-ary $r \in S$ is labeled with a probability formula $F_r(v_1, \ldots, v_n)$ over the symbols in *Pa(r)*.

Given an RBN N and a finite domain D, we define the conditional probability $P^N_D(I(r_i) \mid \mathscr{M}_i)$ by substituting $F_{r_i}(\boldsymbol{d})[\mathscr{M}_i]$ for $P(\boldsymbol{d} \in I(r_i) \mid \mathscr{M}_i)$ in (6). This finally leads to the definition of the semantics of an RBN.

Definition 2.8 Let N be a relational Bayesian network over S, D a finite domain. N defines a probability measure P^N_D on $\mathrm{Mod}_D(S)$ by

$$P^N_D(I(r_1), \ldots, I(r_k)) := \prod_{i=1}^{k} P^N_D(I(r_i) \mid \mathscr{M}_i). \qquad (8)$$

We conclude this section with a very small example that we will refer to later on.

Example 2.9 Let $S = \{r_1, r_2, r_3\}$, where r_1 and r_3 are unary, r_2 is binary. An RBN N over S is defined via the graph in figure 1 (a), and the probability formulas

$$F_{r_1}(v) \equiv 0.4$$
$$F_{r_2}(v, w) \equiv 0.2r_1(v)r_1(w)$$
$$F_{r_3}(v) \equiv \textit{n-o}\{\!|0.3r_2(v, w) \mid w; w \neq v|\!\}$$

Figure 1 (b) shows one representative S-strucutre $\mathscr{M}$ over a domain D of 15 elements. The interpretation of r_1 in $\mathscr{M}$ is delimited in the figure by a dashed line, that of r_3 by a dotted line, and that of r_2 is represented by arrows. The probability of $\mathscr{M}$ is computed according to (8) by computing the three factors $P^N_D(I(r_i) \mid \mathscr{M}_i)$ $(i = 1, 2, 3)$, where $\mathscr{M}_3$ is the $\{r_2\}$-reduct of $\mathscr{M}$, $\mathscr{M}_2$ is the $\{r_1\}$-reduct, and $\mathscr{M}_1$ is just D. For each d_i $(i = 1, \ldots, 15)$ in D we have $F_{r_1}(d_i)[\mathscr{M}_1] = 0.4$, so that according to (6), $P^N_D(I(r_1) \mid \mathscr{M}_1) = 0.4^6 \cdot 0.6^9$. The value of $F_{r_2}(d_i, d_j)[\mathscr{M}_2]$ is 0.2 when $i, j \in \{10, \ldots, 15\}$, and 0 when at least one of i or j is not in $\{10, \ldots, 15\}$. Thus, $P^N_D(I(r_2) \mid \mathscr{M}_2) = 0.2^6 \cdot 0.8^{30}$. The value of $F_{r_3}(d_i)[\mathscr{M}_3]$ is $1 - 0.7^{k_i}$, where k_i is the number of elements $d_j \neq d_i$ with $r_2(d_i, d_j)$. Thus, we get

$F_{r_3}(d_i)[\mathscr{M}_3] = 0$ for $i = 1, \ldots, 9, 11, 12$, $F_{r_3}(d_i)[\mathscr{M}_3] = 1-0.7^1$ for $i = 10, 14, 15$, and $F_{r_3}(d_i)[\mathscr{M}_3] = 1-0.7^3$ for $i = 13$, obtaining $P^N_D(I(r_3) \mid \mathscr{M}_3) = 0.3^2 \cdot 0.7^1 \cdot (1-0.7^3)$. Multiplying the three factors then yields $P^N_D(\mathscr{M})$.

2.2 RECURSIVE NETWORKS

In the distributions P^N_D defined via (6) and (8) strong independence assumptions hold: given the interpretation $\mathscr{M}_i$ of the parent relations of r_i, the events $r_i(\boldsymbol{d})$ and $r_i(\boldsymbol{d}')$ are independent for $\boldsymbol{d} \neq \boldsymbol{d}'$. As discussed in (Jaeger 1997), this is a serious limitation on what kind of probability distributions we are able to define with RBNs. Examples mentioned there of interesting types of relations that require a dependency of r-atoms are symmetric relations ($r(d,e)$ depends on $r(e,d)$), functional relations ($r(d,e)$ depends on $r(d,e')$ for all $e' \neq e$: exactly one of these atoms must be true), and temporal relations ($r(t,d)$ depends on $r(t-1,d)$).

On closer examination it turns out, though, that the assessment given in (Jaeger 1997) of the expressiveness of RBNs with respect to such dependencies is somewhat too pessimistic. While it is true that using an RBN with only one node r, we cannot define a distribution P^N_D on $\mathrm{Mod}_D(r)$ with, e.g., $P^N_D(r(d,e)) = 1/2$, $P^N_D(r(d,e) \leftrightarrow r(e,d)) = 1$ for all d, e, we can define such a distribution by a network containing a second binary relation symbol s, and the two probability formulas

$$\begin{aligned} F_s(v,w) &:\equiv q \\ F_r(v,w) &:\equiv (1-s(v,w))(1-s(w,v)) \end{aligned}$$

where q is such that $(1-q)^2 = 1/2$. In a similar manner, some forms of functional relations can be modeled with RBNs. Still, there are limits to what can be achieved within the framework presented so far. Temporal relations, for example, remain outside its scope.

To increase the expressiveness of RBNs, in (Jaeger 1997) *recursive RBNs* were defined. In a recursive RBN, probability formulas $F_r(\boldsymbol{v})$ are allowed to contain indicator functions of the form $r(\boldsymbol{w})$ in addition to the indicator functions $s(\boldsymbol{w})$ with $s \in Pa(r)$. The set of tuples $\boldsymbol{e}$ for which the evaluation of $F_r(\boldsymbol{d})$ requires the truth value of $r(\boldsymbol{e})$ now must be restricted in a way that makes the definition of $F_r(\boldsymbol{d})$ well-founded. For this purpose, we require that the domain D is equipped with suitable fixed relations or functions that enable us to define a well-founded partial order on $D^{|r|}$. Typical examples of such fixed relations and functions are a total order $<$ on D, or a successor function s. For the purpose of the present paper, we can limit ourselves to the case where D comes equipped with a built-in successor function (on finite domains D we take a successor function s to be undefined on the "last" element of D). The set

$$r\text{-}Pa(\boldsymbol{d}) := \{\boldsymbol{e} \mid F_r(\boldsymbol{d}) \text{ depends on } r(\boldsymbol{e})\}$$

now can be restricted by generalizing the equality constraints in probability formulas to constraints involving s. Here is an example of a probability formula that uses this generalization of the syntax to define a temporal kind of relation:

$$\begin{aligned} &F_r(t,v) :\equiv \\ &\qquad max\{\!|\frac{1}{2}(1-r(t',v)), r(t',v) \mid t'; s(t') = t|\!\}. \quad (9) \end{aligned}$$

For $(d,e) \in D^2$ here r-$Pa((d,e)) = \{(s^{-1}(d), e)\}$ ($= \emptyset$ if d is the "first" element of D). Given the truth value of $r(s^{-1}(d), e)$, the value of $F_r(d,e)$ is 1/2 if $r(s^{-1}(d), e)$ is false, and 1 if $r(s^{-1}(d), e)$ is true. We can think of r as a temporal property of e that becomes true at time d with probability 1/2 if not true already, and then remains true (because we have not introduced any machinery for dealing with multi-sorted domains, here the elements of D have to double as time points and as objects to which we ascribe property r).

As was shown in (Jaeger 1997), given a recursive probability formula $F_r(\boldsymbol{v})$ we can effectively compute a formula r-$pa(\boldsymbol{v}, \boldsymbol{w})$ over s, such that for all $\boldsymbol{d}, \boldsymbol{e} \in D^{|\boldsymbol{v}|}$:

$$\boldsymbol{e} \in r\text{-}Pa(\boldsymbol{d}) \text{ iff } (D,s) \models r\text{-}pa(\boldsymbol{d}, \boldsymbol{e}). \quad (10)$$

A recursive RBN now defines a probability distribution on $\mathrm{Mod}_D(S)$ iff for all $r \in S$ the relation

$$\boldsymbol{e} \prec_r \boldsymbol{d} \;:\Leftrightarrow \boldsymbol{e} \in r\text{-}Pa(\boldsymbol{d}) \quad (11)$$

is well-founded. The resulting distribution P^N_D then still is explicitly defined by (6) and (8); only the terms $P(\boldsymbol{d} \in I(r_i) \mid \mathscr{M}_i)$ in (6) have to be replaced by

$$P(\boldsymbol{d} \in I(r_i) \mid \mathscr{M}_i, I(r_i\text{-}Pa(\boldsymbol{d}))),$$

where $I(r_i\text{-}Pa(\boldsymbol{d}))$ represents an interpretation of r_i restricted to r_i-$Pa(\boldsymbol{d})$.

Given a recursive network N, and a finite domain (D, s) it can be effectively decided whether the relations $\prec_r$ are well-founded, and if so, probabilities $P^N_D(r(\boldsymbol{d}))$ can be computed. Thus, the difference between recursive and non-recursive RBNs, for finite domains is of computational complexity, but not of a fundamental nature. In section 5 it will be shown that on infinite domains this changes drastically.

3 INFINITE DOMAINS: SEMANTICS

Even if the actual domains of random objects we encounter in the real world usually are finite, infinite domains are important for at least two reasons: they can afford conceptual simplicity, and they can be seen as the limiting case for large finite domains.

Conceptual simplicity is a reason for turning to infinite domains in cases where the actual (finite) domain is very large, and does not admit of a manageable description by a finite model. As an example, consider a model for a person's family tree. Even though the whole ancestry of that person, in fact, is finite, when we try to construct a formal model we will most likely end up with certain specifications – such as that everybody has two parents, and no one is ancestor of oneself – that only are satisfiable over an infinite domain of individuals. Hence, an infinite domain here would be a natural basis for, say, a probabilistic model of the propagation of genetic traits. In a similar vein, probabilistic models of language, for example as defined by stochastic context free grammars (see e.g. (Pynadath & Wellman 1996)), are defined on an idealized domain of infinitely many possible words and sentences, even though the collection of all sentences ever uttered is finite.

A somewhat different use for infinite domains is given when the existing finite domain admits of an adequate, manageable model, but the domain is large and difficult to determine exactly (cf. Bacchus et al. (1997)). An example of such an "open" domain is the set of all people that a given patient had contact with throughout his life. Here we may very well be able to specify suitable models for every domain D of finite size n by some generic description. Not knowing the appropriate n, however, rather than experimenting with arbitrarily chosen numbers, we may regard the limiting case of an infinite domain as the canonical approximation to the unknown domain. This, of course, only makes sense when we can apply the model we have designed for finite domains to an infinite domain as well, and when the properties of the model in the infinite case reflect the limiting behavior of large finite models.

Thus, there is an essential difference between the use of infinite models as conceptual idealizations, and as limiting cases: in the first use it is not expected that the infinite model in any way reflects properties of specific finite models. Inferences we draw in the infinite model we can accept without further ado for what they are: statements that are true in the infinite structure that we study for its own sake. In the second use, on the other hand, results inferred from the infinite model only are relevant in conjunction with the knowledge that (approximately) the same results will hold in all sufficiently large finite models.

In this and the following sections we show how to extend the semantics and inference techniques for RBNs to infinite domains. This makes RBNs available as probabilistic models for infinite structures in their first use. In a companion paper (Jaeger 1998) the problem is studied of when these infinite models can be seen as the limiting case of finite models, and a certain subclass of RBNs is identified that define infinite models suitable for employment in the second type of application.

Given an RBN N, and a countably infinite domain D, we need to define the probability distribution P^N_D induced by N on $\mathrm{Mod}_D(S)$. This definition will essentially follow the same pattern as in the case of finite domains. The main task is to substitute suitable continuous concepts where the discrete ones used in the finite case no longer work. In the following, only an overview of the construction of P^N_D is given, leaving out several standard measure theoretic arguments needed to fully justify the construction. It should be pointed out, however, that some of these arguments crucially depend on the countability of D, so that our construction does not carry over to uncountable D.

For simplicity, it is assumed throughout that $D = \omega = \{1, 2, \ldots\}$, and that $\mathit{Pa}(r_i) \subseteq S_i := \{r_1, \ldots, r_{i-1}\}$ holds in N for all i. When interpreted over $D = \omega$, each node r_i in N has the set of possible values $\mathrm{Mod}_\omega(r_i)$, which can be identified with the powerset $\omega^{|r_i|}$, and hence is uncountable. Specifically, the whole space $\mathrm{Mod}_\omega(S)$ is uncountable, and therefore needs to be equipped with a σ-algebra of measurable subsets before probability measures on that space can be defined. For our purpose a suitable σ-algebra $\mathfrak{A}_S$ is canonically defined as follows.

For $1 \leq i \leq k$ let $\mathfrak{E}_i$ contain all subsets $E \subseteq \mathrm{Mod}_\omega(r_i)$ of the form

$$\begin{aligned} E = E((\boldsymbol{d}_j)_j, (\boldsymbol{d}'_h)_h) := \\ \{\mathscr{M} \in \mathrm{Mod}_\omega(r_i) \mid \mathscr{M} \models r_i(\boldsymbol{d}_j), \mathscr{M} \models \neg r_i(\boldsymbol{d}'_h); \\ j = 1, \ldots, l, h = 1, \ldots, m\}, \end{aligned}$$

where $l, m \in \omega$, and $\boldsymbol{d}_j, \boldsymbol{d}'_h \in \omega^{|r_i|}$. Let $\mathfrak{A}_{r_i}$ be the σ-algebra generated by $\mathfrak{E}_i$. Define $\mathfrak{A}_S$ to be the product σ-algebra $\mathfrak{A}_{r_1} \otimes \ldots \otimes \mathfrak{A}_{r_k}$.

It is no longer possible to define conditional probabilities $P(I(r_i) \mid \mathscr{M}_i)$ "pointwise" for each $I(r_i) \in \mathrm{Mod}_\omega(r_i)$ and $\mathscr{M}_i \in \mathrm{Mod}_\omega(\mathit{Pa}(r_i))$, because any specific $I(r_i)$ or $\mathscr{M}_i$ typically will have probability 0. Instead, we have to replace the definition of $P(I(r_i) \mid \mathscr{M}_i)$ in (6) by a definition of *transition probabilities* $K_i(\mathscr{M}_i, A)$ from $\mathscr{M}_i$ to measurable subsets $A \in \mathfrak{A}_{r_i}$. Functions $K(\cdot, \cdot)$ that represent such transition probabilities are known as *stochastic kernels* (see e.g. (Jacobs 1978)).

It is sufficient to define the values $K_i(\mathscr{M}_i, A)$ for $A = E \in \mathfrak{E}_i$. In analogy to (6), for $E = E((\boldsymbol{d}_j)_j, (\boldsymbol{d}'_h)_h) \in \mathfrak{E}_i$, we would like to define

$$K_i(\mathscr{M}_i, E) := \prod_{j=1}^{l} F_{r_i}(\boldsymbol{d}_j)[\mathscr{M}_i] \prod_{h=1}^{m} (1 - F_{r_i}(\boldsymbol{d}'_h)[\mathscr{M}_i]). \quad (12)$$

It can be shown that (12) defines a stochastic kernel from $\mathrm{Mod}_\omega(\mathit{Pa}(r_i))$ to $\mathrm{Mod}_\omega(r_i)$, provided that F_{r_i} only contains combination functions satisfying a certain measurability condition. We do not go into the details here, and only note that virtually every combination function of practical interest will satisfy this condition. We call a probability formula *admissible*, if it only contains combination functions satisfying the measurability condition. Similarly, an RBN is called admissible, if it only contains admissible probability formulas.

When via (12) stochastic kernels $K_1, \ldots, K_k$ have been defined, in analogy to (8), a probability measure P_ω^N on $\mathfrak{A}_S$ is defined by

$$P_\omega^N(A) = \int\int \cdots \int 1_A(I_1, \ldots, I_k) K_k((I_1, \ldots, I_{k-1}), dI_k) \ldots \ldots K_2(I_1, dI_2) K_1(dI_1), \qquad (13)$$

where $A \in \mathfrak{A}_S$, 1_A is the indicator function [1] of A, and the I_i are integration variables ranging over $\text{Mod}_\omega(r_i)$.

4 INFINITE DOMAINS: INFERENCE

In this section the main technical results are derived: it is shown that queries for the probability distributions P_ω^N are decidable.

We first give an outline of the argument by an informal derivation of the probability $P_\omega^N(r_3(d))$ for the network N from example 2.9, and an arbitrary $d \in \omega$. We try to determine $P_\omega^N(r_3(d))$ by reasoning about the form of structures $\mathscr{M} \in \text{Mod}_\omega(S)$ that are "typical" according to P_ω^N. First, we consider the expected interpretation of r_1: since each $d \in \omega$ has a positive probability both of belonging to r_1, and of not belonging to r_1, and since membership in r_1 is determined independently for distinct d, we know by probabilistic 0-1 laws that with probability 1 $\mathscr{M}$ will contain infinitely many elements both within and outside r_1.

Next, we consider the expected interpretation of r_2. For each given element d with $r_1(d)$, and each d' with $r_1(d')$ there is a positive probability that $r_2(d, d')$ holds. Since this is independent for different d', and because, according to our first result, there are infinitely many candidates d', by the same 0-1 laws as above, we know that with probability 1 there will exist infinitely many d' with $r_2(d, d')$. Since there are only countably many different d, this even means that with probability 1 for every d with $r_1(d)$ there will exist infinitely many d' with $r_2(d, d')$. For d with $\neg r_1(d)$, on the other hand, with probability 1 there exist no d' with $r_2(d, d')$. Thus, in a typical structure $\mathscr{M}$ we have $F_{r_3}(d)[\mathscr{M}] = 0$ for d with $\neg r_1(d)$, and $F_{r_3}(d)[\mathscr{M}] = 1$ for d with $r_1(d)$. Hence, for an arbitrary $d \in \omega$: $P_\omega^N(r_3(d)) = P_\omega^N(r_1(d)) = 0.4$.

[1] In the usual measure-theoretic sense, not as appearing in definition 2.3.

4.1 ALMOST SURE PROPERTIES

In the above derivation we have reasoned that with probability 1 structures $\mathscr{M} \in \text{Mod}_\omega(S)$ will possess certain structural properties. Our first step in developing a general and rigorous method for computing probabilities in P_ω^N will be to provide a well-defined class of such structural properties in terms of syntactic characterizations. This is done in definition 4.1 We then, in theorem 4.5, show that each of these structural properties is either satisfied by almost every structure in $\text{Mod}_\omega(S)$, or by almost none. Theorem 4.5 allows us to ignore for the computation of probabilities in P_ω^N all those structures that do not possess the canonical structural properties. This, in conjunction with the fact that in canonical structures $\mathscr{M}$ the functions $F_{r_i}(\cdot)[\mathscr{M}]$ have a very regular behavior (lemma 4.2), allows us to develop an effective decision procedure for P_ω^N in section 4.2.

The following definition introduces (variants of) concepts that are commonly used in finite model theory: types and extension axioms. Intuitively, an S- type $\tau(v_1, \ldots, v_k)$ is a formula that gives an explicit, complete description of an S-structure of size k.

Definition 4.1 Let S be a relational vocabulary. An S-*type* in the variables $\boldsymbol{v} = v_0, \ldots, v_k$ is a maximal consistent conjunction $\tau(\boldsymbol{v})$ of atomic and negated atomic formulas over S in the variables $\boldsymbol{v}$. A type $\tau(\boldsymbol{v})$ is called *proper* if it contains all the formulas $v_i \neq v_j$ $(v_i, v_j \in \boldsymbol{v}, i \neq j)$. A type $\sigma(\boldsymbol{v}, w)$ *extends* the type $\tau(\boldsymbol{v})$, written $\tau \subset \sigma$, if every conjunct of τ is a conjunct in σ.

Let $\tau \subset \sigma$ be proper types. The sentence

$$\forall \boldsymbol{v}(\tau(\boldsymbol{v}) \rightarrow \neg\exists w \sigma(\boldsymbol{v}, w)). \qquad (14)$$

is called a *no-extension axiom* for τ, σ. The set of all sentences

$$\forall \boldsymbol{v}(\tau(\boldsymbol{v}) \rightarrow \exists^{\geq n} w \sigma(\boldsymbol{v}, w)) \quad (n \in \omega) \qquad (15)$$

we call an ω-*extension axiom*, and denote it by the (non first-order) formula $\forall \boldsymbol{v}(\tau(\boldsymbol{v}) \rightarrow \exists^\omega w \sigma(\boldsymbol{v}, w))$ (the quantifier $\exists^\omega$ thus is to be read as "there exist infinitely many"). An *extension theory* is a consistent set Φ_{Ext} of no-extension and ω-extension axioms, s.t. for every pair of proper types $\tau(\boldsymbol{v}) \subset \sigma(\boldsymbol{v}, w)$, Φ_{Ext} contains (14) or (15).

The no-extension axiom (14) is in fact logically equivalent to the simpler sentence $\neg\exists \boldsymbol{v} w \sigma(\boldsymbol{v}, w)$. To enable a more uniform treatment of the two types of axioms, we here use the syntactically more complicated form (14).

Lemma 4.2 Let $F(v_1, \ldots, v_n)$ be an admissible probability formula over S. Let Φ_{Ext} be an extension theory for S. Let $\mathscr{M} \in \text{Mod}_\omega(S)$ with $\mathscr{M} \models \Phi_{\text{Ext}}$. Let $\boldsymbol{d}, \boldsymbol{d}' \in \omega^{|n|}$ so that $\mathscr{M} \models \tau(\boldsymbol{d}) \wedge \tau(\boldsymbol{d}')$ for some type τ. Then $F(\boldsymbol{d})[\mathscr{M}] = F(\boldsymbol{d}')[\mathscr{M}]$.

The lemma is proved using a standard model-theoretic back-and-forth argument, which shows that under the given assumptions there is an automorphism of $\mathscr{M}$ that maps $\boldsymbol{d}$ to $\boldsymbol{d}'$. This means that $\boldsymbol{d}$ and $\boldsymbol{d}'$ satisfy the same first-order formulas in $\mathscr{M}$. The first-order theory of $\boldsymbol{d}$ in $\mathscr{M}$ uniquely determines the value $F(\boldsymbol{d})[\mathscr{M}]$.

The following definition and lemma are merely technical tools that allow us to take somewhat more information about extensions out of an extension theory, than we explicitly put in.

Definition 4.3 Let $\boldsymbol{w} = (w_1, \ldots, w_l)$. Let $\tau(\boldsymbol{v}) \subset \sigma(\boldsymbol{v}, \boldsymbol{w})$ be types. A *generalized extension axiom* is a formula of the form

$$\forall \boldsymbol{v}(\tau(\boldsymbol{v}) \rightarrow Q\boldsymbol{w}\sigma(\boldsymbol{v}, \boldsymbol{w}))$$

where $Q \in \{\neg\exists, \exists^{=1}, \exists^{\omega}\}$. The meaning of $\exists^{=1}$ is to be read as "there exists exactly one tuple", that of $\exists^{\omega}$ as "there exist infinitely many tuples". Specifically, $\exists^{\omega} w_1 w_2 \sigma(\boldsymbol{v}, w_1, w_2)$ already is true when $\exists^{=1} w_1 \exists^{\omega} w_2 \sigma(\boldsymbol{v}, w_1, w_2)$ holds.

For the formulation of the lemma, and also for subsequent use, we introduce the notation $\boldsymbol{w} \sqsubseteq \boldsymbol{v}$ as shorthand for the formula $\wedge_{w \in \boldsymbol{w}} \vee_{v \in \boldsymbol{v}} w = v$.

Lemma 4.4 Let Φ_{Ext} be an S-extension theory. Let $\tau(\boldsymbol{v}) \subset \sigma(\boldsymbol{v}, \boldsymbol{w})$ be S-types. Then

$$\Phi_{\text{Ext}} \models \forall \boldsymbol{v}(\tau(\boldsymbol{v}) \rightarrow Q\boldsymbol{w}\sigma(\boldsymbol{v}, \boldsymbol{w})), \qquad (16)$$

for either $Q = \neg\exists$, $Q = \exists^{=1}$, or $Q = \exists^{\omega}$. The case $Q = \exists^{=1}$ holds exactly when $\sigma(\boldsymbol{v}, \boldsymbol{w}) \models \boldsymbol{w} \sqsubseteq \boldsymbol{v}$.

The proof of the lemma is fairly straightforward by reducing the generalized extension axiom to several no-extension and ω-extension axioms in the sense of definition 4.1. The following is the main theorem in this section.

Theorem 4.5 Let N be an admissible relational Bayesian network over S. For each $i = 1, \ldots, k$ then there exists an extension theory $\Phi^N_{\text{Ext}}(S_i)$ for S_i with $P^N_{\omega}(\Phi^N_{\text{Ext}}(S_i)) = 1$.

Proof: We proceed by induction on $i = 1, \ldots, k$. The base case $i = 1$ is a simpler variation of the induction step, and is here omitted. Thus, let $i > 1$, and assume that $P^N_{\omega}(\Phi^N_{\text{Ext}}(S_{i-1})) = 1$ for some extension theory $\Phi^N_{\text{Ext}}(S_{i-1})$. In the following, for an S_{i-1}-structure $\mathscr{M}$, and an S_i-theory Φ, we write $K(\mathscr{M}, \Phi)$ for $K(\mathscr{M}, R)$, where $R := \{I(r_i) \subseteq \omega^{|r_i|} \mid (\mathscr{M}, I(r_i)) \models \Phi\}$.

To prove the theorem for i, we show that there exists an extension theory $\Phi^N_{\text{Ext}}(S_i)$, such that for every $\mathscr{M} \models \Phi^N_{\text{Ext}}(S_{i-1})$

$$K(\mathscr{M}, \Phi^N_{\text{Ext}}(S_i)) = 1. \qquad (17)$$

To prove (17) it is sufficient to show that for every pair $\tau(\boldsymbol{v}) \subset \sigma(\boldsymbol{v}, w)$ of proper S_i-types we have

$$K(\mathscr{M}, \forall \boldsymbol{v}(\tau(\boldsymbol{v}) \rightarrow Qw\sigma(\boldsymbol{v}, w))) = 1 \qquad (18)$$

for $Q = \neg\exists$, or $Q = \exists^{\omega}$. Since there are only countably many such pairs τ, σ, from (18) we obtain (17) by letting $\Phi^N_{\text{Ext}}(S_i)$ be the collection of all extension axioms for which (18) holds.

To show (18), we first partition $\tau(\boldsymbol{v})$ and $\sigma(\boldsymbol{v}, w)$ into two, respectively four, conjuncts:

$$\begin{aligned} \tau(\boldsymbol{v}) &\equiv \tau^{S_{i-1}}(\boldsymbol{v}) \wedge \tau^{r_i}(\boldsymbol{v}) \\ \sigma(\boldsymbol{v}, w) &\equiv \tau^{S_{i-1}}(\boldsymbol{v}) \wedge \tau^{r_i}(\boldsymbol{v}) \\ &\quad \wedge \sigma^{S_{i-1},w}(\boldsymbol{v}, w) \wedge \sigma^{r_i,w}(\boldsymbol{v}, w) \end{aligned}$$

where $\tau^{S_{i-1}}(\boldsymbol{v})$ contains all conjuncts of $\tau(\boldsymbol{v})$ that are S_{i-1}-literals (including the (in-)equality formulas), $\tau^{r_i}(\boldsymbol{v})$ contains all r_i-literals of $\tau(\boldsymbol{v})$, $\sigma^{S_{i-1},w}(\boldsymbol{v}, w)$ contains all S_{i-1}-literals of $\sigma(\boldsymbol{v}, w)$ that contain the variable w, and $\sigma^{r_i,w}(\boldsymbol{v}, w)$ contains all r_i-literals of $\sigma(\boldsymbol{v}, w)$ containing w.

We now consider the slightly strengthened axiom

$$\begin{aligned} \chi(Q) :\equiv \forall \boldsymbol{v}(\tau^{S_{i-1}}(\boldsymbol{v}) \rightarrow \\ Qw(\sigma^{S_{i-1},w}(\boldsymbol{v}, w) \wedge \sigma^{r_i,w}(\boldsymbol{v}, w))) \end{aligned} \qquad (19)$$

($Q \in \{\neg\exists, \exists^{\omega}\}$), which is equivalent to the finite collection of extension axioms obtained by replacing $\tau^{S_{i-1}}(\boldsymbol{v})$ with $\tau^{S_{i-1}}(\boldsymbol{v}) \wedge \tau^{r_i}(\boldsymbol{v})$ in (19) for all possible choices of $\tau^{r_i}(\boldsymbol{v})$. Hence, when we show

$$K(\mathscr{M}, \chi(Q)) = 1 \qquad (20)$$

for $Q = \neg\exists$, or $Q = \exists^{\omega}$, we have shown (18) for the same Q and all $\tau(\boldsymbol{v}) \supseteq \tau^{S_{i-1}}(\boldsymbol{v})$.

To show (20), first consider the case that

$$\mathscr{M} \models \forall \boldsymbol{v}(\tau^{S_{i-1}}(\boldsymbol{v}) \rightarrow \neg\exists w \sigma^{S_{i-1},w}(\boldsymbol{v}, w)). \qquad (21)$$

In this case (20) holds with $Q = \neg\exists$, and we are done. If, in particular, $\mathscr{M} \models \neg\exists \boldsymbol{v} \tau^{S_{i-1}}(\boldsymbol{v})$, then (20) holds for both $Q = \neg\exists$ and $Q = \exists^{\omega}$.

If (21) does not hold, then because $\mathscr{M} \models \Phi^N_{\text{Ext}}(S_{i-1})$,

$$\mathscr{M} \models \exists^{\omega} \boldsymbol{v} \tau^{S_{i-1}}(\boldsymbol{v}) \qquad (22)$$

(by lemma 4.4), and

$$\mathscr{M} \models \forall \boldsymbol{v}(\tau^{S_{i-1}}(\boldsymbol{v}) \rightarrow \exists^{\omega}(\sigma^{S_{i-1},w}(\boldsymbol{v}, w)). \qquad (23)$$

Now consider a fixed tuple $\boldsymbol{d}$ with $\mathscr{M} \models \tau^{S_{i-1}}(\boldsymbol{d})$. By (23) there exist $e_1, e_2, \ldots \in D$ s.t. $\mathscr{M} \models \sigma^{S_{i-1},w}(\boldsymbol{d}, e_i)$ for $i \in \omega$. In the conjunction $\sigma^{r_i,w}(\boldsymbol{d}, e_i)$ let $\boldsymbol{f}_1, \ldots, \boldsymbol{f}_k$ be the n-tuples over $(\boldsymbol{d}, e_i)$ s.t. $r_i(\boldsymbol{f}_j)$ appears in $\sigma^{r_i,w}(\boldsymbol{d}, e_i)$, and $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_l$ be the tuples for which $\neg r_i(\boldsymbol{g}_h)$ appears in $\sigma^{r_i,w}(\boldsymbol{d}, e_i)$. Then

$$K(\mathscr{M}, \sigma^{r_i,w}(\boldsymbol{d}, e_i)) = \prod_{j=1}^{k} F_{r_i}(\boldsymbol{f}_j)[\mathscr{M}] \prod_{h=1}^{l}(1 - F_{r_i}(\boldsymbol{g}_h)[\mathscr{M}]) =: p. \quad (24)$$

By lemma 4.2, p is independent of the actual choice of $\boldsymbol{d}$ and e_i, as long as $\mathscr{M} \models \tau^{S_{i-1}}(\boldsymbol{d}) \wedge \sigma^{S_{i-1},w}(\boldsymbol{d}, e_i)$.

Again, we now distinguish two cases. First assume that $p = 0$. Then

$$K(\mathscr{M}, \neg\exists w(\sigma^{S_{i-1},w}(\boldsymbol{d}, w) \wedge \sigma^{r_i,w}(\boldsymbol{d}, w)) \geq 1 - \sum_{i \in \omega} K(\mathscr{M}, \sigma^{r_i,w}(\boldsymbol{d}, e_i)) = 1. \quad (25)$$

Now assume $p > 0$. Then

$$K(\mathscr{M}, \exists^{\omega} w(\sigma^{S_{i-1},w}(\boldsymbol{d}, w) \wedge \sigma^{r_i,w}(\boldsymbol{d}, w))) = K(\mathscr{M}, \cap_{j \geq 1} \cup_{i \geq j} \sigma^{r_i,w}(\boldsymbol{d}, e_i)) = 1. \quad (26)$$

The last identity in (26) follows from the Borel-Cantelli lemma, using that for $i \neq j$ $\sigma^{r_i,w}(\boldsymbol{d}, e_i)$ and $\sigma^{r_i,w}(\boldsymbol{d}, e_j)$ are independent in $\mathfrak{A}_{r_i}$ with respect to $K(\mathscr{M}, \cdot)$ (because $\sigma^{r_i,w}(\boldsymbol{d}, e_i)$ and $\sigma^{r_i,w}(\boldsymbol{d}, e_j)$ do not share any common r_i-atom).

By our observation following (24), whether (25) or (26) holds for $\boldsymbol{d}$ does not depend on the particular choice of $\boldsymbol{d}$, as long as $\mathscr{M} \models \tau^{S_{i-1}}(\boldsymbol{d})$. Taking the conjunction over the countably many $\boldsymbol{d}$, we can finally say:

$$K(\mathscr{M}, \chi(\neg\exists)) = 1 \quad \text{(if } p = 0 \text{ in (24)), or}$$
$$K(\mathscr{M}, \chi(\exists^{\omega})) = 1 \quad \text{(if } p > 0 \text{ in (24)).}$$

□

The following lemma is closely related to the fact that in the preceding proof the question of whether to include $\chi(\neg\exists)$ or $\chi(\exists^{\omega})$ into the extension theory for S_i was essentially reduced to the question of whether p in (24) is positive. We now make this reduction somewhat more explicit, by showing that we can "decide" Φ^N_{Ext} via computation of certain probabilities.

Lemma 4.6 Let $N, \Phi^N_{\text{Ext}}(S_i)$ be as in theorem 4.5. Let $\tau(\boldsymbol{v}) \subset \sigma(\boldsymbol{v}, \boldsymbol{w})$ be S_i-types. Let $\boldsymbol{d}, \boldsymbol{e} \subset \omega$ be arbitrary tuples with $\mid \boldsymbol{d} \mid = \mid \boldsymbol{v} \mid$, $\mid \boldsymbol{e} \mid = \mid \boldsymbol{w} \mid$, and such that $\boldsymbol{d}$ satisfies the equality constraints in $\tau(\boldsymbol{v})$, and $(\boldsymbol{d}, \boldsymbol{e})$ satisfies the equality constraints in $\sigma(\boldsymbol{v}, \boldsymbol{w})$. For

$$\lambda(Q) :\equiv \forall \boldsymbol{v}(\tau(\boldsymbol{v}) \to Q\boldsymbol{w}\sigma(\boldsymbol{v}, \boldsymbol{w})) \ (Q \in \{\neg\exists, \exists^{=1}, \exists^{\omega}\})$$

we then have

(a) $\Phi^N_{\text{Ext}} \models \lambda(\exists^{=1})$ iff $\sigma(\boldsymbol{v}, \boldsymbol{w}) \models \boldsymbol{w} \sqsubseteq \boldsymbol{v}$;

(b) $\Phi^N_{\text{Ext}} \models \lambda(\neg\exists)$ iff $P^N_{\omega}(\tau(\boldsymbol{d})) = 0$ or $P^N_{\omega}(\sigma(\boldsymbol{d}, \boldsymbol{e})) = 0$;

(c) $\Phi^N_{\text{Ext}} \models \lambda(\exists^{\omega})$ iff $P^N_{\omega}(\tau(\boldsymbol{d})) = 0$ or $P^N_{\omega}(\sigma(\boldsymbol{d}, \boldsymbol{e})) > 0$.

Proof: Part (a) is already covered by lemma 4.4. For parts (b) and (c) first observe that $\Phi^N_{\text{Ext}}(S_i) \models \lambda(\neg\exists) \wedge \lambda(\exists^{\omega})$ iff $\Phi^N_{\text{Ext}}(S_i) \models \neg\exists\boldsymbol{v}\tau(\boldsymbol{v})$. By symmetry, we have that $P^N_{\omega}(\tau(\boldsymbol{d})) = P^N_{\omega}(\tau(\boldsymbol{d}'))$ for all of the countably many tuples $\boldsymbol{d}'$ that satisfy the equality constraints in $\tau(\boldsymbol{v})$. With $P^N_{\omega}(\exists\boldsymbol{v}\tau(\boldsymbol{v})) = P^N_{\omega}(\cup_{\boldsymbol{d}'}\tau(\boldsymbol{d}'))$ we thus get $P^N_{\omega}(\tau(\boldsymbol{d})) = 0$ iff $P^N_{\omega}(\neg\exists\boldsymbol{v}\tau(\boldsymbol{v})) = 1$ iff $\Phi^N_{\text{Ext}}(S_i) \models \neg\exists\boldsymbol{v}\tau(\boldsymbol{v})$.

By analogous symmetry and countability arguments we get $P^N_{\omega}(\sigma(\boldsymbol{d}, \boldsymbol{e})) > 0$ iff $P^N_{\omega}(\neg\exists\boldsymbol{v}\boldsymbol{w}\sigma(\boldsymbol{v}, \boldsymbol{w})) = 0$ iff $P^N_{\omega}(\lambda(\neg\exists)) = 0$ iff $P^N_{\omega}(\lambda(\exists^{\omega})) = 1$ iff $\Phi^N_{\text{Ext}}(S_i) \models \lambda(\exists^{\omega})$. □

4.2 COMPUTING THE PROBABILITIES

We now show how the insights gained in the preceding section into the structure that, according to P^N_{ω}, an infinite structure $\mathscr{M}$ will almost certainly have enable us to compute probabilities in P^N_{ω}. The central insight is provided by lemma 4.2. According to that lemma, when we evaluate $F_{r_i}(\boldsymbol{d})[\mathscr{M}]$ for some canonical $\mathscr{M}$, we only need to know the S_i-type of $\boldsymbol{d}$. This means that we can in effect replace F_{r_i} by a fundamentally simpler formula $F^*_{r_i}$:

Theorem 4.7 Let N be an admissible relational Bayesian network over $S = \{r_1, \ldots, r_k\}$ with *Pa(r_i)* $\subseteq \{r_1, \ldots, r_{i-1}\}$ $(i = 1, \ldots, k)$. There exists a relational Bayesian network N^* over S with the following properties

(i) *Pa(r_i)* $\subseteq \{r_1, \ldots, r_{i-1}\}$ in N^*.

(ii) Every probability formula $F^*_{r_i}$ in N^* is of the form

$$F^*_{r_i}(\boldsymbol{v}) = \begin{cases} q_1 & \tau_1(\boldsymbol{v}) \\ \vdots & \\ q_m & \tau_m(\boldsymbol{v}) \end{cases} \quad (27)$$

with $q_h \in [0, 1]$ $(h = 1, \ldots, m)$, and the $\tau_h(\boldsymbol{v})$ are a complete list of the *Pa(r_i)*-types of $\boldsymbol{v}$.

(iii) $P^{N^*}_{\omega} = P^N_{\omega}$.

Given N, the network N^* can be computed effectively.

The salient feature of the probability formulas $F^*_{r_i}$ is that they are essentially combination function free (even though, in proper syntax, the distinction by cases in (27) would by encoded by some "benign" combination functions). Specifically, the formula $F^*_{r_i}(\boldsymbol{v})$ does not contain any variables $\boldsymbol{w}$

other than $\boldsymbol{v}$. In that sense, theorem 4.7 can be understood as a kind of quantifier elimination result, that allows us to get rid of the quantification over $\boldsymbol{w}$ by combination functions *comb*$\{\!|\cdot \mid \boldsymbol{w};\cdot|\!\}$. In practical terms, this means that the probability of a statement $\phi(\boldsymbol{d})$ can be computed without considering any elements other than $\boldsymbol{d}$. This is spelled out in the following corollary, which is of central interest in itself, and also plays an important role in the proof of the theorem.

Corollary 4.8 Let N be an admissible relational Bayesian network. Let $\phi(\boldsymbol{d})$ be a boolean expression over ground S-atoms. The probability $P^N_\omega(\phi(\boldsymbol{d}))$ then is computable.

Proof: First observe that the computation of $P^N_\omega(\phi(\boldsymbol{d}))$ can be reduced to finitely many computations of probabilities $P^N_\omega(\sigma(\boldsymbol{d}))$ for S-types $\sigma(\boldsymbol{v})$. The S-type $\sigma(\boldsymbol{v})$ can be written as a conjunction $\wedge_{i=1}^k \sigma_i(\boldsymbol{v})$ of S_i-types σ_i. For $\boldsymbol{d}' \subseteq \boldsymbol{d}$ let $\sigma_i(\boldsymbol{d}')$ be the S_i-type of $\boldsymbol{d}'$ implied by $\sigma(\boldsymbol{d})$. Then

$$
\begin{aligned}
&P^N_\omega(\sigma(\boldsymbol{d})) = \\
&\prod_{i=1}^{k} P^N_\omega(\sigma_i(\boldsymbol{d}) \mid \sigma_{i-1}(\boldsymbol{d})) = \\
&\quad\prod_{i=1}^{k} \prod_{\{\boldsymbol{d}'\subseteq \boldsymbol{d} \mid \sigma_i(\boldsymbol{d})\models r_i(\boldsymbol{d}')\}} F_{r_i}(\boldsymbol{d}')[\sigma_{i-1}(\boldsymbol{d}')] \cdot \\
&\qquad\qquad \prod_{\{\boldsymbol{d}'\subseteq \boldsymbol{d} \mid \sigma_i(\boldsymbol{d})\models \neg r_i(\boldsymbol{d}')\}} (1 - F_{r_i}(\boldsymbol{d}')[\sigma_{i-1}(\boldsymbol{d}')]).
\end{aligned}
$$

The terms $F_{r_i}(\boldsymbol{d}')[\sigma_{i-1}(\boldsymbol{d}')]$, here stand for $F_{r_i}(\boldsymbol{d}')[\mathscr{M}]$ for an arbitrary $\mathscr{M} \models \Phi^N_{\text{Ext}} \wedge \sigma_{i-1}(\boldsymbol{d}')$. Using the network N^* we can determine $q_h = F_{r_i}(\boldsymbol{d}')[\sigma_{i-1}(\boldsymbol{d}')]$ by looking up the *Pa(*r_i*)*-type τ_h that is implied by σ_{i-1}. □

Proof of theorem 4.7: We show how to effectively construct probability formulas $F^*_{r_i}$ over $\{r_1, \ldots, r_{i-1}\}$ such that for all $\mathscr{M} \models \Phi^N_{\text{Ext}}(S_{i-1})$, and all $\boldsymbol{d} \in \omega^{|r_i|}$ we get

$$F_{r_i}(\boldsymbol{d})[\mathscr{M}] = F^*_{r_i}(\boldsymbol{d})[\mathscr{M}]. \tag{28}$$

With theorem 4.5 it then follows that the network N^* defined by the probability formulas $F^*_{r_i}$ defines the same distribution as the original network N.

For the construction of $F^*_{r_i}$ assume that $F^*_1, \ldots, F^*_{r_{i-1}}$ already have been constructed (the base case F^*_1, again, is a simpler variation of the induction step, and is here omitted). We define $F^*_{r_i}$ by induction on the structure of F_{r_i}.

The first three cases – F_{r_i} being a constant, an indicator function, or a convex combination are quite straightforward. We therefore turn directly to the case $F_{r_i}(\boldsymbol{v}) \equiv$ *comb*$\{\!|F(\boldsymbol{v},\boldsymbol{w}) \mid \boldsymbol{w}; c(\boldsymbol{v},\boldsymbol{w})|\!\}$ (to simplify matters slightly, we here only consider the case of a single probability formula within the combination function. The extension to the general case is straightforward). By induction hypothesis, we can assume that $F(\boldsymbol{v},\boldsymbol{w})$ is in the form (27) with values $p_1, \ldots, p_l$ for types $\sigma_1(\boldsymbol{v},\boldsymbol{w}), \ldots, \sigma_l(\boldsymbol{v},\boldsymbol{w})$ in some vocabulary $S^* \subseteq S_{i-1}$. We now define *Pa(*r_i*)* to be the smallest subset of S_{i-1} that contains S^*, and for which Φ^N_{Ext} provides a complete set of extension axioms. In concrete terms, this is S_j for $j := \mathit{max}\{h \mid r_h \in S^*\}$.

Now, let $\tau_h(\boldsymbol{v})$ be a *Pa(*r_i*)*$-type$. We need to find the value q_h such that

$$q_h = F_{r_i}(\boldsymbol{d})[\mathscr{M}] \tag{29}$$

for all $\mathscr{M} \models \Phi_{\text{Ext}}$ and $\boldsymbol{d} \in \omega^{|r_i|}$ with $\mathscr{M} \models \tau_h(\boldsymbol{d})$. By lemma 4.2 we know that q_h only depends on τ_h, but not on $\boldsymbol{d}$ or $\mathscr{M}$ (note that, while F_{r_i} is an S^*- probability formula, we really do need to fix the *Pa(*r_i*)*$[\supseteq S^*]$-type of $\boldsymbol{d}$ in order to insure this invariance, because Φ^N_{Ext} usually will not include an extension theory for S^*).

The first case to be considered for the computation of q_h is whether $\Phi^N_{\text{Ext}} \models \neg\exists \boldsymbol{v}\tau_h(\boldsymbol{v})$. In that case an arbitrarily chosen value for q_h will satisfy (29). By lemma 4.6, we can determine whether $\Phi^N_{\text{Ext}} \models \neg\exists \boldsymbol{v}\tau_h(\boldsymbol{v})$ by computing $P^N_\omega(\tau_h(\boldsymbol{d}))$ for some test tuple $\boldsymbol{d}$. This probability is determined by the subnetwork N_{i-1} containing all the S_{i-1}-nodes. By induction hypothesis (for the outer induction on the network structure), N^*_{i-1} already has been constructed, so that by corollary 4.8, $P^N_\omega(\tau_h(\boldsymbol{d}))$ can be computed. If $P^N_\omega(\tau_h(\boldsymbol{d})) = 0$, we set, e.g., $q_h := 0$, and are done with τ_h.

If $P^N_\omega(\tau_h(\boldsymbol{d})) > 0$ we continue as follows. By definition, $q_h = F_{r_i}(\boldsymbol{d})[\mathscr{M}] =$ *comb* A for

$$A = \{\!|F(\boldsymbol{d},\boldsymbol{e}) \mid \boldsymbol{e}; c(\boldsymbol{d},\boldsymbol{e})|\!\}$$

(with $\boldsymbol{d}$ and $\mathscr{M}$ as in (29)). Thus, we need to determine A. We do this by computing for each *Pa(*r_i*)*-type $\rho(\boldsymbol{v},\boldsymbol{w})$ the contribution of those elements $\boldsymbol{e}$ to A that satisfy $\mathscr{M} \models \rho(\boldsymbol{d},\boldsymbol{e})$. Since $\boldsymbol{d}$ was assumed to be of type τ_h, we only need to consider types ρ that extend τ_h. Also, because of the restriction to $\boldsymbol{e}$ with $c(\boldsymbol{d},\boldsymbol{e})$, only types $\rho(\boldsymbol{v},\boldsymbol{w})$ with $\rho(\boldsymbol{v},\boldsymbol{w}) \models c(\boldsymbol{v},\boldsymbol{w})$ are relevant (note that either $\rho(\boldsymbol{v},\boldsymbol{w}) \models c(\boldsymbol{v},\boldsymbol{w})$, or $\rho(\boldsymbol{v},\boldsymbol{w}) \models \neg c(\boldsymbol{v},\boldsymbol{w})$). There is at most one such type $\rho_0(\boldsymbol{v},\boldsymbol{w})$ with $\rho_0(\boldsymbol{v},\boldsymbol{w}) \models \boldsymbol{w} \sqsubseteq \boldsymbol{v}$. For that type, there exists exactly one $\boldsymbol{e}_0$ with $\rho_0(\boldsymbol{d},\boldsymbol{e}_0)$, and ρ_0 contributes a single copy of $F(\boldsymbol{d},\boldsymbol{e}_0)[\mathscr{M}]$ to A. The value of $F(\boldsymbol{d},\boldsymbol{e}_0)[\mathscr{M}]$ is given in our representation of F^* by p_j corresponding to the S^*-type $\sigma_j(\boldsymbol{v},\boldsymbol{w}) = \rho_0(\boldsymbol{v},\boldsymbol{w})\restriction S^*$.

For $\rho \not\models \boldsymbol{w} \sqsubseteq \boldsymbol{v}$, by lemma 4.6, we have that $\mathscr{M} \models \neg\exists \boldsymbol{w}\rho(\boldsymbol{d},\boldsymbol{w})$ iff $P^N_\omega(\rho(\boldsymbol{d},\boldsymbol{e})) = 0$, and $\mathscr{M} \models \exists^\omega \boldsymbol{w}\rho(\boldsymbol{d},\boldsymbol{w})$ iff $P^N_\omega(\rho(\boldsymbol{d},\boldsymbol{e})) > 0$ for any test tuple $\boldsymbol{e}$. As above, the probabilities on the right hand sides of these equivalences can be computed in the already constructed network N^*_{i-1}. If $P^N_\omega(\rho(\boldsymbol{d},\boldsymbol{e})) > 0$, the type ρ contributes to A ω copies of the value p_j assigned by F^* to arguments of type $\rho\restriction S^*$. If $P^N_\omega(\rho(\boldsymbol{d},\boldsymbol{e})) = 0$, the type ρ does not contribute to A.

After performing these computations for all relevant types ρ, we obtain a representation of A of the form $A = \{\!|p_{h_1} : \lambda_1, \ldots, p_{h_m} : \lambda_m|\!\}$. where the $p_{h_j} \in \{p_1, \ldots, p_l\}$, and $\lambda_j \in \{1, \omega\}$ with $\lambda_j = 1$ for at most one j. Finally, we compute $q_h :=$ *comb* A □

If we apply the transformation here described to our network from example 2.9, the probability formulas of r_1 and r_2 remain essentially unchanged. In the construction of $F^*_{r_3}(v)$ we will first obtain the new parent-set *Pa(r₃)* $= \{r_1, r_2\}$. The values q_j in the representation (27) are computed as $q_j = 1$ for $\{r_1, r_2\}$-types $\tau_j(v)$ with $\tau_j(v) \models r_1(v)$, and $q_j = 0$ for $\tau_j(v) \models \neg r_1(v)$. This means that in an additional step we can simplify $F^*_{r_3}$ to

$$F^*_{r_3}(v) = \begin{cases} 1 & r_1(v) \\ 0 & \neg r_1(v). \end{cases}$$

This illustrates that the transformation $N \mapsto N^*$ not only replaces probability formulas, but in the process also changes the underlying network structure.

5 Infinite Domains: Recursive Networks

We now turn to the question of what part of the results of sections 3 and 4 carry over to recursive RBNs. As might be expected, in the case $D = \omega$ introduction of recursion causes much more fundamental problems than was the case for finite domains.

The first problem we are faced with is the existence of semantics for a recursive RBN. Recall that we identified as one main advantage of (non-recursive) RBNs their semantic clarity, particularly the fact that simply adhering to the syntactic rules for the construction of an RBN guarantees the user that exactly one distribution P^N_D will be defined. This is true both for finite and infinite D. Turning to recursive RBNs, in section 2.2 we found that for finite domains the existence of semantics no longer is guaranteed, but that a decision procedure for the well-definedness of P^N_D exists in checking the well-foundedness of the relations $\prec_r$ defined by formulas *r-pa*$(\boldsymbol{v}, \boldsymbol{w})$. One can show that in the case $D = (\omega, s)$ well-foundedness of $\prec_r$ on $\omega^{|r|}$ still enables us to define P^N_ω (though the construction is more complicated than the one in section 3). It is unclear, however, that it is possible, in general, to decide whether a given formula *r-pa*$(\boldsymbol{v}, \boldsymbol{w})$ defines a well-founded relation $\prec_r$ on $\omega^{|r|}$. While the problem of decidability of well-foundedness of $\prec_r$ is still open, the following theorem tells us that even in the (unlikely) event of a positive solution, not too much would be gained, because the problem of ultimate interest, computation of probabilities, will still be unsolvable.

Theorem 5.1 There does not exist an algorithm that, given a recursive RBN N over S with defined semantics P^N_ω as input, enumerates all probabilities $P^N_\omega(r(\boldsymbol{d}))$ for $r \in S$, $\boldsymbol{d} \in \omega^{|r|}$.

The theorem can be paraphrased as: there does not exist a sound and complete proof system for recursive RBNs with well-defined semantics over infinite domains.

Proof: We reduce the validity problem of first-order sentences in arithmetic to the computation of probabilities $P^N_\omega(r(\boldsymbol{d}))$. In order to fit into our relational framework, we here code sentences in arithmetic by using a relational vocabulary with unary relation symbols r_0, r_1 (rather than constant symbols 0,1), and ternary relation symbols $r_+, r_\cdot$ (rather than binary function symbols $+, \cdot$). The standard model $\mathcal{N}$ for the vocabulary $S_{\text{ar}} := \{r_0, r_1, r_+, r_\cdot\}$ then is ω with $I(r_0) = \{0\}$, $I(r_1) = \{1\}$, $I(r_+) = \{(d, e, f) \mid d + e = f\}$, and $I(r_\cdot) = \{(d, e, f) \mid d \cdot e = f\}$.

Given an S_{ar}-sentence ϕ, we construct a network N_ϕ containing a unary relation r_ϕ, so that P^N_ω is defined, and $P^N_\omega(r_\phi(1)) = 1$ iff $\mathcal{N} \models \phi$. The non-enumerability of P^N_ω then follows from the non-enumerability of the theory of $\mathcal{N}$.

N_ϕ is constructed as follows. We first define a network N_{ar} over S_{ar} that assigns probability 1 to $\mathcal{N}$. This is achieved by the probability formulas

$$F_{r_0}(v) \equiv 1 - \textit{max}\{\!|1 \mid w; s(w) = v|\!\}$$
$$F_{r_+}(v, w, z) \equiv \begin{cases} 1 & (r_0(w) \wedge v = z) \vee \\ & r_+(v, s^{-1}(w), s^{-1}(z)) \\ 0 & \text{else} \end{cases}$$

and similar formulas for F_{r_1} and $F_{r_\cdot}$. Here a somewhat loose syntax has been used for F_{r_+}. It is quite straightforward, though, to transform it into a probability formula proper. The dependency relations $\prec_{r_+}$ and $\prec_{r_\cdot}$ are well-founded, and $P^{N_{\text{ar}}}_\omega(\mathcal{N}) = 1$, as desired.

Given the S_{ar}-sentence ϕ, according to lemma 2.6, we can define a (non-recursive) probability formula $F_\phi(v)$, so that for all $d \in \omega$: $F_\phi(d)[\mathcal{N}] = 1$ iff $\mathcal{N} \models \phi$; $F_\phi(d)[\mathcal{N}] = 0$ else. Adding a node r_ϕ labeled with F_ϕ to the network N_{ar} yields a network N_ϕ that, if $\mathcal{N} \models \phi$, places probability 1 on the r_ϕ-extension $\mathcal{N}_\phi$ of $\mathcal{N}$ with $\mathcal{N}_\phi \models \forall v r_\phi(v)$, and if $\mathcal{N} \models \neg\phi$, places probability 1 on the r_ϕ-extension $\mathcal{N}_{\neg\phi}$ of $\mathcal{N}$ with $\mathcal{N}_{\neg\phi} \models \forall v \neg r_\phi(v)$. Hence, $P^{N_\phi}_\omega(r_\phi(1)) = 1$ iff $\mathcal{N} \models \phi$. □

While theorem 5.1 is formulated as a result for recursive RBNs, arguments similar to the one used in its proof clearly could be applied to other probabilistic representation systems for distributions over infinite, structured domains. Particularly systems like that of Ngo and Haddawy (1995) that use Herbrand universes as underlying domains (which in the special case of a single constant and a single unary function symbol is essentially the same as (ω, s)), provide a potential basis for carrying out the same argument. Exact results along these lines, however, first require a more detailed analysis of the first-order reasoning capabilities within these systems, particularly of the question whether they give rise to analogues of our lemma 2.6.

6 RELATED WORK AND CONCLUSION

Apart from the work by Ngo and Haddawy (1995) already mentioned, examples of previously proposed probabilistic representation and inference systems over infinite domains include work on probabilistic context free grammars (Pynadath & Wellman 1996) and stochastic programs (Koller, McAllester & Pfeffer 1997). In both these frameworks, distributions are defined over richly structured domains (labeled trees, essentially). These works differ in their semantic intent somewhat from the one of Ngo and Haddawy, and the one presented here: they, like standard Bayesian networks, are models of attributes of randomly sampled individuals, only that individuals now come from an infinite domain of distinguishable objects. This is to be contrasted with our objective of modeling random relations between multiple objects (taken from a uniform domain). On a sufficiently high level of abstraction, of course, this distinction is not very strict, since in our approach we can also view a relation between elements of the domain as an attribute of a randomly sampled complete structure.

For none of the frameworks mentioned here, decision procedures for general, complex queries have been given (in the particular rich system of Koller et al. (1997) it is moreover clear that none can exist). Main objective of the present paper has been to show that in the case of RBNs such a decision procedure exists. This was achieved by showing that infinite random structures, as generated by an RBN, with probability one will possess certain structural properties, and that therefore every distribution defined over ω by a network N can also be defined by a network N^* of a particularly simple form. The main, nontrivial, component of inference from N for domain ω is the transformation of N to N^*. Once N^* has been determined, computation of $P_\omega^{N^*}(\phi(\boldsymbol{d}))$ for some query $\phi(\boldsymbol{d})$ proceeds without any reference to the underlying domain.

Acknowledgments

Most of this work was conducted during a visit to Stanford University made possible by Daphne Koller. This work was funded in part by DARPA contract DACA76-93-C-0025, under subcontract to Information Extraction and Transport, Inc.

References

Bacchus, F., Grove, A. J., Halpern, J. Y. & Koller, D. (1997), 'From statistical knowledge bases to degrees of belief', *Artificial Intelligence* **87**, 75–143.

Breese, J. S. (1992), 'Construction of belief and decision networks', *Computational Intelligence* **8**(4), 624–647.

Haddawy, P. (1994), Generating bayesian networks from probability logic knowledge bases, *in* 'Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence'.

Jacobs, K. (1978), *Measure and Integral*, New York, Academic Press.

Jaeger, M. (1997), Relational bayesian networks, *in* 'Proceedings of UAI-97'.

Jaeger, M. (1998), Convergence results for relational bayesian networks, *in* 'Proceedings of the Thirteenth Annual IEEE Symposium on Logic in Computer Science (LICS-98)'. To appear.

Koller, D., McAllester, D. & Pfeffer, A. (1997), Effective bayesian inference for stochastic programs, *in* 'Proceedings of the 14th National Conference on Artificial Intelligence (AAAI-97)', pp. 740–747.

Ngo, L. & Haddawy, P. (1995), Probabilistic logic programming and bayesian networks, *in* 'Algorithms, Concurrency and Knowledge (Proceedings ACSC95)', Springer Lecture Notes in Computer Science 1023, pp. 286–300.

Pearl, J. (1988), *Probabilistic reasoning in intelligent systems : networks of plausible inference*, The Morgan Kaufmann series in representation and reasoning, rev. 2nd pr. edn, Morgan Kaufmann, San Mateo, CA.

Pynadath, D. & Wellman, M. (1996), Generalized queries on probabilistic context-free grammars, *in* 'Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96)'.

Wellman, M. P., Breese, J. S. & Goldman, R. P. (1992), 'From knowledge bases to decision models', *The Knowledge Engineering Review* **7**(1), 35–53.