

On Prototypical Indifference and Lifted Inference in Relational Probabilistic Conditional Logic

Matthias Thimm

Technische Universität Dortmund, Germany

Abstract. Semantics for formal models of probabilistic reasoning rely on probability functions that are defined on the interpretations of the underlying classical logic. When this underlying logic is of relational nature, i. e. a fragment of first-order logic, then the space needed for representing these probability functions explicitly is exponential in both the number of predicates and the number of domain elements. Consequently, probabilistic reasoning becomes a demanding task. Here, we investigate *lifted inference* in the context of explicit model representation with respect to an inference operator that satisfies *prototypical indifference*, i. e. an inference operator that is indifferent about individuals for which the same information is represented. As reasoning based on the principle of maximum entropy satisfies this property we exemplify our ideas by compactly characterizing the maximum entropy model of a probabilistic knowledge base in a relational probabilistic conditional logic. Our results show that lifted inference is no longer exponential in the number of domain elements when we restrict the language to unary predicates but is still infeasible for the general case.

1 Introduction

Applying probabilistic reasoning to relational representations of knowledge is a topic that has been mostly investigated within the fields of *statistical relational learning* and *probabilistic inductive logic programming* [3]. Those areas have put forth a variety of approaches that deal with combining traditional probabilistic models of knowledge like Bayes nets or Markov nets [10] with first-order logic, see e. g. Bayesian logic programs (BLPs) [3, Ch. 10] and Markov logic networks (MLNs) [3, Ch. 12]. Those frameworks employ knowledge-based model construction techniques [16] to reduce the problem of probabilistic reasoning in a relational context to probabilistic reasoning in a propositional context by appropriately grounding the parts of the knowledge base that are needed for answering a particular query.

In this paper we continue work on *relational probabilistic conditional logic* (RPCL) [6,15] which is a formalism for relational probabilistic knowledge representation that is apt for default reasoning as well. In RPCL uncertain knowledge is represented using probabilistic conditionals, i. e. *if-then* rules. Consider the following conditionals which represent both generic and specific rules of how

elephants like their keepers (the example is inspired by [2]):

$$r_1 =_{def} (\text{likes}(X, Y) \mid \text{elephant}(X) \wedge \text{keeper}(Y))[0.6]$$

$$r_2 =_{def} (\text{likes}(X, \text{fred}) \mid \text{elephant}(X) \wedge \text{keeper}(\text{fred}))[0.4]$$

$$r_3 =_{def} (\text{likes}(\text{clyde}, \text{fred}) \mid \text{elephant}(\text{clyde}) \wedge \text{keeper}(\text{fred}))[0.7]$$

In many approaches to statistical relational learning such as BLPs relational rules are grounded, and the probability is attached to each instance. There, r_1 becomes the set $\{(\text{likes}(a, b) \mid \text{elephant}(a) \wedge \text{keeper}(b))[0.6] \mid a, b \in U\}$ where U is some pool of constant symbols. As one can see, using naive grounding approaches renders the set of probabilistic conditionals from above inconsistent as there are instances of r_1 which contradict instances of r_2 and r_3 . In [6] two novel semantics for relational probabilistic conditionals are introduced that avoid this problem. Further, by employing the principle of maximum entropy [9] one obtains a commonsense reasoning behavior [15].

Probabilistic reasoning in relational domains is, in general, a demanding task and there has been some efforts to speed up inference by exploiting structural equivalence in probabilistic knowledge [11]. This so-called *lifted inference* has been applied to e. g. parametrized belief networks and performed well in empirical experiments, cf. [13,8]. In this paper, we investigate *lifted inference* in RPCL. Our approach relies on the property of *prototypical indifference* which is satisfied by the maximum entropy approaches proposed in [6,15]. Basically, this property states that if a knowledge base \mathcal{R} contains exactly the same information for constants c_1 and c_2 then reasoning with \mathcal{R} is indifferent with respect to c_1 and c_2 . Consequently, the maximum entropy models of [6,15] carry a lot of redundant information. We introduce *condensed probability functions* as a compact way to represent those probability functions. Condensed probability functions are defined on *reference worlds* which subsume a whole set of first-order interpretations that model the same situation modulo exchanging equivalent constants. Using reference worlds and condensed probability functions we rephrase the maximum entropy models of [6,15] in a computationally feasible way.

The rest of this paper is organized as follows. In Section 2 we briefly review the semantical and inferential approaches of [6,15]. In Section 3 we introduce condensed probability functions as a compact way to represent prototypically uniform probability functions. Afterwards, we propose our approach to lifted inference in Section 4 and analyze its advantages in Section 5. In Section 6 we briefly discuss the issue of extending our approach to non-unary languages. In Section 7 we review related work and in Section 8 we conclude. All proofs of technical results can be found in an online appendix¹.

2 Relational Probabilistic Conditional Logic and Inductive Reasoning

In the following, we give a brief overview on the syntax of *relational probabilistic conditional logic* (RPCL) and *averaging* and *aggregating semantics*, see [6] for a

¹ http://ls1-www.cs.tu-dortmund.de/~thimm/misc/thimm_lifted_dkb11_proofs.pdf

discussion. We consider only a fragment of a first-order language, so let Σ be a first-order signature consisting of a finite set of predicate symbols and without functions of arity greater than zero. We also assume that Σ contains some fixed and finite set of constant symbols U_Σ , i. e. functions of arity zero. An *atom* is a predicate together with some terms (variables or constant symbols), e. g., if $a/3$ is a predicate of arity three, $\mathbf{a}, \mathbf{b} \in U_\Sigma$ and \mathbf{X} is a variable then $a(\mathbf{a}, \mathbf{X}, \mathbf{b})$ is an atom. Let \mathcal{L}_Σ be the corresponding first-order language over the signature Σ that is generated in the usual way using negation, conjunction, and disjunction, but without quantifiers. If appropriate we abbreviate conjunctions $\phi \wedge \psi$ by $\phi\psi$ and negation $\neg\psi$ by $\bar{\psi}$. We denote constants with a beginning lowercase, variables with a beginning uppercase letter, and vectors of these with $\vec{\mathbf{a}}$ resp. $\vec{\mathbf{X}}$.

The central notion of RPCL is the probabilistic conditional, see also [12].

Definition 1. *Let $\psi, \phi \in \mathcal{L}_\Sigma$ be some formulas and $d \in [0, 1]$. Then $(\psi | \phi)[d]$ is called a probabilistic conditional.*

A probabilistic conditional $(\psi | \phi)[d]$ is meant to represent the uncertain rule “if ϕ then ψ with probability d ”. If ϕ is tautological, i. e., if $\phi \equiv \top$, we abbreviate $(\psi | \phi)[d]$ by $(\psi)[d]$ and call $(\psi)[d]$ a *probabilistic fact*. A *knowledge base* \mathcal{R} is a finite set of probabilistic conditionals.

For a formula $\phi \in \mathcal{L}_\Sigma$ let $\text{Const}(\phi) \subseteq U_\Sigma$ denote the set of constant symbols appearing in ϕ and let $\text{Var}(\psi)$ denote the set of variable symbols appearing in ψ . The operators $\text{Const}(\cdot)$ and $\text{Var}(\cdot)$ are extended to probabilistic conditionals and knowledge bases in the usual way. Let now x be either a formula, a probabilistic conditional, or a knowledge base. If $\text{Var}(x) = \emptyset$ then x is called *ground*. Furthermore, let $\text{gnd}_\Sigma(x)$ denote the set of ground instances of x with respect to the constant symbols in U_Σ . For example, if $U_\Sigma = \{c_1, c_2, c_3\}$ and $a(X) \in \mathcal{L}_\Sigma$ then $\text{gnd}_\Sigma(a(X)) = \{a(c_1), a(c_2), a(c_3)\}$. Note that for ground x we have $\text{gnd}_\Sigma(x) = \{x\}$.

In order to interpret the classical formulas within conditionals we use *Herbrand interpretations*, i. e. sets of ground atoms of Σ . Let $\Omega(\Sigma)$ denote the set of all Herbrand interpretations for the signature Σ . If $\phi \in \mathcal{L}_\Sigma$ is ground then $\omega \in \Omega(\Sigma)$ *satisfies* ϕ , denoted by $\omega \models^F \phi$, by the usual definition. Note that every $\omega \in \Omega(\Sigma)$ is finite and $\Omega(\Sigma)$ is finite as well as Σ contains only finitely many predicate and constant symbols. Semantics are given to relational probabilistic conditionals by means of probability functions $P : \Omega(\Sigma) \rightarrow [0, 1]$, so let $\mathcal{P}^F(\Sigma)$ denote the set of all these probability functions. A probability function $P \in \mathcal{P}^F(\Sigma)$ is extended to ground formulas $\phi \in \mathcal{L}_\Sigma$ via

$$P(\phi) =_{\text{def}} \sum_{\omega \in \Omega(\Sigma), \omega \models^F \phi} P(\omega) \quad . \quad (1)$$

The approach of *averaging semantics* interprets a probabilistic conditional $r = (\psi | \phi)[d]$ with variables by imposing that the average conditional probability of the instances of $(\psi | \phi)[d]$ matches d . For a probability function $P \in \mathcal{P}^F(\Sigma)$ we abbreviate with $\text{gnd}_\Sigma^P(r) =_{\text{def}} \{(\phi' | \psi')[d] \in \text{gnd}_\Sigma(r) \mid P(\psi') > 0\}$ the set of ground instances of r for which the premise has a non-zero probability in P .

Then P \emptyset -satisfies $r = (\psi | \phi)[d]$ ($P \models_{\emptyset}^{pr} r$) iff $|\mathbf{gnd}_{\Sigma}^P(r)| > 0$ and

$$\frac{\sum_{(\psi' | \phi')[d] \in \mathbf{gnd}_{\Sigma}^P((\psi | \phi)[d])} P(\psi' | \phi')}{|\mathbf{gnd}_{\Sigma}^P((\psi | \phi)[d])|} = d \quad . \quad (2)$$

The interpretation behind the above equation is that a probability function P \emptyset -satisfies a probabilistic conditional $(\psi | \phi)[d]$ if the average of probabilities of the individual instances of $(\psi | \phi)[d]$ is d . By considering only those ground instances where the premise has probability greater zero we average only over the probabilities of ground instances that are relevant for the open conditional.

Example 1. Consider the probabilistic conditional $r = (b(X) | a(X))[0.7]$ and $U_{\Sigma} = \{c_1, c_2, c_3\}$. Let P be a probability function with $P(a(c_1)) > 0$, $P(a(c_2)) > 0$, and $P(a(c_3)) > 0$. If e.g. $P(b(c_1) | a(c_1)) = 0.9$, $P(b(c_2) | a(c_2)) = P(b(c_3) | a(c_3)) = 0.6$ then $P \models_{\emptyset}^{pr} r$.

The similar approach of *aggregating semantics* is defined as follows². A probability function P \odot -satisfies a probabilistic conditional $r = (\psi | \phi)[d]$ ($P \models_{\odot}^{pr} r$) iff $\sum_{(\psi' | \phi')[d] \in \mathbf{gnd}_{\Sigma}^P((\psi | \phi)[d])} P(\phi') > 0$ and

$$\frac{\sum_{(\psi' | \phi')[d] \in \mathbf{gnd}_{\Sigma}^P((\psi | \phi)[d])} P(\psi' | \phi')}{\sum_{(\psi' | \phi')[d] \in \mathbf{gnd}_{\Sigma}^P((\psi | \phi)[d])} P(\phi')} = d \quad .$$

For a knowledge base \mathcal{R} it holds that $P \models_{\emptyset}^{pr} \mathcal{R}$ ($P \models_{\odot}^{pr} \mathcal{R}$) iff $P \models_{\emptyset}^{pr} r$ ($P \models_{\odot}^{pr} r$) for all $r \in \mathcal{R}$. A knowledge base \mathcal{R} is \emptyset -consistent (\odot -consistent) if there is a probability function P with $P \models_{\emptyset}^{pr} \mathcal{R}$ ($P \models_{\odot}^{pr} \mathcal{R}$). Both averaging and aggregating semantics are extensions of the standard semantics for propositional probabilistic conditional logic. More precisely, if $r = (\psi | \phi)[d]$ is a ground probabilistic conditional, i.e. $\mathbf{Var}(r) = \emptyset$, then $P \models_{\emptyset}^{pr} r$ iff $P \models_{\odot}^{pr} r$ iff $P(\phi) > 0$ and $P(\psi | \phi) = P(\psi\phi)/P(\phi) = d$. However, the semantics are quite different in general, see [14] for a discussion.

In the following, let $\circ \in \{\emptyset, \odot\}$ be one of the semantics presented above. One can define a model-based inductive reasoning operator \mathcal{I}_{\circ} —which maps a knowledge base \mathcal{R} onto a “suitable” probability function $\mathcal{I}_{\circ}(\mathcal{R})$ with $\mathcal{I}_{\circ}(\mathcal{R}) \models_{\circ}^{pr} \mathcal{R}$ —as follows, cf. [9]. Let the *entropy* $H(P)$ of a probability function $P \in \mathcal{P}^F(\Sigma)$ be defined via $H(P) = -\sum_{\omega \in \Omega(\Sigma)} P(\omega) \text{ld} P(\omega)$.³ The entropy measures the amount of indeterminateness of a probability function P . By selecting a model of a knowledge base \mathcal{R} that has maximal entropy one gets a probability function that both satisfies all conditionals in \mathcal{R} and adds as less additional information (in the information-theoretic sense) as possible [9].

Definition 2. *Let \mathcal{R} be \circ -consistent. Then the maximum entropy model $\mathcal{I}_{\circ}(\mathcal{R})$ of \mathcal{R} is defined via*

$$\mathcal{I}_{\circ}(\mathcal{R}) = \arg \max_{P \models_{\circ}^{pr} \mathcal{R}} H(P) \quad . \quad (3)$$

² For a justification for the aggregating semantics see [6,14].

³ $\text{ld} x$ is the binary logarithm of x with $0 \text{ld} 0 = 0$.

Note that $\mathcal{I}_\emptyset(\mathcal{R})$ has not yet proven to be well-defined in general, see [14] for a discussion. But—as this issue is not the topic of the current work—we assume in the following that (3) is always well-defined.

Inference via \mathcal{I}_\circ satisfies a series of rationality postulates such as the System P properties [15] and several properties for relational probabilistic reasoning [14]. One of this properties is *prototypical indifference* which can be exploited for our purpose of lifted inference. In order to state this property we need some further notation. If x is either a formula, a probabilistic conditional, or a knowledge base and $c_1, c_2 \in U_\Sigma$ then $x[c_1 \leftrightarrow c_2]$ is the same as x except that every occurrence of c_1 is replaced with c_2 and vice versa.

Definition 3. *Let \mathcal{R} be a knowledge base. The constants $c_1, c_2 \in U_\Sigma$ are \mathcal{R} -equivalent ($c_1 \equiv_{\mathcal{R}} c_2$) iff $\mathcal{R} = \mathcal{R}[c_1 \leftrightarrow c_2]$.*

Observe that $\equiv_{\mathcal{R}}$ is indeed an equivalence relation. Two \mathcal{R} -equivalent constants c_1 and c_2 are indistinguishable with respect to knowledge base \mathcal{R} . That is, \mathcal{R} models exactly the same knowledge on both c_1 and c_2 . Also note that every two $c_1, c_2 \in U_\Sigma$ with $c_1, c_2 \notin \text{Const}(\mathcal{R})$ are \mathcal{R} -equivalent.

Definition 4. *A set $S = \{c' \mid c' \equiv_{\mathcal{R}} c\} \subseteq U_\Sigma$ for $c \in U_\Sigma$ is called \mathcal{R} -equivalence class and $\mathfrak{S}(\mathcal{R})$ is the set of all \mathcal{R} -equivalence classes.*

In [6] it has been shown that \mathcal{I}_\circ satisfies the following property of *prototypical indifference*.

Theorem 1 (Prototypical indifference). *If \mathcal{R} is \circ -consistent then $c_1 \equiv_{\mathcal{R}} c_2$ implies $\mathcal{I}_\circ(\mathcal{R})(\psi) = \mathcal{I}_\circ(\mathcal{R})(\psi[c_1 \leftrightarrow c_2])$ for every ground formula ψ .*

The above theorem implies that the probability function $\mathcal{I}_\circ(\mathcal{R})$ carries a lot of redundant information. Consider the following example.

Example 2. Let $U_\Sigma =_{\text{def}} \{\text{tweety}, \text{huey}, \text{dewey}, \text{louie}\}$ be a set of constant symbols and let $\mathcal{R}_{\text{birds}} =_{\text{def}} \{(\text{flies}(\text{X}))[0.8], (\text{flies}(\text{tweety}))[0.3]\}$ be a knowledge base stating that 80 % of all birds fly and that Tweety flies only up to a degree of belief of 0.3. Consider now the probability function $P^* = \mathcal{I}_\circ(\mathcal{R}_{\text{birds}})$ which is defined on the set of Herbrand interpretations $\Omega(\Sigma) = \{\omega_0, \dots, \omega_{15}\}$ with e. g.

$$\begin{aligned} \omega_5 &=_{\text{def}} \{\text{flies}(\text{tweety}), \text{flies}(\text{huey})\} & \omega_6 &=_{\text{def}} \{\text{flies}(\text{tweety}), \text{flies}(\text{dewey})\} \\ \omega_7 &=_{\text{def}} \{\text{flies}(\text{tweety}), \text{flies}(\text{louie})\} & \omega_8 &=_{\text{def}} \{\text{flies}(\text{huey}), \text{flies}(\text{dewey})\} \\ \omega_9 &=_{\text{def}} \{\text{flies}(\text{huey}), \text{flies}(\text{louie})\} & \omega_{10} &=_{\text{def}} \{\text{flies}(\text{dewey}), \text{flies}(\text{louie})\} \end{aligned}$$

The \mathcal{R} -equivalence classes $\mathfrak{S}(\mathcal{R}) = \{S_1, S_2\}$ of \mathcal{R} are given by $S_1 = \{\text{tweety}\}$ and $S_2 = \{\text{huey}, \text{dewey}, \text{louie}\}$ and due to Theorem 1 it follows that e. g. $P^*(\psi) = P^*(\psi[\text{huey} \leftrightarrow \text{dewey}])$ for every ground sentence ψ . In particular, as every $\omega \in \Omega(\Sigma)$ can be understood as a ground conjunction we obtain $P^*(\omega_5) = P^*(\omega_6) = P^*(\omega_7)$. Therefore, it suffices to represent P^* by only eight Herbrand interpretations as the other eight contain only redundant information.

For the rest of this paper we elaborate on the idea suggested in the above example.

3 Condensed Probability Functions

In the previous section the notion of \mathcal{R} -equivalence has been introduced as a relation among constant symbols, cf. Definition 3. We can generalize this relation to be applicable on Herbrand interpretations as follows. Let $\mathfrak{S}(\mathcal{R}) = \{S_1, \dots, S_n\}$.

Definition 5. *Let $\omega_1, \omega_2 \in \Omega(\Sigma)$. Then ω_1 and ω_2 are \mathcal{R} -equivalent, denoted by $\omega_1 \equiv_{\mathcal{R}} \omega_2$, if there is some $G \in \mathbb{N}$ and a set $T = \{(c_1^1, c_2^1), \dots, (c_1^G, c_2^G)\} \subseteq S_1 \times S_1 \cup \dots \cup S_n \times S_n$ such that $\omega_1 = \omega_2[c_1^1 \leftrightarrow c_2^1] \dots [c_1^G \leftrightarrow c_2^G]$.*

Basically, ω_1 and ω_2 are \mathcal{R} -equivalent if we can permute elements within each \mathcal{R} -equivalence class such that ω_2 becomes ω_1 , e. g. in Example 2 we have $\omega_5 \equiv_{\mathcal{R}_{\text{birds}}} \omega_6 \equiv_{\mathcal{R}_{\text{birds}}} \omega_7$. It is also easy to see that $\equiv_{\mathcal{R}}$ is an equivalence relation and, therefore, both the \mathcal{R} -equivalence class $[\omega] =_{\text{def}} \{\omega' \in \Omega(\Sigma) \mid \omega \equiv_{\mathcal{R}} \omega'\}$ and the quotient set $\Omega(\Sigma)/\equiv_{\mathcal{R}} =_{\text{def}} \{[\omega] \mid \omega \in \Omega(\Sigma)\}$ are well-defined.

Proposition 1. *If $\omega_1 \equiv_{\mathcal{R}} \omega_2$ then $\mathcal{I}_o(\omega_1) = \mathcal{I}_o(\omega_2)$.*

The above proposition states that the probability function \mathcal{I}_o carries a lot of redundant information stemming from the \mathcal{R} -equivalence of certain $\omega \in \Omega(\Sigma)$. In the following, we exploit this observation by using $\Omega(\Sigma)/\equiv_{\mathcal{R}}$ instead of $\Omega(\Sigma)$ for redefining \mathcal{I}_o . To do so, we go on by developing a method that enumerates the elements of $\Omega(\Sigma)/\equiv_{\mathcal{R}}$ in an effective way.

For the rest of this section we restrain our attention to signatures containing only unary predicates. Therefore, let $Pred =_{\text{def}} \{p_1, \dots, p_P\}$ be the set of unary predicates of Σ . We discuss the issue of generalizing our approach in Section 6.

Definition 6. *A truth configuration t for $Pred$ is an expression $t =_{\text{def}} \dot{p}_1 \dots \dot{p}_P$ with $\dot{p}_i \in \{p_i, \bar{p}_i\}$ for $i = 1, \dots, P$. Let Θ denote the set of all truth configurations.*

A truth configuration is meant to characterize the state of a constant c in some interpretation as it enumerates which predicates apply for c and which do not. For a constant c and a truth configuration $t = \dot{p}_1 \dots \dot{p}_P$ define $t^\wedge(c) =_{\text{def}} \dot{p}_1(c) \wedge \dots \wedge \dot{p}_P(c)$. Furthermore, for a ground sentence ϕ and constants c_1, \dots, c_n let

$$\begin{aligned} \Theta(\phi, c_1) &=_{\text{def}} \{t \in \Theta \mid t^\wedge(c_1) \wedge \phi \not\models^F \perp\} \\ \Theta(\phi, c_1, \dots, c_n) &=_{\text{def}} \Theta(\phi, c_1) \times \dots \times \Theta(\phi, c_n) \quad . \end{aligned}$$

The set $\Theta(\phi, c_1)$ contains all those truth configurations t for a constant c_1 that are compatible with some sentence ϕ . The set $\Theta(\phi, c_1, \dots, c_n)$ extends this notion to tuples of constants.

Example 3. Let $Pred =_{\text{def}} \{p_1/1, p_2/1\}$ and let $\psi =_{\text{def}} p_1(c) \wedge (p_2(c) \vee p_2(d))$. Then it holds that $\Theta(\psi, c) = \{p_1 p_2, p_1 \bar{p}_2\}$.

Definition 7. *An instance assignment I is a function $I : \mathfrak{S}(\mathcal{R}) \rightarrow \mathbb{N}_0$ with $I(S_i) \leq |S_i|$ for all $i = 1, \dots, n$. Let \mathfrak{I} denote the set of all instance assignments.*

An instance assignment I assigns to each \mathcal{R} -equivalence class the number of constants that are part of the current instance, see below.

Definition 8. A reference world $\hat{\omega}$ is a function $\hat{\omega} : \Theta \rightarrow \mathcal{I}$ that satisfies

$$\sum_{t \in \Theta} \hat{\omega}(t)(S_i) = |S_i| \quad (\text{for all } i = 1, \dots, n) \quad . \quad (4)$$

Let $\hat{\Omega}$ be the set of all reference worlds.

Basically, a reference world is a function that maps a truth configuration to the number of constants of each \mathcal{R} -equivalence class that satisfy this truth configuration. As we show later, a reference world is a compact representation of $[\omega]$ for some $\omega \in \Omega(\Sigma)$.

Example 4. We continue Example 2. The set of truth configurations $\Theta = \{t_1, t_2\}$ with respect to Σ and $\mathcal{R}_{\text{birds}}$ is given via $t_1 = \text{flies}$ and $t_2 = \overline{\text{flies}}$. Consider $I, I' \in \mathcal{I}$ with $I(S_1) = 0$, $I(S_2) = 2$, $I'(S_1) = 1$, $I'(S_2) = 1$ and $\hat{\omega} \in \hat{\Omega}$ with $\hat{\omega}(t_1) = I$ and $\hat{\omega}(t_2) = I'$. The intuitive description of $\hat{\omega}$ is that $\hat{\omega}$ represents a state where the one element of S_1 does not fly and two elements of S_2 do fly.

In the following we show that $\hat{\Omega}$ is indeed a characterization of the quotient set $\Omega(\Sigma)/\equiv_{\mathcal{R}}$. For that, consider the following definition.

Definition 9. The equivalence mapping κ is the function $\kappa : \Omega(\Sigma) \rightarrow \hat{\Omega}$ defined as $\kappa(\omega) =_{\text{def}} \hat{\omega}$ with $\hat{\omega}(\dot{p}_1 \dots \dot{p}_P)(S_i) =_{\text{def}} |\{\mathbf{c} \in S_i \mid \omega \models^F \dot{p}_1(\mathbf{c}) \wedge \dots \wedge \dot{p}_P(\mathbf{c})\}|$ for every $\dot{p}_1 \dots \dot{p}_P \in \Theta$ and $i = 1, \dots, n$.

The function κ maps a $\omega \in \Omega(\Sigma)$ onto a reference world $\hat{\omega} \in \hat{\Omega}$ with the intended meaning that $\kappa(\omega)$ is the (unique) reference world that represents ω . It holds that $\kappa(\omega) = \hat{\omega}$ whenever $\hat{\omega}$ assigns the same number of elements of an \mathcal{R} -equivalence class S_i to some truth configuration t as ω contains specific instances of this truth configuration for elements in S_i . Also note that κ is surjective.

Let the *span number* $\rho_{\hat{\omega}}$ of a reference world $\hat{\omega} \in \hat{\Omega}$ be defined as⁴

$$\rho_{\hat{\omega}} =_{\text{def}} \prod_{i=1}^n \binom{|S_i|}{\hat{\omega}(t_1)(S_i), \dots, \hat{\omega}(t_T)(S_i)}$$

with $\Theta = \{t_1, \dots, t_T\}$. Note that $\rho_{\hat{\omega}}$ is well-defined as $\hat{\omega}(t_1)(S_i) + \dots + \hat{\omega}(t_T)(S_i) = |S_i|$ for every $\hat{\omega}$. The span number of $\hat{\omega}$ is exactly the number of Herbrand interpretations that are subsumed by $\hat{\omega}$.

Proposition 2. It holds that $|\kappa^{-1}(\hat{\omega})| = \rho_{\hat{\omega}}$ for every $\hat{\omega} \in \hat{\Omega}$.

The following proposition states that $\hat{\Omega}$ indeed characterizes $\Omega(\Sigma)/\equiv_{\mathcal{R}}$.

Proposition 3. The function $\iota : \Omega(\Sigma)/\equiv_{\mathcal{R}} \rightarrow \hat{\Omega}$ with $\iota([\omega]) = \kappa(\omega)$ is a bijection.

After having established the equivalence of $\Omega(\Sigma)/\equiv_{\mathcal{R}}$ and $\hat{\Omega}$ we now turn to the issue of representing \mathcal{I}_{\circ} on the basis of $\hat{\Omega}$.

⁴ $\binom{n}{k_1, \dots, k_r} =_{\text{def}} \frac{n!}{k_1! \dots k_r!}$ is the *multinomial coefficient* indexed by n and k_1, \dots, k_r with $n = k_1 + \dots + k_r$ and $\binom{n}{k_1, \dots, k_r} =_{\text{def}} 0$ if any $k_i < 0$ for $i = 1, \dots, r$.

Definition 10. A probability function $P : \Omega(\Sigma) \rightarrow [0, 1]$ is called prototypically uniform wrt. \mathcal{R} iff for $\omega_1, \omega_2 \in \Omega(\Sigma)$ with $\omega_1 \equiv_{\mathcal{R}} \omega_2$ it follows that $P(\omega_1) = P(\omega_2)$.

Note that $\mathcal{I}_o(\mathcal{R})$ is prototypically uniform wrt. \mathcal{R} . Prototypically uniform probability functions can be be concisely represented as follows.

Definition 11. Let P be a probability function $P : \Omega(\Sigma) \rightarrow [0, 1]$ that is prototypically uniform wrt. \mathcal{R} . Then the condensed probability function \hat{P} for P is the probability function $\hat{P} : \hat{\Omega} \rightarrow [0, 1]$ defined via $\hat{P}(\hat{\omega}) =_{def} P(\omega)$ for some ω with $\kappa(\omega) = \hat{\omega}$ and for all $\hat{\omega} \in \hat{\Omega}$. Let $\hat{\mathcal{P}}$ denote the set of all condensed probability functions.

As $\kappa(\omega_1) = \kappa(\omega_2)$ implies $P(\omega_1) = P(\omega_2)$ for prototypically uniform P the function \hat{P} is well-defined. It also holds that the mapping between prototypically uniform probability functions and condensed probability functions is bijective.

Proposition 4. Let P_1, P_2 be prototypically uniform probability functions wrt. \mathcal{R} . It holds that $P_1 = P_2$ iff $\hat{P}_1 = \hat{P}_2$.

For a prototypically uniform probability function P , its condensed probability function \hat{P} , and a ground sentence ψ it follows directly by definition that

$$\hat{P}(\psi) =_{def} P(\psi) = \sum_{\omega \in \Omega(\Sigma), \omega \models^F \psi} \hat{P}(\kappa(\omega)) \quad . \quad (5)$$

As one can see, one can determine the probability of any ground sentence using \hat{P} instead of P . However, the sum in the above equation still considers every $\omega \in \Omega(\Sigma)$. In the next section we consider the question of how to determine the probability of ψ *without* considering $\Omega(\Sigma)$ but only $\hat{\Omega}$ instead.

4 Lifted Inference

Looking closer at Equation (5) one can see that the probability of a $\hat{\omega} \in \hat{\Omega}$ may occur more than once within the sum as for different $\omega, \omega' \in \Omega(\Sigma)$ with $\omega \models^F \psi$ and $\omega' \models^F \psi$ it may hold that $\kappa(\omega) = \kappa(\omega')$. Therefore, (5) can be rewritten to

$$\hat{P}(\psi) = \sum_{\hat{\omega} \in \hat{\Omega}} A(\hat{\omega}, \psi) \hat{P}(\hat{\omega}) \quad . \quad (6)$$

with $A(\hat{\omega}, \psi) = |\{\omega \in \Omega(\Sigma) \mid \kappa(\omega) = \hat{\omega} \wedge \omega \models^F \psi\}| \in \mathbb{N}_0$, i.e., $A(\hat{\omega}, \psi)$ is the number of $\omega \in \Omega(\Sigma)$ in (5) that satisfy ψ and are mapped by κ to $\hat{\omega}$. Note, however, that determining $A(\hat{\omega}, \psi)$ by its definition above still requires considering all $\omega \in \Omega(\Sigma)$. By exploiting combinatorial patterns within the structure of $\Omega(\Sigma)$ we can avoid considering $\Omega(\Sigma)$ as a whole and characterize $A(\hat{\omega}, \psi)$ as follows.

Proposition 5. Let ψ be a conjunction of ground literals, let $\text{Const}(\psi) = \{c_1, \dots, c_m\}$, and let $\Theta = \{t_1, \dots, t_T\}$. Then

$$A(\hat{\omega}, \psi) = \sum_{(t'_1, \dots, t'_m) \in \Theta(\psi, c_1, \dots, c_m)} \prod_{i=1}^n \binom{|S_i \setminus \text{Const}(\psi)|}{\alpha_i^{t'_1}(t'_1, \dots, t'_m), \dots, \alpha_i^{t'_T}(t'_1, \dots, t'_m)}$$

with $\alpha_i^t(t'_1, \dots, t'_m) =_{def} \hat{\omega}(t)(S_i) - |\{k \mid t'_k = t \wedge c_k \in S_i\}|$.

Note that there is no more reference to $\Omega(\Sigma)$ in the above characterization of $A(\hat{\omega}, \psi)$.

In order to determine $\hat{P}(\psi)$ for an arbitrary ground sentence ψ remember that ψ can be rewritten to be in disjunctive normal form. Assume ψ to be in disjunctive normal form and let $c(\psi)$ denote the set of conjuncts of ψ . Then we can write

$$\hat{P}(\psi) = \sum_{\psi' \in c(\psi)} \hat{P}(\psi') - \sum_{(\psi', \psi'') \in c(\psi)^2, \psi' \neq \psi''} \hat{P}(\psi' \wedge \psi'') \quad .$$

As for \hat{P} , for every $\psi', \psi'' \in c(\psi)$ the terms $\hat{P}(\psi')$ and $\hat{P}(\psi' \wedge \psi'')$ are well-defined by Equation (6) and Proposition 5.

So far, we have shown how that \hat{P}^* compactly represents $P^* = \mathcal{I}_o(\mathcal{R})$ and that \hat{P}^* can be used for reasoning just as P^* . Nonetheless, in order to determine \hat{P}^* one needs to compute P^* first using Equation (3). In the following, we show that we can modify (3) in a straightforward fashion to determine \hat{P}^* directly. Note, although the approach of condensed probability distributions is applicable to any inductive inference mechanism that obeys *prototypical indifference* we restrain our attention to \mathcal{I}_o .

For a condensed probability function \hat{P} we define the *entropy* $H(\hat{P})$ of \hat{P} to be the entropy of P , i. e. $H(\hat{P}) =_{def} H(P)$, which is equivalent to

$$H(\hat{P}) = - \sum_{\hat{\omega} \in \hat{\Omega}} \rho_{\hat{\omega}} \hat{P}(\hat{\omega}) \text{ld } \hat{P}(\hat{\omega})$$

and thus can be determined by just considering $\hat{\Omega}$.

Proposition 6. *Let \mathcal{S} be a set of prototypical uniform probability functions wrt. \mathcal{R} and*

$$\hat{\mathcal{S}} =_{def} \{ \hat{P} \mid P \in \mathcal{S} \} \quad .$$

If the probability function $P_1 = \arg \max_{P \in \mathcal{S}} H(P)$ is uniquely determined so is $\hat{Q} = \arg \max_{\hat{P} \in \hat{\mathcal{S}}} H(\hat{P})$ and it holds that $\hat{Q} = \hat{P}_1$.

Proposition 7. *Let $\mathcal{S} =_{def} \{ P \mid P \models_{\circ}^{pr} \mathcal{R} \}$ and let $\mathcal{S}' \subseteq \mathcal{S}$ be its subset of prototypical uniform probability functions with respect to \mathcal{R} . If $\arg \max_{P \in \mathcal{S}} H(P)$ is uniquely determined then it holds that*

$$\arg \max_{P \in \mathcal{S}'} H(P) = \arg \max_{P \in \mathcal{S}} H(P) \quad .$$

The implications of the above two propositions are as follows. Instead of determining first $P^* = \mathcal{I}_o(\mathcal{R})$ via (3) and then determining \hat{P}^* we can directly determine \hat{P}^* by rewriting (3) to

$$\widehat{\mathcal{I}_o(\mathcal{R})} = \arg \max_{\hat{P} \in \hat{\mathcal{P}} \text{ and } \hat{P} \models_{\circ}^{pr} \mathcal{R}} H(P) \quad (7)$$

Note that $\hat{P} \models_{\circ}^{pr} \mathcal{R}$ can be checked directly for \hat{P} by employing Equation (6) and Proposition 5.

5 Analysis

We now analyze the computational benefits of using \hat{P}^* instead of $P^* = \mathcal{I}_c(\mathcal{R})$. In particular, we are interested in the question how the cardinality of $\hat{\Omega}$ compares to the cardinality of $\Omega(\Sigma)$ with respect to the number of constants $|U_\Sigma|$ considered. It is easy to see that $|\Omega(\Sigma)| = 2^{|U_\Sigma||Pred|}$ (remember that $Pred$ is the set of (unary) predicates in Σ) and therefore the space needed to represent P^* is exponential in both $|U_\Sigma|$ and $|Pred|$. We do not expect to avoid an exponential blow-up in the number of predicates in $Pred$ but we show that $|\hat{\Omega}|$ is not exponential in $|U_\Sigma|$ any more. Remember that each $\hat{\omega} \in \hat{\Omega}$ satisfies

$$\sum_{t \in \Theta} \hat{\omega}(t)(S_i) = |S_i| \quad (\text{for all } i = 1, \dots, n) \quad .$$

This means, that for each $\hat{\omega}$ the constants of each S_i are distributed among the truth configurations in Θ . Note that $|\Theta| = 2^{|Pred|}$. A distribution of constants of S_i among Θ can be combined with any distribution of constants of S_j for every $i \neq j$, yielding a single reference world $\hat{\omega}$. In order to count the number of reference worlds we need to multiply the number of combinations one can distribute the constants of S_i onto the truth configurations in Θ with the number of combinations for every other S_j ($i \neq j$). Then we get

$$|\hat{\Omega}| = \prod_{i=1}^n |\{(l_1, \dots, l_{2^{|Pred|}}) \in \mathbb{N}_0^{2^{|Pred|}} \mid l_1 + \dots + l_{2^{|Pred|}} = |S_i|\}| \quad . \quad (8)$$

Each factor in the product of the above equation represents the number of combinations the constants of a single \mathcal{R} -equivalence class can be distributed among the possible truth configurations in Θ . The condition $l_1 + \dots + l_{2^{|Pred|}} = |S_i|$ ensures that each constant is exactly assigned one truth configuration in every combination. Still, Equation (8) gives no direct hint on the space needed to represent $\hat{\Omega}$ in terms of $|U_\Sigma|$ and $|Pred|$. But it is possible to rewrite (8) as follows.

Definition 12. *The cardinality generator g_c is the function $g_c : \mathbb{N}_0^2 \rightarrow \mathbb{N}_0$ defined via*

$$g_c(n_1, n_2) =_{def} \begin{cases} \sum_{i=0}^{n_2} g_c(n_1 - 1, n_2 - i) & \text{if } n_2 > 0 \text{ and } n_1 > 0 \\ 1 & \text{if } n_2 = 0 \\ 0 & \text{otherwise} \end{cases} \quad .$$

The intuition behind using g_c to enumerate the number of reference worlds is as follows. The first argument of g_c is meant to represent the number of truth configurations and the second the number of constants in an \mathcal{R} -equivalence class. By defining $g_c(n_1, n_2) = g_c(n_1 - 1, 0) + \dots + g_c(n_1 - 1, n_2)$ we say that the number of combinations to distribute n_2 constants on n_1 truth configuration is equal to the number of combinations to distribute zero constants on $n_1 - 1$ truth configurations plus the number of combinations to distribute one constant on $n_1 - 1$ truth configurations, and so on. The first case describes a setting where we assign all

n_2 constants to the n_1 th truth configuration and as there are no remaining constants left this amounts to the number of $g_c(n_1 - 1, 0)$ remaining combinations. The second case describes a setting where we assign $n_2 - 1$ constants to the n_1 th truth configuration and the remaining single constant to the remaining $n_1 - 1$ truth configurations. The final case describes the setting of assigning no constant to the n_1 th truth configuration and the remaining n_2 constants to the remaining $n_1 - 1$ truth configurations. Consider $g_c(1, 3)$ as the number of combinations to distribute three constants on one truth configuration. Applying the first case of the definition of g_c yields $g_c(1, 3) = g_c(0, 0) + g_c(0, 1) + g_c(0, 2) + g_c(0, 3)$ and therefore the number of combinations to distribute three constants on one truth configuration is to assign all three constants to the one truth configuration, or to assign zero, one, or two to it. Obviously, the latter cases are not valid and the only valid assignment is that three constants are assigned to the one truth configuration. Due to the third case in the definition of g_c the terms $g_c(0, 1)$, $g_c(0, 2)$, and $g_c(0, 3)$ are set to zero.

Proposition 8. *It holds that*

$$|\hat{\Omega}| = \prod_{i=1}^n g_c(2^{|Pred|}, |S_i|) \quad . \quad (9)$$

Still, Equation (9) does not allow to get an idea of the size of $|\hat{\Omega}|$. However, the function g_c can be bounded from above as follows.

Lemma 1. *It holds that $g_c(n_1, n_2) \leq (n_2 + 1)^{n_1}$ for every $n_1, n_2 \in \mathbb{N}_0$.*

Theorem 2. *It holds that $|\hat{\Omega}| \leq (|\text{Const}(\mathcal{R})| + 1)(|U_\Sigma| + 1)^{2^{|Pred|}}$.*

The obvious observation to be made when comparing $|\Omega(\Sigma)|$ to the upper bound of $|\hat{\Omega}|$ is that the latter is not exponential in the number of constants $|U_\Sigma|$. But note that the complexity increases with respect to $|Pred|$. While $|\Omega(\Sigma)|$ is exponential in $|Pred|$, the above bound for $|\hat{\Omega}|$ is exponential in $2^{|Pred|}$. However, we believe that this is due to a very coarse estimation in Lemma 1. Experiments suggest that g_c can be much better estimated.

Conjecture 1. It holds that $g_c(n_1, n_2) \leq (n_2 + 1)^{2^{\text{ld } n_1}}$ (with $\text{ld } 0 = 0$).

The above conjecture would result in an upper bound of $(|\text{Const}(\mathcal{R})| + 1)(|U_\Sigma| + 1)^{2^{|Pred|}}$ which is far more beneficial than the result of Theorem 2. However, until now no formal proof for the above conjecture has been found.

Table 1 shows some exemplary cardinalities of $\Omega(\Sigma)$ and $\hat{\Omega}$ for different values of $|U_\Sigma|$ and $|Pred|$. The knowledge base \mathcal{R} used to determine the \mathcal{R} -equivalences classes in $\hat{\Omega}$ mentions a single constant yielding $\mathfrak{S}(\mathcal{R}) = \{\{c\}, U_\Sigma \setminus \{c\}\}$ for $\text{Const}(\mathcal{R}) = \{c\}$. Table 1 shows that especially for this kind of scenarios employing $\hat{\Omega}$ rather than $\Omega(\Sigma, D)$ is computationally beneficial. The numbers in Table 1 also justify the belief in Conjecture 1.

6 Generalizing Lifted Inference

In contrast to the case without non-unary predicates there is no simple and compact representation of $\Omega(\Sigma)/_{\equiv_{\mathcal{R}}}$ if Σ contains at least one non-unary predicate

$ Pred $	$ U_{\Sigma} $	$ \Omega(\Sigma) $	$ \hat{\Omega} $	$ Pred $	$ U_{\Sigma} $	$ \Omega(\Sigma) $	$ \hat{\Omega} $
1	2	4	4	2	2	16	16
1	8	256	16	2	4	256	64
1	32	4294967296	64	2	8	65536	256

Table 1. Comparison of $|\Omega(\Sigma)|$ and $|\hat{\Omega}|$ with respect to a knowledge base \mathcal{R} with $|\text{Const}(\mathcal{R})| = 1$

and, in particular, no compact way to enumerate the elements of $\Omega(\Sigma)/_{\equiv_{\mathcal{R}}}$. Consider a predicate $p/2$ and \mathcal{R} -equivalence classes S_1 and S_2 . Then there are six different instantiations of p that have to be considered as essentially different with respect to \mathcal{R} -equivalence. For constants $c_1 \in S_1$ and $c_2 \in S_2$ we have the variants $p(c_1, c_2)$ and $p(c_2, c_1)$; for $c_1 \in S_i$ we have $p(c_1, c_1)$ for $i = 1, 2$; for $c_1, c_2 \in S_i$ with $c_1 \neq c_2$ we have $p(c_1, c_2)$ for $i = 1, 2$. An extended notion of truth configuration must adhere to this combinatorial observation and also take the relations into account that arise by transitivity. In the unary case, we used truth configurations to be able to enumerate the elements of $\Omega(\Sigma)/_{\equiv_{\mathcal{R}}}$ in an effective way without considering $\Omega(\Sigma)$ itself. In the non-unary case there seems to be no simple way to extend the concept of truth configuration. This observation has also been made by Grove et. al. in [5] when they attempted to generalize the notion of entropy of an interpretation to non-unary languages, see [5] on page 67 for a discussion.

However, the approach of lifted inference developed in this chapter can be applied for non-unary languages by determining first $\Omega(\Sigma)$ and afterwards (by pair-wise comparisons) merge \mathcal{R} -equivalent interpretations to reference worlds (yielding the quotient set $\Omega(\Sigma)/_{\equiv_{\mathcal{R}}}$). Note that we lose the computational advantage of avoiding to consider the full set $\Omega(\Sigma)$ in this approach. It is also questionable whether using $\Omega(\Sigma)/_{\equiv_{\mathcal{R}}}$ instead of $\Omega(\Sigma)$ for inference is beneficial. Table 2 shows the cardinalities of both $\Omega(\Sigma)$ and $\Omega(\Sigma)/_{\equiv_{\mathcal{R}}}$, depending on the size of U_{Σ} and with respect to a signature containing a single binary predicate and a knowledge base \mathcal{R} with $\text{Const}(\mathcal{R}) = \emptyset$. As \mathcal{R} mentions no constants there is only one single \mathcal{R} -equivalence class which makes this scenario the simplest imaginable. Nonetheless, the cardinality of $\Omega(\Sigma)/_{\equiv_{\mathcal{R}}}$ —although being significantly smaller than the cardinality of $\Omega(\Sigma)$ —still seems to grow exponentially in the number of constants considered. Until now, no formal proofs for lower or upper bounds on the growing behavior of $|\Omega(\Sigma)/_{\equiv_{\mathcal{R}}}|$ have been found. However, Table 2 gives reason to believe that there is no polynomial upper bound for $|\Omega(\Sigma)/_{\equiv_{\mathcal{R}}}|$ in $|U_{\Sigma}|$. As a consequence, lifted inference in RPCL can be doubted to be beneficial at all for non-unary languages.

7 Related Work

The notion of *lifted inference* used in this paper has been adopted from the works [11,13,8] which also use this notion to describe effective reasoning procedures for

$ U_{\Sigma} $	$ \Omega(\Sigma) $	$ \Omega(\Sigma)/\equiv_{\mathcal{R}} $	$ U_{\Sigma} $	$ \Omega(\Sigma) $	$ \Omega(\Sigma)/\equiv_{\mathcal{R}} $
1	2	2	3	512	244
2	16	10	4	65536	12235

Table 2. Comparison of $|\Omega(\Sigma)|$ and $|\Omega(\Sigma)/\equiv_{\mathcal{R}}|$ with respect to a signature that contains a single binary predicate and a knowledge base \mathcal{R} with $\text{Const}(\mathcal{R}) = \emptyset$

relational probabilistic knowledge, see also [1,4] for some recent work.. Although the knowledge representation formalisms of those approaches differ to our approach, the motivation and ideas of those approaches are similar to ours. The work [13]—which extends work begun in [11]—develops an algorithm for lifted probabilistic inference in parametrized belief networks. The basic idea of [11,13] is the observation that in order to determine the probability of some query the information used to infer the probability can be partitioned with respect to the information we have for specific individuals. This approach uses the technique of *variable elimination* to simplify computation of probabilities with respect to equivalencies of undistinguishable constants. We do not give a formal description of the algorithms developed in [11,13] but rather give an idea of the approach by means of an example. Consider the clause $c =_{def} (p(\mathbf{X}) \mid q(\mathbf{X}, \mathbf{Y}), r(\mathbf{Y}))$ and a function cpd_c (conditional probability distribution) which maps each possible truth configuration to a probability, e. g. $\text{cpd}_c(\text{true}, \text{true}, \text{false}) = 0.7$ states that the probability of observing $p(\mathbf{c}_1)$ given that $q(\mathbf{c}_1, \mathbf{c}_2)$ is true and $r(\mathbf{c}_2)$ is false is 0.7 (for all constant symbols $\mathbf{c}_1, \mathbf{c}_2$). Note that c can be instantiated using different assignments for \mathbf{Y} but with the same \mathbf{X} . In this case, one can employ a combining rule such as *noisy-or* [10] to aggregate probabilities, i. e., the *noisy-or* of two probabilities p_1 and p_2 is defined as $1 - (1 - p_1)(1 - p_2)$. Let now $E_{n,m} =_{def} \{q(\mathbf{c}, \mathbf{d}_1), \dots, q(\mathbf{c}, \mathbf{d}_{n+m})\} \cup \{r(\mathbf{d}_1), \dots, r(\mathbf{d}_n), \neg r(\mathbf{d}_{n+1}), \dots, \neg r(\mathbf{d}_{n+m})\}$ be some observed evidence with $n, m \in \mathbb{N}$ and consider determining the probability $P(p(\mathbf{c}) \mid E)$. In e. g. ordinary BLPs [3, Ch. 10] one has to instantiate a ground Bayesian network for the node $p(\mathbf{c})$ with parents $q(\mathbf{d}_1), \dots, q(\mathbf{d}_{n+m}), r(\mathbf{c}, \mathbf{d}_1), \dots, r(\mathbf{c}, \mathbf{d}_{n+m})$, and combine the probabilities using *noisy-or*. This amounts to

$$P(p(\mathbf{c}) \mid E) = 1 - (1 - P(p(\mathbf{c}) \mid q(\mathbf{c}, \mathbf{d}_1), r(\mathbf{d}_1))) \cdot \dots \cdot (1 - P(p(\mathbf{c}) \mid q(\mathbf{c}, \mathbf{d}_{n+m}), r(\mathbf{d}_{n+m}))) \quad .$$

Note that we have the same information for the constant symbols $\mathbf{d}_1, \dots, \mathbf{d}_n$ and $\mathbf{d}_{n+1}, \dots, \mathbf{d}_m$, respectively. It follows that

$$\begin{aligned} P(p(\mathbf{c}) \mid q(\mathbf{c}, \mathbf{d}_1), r(\mathbf{d}_1)) &= \dots = P(p(\mathbf{c}) \mid q(\mathbf{c}, \mathbf{d}_n), r(\mathbf{d}_n)) \\ &= \text{cpd}_c(\text{true}, \text{true}, \text{true}) \\ P(p(\mathbf{c}) \mid q(\mathbf{c}, \mathbf{d}_{n+1}), r(\mathbf{d}_{n+1})) &= \dots = P(p(\mathbf{c}) \mid q(\mathbf{c}, \mathbf{d}_{n+m}), r(\mathbf{d}_{n+m})) \\ &= \text{cpd}_c(\text{true}, \text{true}, \text{false}) \end{aligned}$$

and therefore

$$P(p(\mathbf{c}) \mid E) = 1 - (1 - \text{cpd}_c(\text{true}, \text{true}, \text{true}))^n (1 - \text{cpd}_c(\text{true}, \text{true}, \text{false}))^m \quad .$$

As one can see, we can avoid grounding the full BLP by just considering *prototypical* groundings for c . In [11,13] this idea is elaborated and a series of algorithms is developed that apply this approach to general parametrized belief networks (or BLPs). Obviously, the ideas of [11,13] are very similar to ours and differences lie mainly on the framework used for knowledge representation and the technical implementation. The work [11] uses parametrized belief networks and inference bases on Bayesian networks and [13] uses a framework similar to MLNs [3, Ch. 12]. However, note that both formalisms are first-order extensions of probabilistic networks but we use RPCL and inference based on the principle of maximum entropy. Furthermore, we developed an explicit computational model for representing prototypical uniform probability functions and showed that the use of this model is beneficial in terms of computational complexity. In [11] no hints on the computational advantages of applying first-order variable elimination are given but [13] gives an experimental evaluation that resembles our observations from Conjecture 1.

8 Summary and Conclusion

We developed a computational account for effective probabilistic inference with relational probabilistic conditionals. In particular, we introduced the notions of reference worlds and condensed probability functions which allow for a compact representation of probability functions that arise from the application of inference operators satisfying *prototypical indifference*. Condensed probability functions are defined on the set of reference worlds and exhibit the same reasoning behavior as the original probability functions, given that those are indifferent with respect to constants from the same \mathcal{R} -equivalence class. Furthermore, we showed that the inference operators from [6] can be modified in order to compute the condensed maximum entropy function in a single step without considering the Herbrand interpretations at all. We analyzed the computational benefits of our approach and concluded that we avoid the exponential blow-up in the number of constants that have to be considered. Our approach is—using the given formalization—only applicable for unary languages and we briefly discussed the issues that arise when considering non-unary languages.

The approach developed in this paper gives some first directions for efficient implementation of reasoning based on the principle of maximum entropy. However, the work reported is only a first step towards this goal and suffers from two major discrepancies. Firstly, we restricted lifted inference to the case of unary languages which, in practice, is a demand that cannot be easily fulfilled. One of the main advantages of first-order extensions of probabilistic reasoning is the capability to reason over relations. However, note that even by restricting attention to unary languages we do not get the equivalence to propositional probabilistic models due to our semantical notions. For example, the knowledge base $\mathcal{R} =_{def} \{(\textit{flies}(\mathbf{X}))[\textit{0.9}], (\textit{flies}(\textit{tweety}))[\textit{0.3}]\}$ cannot be represented using a propositional probabilistic model that exhibits the same inference behavior. Secondly, in order to determine the (condensed) maximum entropy model of a knowledge base \mathcal{R} we have to solve a complex optimization problem. However,

there are approaches to avoid solving problems like (3) for the propositional case. For example, in [7] an approximate algorithm for computing the maximum entropy model for propositional probabilistic conditional logic is developed. The algorithm of [7] benefits from several characteristic properties of the maximum entropy model in the propositional case and it is to investigate if these properties (or similar ones) can be found for our semantical approaches which may lead to the development of algorithms for approximate inference in relational probabilistic conditional logic.

Acknowledgements. The research reported here was partially supported by the Deutsche Forschungsgemeinschaft (grant KE 1413/2-1).

References

1. van den Broeck, G., Taghipour, N., Meert, W., Davis, J., Raedt, L.D.: Lifted Probabilistic Inference by First-Order Knowledge Compilation. In: Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI'11). pp. 2178–2185 (2011)
2. Delgrande, J.P.: On first-order conditional logics. *Artificial Intelligence* 105(1-2), 105–137 (1998)
3. Getoor, L., Taskar, B. (eds.): *Introduction to Statistical Relational Learning*. MIT Press (2007)
4. Gogate, V., Domingos, P.: Probabilistic theorem proving. In: Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence (UAI'11) (2011)
5. Grove, A.J., Halpern, J.Y., Koller, D.: Random worlds and maximum entropy. *Journal of Artificial Intelligence Research (JAIR)* 2, 33–88 (1994)
6. Kern-Isberner, G., Thimm, M.: Novel semantical approaches to relational probabilistic conditionals. In: Proceedings of KR'10 (2010)
7. Meyer, C.H.: *Korrekt Schliessen bei unvollständiger Information*. Ph.D. thesis, FernUniversität in Hagen, Germany (1997)
8. Milch, B., Zettlemoyer, L.S., Kersting, K., Haimes, M., Kaelbling, L.P.: Lifted Probabilistic Inference with Counting Formulas. In: Proceedings of AAAI'08 (2008)
9. Paris, J.B.: *The Uncertain Reasoner's Companion – A Mathematical Perspective*. Cambridge University Press (1994)
10. Pearl, J.: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann (1998)
11. Poole, D.: First-Order Probabilistic Inference. In: Proceedings of IJCAI'03. pp. 985–991. Morgan Kaufmann (2003)
12. Rödder, W.: Conditional logic and the principle of entropy. *Artificial Intelligence* 117, 83–106 (2000)
13. de Salvo Braz, R., Amir, E., Roth, D.: Lifted First-Order Probabilistic Inference. In: Proceedings of IJCAI'05. pp. 1319–1325 (2005)
14. Thimm, M.: *Probabilistic Reasoning with Incomplete and Inconsistent Beliefs*. Ph.D. thesis, Technische Universität Dortmund, Germany (2011), submitted
15. Thimm, M., Kern-Isberner, G., Fisseler, J.: Relational probabilistic conditional reasoning at maximum entropy. In: Proceedings of ECSQARU'11 (2011)
16. Wellman, M.P., Breese, J.S., Goldman, R.P.: From Knowledge Bases to Decision Models. *The Knowledge Engineering Review* 7(1), 35–53 (1992)