Representing Statistical Information and Degrees of Belief in First-Order Probabilistic Conditional Logic

Matthias Thimm

Information Engineering Group, Department of Computer Science,
Technische Universität Dortmund, Germany

Abstract. Employing maximum entropy methods on probabilistic conditional logic has proven to be a useful approach for commonsense reasoning. Yet, the expressive power of this logic and similar formalisms is limited due to their foundations on propositional logic and in the past few years a lot of proposals have been made for probabilistic reasoning in relational settings. Most of these proposals rely on extensions of traditional graph-based probabilistic models like Bayes nets or Markov nets whereas probabilistic conditional logic does not presuppose any graphical structure underlying the model to be represented. In this paper we take an approach of lifting maximum entropy methods to the relational case by using a first-order version of probabilistic conditional logic. Furthermore, we take a specific focus on representing relational probabilistic knowledge by differentiating between different intuitions on relational probabilistic conditionals, namely between statistical interpretations and interpretations on degrees of belief. We develop a list of desirable properties on an inference procedure that supports these different interpretations and propose a specific inference procedure that fulfills these properties. We furthermore discuss related work and give some hints on future research.

1 Introduction

Applying probabilistic reasoning to relational representations of knowledge is a very active and controversy research area. In the past few years the fields of *probabilistic inductive logic programming* and *statistical relational learning* put forth a lot of proposals that deal with combining traditional probabilistic models of knowledge like Bayes nets or Markov nets [1] with first-order logic, see [2, 3] for some excellent surveys. For example, two of the most prominent approaches for extending propositional approaches to the relational case are Bayesian logic programs [4] and Markov logic networks [5]. While Bayesian logic programs extend Bayes nets using a logic programming language Markov logic networks extend Markov nets using a restricted form of first-order logic. Both frameworks use knowledge-based model construction techniques [6, 7] to reduce the problem of probabilistic reasoning in a relational context to probabilistic reasoning in a propositional context. In both frameworks—and also in most other approaches—this is done by appropriately grounding the parts of the knowledge base that are needed for answering a particular query and treating this grounded parts as a propositional knowledge base. While most approaches to relational probabilistic reasoning employ graphical models for probabilistic reasoning, in this paper we take another direction by lifting probabilistic conditional logic [8, 9] to the first-order case and applying maximum entropy methods [10–12] for reasoning.

In (propositional) probabilistic conditional logic knowledge is captured using conditionals of the form $(\phi \,|\, \psi)[\alpha]$ with some formulas ϕ, ψ of a given propositional language and $\alpha \in [0, 1]$. A probabilistic conditional of this form partially describes an (unknown) probability distribution P^* by stating that $P^*(\phi \,|\, \psi) = \alpha$ holds. In contrast to Bayes nets probabilistic conditional logic does not demand to fully describe a probability distribution but only to state constraints on it. On the one hand this is of great advantage because normally the knowledge engineer cannot fully specify a probability distribution for the problem area at hand. For example, if one has to represent probabilistic information on the relationships between symptoms and diseases then (usually) one can specify the probability of a specific disease given that a specific symptom is present but not if the symptom is not present. Probabilistic conditional logic avoids such problems by allowing to only partially specify a probability distribution. On the other hand, an incomplete specification of the problem area may lead to inconclusive inferences because there may be multiple probability distributions that satisfy the specified knowledge. The naïve approach to reason in probabilistic conditional logic is to compute upper and lower bounds for specific queries by consulting every probability distribution that is a model of the given knowledge base. While this skeptical form of reasoning may be appropriate for some applications, usually the inferences of this approach tend to be too weak to be meaningful. As a credulous alternative, one can select a specific probability distribution from the models of the knowledge base and do reasoning by just using this probability distribution. A reasonable choice for such a model is the one probability distribution with maximum entropy [10–12]. This probability distribution satisfies several desirable properties for commonsense reasoning and is uniquely determined among the probability distributions that satisfy a given set of probabilistic conditionals, see [10–12] for the theoretical foundations.

While applying maximum entropy methods in a direct fashion onto a first-order probabilistic conditional logic has already been investigated for example in [13–15], in this paper we take a special focus on relational probabilistic knowledge representation, namely the differentiation between statistical information and degrees of belief [16–18]. When considering conditionals, the modeled knowledge becomes ambiguous by introducing variables. Consider the following example inspired by [19]:

$$(likes(X, Y) \,|\, elephant(X) \wedge keeper(Y))[0.8]$$
$$(likes(X, fred) \,|\, elephant(X) \wedge keeper(fred))[0.4]$$
$$(likes(clyde, fred) \,|\, elephant(clyde) \wedge keeper(fred))[0.6]$$

The first conditional represents the information that with a probability of 0.8 a (typical) elephant likes his (typical) keeper. The second conditional states that a (typical) elephant likes the keeper $fred$ with a probability of 0.4 and the third conditional states that the elephant *clyde* likes the keeper $fred$ with a probability of 0.6. From a commonsensical point of view this knowledge base makes perfect sense. Given an adequate population of elephants and keepers this knowledge base says that typically an elephant likes his keeper, $fred$ is an exception and mostly unpopular, but *clyde* likes $fred$ still a bit more. But when treating the first two conditionals as schemas for their propositional instantiations (given a finite universe) then the grounded knowledge base becomes inconsistent because there are instantiations of the first two conditionals that are in direct conflict with each other and with the third conditional. The problem of inconsistency arises when treating conditionals like the first one as schemas for conditionals on the degrees of belief. But presumably what one really want to model when representing conditionals of this form is some kind of statistical information or maybe a default rule [19]. In the example above, the first conditional describes some form of statistical distribution on the all pairs of elephants and keepers and the second conditional describes a distribution on all elephants. In contrast to the first two conditionals the third conditional does not mention any variable. In fact, it mentions only ground instances regarding the constants *clyde* and $fred$ thus describing a *degree of belief* on the truth-value of $likes(clyde, fred)$ given that *clyde* is an elephant and $fred$ is a keeper. As a consequence, the knowledge represented by the third conditional describes some belief on the distribution of possible worlds rather than on the individuals of the universe. In this paper, we argue that an explicit differentiation of this two types of knowledge is important in order to reason with relational probabilistic knowledge bases.

The rest of this paper is organized as follows. In the following Section 2 we give a brief overview on (propositional) probabilistic conditional logic and continue in Section 3 with syntax and semantics of its extension to the first-order case. In Section 4 we propose some properties a reasonable inference mechanism should fulfill in order to interpret a relational probabilistic knowledge base in the sense described above and present our approach for an inference mechanism afterwards. In Section 5 we discuss our approach and review related work. In Section 6 we conclude.

2 Preliminaries

Before introducing probabilistic conditional logic for a relational language we begin by giving an overview on (propositional) probabilistic conditional logic. We extend this framework to the relational case in the subsequent section. But first, we consider a framework of propositional variables. Let $\mathcal{V} = \{V_1, \ldots, V_n\}$ be a set of propositional variables with finite domains $\mathsf{Dom}(V_1), \ldots, \mathsf{Dom}(V_n)$. An expression of the form $V_i = v_i$ is called a *literal* if v_i is in the domain of V_i, i. e. $v_i \in \mathsf{Dom}(V_i)$. The language $\mathcal{L}_{\mathcal{V}}$ is generated using the connectives $\neg$, $\wedge$, and $\vee$ on the literals in $\mathcal{V}$ in the usual way. For arbitrary formulas ϕ, ψ we abbreviate conjunctions $\phi \wedge \psi$ by $\phi\psi$ and negation $\neg\phi$ by overlining $\overline{\phi}$. If V is a binary variable, i. e., it is $\mathsf{Dom}(V) = \{\mathsf{true}, \mathsf{false}\}$, we abbreviate $V = \mathsf{true}$ by just V and $V = \mathsf{false}$ by $\overline{V}$. We write $\top$ for tautological formulas, e. g. $\phi \vee \overline{\phi} \equiv \top$. A *possible world* (interpretation) assigns to each variable $V_i \in \mathcal{V}$ a value in $\mathsf{Dom}(V_i)$. If ω is a possible world, then $\omega \models (V_i = v_i)$ if and only if ω assigns v_i to V_i. For an arbitrary formula ϕ the expression $\omega \models \phi$ evaluates in the usual way. Let $\Omega_{\mathcal{V}}$ be the set of all possible worlds of $\mathcal{L}_{\mathcal{V}}$.

Propositional probabilistic knowledge bases are build using propositional probabilistic conditionals, that impose certain restrictions on the conditional probabilities of the models of the knowledge base. A *(propositional) probabilistic conditional* r is an expression of the form $r = (\phi \,|\, \psi)[\alpha]$ with formulas ϕ, ψ and $\alpha \in [0, 1]$. If $\psi \equiv \top$ we write $(\phi)[\alpha]$ instead of $(\phi \,|\, \top)[\alpha]$. A set of probabilistic conditionals $R = \{r_1, \ldots, r_m\}$ is called a *(propositional) knowledge base*. The models of a knowledge base R are the probability distributions $P : \Omega_{\mathcal{V}} \rightarrow [0, 1]$ that fulfill all restrictions on the conditional probabilities imposed by the probabilistic conditionals in R. More specifically, a probability distribution $P : \Omega_{\mathcal{V}} \rightarrow [0, 1]$ is a model for a knowledge base R, written $P \models R$, if and only if $P \models r$ for every $r \in R$. That is

$$\begin{aligned} P \models (\phi \,|\, \psi)[\alpha] \; &:\Longleftrightarrow \; P(\phi \,|\, \psi) = \alpha \text{ and } P(\psi) > 0 \\ &\Longleftrightarrow \; \frac{P(\phi\psi)}{P(\psi)} = \alpha \text{ and } P(\psi) > 0 \end{aligned}$$

with

$$P(\phi) = \sum_{\omega \in \Omega_{\mathcal{V}}, \omega \models \phi} P_R(\omega) \quad .$$

A knowledge base R made of probabilistic conditionals describes incomplete knowledge. Usually, one is interested in performing inductive representation techniques and thus in computing a single probability distribution that describes R best and gives a complete description of the problem area at hand. This can be done using methods based on maximum entropy, which feature several nice properties [11, 10, 12, 20]. The entropy $H(P)$ of a probability distribution P is defined as

$$H(P) = -\sum_{\omega \in \Omega_{\mathcal{V}}} P(\omega) \log P(\omega)$$

and measures the amount of indeterminateness inherent in P. By selecting the probability distribution P^* among all probability distributions that satisfy a given knowledge base R, i. e. by computing the solution to the optimization problem

$$P^* := \mathsf{ME}(R) = \arg \max_{P \models R} H(P) \quad ,$$

we get the one probability distribution that satisfies R and adds as little information as necessary.

3 Syntax and Semantics of First-Order Conditional Logic

In the following we give an extension of probabilistic conditional logic to the relational case similar as in [21, 14]. To simplify presentation we use the same names for logical constructs as in propositional conditional language, e. g. we will refer to *relational probabilistic conditionals* just by *probabilistic conditionals*.

Let $\mathcal{L}_D$ be a first-order language with a fixed finite universe (domain) D without quantifiers and functions. We denote variables with a beginning uppercase, constants with a beginning lowercase letter, and vectors of these in boldface. We use greek letters ϕ and ψ for formulas and ξ for sentences (formulas with no free variables). For a first-order formula ϕ let $\text{FREE}(\phi)$ denote the set of (free) variables appearing in ϕ. We will write $\phi(\boldsymbol{X})$ to explicitly name the variables $\boldsymbol{X}$ in ϕ and we denote by $\phi(\boldsymbol{c})$ the grounded instance of ϕ with constants $\boldsymbol{c}$. Let $\mathsf{ground}_C(\phi)$ denote the set of grounded instances of ϕ with respect to a set of constants C.

Definition 1 (Probabilistic Conditional). *An expression of the form*

$$(\phi \mid \psi)[\alpha]$$

with first-order formulas ϕ, ψ (not necessarily ground) and a real $\alpha \in [0,1]$ is called a probabilistic conditional. *A probabilistic conditional $(\phi \mid \psi)[\alpha]$ is* ground *if $\text{FREE}(\phi) = \text{FREE}(\psi) = \emptyset$.*

As above we abbreviate $(\phi \mid \top)[\alpha]$ by $(\phi)[\alpha]$. For a probabilistic conditional $(\phi \mid \psi)[\alpha]$ let $\mathsf{ground}_C((\phi \mid \psi)[\alpha])$ denote the set of all grounded probabilistic conditionals of $(\phi \mid \psi)[\alpha]$ with respect to the set of constants C.

Definition 2 (Knowledge base). *A finite set R of probabilistic conditionals is called a* knowledge base. *A knowledge base R is* ground *if every probabilistic conditional in R is ground. Let $\mathfrak{R}$ denote the set of knowledge bases and $\mathfrak{R}_P \subseteq \mathfrak{R}$ the set of ground knowledge bases.*

Remark 1. Bear in mind that a ground knowledge base $R \in \mathfrak{R}_P$ is equivalent to a propositional knowledge base R' by interpreting ground atoms in R as ordinary propositional atoms. For the rest of this paper we treat ground relational knowledge bases and propositional knowledge bases interchangeably.

The informal semantics of a probabilistic conditional $(\phi \mid \psi)[\alpha]$ are as follows.

- If $\text{FREE}(\phi\psi) = \emptyset$ we interpret $(\phi \mid \psi)[\alpha]$ as an *uncertainty assessment* over the possible worlds as in propositional probabilistic conditional logic, thus specifying a degree of belief on a conditional probability.
- If $\text{FREE}(\phi\psi) \neq \emptyset$ we interpret $(\phi \mid \psi)[\alpha]$ as a *statistical assessment* stating that in the actual world a portion α of all ψ's are ϕ's.

We illustrate our intuition behind this informal semantics by means of an example.

Example 1. Consider again the scenario from the introduction. Let R be a knowledge base given as follows.

$$\begin{aligned} R = \quad \{ \quad & (likes(X,Y) \,|\, elephant(X) \wedge keeper(Y))[0.8] \\ & (likes(X, fred) \,|\, elephant(X) \wedge keeper(fred))[0.4] \\ & (likes(clyde, fred) \,|\, elephant(clyde) \wedge keeper(fred))[0.6] \quad \} \quad . \end{aligned}$$

In R we can assign the following informal meanings to the individual probabilistic conditionals:

- $(likes(clyde, fred) \,|\, elephant(clyde) \wedge keeper(fred))[0.6]$
  This conditional states that our subjective degree of belief of *clyde* liking *fred* is 0.6. So, if we know that *clyde* is an elephant and *fred* is a keeper we expect in 60 % of all occasions that *clyde* likes *fred*.
- $(likes(X, fred) \,|\, elephant(X) \wedge keeper(fred))[0.4]X$
  This conditional states that we expect 40 % of all elephants to like *fred*.
- $(likes(X,Y) \,|\, elephant(X) \wedge keeper(Y))[0.8]X,Y$
  This conditional states that we expect for 80 % of all elephant-keeper combinations that the elephant likes the keeper.

For a knowledge base R we denote by $\mathsf{Bel}(R)$ its projection on $\mathfrak{R}_P$, i. e., it is $\mathsf{Bel}(R) = \{r \in R \,|\, \text{FREE}(r) = \emptyset\}$. In other words, $\mathsf{Bel}(R)$ contains all uncertainty assessments. Analogously, let $\mathsf{Stat}(R) = R \backslash \mathsf{Bel}(R)$ denote the set of all statistical assessments of R.

Formal semantics for first-order conditional logic are given by probability distributions. The probability distributions under consideration are defined over the possible worlds of the given first-order language $\mathcal{L}_D$. A possible world ω for $\mathcal{L}$ is a tuple $\omega = \langle D, I \rangle$ with domain D and interpretation I which maps in the usual way constants to domain elements, unary predicate symbols to subsets of D and so on. As a simplification we interpret constants by themselves, i. e., for any constant c it is $I(c) = c$ in any possible world ω. As D is fixed for all possible worlds we will identify $\omega = \langle D, I \rangle$ with I when appropriate. Let Ω_D be the set of all these possible worlds with domain D and so we are interested in probability distributions $P : \Omega_D \to [0, 1]$. Let Prob_D be the set of probability distributions for domain D. $P \in \mathsf{Prob}_D$ is extended on first-order sentences (ground formulas) ξ by

$$P(\xi) = \sum_{\omega \models \xi} P(\omega) \quad .$$

Interpreting uncertainty assessments with probability distributions can be done analogously like in the propositional case. The problem at hand arises when considering statistical assessments like

$$c = (likes(X, fred) \,|\, elephant(X) \wedge keeper(fred))[0.5] \quad .$$

What is an appropriate satisfaction relation $\models^{cp}$ such that for a probability distribution P the statement $P \models^{cp} c$ describes our intuition on statistical assessments described above? We propose a new satisfaction relation $\models^{cp}_D$ on probabilistic conditionals that specifies when a probability distribution $P \in \mathsf{Prob}_D$

satisfies a given probabilistic conditional r. For the case of an uncertainty assessment $(\phi(\boldsymbol{c}) \mid \psi(\boldsymbol{c}))$, we define the satisfaction relation $\models^{cp}_D$ through

$$P \models^{cp}_D (\phi(\boldsymbol{c}) \mid \psi(\boldsymbol{c}))[\alpha] \quad :\Leftrightarrow \quad P((\phi(\boldsymbol{c}) \mid \psi(\boldsymbol{c}))) = \alpha \tag{1}$$

as in the propositional case. For a statistical assessment $(\phi(\boldsymbol{X}) \mid \psi(\boldsymbol{X}))[\alpha]$, we say that a probability distribution P satisfies $(\phi(\boldsymbol{X}) \mid \psi(\boldsymbol{X}))[\alpha]$ if the average of the conditional probabilities of all instantiations of $(\phi(\boldsymbol{X}) \mid \psi(\boldsymbol{X}))[\alpha]$ is α. So it is $P \models^{cp}_D (\phi(\boldsymbol{X}) \mid \psi(\boldsymbol{X}))[\alpha]$ if and only if

$$\frac{\sum_{(\phi(\boldsymbol{c}) \mid \psi(\boldsymbol{c})) \in \mathsf{ground}_D((\phi(\boldsymbol{X}) \mid \psi(\boldsymbol{X})))} P(\phi(\boldsymbol{c}) \mid \psi(\boldsymbol{c}))}{|\mathsf{ground}_D(\phi(\boldsymbol{X}) \mid \psi(\boldsymbol{X}))|} = \alpha \quad . \tag{2}$$

Notice, that Equation (2) also subsumes the case of uncertainty assessments in Equation (1) as a special case. As usual, a probability distribution P satisfies a knowledge base R, denoted $P \models^{cp}_D R$, if P satisfies every probabilistic conditional $r \in R$. We say that R is consistent iff there is at least on P with $P \models^{cp}_D R$, otherwise R is inconsistent.

4 Inference in First-Order Conditional Logic

We are interested in finding a "good" probability distribution P that satisfies all probabilistic conditionals of a given knowledge base R. More specifically, we are interested in a function $\mathsf{SRME}(R)$ (Statistical relational maximum entropy) that takes a knowledge base R and gives a probability distribution $P = \mathsf{SRME}(R)$ as output such that P describes R "best" in a commonsensical manner. In the following we state some properties on the operator SRME that derive from our intuition and afterwards describe such a function that fulfills these properties.

4.1 Desirable Properties

When considering knowledge bases like the one in Example 1 we want to be able to name a single probability distribution P that is the "best" model of R. Taking a naïve approach by grounding all conditionals in R universally and taking this grounding R' as a propositional knowledge base, we can not determine any probability distribution that satisfies R' due to its inherent inconsistency [15]. So our first demand on an appropriate operator SRME is its well-definedness. In the following, let $\mathsf{SRME} : \mathfrak{R} \rightarrow \mathsf{Prob}_D$ be an operator that maps a knowledge base $R \in \mathfrak{R}$ onto a probability distribution $P = \mathsf{SRME}(R) \in \mathsf{Prob}_D$ such that P commonsensical describes R.

(Well-Definedness) If R is a consistent then $\mathsf{SRME}(R)$ is well-defined.

We need some further notation to go on. For a formula ϕ let $\phi[d/c]$ denote the formula that is the same as ϕ except that every occurrence of the term c (either a variable or a constant) is substituted with the term d. More generally, let $\phi[d_1/c_1, \ldots, d_n/c_n]$ denote the formula that is the same as ϕ except that every occurrence of c_i is substituted with d_i for $1 \leq i \leq n$ simultaneously. The substitution operator $[\cdot]$ is extended on sets of formulas and conditionals in the usual way.

When considering knowledge bases based on a relational language the beliefs one gains on specific individuals is of special interest. An important demand to made is that the information one gains for different individuals is the same when these individuals are indistinguishable. More specifically, if the explicit information encoded in R for two different individuals $c_1, c_2 \in D$ is the same the probability distribution P should treat them indistinguishable. We can formalizing this indistinguishable property by introducing an equivalence relation on constants.

Definition 3 (Syntactical Equivalence). *Let R be a knowledge base. The constants $c_1, c_2 \in D$ are* syntactical equivalent*, denoted by $c_1 \equiv_R c_2$, if and only if $R = R[a_1/a_2, a_2/a_1]$.*

Observe that $\equiv_R$ is indeed an equivalence relation, i. e., it is reflexive, transitive, and symmetric. The equivalence classes of $\equiv_R$ are called *R-equivalence classes* and the set of all R-equivalence classes is denoted by $\mathcal{S}_R$. Note, that the notion of syntactical equivalence bears a resemblance with the notion of *reference classes* [17] but on a pure syntactical level.

Using syntactical equivalence we can state our demand for equal treatment of indistinguishable individuals as follows.

(Prototypical Indifference) Let R be a knowledge base and ξ a ground sentence. For any $c_1, c_2 \in D$ with $c_1 \equiv_R c_2$ it is

$$\mathsf{SRME}(R)(\xi) = \mathsf{SRME}(R)(\xi[c_1/c_2, c_2/c_1]) \quad .$$

From (Prototypical Indifference) some generalizations follow naturally.

Proposition 1. *Let* SRME *satisfy (Prototypical Indifference).*

1. *Let R be a knowledge base and ξ_1, ξ_2 be two ground sentences. For $c_1, c_2 \in D$ with $c_1 \equiv_R c_2$ it holds*

$$\mathsf{SRME}(R)(\xi_1 \mid \xi_2) = \mathsf{SRME}(R)(\xi_1[c_1/c_2, c_2/c_1] \mid \xi_2[c_1/c_2, c_2/c_1]) \quad .$$

2. *Let $S \in \mathcal{S}_R$, $c_1, \ldots, c_n \in S$, and $\sigma : S \rightarrow S$ a permutation on S, i. e. a bijective function on S. Then it holds for a ground sentence ξ*

$$\mathsf{SRME}(R)(\xi) = \mathsf{SRME}(R)(\xi[\sigma(c_1)/c_1, \ldots, \sigma(c_n)/c_n]) \quad .$$

Proof.

1. *Because of (Prototypical Indifference) it holds directly*

$$\mathsf{SRME}(R)(\xi_2) = \mathsf{SRME}(R)(\xi_2[c_1/c_2, c_2/c_1]) \quad \textit{and}$$
$$\mathsf{SRME}(R)(\xi_1 \wedge \xi_2) = \mathsf{SRME}(R)((\xi_1 \wedge \xi_2)[c_1/c_2, c_2/c_1])$$

and hence

$$\begin{aligned}\mathsf{SRME}(R)(\xi_1 \mid \xi_2) &= \frac{\mathsf{SRME}(R)(\xi_1 \wedge \xi_2)}{\mathsf{SRME}(R)(\xi_2)} \\ &= \frac{\mathsf{SRME}(R)(\xi_1 \wedge \xi_2[c_1/c_2, c_2/c_1])}{\mathsf{SRME}(R)(\xi_2[c_1/c_2, c_2/c_1])} \\ &= \mathsf{SRME}(R)(\xi_1[c_1/c_2, c_2/c_1] \mid \xi_2[c_1/c_2, c_2/c_1])\end{aligned}$$

due to $(\xi_1 \wedge \xi_2)[x_i/y_i]_{i=1,\ldots,n} = \xi_1[x_i/y_i]_{i=1,\ldots,n} \wedge \xi_2[x_i/y_i]_{i=1,\ldots,n}$.

2. *This follows from the fact that every permutation can be represented as a product of transpositions [22], i. e. permutations that exactly transpose two elements. Let* $\sigma_1, \ldots, \sigma_m$ *be these transpositions of* σ *and let* $\sigma_{1\ldots i} = \sigma_i \circ \ldots \circ \sigma_1$ *for* $i = 1, \ldots, m$. *Note, that* $\sigma_{1\ldots 1} = \sigma_1$ *and* $\sigma_{1\ldots m} = \sigma$. *Due to (Prototypical Indifference) it holds*

$$\mathsf{SRME}(R)(\xi) = \mathsf{SRME}(R)(\xi[\sigma_1(c_1)/c_1, \ldots, \sigma_1(c_n)/c_n])$$

and for any $i = 2\ldots, m$ *it holds*

$$\begin{aligned}&\mathsf{SRME}(R)(\xi[\sigma_{1\ldots i-1}(c_1)/c_1, \ldots, \sigma_{1\ldots i-1}(c_n)/c_n]) \\ =\ &\mathsf{SRME}(R)(\xi[\sigma_{1\ldots i}(c_1)/c_1, \ldots, \sigma_{1\ldots i}(c_n)/c_n]) \quad .\end{aligned}$$

Via transitivity and $\sigma_{1\ldots m} = \sigma$ *it follows*

$$\mathsf{SRME}(R)(\xi) = \mathsf{SRME}(R)(\xi[\sigma(c_1)/c_1, \ldots, \sigma(c_n)/c_n]) \quad .$$

□

Another aspect that should be satisfied by the operation SRME is some form of compatibility to the propositional case. For (relational) knowledge bases that are equivalent to propositional knowledge bases, i. e., ground knowledge bases, the operation SRME should coincide with the ME operator on propositional knowledge bases, cf. Section 2.

(Compatibility I) Let R be a ground knowledge base. If ξ is a ground sentence then it is $\mathsf{ME}(R)(\xi) = \mathsf{SRME}(R)(\xi)$.

Moreover, as the uncertainty assessments of a knowledge base R describe "strict" uncertain knowledge the probability distribution $\mathsf{SRME}(R)$ should reflect this knowledge faithfully.

(Compatibility II) Let R be a knowledge base. If $(\phi(\boldsymbol{c}) \mid \psi(\boldsymbol{c}))[\alpha] \in R$ is an uncertainty assessment it is

$$\mathsf{SRME}(R)(\phi(\boldsymbol{c}) \mid \psi(\boldsymbol{c})) = \alpha \quad .$$

So far, we have not taken into account the intention for representing statistical assessments. Given a statistical assessment $r = (\phi(\boldsymbol{X}) \mid \psi(\boldsymbol{X}))[\alpha]$ our intention in representing r in a knowledge base R is that for every instantiation $r' = (\phi(\boldsymbol{c}) \mid \psi(\boldsymbol{c}))[\alpha]$ of r the conditional probability of $\phi(\boldsymbol{c})$ given $\psi(\boldsymbol{c})$ "should" be α. But how do we capture this intention? Surely, we cannot guarantee that every possible instantiation r' of r will conform to a strict interpretation of this demand. This follows mainly from the fact, that using uncertainty assessments we should be able to give exceptions to this rule, cf. Example 1. What we are really want to describe when representing a statistical assessment r is that *given an adequate large domain* the conditional probability of the bigger part of the interpretations (neglecting exceptions) will converge towards α. This behavior resembles the intuition behind the "Law of Large Numbers" [23].

(Convergence) Let $R_1, R_2, \ldots$ be knowledge bases on $\mathcal{L}_{D_1}, \mathcal{L}_{D_2}, \ldots$ with $R_1 = R_2 = \ldots$ and $D_1 \subset D_2 \subset \ldots$ (for $i \in \mathbb{N}^+$). For a statistical assessment $r = (\phi(\boldsymbol{X}) \mid \psi(\boldsymbol{X}))[\alpha] \in R_1$ let $r' = (\phi(\boldsymbol{c}) \mid \psi(\boldsymbol{c}))[\alpha]$ be a proper instantiation of r with constants $\boldsymbol{c}$ that do not appear in R_1. For any such r and r' it is

$$\lim_{i \to \infty} \mathsf{SRME}(R_i)(\phi(\boldsymbol{c}) \mid \psi(\boldsymbol{c})) = \alpha$$

Another aspect of statistical assessments is their capability to comprehend for exceptions. Usually, statistical assessments are defined to model some kind of *expected value* over the set of instantiations. As such, if the probability of one instantiation of a statistical assessment lies below the value of the statement there has to be another instantiation with a probability higher than the value of the statement in order to compensate for the other exception (remember that the domain D is assumed to be finite).

(Compensation) Let R be a knowledge base and $(\phi(\boldsymbol{X}) \mid \psi(\boldsymbol{X}))[\alpha] \in R$ a statistical assessment with $\alpha \in (0, 1)$ (the open interval). If $\boldsymbol{c_1}$ is a vector of constants such that $\mathsf{SRME}(R)(\phi(\boldsymbol{c_1}) \mid \psi(\boldsymbol{c_1})) < \alpha$ then there is another vector of constants $\boldsymbol{c_2}$ with $\mathsf{SRME}(R)(\phi(\boldsymbol{c_2}) \mid \psi(\boldsymbol{c_2})) > \alpha$.

4.2 Statistical Relational Maximum Entropy

In the following we define a function $\mathsf{SRME}_1 : \mathfrak{R} \to \mathsf{Prob}_D$ that fulfills the desired properties defined in the previous section. We define the function $\mathsf{SRME}_1(R)$ using the the proposed semantics $\models^{cp}_D$ analogously like in the propositional case by selecting a probability distribution with maximum entropy among all probability distributions that satisfy R.

$$\mathsf{SRME}_1(R) = \arg \max_{P \models^{cp}_D R} - \sum_{\omega \in \Omega} P(\omega) \log P(\omega) \tag{3}$$

Remark 2. Note, that it seems that the optimization problem defined by Equation (3) is (in general) not uniquely solvable because the set of probability distributions defined by Equation (2) is non-convex. However, preliminary experimental results indicate that the probability distribution in Equation (3) is uniquely determined. As the formal proof has yet to be made, we assume for the rest of this paper that in Equation (3) an arbitrary probability distribution with maximum entropy will be chosen. Nonetheless, if R is consistent there is at least one probability distribution with maximum entropy that can be chosen in Equation (3).

Example 2. We continue Example 1. Let $\mathcal{L}_D$ be a first-order language with predicates $elephant/1$, $keeper/1$, and $likes/2$ and domain $D = \{clyde, dumbo, catty, giddy, fred, dave\}$. Let R be given by

$$(elephant(clyde))[1] \tag{4}$$
$$(elephant(dumbo))[1] \tag{5}$$
$$(elephant(catty))[1] \tag{6}$$
$$(elephant(giddy))[1] \tag{7}$$
$$(keeper(fred))[1] \tag{8}$$
$$(keeper(dave))[1] \tag{9}$$
$$(likes(X, Y) \mid elephant(X) \land keeper(Y))[0.5] \tag{10}$$
$$(likes(X, fred) \mid elephant(X) \land keeper(fred))[0.25] \tag{11}$$
$$(likes(clyde, fred) \mid elephant(clyde) \land keeper(fred))[0.1] \tag{12}$$

Here, the conditional (10) states that in 50 % of the elephant/keeper combinations the elephant likes the keeper, conditional (11) states, that 25 % of the elephants like $fred$ and conditional (12) states that the probability of $clyde$ liking $fred$ is 0.1. In the following we give the probabilities of several instantiations of $likes$ in $\mathsf{SRME}_1(R)$. Notice, how the probabilities of the instantiations of the conditionals (10) and (11) change in order to compensate for the exceptional instantiations involving $clyde$ and $fred$.

$$\begin{aligned}
\mathsf{SRME}_1(R)(likes(clyde, dave)) &= 0.75\\
\mathsf{SRME}_1(R)(likes(dumbo, dave)) &= 0.75\\
\mathsf{SRME}_1(R)(likes(catty, dave)) &= 0.75\\
\mathsf{SRME}_1(R)(likes(giddy, dave)) &= 0.75\\
\mathsf{SRME}_1(R)(likes(clyde, fred)) &= 0.1\\
\mathsf{SRME}_1(R)(likes(dumbo, fred)) &= 0.3\\
\mathsf{SRME}_1(R)(likes(catty, fred) &= 0.3\\
\mathsf{SRME}_1(R)(likes(giddy, fred)) &= 0.3
\end{aligned}$$

As there is no additional information on the elephants in R except $clyde$ they have to be treated in the same manner. Due to conditional (11) every elephant is equally likely to like $fred$ with a probability 0.3 and due to conditional (12) $clyde$ likes fred with probability 0.1. Due to conditional (10) the total percentage of $like$ relations that hold have to be 50 %. This information increases the probability of the elephants liking $dave$ accordingly to 75 %.

Considering the comments in Remark 2 we first state the following conjecture.

Conjecture 1. SRME_1 satisfies (Well-Definedness).

In the following we give some theoretical results mostly in form of proof sketches that show that the proposed operator SRME_1 indeed fulfills the desired properties discussed in Section 4.1.

Proposition 2. SRME_1 *satisfies (Prototypical Indifference).*

Proof. (Sketch) This is ensured by selecting in Equation (3) a probability distribution with maximum entropy. Suppose $(\phi(\boldsymbol{X}) \,|\, \psi(\boldsymbol{X}))[\alpha] \in R$ *is a statistical assessment,* $\boldsymbol{c_1}, \boldsymbol{c_2}$ *vectors of constants that only differ in constants that do not appear in* R*. If* $p_1 = \mathsf{SRME}_1(R)(\phi(\boldsymbol{c_1}) \,|\, \psi(\boldsymbol{c_1})) \neq \mathsf{SRME}_1(R)(\phi(\boldsymbol{c_2}) \,|\, \psi(\boldsymbol{c_2})) = p_2$*, then the probability distribution* P *with*

$$P(\phi(\boldsymbol{c_1}) \,|\, \psi(\boldsymbol{c_1})) = P(\phi(\boldsymbol{c_2}) \,|\, \psi(\boldsymbol{c_2})) = \frac{p_1 + p_2}{2}$$

yields a higher entropy than $\mathsf{SRME}_1(R)$ *but still fulfills Equation (2).* □

Proposition 3. SRME_1 *satisfies (Compatibility I).*

Proof. Let R *be a ground knowledge base. Now, only Equation (1) is used for determining the space of probability distributions, so* $\models^{cp}_{D}$ *is equivalent to* $\models$ *in the propositional case, cf. Section 2. Then Equation (3) also becomes equivalent to the propositional case and it is* $\mathsf{ME}(R')(\xi) = \mathsf{SRME}_1(R)(\xi)$ *for any ground sentence* ξ*.* □

Proposition 4. SRME_1 *satisfies (Compatibility II).*

Proof. This is ensured by Equation (1). □

Proposition 5. SRME_1 *satisfies (Convergence).*

Proof. (Sketch) This property follows from (Prototypical Indifference). When the number of constants grows towards infinity, most of the instantiations of a statistical assessment have the same probability in $\mathsf{SRME}_1(R)$ *and in order to fulfill Equation (1) these probabilities must converge to* α*.* □

Proposition 6. SRME_1 *satisfies (Compensation).*

Proof. Let R *be a knowledge base and* $(\phi(\boldsymbol{X}) \,|\, \psi(\boldsymbol{X}))[\alpha] \in R$ *a statistical assessment with* $\alpha \in (0, 1)$*. Suppose*

$$\mathsf{SRME}_1(R)(\phi(\boldsymbol{c}) \mid \psi(\boldsymbol{c})) < \alpha$$

for all $(\phi(\boldsymbol{c}) \,|\, \psi(\boldsymbol{c}))[\alpha] \in \mathsf{ground}_D(\phi(\boldsymbol{X}) \mid \psi(\boldsymbol{X}))$*. Then (for finite* D*) it is*

$$\begin{aligned} &\frac{\sum_{(\phi(\boldsymbol{c})|\psi(\boldsymbol{c}))\in\mathsf{ground}_D((\phi(\boldsymbol{X})|\psi(\boldsymbol{X})))} P(\phi(\boldsymbol{c}) \mid \psi(\boldsymbol{c}))}{|\mathsf{ground}_D(\phi(\boldsymbol{X}) \mid \psi(\boldsymbol{X}))|} \\ &< \frac{\alpha \cdot |\mathsf{ground}_D(\phi(\boldsymbol{X}) \mid \psi(\boldsymbol{X}))|}{|\mathsf{ground}_D(\phi(\boldsymbol{X}) \mid \psi(\boldsymbol{X}))|} \\ &= \alpha \end{aligned}$$

contradicting $\mathsf{SRME}_1(R) \models^{cp}_{D} R$*.* □

5 Discussion and Related Work

The work discussed in this paper is at a preliminary stage and more investigations and experiments have to be undertaken in order to demonstrate usability and usefulness of the proposed approach. To this end the KREATE project[1] investigates different approaches for combining relational representations and probabilistic reasoning under maximum entropy. Within this project there are some approaches of applying maximum entropy methods to relational knowledge bases without taking into account statistical information explicitly. Loh [15] and Fisseler [14] both employ the principle of maximum entropy directly on a grounded version of the relational knowledge base. In [15] inconsistencies in the grounded knowledge base are handled by removing contradictory instances of the individual conditionals. This results in a consistent propositional knowledge base for which the probability distribution with maximum entropy can be computed as in the propositional case. Consequently, these approaches treat conditionals with variables as schemas for their instances and thus use only an interpretation of conditionals based on the degree of belief. Nevertheless, it seems that even these approaches satisfy all properties discussed in this paper except (Compensation). So it seems reasonable to assume that the list of properties is incomplete for describing the intuition behind statistical assessments as well as uncertainty assessments. For future work we plan to investigate these properties in more depth and find new properties that characterize this intuition.

6 Summary

In this paper we investigated relational probabilistic reasoning from the point of view of maximum entropy methods and by taking into account the differences of statistical information and degrees of belief. We defined common sense properties for inference in first-order probabilistic conditional logic that represent this distinction and proposed an inference operator that fulfills this properties. Finally, we closed with some discussions.

As mentioned above the work reported here is at a preliminary stage and further investigations of the topic are mandatory. A comprehensive comparison of our approach and the approaches discussed in the previous section is part of current research.

Acknowledgements. The author thanks the reviewers for their helpful comments to improve the original version of this paper. Special thanks go to Gabriele Kern-Isberner for discussions and helpful suggestions. The research reported here was partially supported by the Deutsche Forschungsgemeinschaft (grant KE 1413/2-1).

[1] `http://www.fernuni-hagen.de/wbs/research/kreate/index.html`

References

1. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann (1998)
2. De Raedt, L., Kersting, K.: Probabilistic inductive logic programming. In Raedt, L.D., Kersting, K., Landwehr, N., Muggleton, S., Chen, J., eds.: Probabilistic Inductive Logic Programming. Volume 4911 of Lecture Notes in Computer Science. Springer (2008) 1–27
3. Cussens, J.: Logic-based Formalisms for Statistical Relational Learning. In Getoor, L., Taskar, B., eds.: Introduction to Statistical Relational Learning. MIT Press, Cambridge, MA (2007)
4. Kersting, K., De Raedt, L.: Bayesian Logic Programming: Theory and Tool. In Getoor, L., Taskar, B., eds.: An Introduction to Statistical Relational Learning. MIT Press (2005)
5. Domingos, P., Richardson, M.: Markov Logic: A Unifying Framework for Statistical Relational Learning. In: Proceedings of the ICML-2004 Workshop on Statistical Relational Learning and its Connections to Other Fields, Banff, Canada (2004) 49–54
6. Wellman, M.P., Breese, J.S., Goldman, R.P.: From Knowledge Bases to Decision Models. The Knowledge Engineering Review **7**(1) (1992) 35–53
7. Breese, J.S.: Construction of Belief and Decision Networks. Computational Intelligence **8**(4) (1992) 624–647
8. Nute, D., Cross, C.: Conditional logic. In Gabbay, D., Guenther, F., eds.: Handbook of Philosophical Logic. Volume 4. Kluwer (2002) 1–98
9. Benferhat, S., Dubois, D., Prade, H.: Possibilistic and Standard Probabilistic Semantics of Conditional Knowledge Bases. Journal of Logic and Computation **9**(6) (1999) 873–895
10. Grove, A.J., Halpern, J.Y., Koller, D.: Random Worlds and Maximum Entropy. Journal of Artificial Intelligence Research (JAIR) **2** (1994) 33–88
11. Paris, J.: The Uncertain Reasoner's Companion: A Mathematical Perspective. Cambridge University Press (1994)
12. Kern-Isberner, G.: Conditionals in Nonmonotonic Reasoning and Belief Revision. Number 2087 in Lecture Notes in Computer Science. Springer (2001)
13. Lukasiewicz, T., Kern-Isberner, G.: Probabilistic Logic Programming under Maximum Entropy. In: Proceedings ECSQARU-99. Volume 1638., Springer Lecture Notes in Artificial Intelligence (1999) 279–292
14. Fisseler, J.: Learning and Modeling with Probabilistic Conditional Logic. PhD thesis, FernUniversität in Hagen, Germany (to appear 2009)
15. Loh, S.: Relational Probabilistic Inference Based on Maximum Entropy. Master's thesis, Technische Universität Dortmund (to appear 2009)
16. Halpern, J.Y.: An Analysis of First-Order Logics of Probability. Artificial Intelligence **46** (1990) 311–350
17. Bacchus, F., Grove, A.J., Halpern, J.Y., Koller, D.: From Statistical Knowledge Bases to Degrees of Belief. Artificial Intelligence **87**(1–2) (1996) 75–143
18. Friedman, N., Halpern, J.Y., Koller, D.: First-order conditional logic revisited. In: AAAI/IAAI, Vol. 2. (1996) 1305–1312
19. Delgrande, J.P.: On First-Order Conditional Logics. Artificial Intelligence **105**(1–2) (1998) 105–137
20. Rödder, W., Meyer, C.H.: Coherent Knowledge Processing at Maximum Entropy by SPIRIT. In: Proc. of the Twelfth Conf. on Uncertainty in Artificial Intelligence. (1996) 470–476

21. Fisseler, J.: Toward Markov Logic with Conditional Probabilities. In Wilson, D.C., Lane, H.C., eds.: Proceedings of the Twenty-First International FLAIRS Conference, AAAI Press (2008) 643–648
22. Beachy, J.A., Blair, W.D.: Abstract Algebra. third edn. Waveland Press, Inc., Long Grove, Illinois, USA (2005)
23. Bernoulli, J.: Usum & Applicationem Praecedentis Doctrinae in Civilibus, Moralibus & Oeconomicis. In: Ars Conjectandi. (1713)