

Article

Relational Probabilistic Conditionals and Their Instantiations under Maximum Entropy Semantics for First-Order Knowledge Bases

Christoph Beierle ^{1,*}, Marc Finthammer ¹ and Gabriele Kern-Isberner ²

¹ Faculty of Mathematics and Computer Science, University of Hagen, 58084 Hagen, Germany;
E-Mail: marc@finthammer.de

² Department of Computer Science, University of Technology Dortmund, 44227 Dortmund, Germany;
E-Mail: gabriele.kern-isberner@cs.uni-dortmund.de

* Author to whom correspondence should be addressed; E-Mail: beierle@fernuni-hagen.de;
Tel.: +49-2331-9874293; Fax: +49-2331-9874288.

Academic Editors: Juergen Landes and Jon Williamson

Received: 26 December 2014 / Accepted: 9 February 2015 / Published: 13 February 2015

Abstract: For conditional probabilistic knowledge bases with conditionals based on propositional logic, the principle of maximum entropy (ME) is well-established, determining a unique model inductively completing the explicitly given knowledge. On the other hand, there is no general agreement on how to extend the ME principle to relational conditionals containing free variables. In this paper, we focus on two approaches to ME semantics that have been developed for first-order knowledge bases: aggregating semantics and a grounding semantics. Since they use different variants of conditionals, we define the logic PCI, which covers both approaches as special cases and provides a framework where the effects of both approaches can be studied in detail. While the ME models under PCI-grounding and PCI-aggregating semantics are different in general, we point out that parametric uniformity of a knowledge base ensures that both semantics coincide. Using some concrete knowledge bases, we illustrate the differences and common features of both approaches, looking in particular at the ground instances of the given conditionals.

Keywords: conditional logic; probabilistic logic; maximum entropy; relational conditional; first-order knowledge base; instantiation restriction; grounding semantics; aggregating semantics; parametric uniformity; maximum entropy model

1. Introduction

Probabilistic conditional knowledge bases containing conditionals of the form $(B|A)[d]$ with the reading “if A , then B with probability d ” are a powerful means for knowledge representation and reasoning when uncertainty is involved [1,2]. If A and B are propositional formulas over a propositional alphabet Σ , possible worlds correspond to elementary conjunctions over Σ , where an elementary conjunction is a conjunction containing every element of Σ exactly once, either in non-negated or in negated form. A possible worlds semantics is given by probability distributions over the set of possible worlds, and a probability distribution P satisfies $(B|A)[d]$ if for the conditional probability $P(B|A) = \frac{P(A \wedge B)}{P(A)}$ the relation $P(B|A) = d$ holds. For a knowledge base \mathcal{R} consisting of a set of propositional conditionals, P is a model of \mathcal{R} if P satisfies each conditional in \mathcal{R} . The principle of maximum entropy (ME principle) is a well-established concept for choosing the uniquely determined model of \mathcal{R} having maximum entropy. This model is the most unbiased model of \mathcal{R} in the sense that it completes the knowledge given by \mathcal{R} inductively but adds as little additional information as possible [3–9].

While for a set of propositional conditionals there is a general agreement about its ME model, the situation changes when the conditionals are built over a relational first-order language. As an illustration, consider the following example.

Example 1 (Elephant Keeper). *The elephant keeper example, adapted from [10,11], models the relationships among elephants in a zoo and their keepers. Elephants usually like their keepers, except for keeper Fred. However, elephant Clyde gets along with everyone, and therefore he also likes Fred. The knowledge base \mathcal{R}_{EK} consists of the following conditionals:*

$$\begin{aligned} ek_1 &: (\text{likes}(E, K) \mid \text{elephant}(E), \text{keeper}(K))[0.9] \\ ek_2 &: (\text{likes}(E, \text{fred}) \mid \text{elephant}(E), \text{keeper}(\text{fred}))[0.05] \\ ek_3 &: (\text{likes}(\text{clyde}, \text{fred}) \mid \text{elephant}(\text{clyde}), \text{keeper}(\text{fred}))[0.85] \end{aligned}$$

Conditional ek_1 models statistical knowledge about the general relationship between elephants and their keepers, whereas conditional ek_2 represents knowledge about the exceptional keeper Fred and his relationship to elephants in general. Conditional ek_3 models the subjective belief about the relationship between the elephant Clyde and keeper Fred. From a common sense point of view, the knowledge base \mathcal{R}_{EK} makes perfect sense: conditional ek_2 is an exception from ek_1 , and ek_3 is an exception from ek_2 .

When trying to extend the ME principle from the propositional case to such a relational setting, a central question is how to interpret the free variables occurring in a conditional. For instance, note that a straightforward complete grounding of \mathcal{R}_{EK} yields a grounded knowledge base that can be viewed as a propositional knowledge base. However, this grounded knowledge is inconsistent since it contains both $(\text{likes}(\text{clyde}, \text{fred}) \mid \text{elephant}(\text{clyde}), \text{keeper}(\text{fred}))[0.9]$ and $(\text{likes}(\text{clyde}, \text{fred}) \mid \text{elephant}(\text{clyde}), \text{keeper}(\text{fred}))[0.05]$, and no probability distribution P can satisfy both $P(\text{likes}(\text{clyde}, \text{fred}) \mid \text{elephant}(\text{clyde}), \text{keeper}(\text{fred})) = 0.9$ and $P(\text{likes}(\text{clyde}, \text{fred}) \mid \text{elephant}(\text{clyde}), \text{keeper}(\text{fred})) = 0.05$.

Thus, when extending the ME principle to the relational case with free variables as in \mathcal{R}_{EK} , the exact role of the variables has to be specified. There are various approaches dealing with a combination of

probabilities with a first-order language (e.g., [12,13]); a comparison and evaluation of some approaches is given [14]). In the following, we focus on two semantics that both employ the principle of maximum entropy for probabilistic relational conditionals, the *aggregation semantics* [15] proposed by Kern-Isberner and the logic FO-PCL [16] elaborated by Fisseler. While both approaches are related in the sense that they refer to a set of constants when interpreting the variables in the conditionals, there is also a major difference. FO-PCL requires all groundings of a conditional to have the same probability d given in the conditional, and in general, FO-PCL needs to restrict the possible instantiations for the variables occurring in a conditional by providing constraint formulas like $U \neq V$ or $U \neq a$ in order to avoid inconsistencies. On the other hand, under aggregation semantics the grounded instances may have distinct probabilities as long as they aggregate to the given probability d , and aggregation semantics is defined only for conditionals without constraint formulas.

In this paper, a logical framework PCI extending aggregation semantics to conditionals with instantiation restrictions and also providing a grounding semantics is proposed. From a knowledge representation point of view, this provides greater flexibility, e.g., when expressing knowledge about individuals known to be exceptional with respect to some relationship. We show that both the aggregation semantics of [15] and the semantics of FO-PCL [16] come out as special cases of PCI, thereby also helping to clarify the relationship between the two approaches. Moreover, we investigate the ME models under PCI-grounding and PCI-aggregating semantics, which are different in general, and we give a condition on knowledge bases ensuring that both ME semantics coincide.

This paper is a revised and extended version of [17] and is organized as follows. In Section 2, we very briefly recall the background of FO-PCL and aggregation semantics. In Section 3, the logic framework PCI is developed and two alternative satisfaction relations for grounding and aggregating semantics are defined for PCI by extending the corresponding notions of [15,16]. In Section 4, the maximum entropy principle is employed with respect to these satisfaction relations; we show that the resulting semantics coincide for knowledge bases that are *parametrically uniform* [11,16]. In Section 5, we present and discuss ME distributions for some concrete knowledge bases both under PCI-grounding and PCI-aggregating semantics, and point out their differences and common features, covering in particular the groundings of the given conditionals. Finally, in Section 6 we conclude and point out further work.

2. Background: FO-PCL and Aggregation Semantics

As already pointed out in Section 1, simply grounding a relational knowledge base \mathcal{R} easily leads to inconsistency. Therefore, the logic FO-PCL [11,16] employs instantiation restrictions for the free variables of a conditional. An FO-PCL conditional has additionally a constraint formula determining the *admissible instantiations* of free variables, and the grounding semantics of FO-PCL requires that all admissible ground instances of a conditional c must have the probability given by c .

Example 2 (Elephant Keeper with instantiation restrictions). *In FO-PCL, adding $K \neq fred$ to conditional ek_1 and $E \neq clyde$ to conditional ek_2 in \mathcal{R}_{EK} yields the knowledge base \mathcal{R}'_{EK} with:*

$$\begin{aligned} ek'_1 &: \langle (\text{likes}(E, K) \mid \text{elephant}(E), \text{keeper}(K))[0.9], K \neq fred) \rangle \\ ek'_2 &: \langle (\text{likes}(E, fred) \mid \text{elephant}(E), \text{keeper}(fred))[0.05], E \neq clyde) \rangle \\ ek'_3 &: \langle (\text{likes}(clyde, fred) \mid \text{elephant}(clyde), \text{keeper}(fred))[0.85], \top) \rangle \end{aligned}$$

Note that, e.g., the ground instance $(likes(clyde, fred)|elephant(clyde), keeper(fred))[0.05]$ of conditional ek'_2 is not admissible, and that the set of admissible ground instances of \mathcal{R}'_{EK} is indeed consistent under probabilistic semantics for propositional knowledge bases as considered, e.g., in [8,18].

Thus, under FO-PCL semantics, \mathcal{R}'_{EK} is consistent, where a probability distribution P satisfies an FO-PCL conditional r , denoted by $P \models_{\text{fopcl}} r$, iff all admissible ground instances of r have the probability specified by r .

In contrast, the aggregation semantics, as given in [15], does not consider instantiation restrictions, since its satisfaction relation (in this paper denoted by \models_{\odot}^{no-ir} to indicate *no instantiation restriction*), is less strict with respect to probabilities of ground instances: $P \models_{\odot}^{no-ir} (B|A)[d]$ iff the quotient of the sum of all probabilities $P(B_i \wedge A_i)$ and the sum of $P(A_i)$ is d , where $(B_1|A_1), \dots, (B_n|A_n)$ are the ground instances of $(B|A)$. In this way, the aggregation semantics is capable of balancing the probabilities of ground instances, resulting in greater flexibility and higher tolerance with respect to consistency issues. Provided that there are enough individuals so that the corresponding aggregation over all probabilities is possible, the knowledge base \mathcal{R}_{EK} that is inconsistent under FO-PCL semantics is consistent under aggregation semantics.

3. PCI Logic

The logical framework PCI (*probabilistic conditionals with instantiation restrictions*) uses probabilistic conditionals with and without instantiation restrictions and provides different options for a satisfaction relation. The syntax of PCI given in [19] uses the syntax of FO-PCL [11,16]. In the following, we will precisely state the formal relationship among \models_{\odot}^{no-ir} , \models_{fopcl} , and the satisfaction relations offered by PCI.

As FO-PCL, PCI uses function-free, sorted signatures of the form $\Sigma = (\mathcal{S}, \mathcal{D}, Pred)$. In a PCI-signature $\Sigma = (\mathcal{S}, \mathcal{D}, Pred)$, $\mathcal{S} = \{s_1, \dots, s_k\}$ is a set of sort names or just sorts. The set \mathcal{D} is a finite set of constants symbols where each $d \in \mathcal{D}$ has a unique sort $s \in \mathcal{S}$. With $\mathcal{D}^{(s)}$ we denote the set of all constants having sort s ; thus $\mathcal{D} = \bigcup_{s \in \mathcal{S}} \mathcal{D}^{(s)}$ is a set being the union of (disjoint) sets of sorted constant symbols. $Pred$ is a set of predicate symbols, each having a particular number of arguments. If $p \in Pred$ is a predicate taking n arguments, each argument position i must be filled with a constant or variable of a specific sort s_i . Thus, each $p \in Pred$ comes with an arity of the form $s_1 \times \dots \times s_n \in \mathcal{S}^n$ indicating the required sorts for the arguments. Variables \mathcal{V} also have a unique sort, and all formulas and variable substitutions must obey the obvious sort restrictions. In the following, we will adopt the unique names assumption, *i.e.*, different constants denote different elements. The set of all terms is defined as $Term_{\Sigma} := \mathcal{V} \cup \mathcal{D}$. Let \mathcal{L}_{Σ} be the set of quantifier-free first-order formulas defined over Σ and \mathcal{V} in the usual way.

Definition 1 (Instantiation Restriction). *An instantiation restriction is a conjunction of inequality atoms of the form $t_1 \neq t_2$ with $t_1, t_2 \in Term_{\Sigma}$. The set of all instantiation restriction is denoted by \mathcal{C}_{Σ} .*

Since an instantiation restriction may be a conjunction of inequality atoms, we can express that a conditional has multiple restrictions, e.g., by stating $E \neq clyde \wedge K \neq fred$.

Definition 2 (q-, p-, r-Conditional). Let $A, B \in \mathcal{L}_\Sigma$ be quantifier-free first-order formulas over Σ and \mathcal{V} .

1. $(B|A)$ is called a qualitative conditional (or just q-conditional). Note that A is the antecedence and B the consequence of the qualitative conditional. The set of all qualitative conditionals over \mathcal{L}_Σ is denoted by $(\mathcal{L}_\Sigma|\mathcal{L}_\Sigma)$.
2. Let $(B|A) \in (\mathcal{L}_\Sigma|\mathcal{L}_\Sigma)$ be a qualitative conditional and let $d \in [0, 1]$ be a real value. Here $(B|A)[d]$ is called a probabilistic conditional (or just p-conditional) with probability d . The set of all probabilistic conditionals over \mathcal{L}_Σ is denoted by $(\mathcal{L}_\Sigma|\mathcal{L}_\Sigma)^{prob}$.
3. Let $(B|A)[d] \in (\mathcal{L}_\Sigma|\mathcal{L}_\Sigma)^{prob}$ be a probabilistic conditional and let $C \in \mathcal{C}_\Sigma$ be an instantiation restriction. In addition, $\langle (B|A)[d], C \rangle$ is called an instantiation restricted conditional (or just r-conditional). The set of all instantiation restricted conditionals over \mathcal{L}_Σ is denoted by $(\mathcal{L}_\Sigma|\mathcal{L}_\Sigma)_{\mathcal{C}_\Sigma}^{prob}$.

Instantiation restricted qualitative conditionals are defined analogously. If it is clear from the context, we may omit *qualitative*, *probabilistic*, and *instantiation restricted* and just use the term *conditional*.

Definition 3 (PCI knowledge base). A pair (Σ, \mathcal{R}) consisting of a PCI signature $\Sigma = (\mathcal{S}, \mathcal{D}, \text{Pred})$ and a set of instantiation restricted conditionals $\mathcal{R} = \{r_1, \dots, r_m\}$ with $r_i \in (\mathcal{L}_\Sigma|\mathcal{L}_\Sigma)_{\mathcal{C}_\Sigma}^{prob}$ is called a PCI knowledge base.

For an instantiation restricted conditional $r = \langle (B|A)[d], C \rangle$, $\Theta_\Sigma(r)$ denotes the set of all ground substitutions with respect to the variables in r . A ground substitution $\theta \in \Theta_\Sigma(r)$ is applied to the formulas A , B and C in the usual way, i.e., each variable is replaced by a certain constant according to the mapping $\theta = \{v_1/c_1, \dots, v_l/c_l\}$ with $v_i \in \mathcal{V}$, $c_i \in \mathcal{D}$, $1 \leq i \leq l$. Therefore, $\theta(A)$, $\theta(B)$, and $\theta(C)$ are ground formulas and we have $\theta((B|A)) := (\theta(B)|\theta(A))$.

Given a ground substitution θ over the variables occurring in an instantiation restriction $C \in \mathcal{C}_\Sigma$, the evaluation of C under θ , denoted by $\llbracket C \rrbracket_\theta$, yields *true* iff $\theta(t_1)$ and $\theta(t_2)$ are different constants for all $t_1 \neq t_2 \in C$.

Definition 4 (Admissible Ground Substitutions and Instances). Let $\Sigma = (\mathcal{S}, \mathcal{D}, \text{Pred})$ be a many-sorted signature and let $r = \langle (B|A)[d], C \rangle \in (\mathcal{L}_\Sigma|\mathcal{L}_\Sigma)_{\mathcal{C}_\Sigma}^{prob}$ be an instantiation restricted conditional. The set of admissible ground substitutions of r is defined as

$$\Theta_\Sigma^{adm}(r) := \{\theta \in \Theta_\Sigma(r) \mid \llbracket C \rrbracket_\theta = \text{true}\}$$

The set of admissible ground instances of r is defined as

$$gnd_\Sigma(r) := \{\theta(B|A)[d] \mid \theta \in \Theta_\Sigma^{adm}(r)\}$$

In the following, when we talk about the ground instances of a conditional, we will always refer to its admissible ground instances.

As for an FO-PCL knowledge base [11], for a PCI knowledge base (Σ, \mathcal{R}) we define the *Herbrand base* $\mathcal{H}(\mathcal{R})$ as the set of all ground atoms in all $gnd_\Sigma(r_i)$ with $r_i \in \mathcal{R}$. Every subset $\omega \subseteq \mathcal{H}(\mathcal{R})$ is a *Herbrand interpretation*, defining a logical semantics for \mathcal{R} . The set $\Omega_\Sigma := \{\omega \mid \omega \subseteq \mathcal{H}(\mathcal{R})\}$ denotes the set of all Herbrand interpretations. Herbrand interpretations are also called *possible worlds*.

Definition 5 (PCI Interpretation). *The probabilistic semantics of (Σ, \mathcal{R}) is a possible worlds semantics [12] where the ground atoms in $\mathcal{H}(\mathcal{R})$ are binary random variables. A PCI interpretation P of a knowledge base (Σ, \mathcal{R}) is thus a probability distribution $P : \Omega_\Sigma \rightarrow [0, 1]$. The set of all probability distributions over Ω_Σ is denoted by $\mathcal{P}_{\Omega_\Sigma}$ or just by \mathcal{P}_Ω .*

The PCI framework offers two different satisfaction relations: $\models_{\Delta}^{\text{pci}}$ is based on grounding as in FO-PCL, and $\models_{\otimes}^{\text{pci}}$ extends aggregation semantics to r-conditionals.

Definition 6 (PCI Satisfaction Relations). *Let $P \in \mathcal{P}_\Omega$ and let $\langle (B|A)[d], C \rangle \in (\mathcal{L}_\Sigma | \mathcal{L}_\Sigma)_{C_\Sigma}^{\text{prob}}$ be an r-conditional with $\sum_{\theta \in \Theta_\Sigma^{\text{adm}}(\langle (B|A)[d], C \rangle)} P(\theta(A)) > 0$. The two PCI satisfaction relations $\models_{\Delta}^{\text{pci}}$ and $\models_{\otimes}^{\text{pci}}$ are defined by:*

$$P \models_{\Delta}^{\text{pci}} \langle (B|A)[d], C \rangle \text{ iff } \frac{P(\theta(A \wedge B))}{P(\theta(A))} = d \quad \text{for all } \theta \in \Theta_\Sigma^{\text{adm}}(\langle (B|A)[d], C \rangle) \tag{1}$$

$$P \models_{\otimes}^{\text{pci}} \langle (B|A)[d], C \rangle \text{ iff } \frac{\sum_{\theta \in \Theta_\Sigma^{\text{adm}}(\langle (B|A)[d], C \rangle)} P(\theta(A \wedge B))}{\sum_{\theta \in \Theta_\Sigma^{\text{adm}}(\langle (B|A)[d], C \rangle)} P(\theta(A))} = d \tag{2}$$

We say that P satisfies $\langle (B|A)[d], C \rangle$ under PCI-grounding semantics iff $P \models_{\Delta}^{\text{pci}} \langle (B|A)[d], C \rangle$. Correspondingly, P satisfies $\langle (B|A)[d], C \rangle$ under PCI-aggregation semantics iff $P \models_{\otimes}^{\text{pci}} \langle (B|A)[d], C \rangle$.

As usual, the satisfaction relations $\models_{\star}^{\text{pci}}$ with $\star \in \{\Delta, \otimes\}$ are extended to a set of conditionals \mathcal{R} by defining

$$P \models_{\star}^{\text{pci}} \mathcal{R} \text{ iff } P \models_{\star}^{\text{pci}} r \text{ for all } r \in \mathcal{R}.$$

The following proposition states that PCI properly captures both the instantiation-based semantics \models_{fopcl} of FO-PCL [11] and the aggregation semantics $\models_{\odot}^{\text{no-ir}}$ of [15] (cf. Section 2).

Proposition 1 (PCI captures FO-PCL and aggregation semantics [19]). *Let $\langle (B|A)[d], C \rangle$ be an r-conditional and let $(B|A)[d]$ be a p-conditional, respectively. Then the following holds:*

$$P \models_{\Delta}^{\text{pci}} \langle (B|A)[d], C \rangle \text{ iff } P \models_{\text{fopcl}} \langle (B|A)[d], C \rangle \tag{3}$$

$$P \models_{\otimes}^{\text{pci}} \langle (B|A)[d], \top \rangle \text{ iff } P \models_{\odot}^{\text{no-ir}} (B|A)[d] \tag{4}$$

4. PCI Logic and Maximum Entropy Semantics

If a knowledge base \mathcal{R} is consistent, there are usually many different models satisfying \mathcal{R} . The principle of maximum entropy chooses the unique distribution that has maximum entropy among all distributions satisfying a knowledge base \mathcal{R} [5,8]. Applying this principle to the PCI satisfaction relations $\models_{\Delta}^{\text{pci}}$ and $\models_{\otimes}^{\text{pci}}$ yields

$$P_{\mathcal{R}}^{\text{ME}\star} = \arg \max_{P \in \mathcal{P}_\Omega : P \models_{\star}^{\text{pci}} \mathcal{R}} H(P) \tag{5}$$

with \star being Δ or \otimes , and where

$$H(P) = - \sum_{\omega \in \Omega} P(\omega) \log P(\omega)$$

is the *entropy* of a probability distribution P .

Example 3 (Misanthrope). *The knowledge base $\mathcal{R}_{MI} = \{R_1, R_2\}$, adapted from [11], models friendship relations within a group of people with one exceptional member, a misanthrope. In general, if a person V likes another person U , then it is very likely that U likes V , too. However, there is one person, the misanthrope, who generally does not like other people:*

$$R_1: \langle (\text{likes}(U, V) | \text{likes}(V, U)) [0.9], U \neq V \rangle$$

$$R_2: \langle (\text{likes}(a, V) | \top) [0.05], V \neq a \rangle$$

Within the PCI framework, consider \mathcal{R}_{MI} together with constants $\mathcal{D} = \{a, b, c\}$ and the corresponding ME distributions $ME_{\Delta}(\mathcal{R}_{MI})$ and $ME_{\otimes}(\mathcal{R}_{MI})$ under PCI-grounding and PCI-aggregation semantics, respectively.

Under $ME_{\Delta}(\mathcal{R}_{MI})$, all six ground conditionals emerging from R_1 have probability 0.9, for instance, $ME_{\Delta}(\mathcal{R}_{MI})(\text{likes}(a, b) | \text{likes}(b, a)) = 0.9$.

On the other hand, for the distribution $ME_{\otimes}(\mathcal{R}_{MI})$, we have $ME_{\otimes}(\mathcal{R}_{MI})(\text{likes}(a, b) | \text{likes}(b, a)) = 0.46016768$ and $ME_{\otimes}(\mathcal{R}_{MI})(\text{likes}(a, c) | \text{likes}(c, a)) = 0.46016768$, while the other four ground conditionals resulting from R_1 have probability 0.96674480.

Example 3 shows that in general the ME model under PCI-grounding semantics of a knowledge base \mathcal{R} differs from its ME model under PCI-aggregation semantics. However, if \mathcal{R} is *parametrically uniform* [11,16], the situation changes. Parametric uniformity of a knowledge base \mathcal{R} is introduced in [11] and refers to the fact that the ME distribution under FO-PCL (or PCI-grounding) semantics satisfying a set of m ground conditionals can be represented by a set of just m optimization parameters. A relational knowledge base \mathcal{R} is parametrically uniform iff for every conditional $r \in \mathcal{R}$, all ground instances of r have the same optimization parameter (see [11,16] for details). For instance, the knowledge base \mathcal{R}'_{EK} from Example 2 is parametrically uniform, while the knowledge base \mathcal{R}_{MI} from Example 3 is not parametrically uniform. Thus, if \mathcal{R} is parametrically uniform, just *one* optimization parameter for each conditional $r \in \mathcal{R}$ instead of one optimization parameter for *each ground instance* of r has to be computed; this can be exploited when computing the ME distribution [17]. In [20], a set of transformation rules is developed that transforms any consistent knowledge base \mathcal{R} into a knowledge base \mathcal{R}' such that \mathcal{R} and \mathcal{R}' have the same ME model under grounding semantics and \mathcal{R}' is parametrically uniform.

Using the PCI framework providing both grounding and aggregating semantics for conditionals with instantiation restrictions, the ME models for PCI-grounding and PCI-aggregation semantics coincide if \mathcal{R} is parametrically uniform.

Proposition 2 ([19]). *Let \mathcal{R} be a PCI knowledge base. If \mathcal{R} is parametrically uniform, then $ME_{\Delta}(\mathcal{R}) = ME_{\otimes}(\mathcal{R})$.*

Thus, while in general $ME_{\Delta}(\mathcal{R}) \neq ME_{\otimes}(\mathcal{R})$, parametric uniformity of \mathcal{R} ensures that $ME_{\Delta}(\mathcal{R}) = ME_{\otimes}(\mathcal{R})$.

5. Computation and Comparison of Maximum Entropy Distributions

In Example 3 we already presented some concrete probability values for ME distributions. We will now look into more details of the ME distributions obtained from both PCI-grounding and PCI-aggregation semantics. In particular, we will illustrate how the ME distribution for PCI-grounding and PCI-aggregation semantics evolve when transforming a knowledge base that is not parametrically uniform into a knowledge base that is parametrically uniform.

5.1. Achieving Parametric Uniformity

While transforming a knowledge base into one that is parametrically uniform [11] does not change its ME model under (FO-PCL or PCI) grounding semantics, it allows for a simpler ME model computation [17]. In [20], a set of transformation rules \mathcal{PU} is presented allowing to transform any consistent knowledge base \mathcal{R} into a parametrically uniform knowledge base $\mathcal{PU}(\mathcal{R})$ with the same maximum entropy model under grounding semantics. An implementation of \mathcal{PU} [21] is available within the KREATOR environment (KREATOR can be found at <http://kreator-ide.sourceforge.net/>), an integrated development environment for relational probabilistic logic [22]. The *CSPU (Conditional Structures and Parametric Uniformity)* component [23] of KREATOR generates \mathcal{PU} transformation protocols, and a part the protocol for the misanthrope knowledge base \mathcal{R}_{MI} from Example 3 is shown in Figure 1. For details of the \mathcal{PU} transformation rules we refer to [20]; we just remark here that \mathcal{PU} stepwise removes all *interactions* among the conditionals where an interaction in a knowledge \mathcal{R} base indicates that \mathcal{R} is not parametrically uniform [20]. In each \mathcal{PU} transformation step, one conditional R is replaced by two conditionals R_1, R_2 originating from R . Table 1 illustrates how \mathcal{R}_{MI} evolves from $\mathcal{R}_{MI} = \mathcal{R}_1$ to \mathcal{R}_2 and from \mathcal{R}_2 to $\mathcal{R}_3 = \mathcal{PU}(\mathcal{R}_{MI})$.

5.2. Maximum Entropy Distributions for Grounding and Aggregation Semantics

Using KREATOR we computed the ME distributions for the three knowledge bases $\mathcal{R}_1, \mathcal{R}_2$, and \mathcal{R}_3 involved in the \mathcal{PU} transformation of \mathcal{R}_{MI} for both PCI-grounding and PCI-aggregation semantics. For all admissible ground instances of the conditionals occurring in $\mathcal{R}_1, \mathcal{R}_2$ and \mathcal{R}_3 , we computed their probability under the ME distributions for PCI-grounding and PCI-aggregation semantics. The results are shown in Table 2, using the abbreviation $l(x, y)$ for *likes*(x, y).

There are three pairwise different ME distributions (*i.e.*, $ME_{\otimes}(\mathcal{R}_1), ME_{\otimes}(\mathcal{R}_2), ME_{\otimes}(\mathcal{R}_3)$) under PCI-aggregation semantics for the three pairwise different knowledge bases $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3$. On the other hand $ME_{\Delta}(\mathcal{R}_1) = ME_{\Delta}(\mathcal{R}_2) = ME_{\Delta}(\mathcal{R}_3) = ME_{\otimes}(\mathcal{R}_3)$ holds since the \mathcal{PU} transformation process does not change the maximum entropy model under PCI-grounding semantics and because \mathcal{R}_3 is parametrically uniform.

Transformation Steps to $PU(\mathcal{R})$

1. The conditional

$\langle (\text{likes}(U, V) \mid \text{likes}(V, U)) [0.9], U \neq V \rangle$

is substituted by

$\langle (\text{likes}(U, a) \mid \text{likes}(a, U)) [0.9], U \neq a \rangle$

$\langle (\text{likes}(U, V) \mid \text{likes}(V, U)) [0.9], U \neq V, V \neq a \rangle$

Transformation of (\mathcal{R}_1)

$\langle (\text{likes}(U, V) \mid \text{likes}(V, U)) [0.9], U \neq V \rangle$

$\langle (\text{likes}(a, V)) [0.05], V \neq a \rangle$

to (\mathcal{R}_2)

$\langle (\text{likes}(U, a) \mid \text{likes}(a, U)) [0.9], U \neq a \rangle$

$\langle (\text{likes}(U, V) \mid \text{likes}(V, U)) [0.9], U \neq V, V \neq a \rangle$

$\langle (\text{likes}(a, V)) [0.05], V \neq a \rangle$

Explanation

Interaction Typ INTER1

• $R_l = \langle (\dots \text{likes}(a, V) \dots) [\dots], C_{R_l} \rangle$ and

$R_r = \langle (\dots \text{likes}(V, U) \dots) [\dots], C_{R_r} \rangle$ and $C_{R_r} \neq V \neq a$

• $R_l = \langle (\dots \text{likes}(a, V) \dots) [\dots], C_{R_l} \rangle$ and $C_{R_l} \neq V \neq a$

$R_r = \langle (\dots \text{likes}(U, V) \dots) [\dots], C_{R_r} \rangle$ and $C_{R_r} \neq V \neq a$

Substitution: $V \rightarrow a$

2. The conditional

$\langle (\text{likes}(U, V) \mid \text{likes}(V, U)) [0.9], U \neq V, V \neq a \rangle$

is substituted by

$\langle (\text{likes}(a, V) \mid \text{likes}(V, a)) [0.9], a \neq V \rangle$

$\langle (\text{likes}(U, V) \mid \text{likes}(V, U)) [0.9], U \neq V, V \neq a, U \neq a \rangle$

Transformation of (\mathcal{R}_2)

$\langle (\text{likes}(U, a) \mid \text{likes}(a, U)) [0.9], U \neq a \rangle$

$\langle (\text{likes}(U, V) \mid \text{likes}(V, U)) [0.9], U \neq V, V \neq a \rangle$

$\langle (\text{likes}(a, V)) [0.05], V \neq a \rangle$

to (\mathcal{R}_3)

$\langle (\text{likes}(U, a) \mid \text{likes}(a, U)) [0.9], U \neq a \rangle$

$\langle (\text{likes}(a, V) \mid \text{likes}(V, a)) [0.9], a \neq V \rangle$

$\langle (\text{likes}(U, V) \mid \text{likes}(V, U)) [0.9], U \neq V, V \neq a, U \neq a \rangle$

$\langle (\text{likes}(a, V)) [0.05], V \neq a \rangle$

Explanation

Interaction Typ INTER1

• $R_l = \langle (\dots \text{likes}(a, V) \dots) [\dots], C_{R_l} \rangle$ and

$R_r = \langle (\dots \text{likes}(U, V) \dots) [\dots], C_{R_r} \rangle$ and $C_{R_r} \neq U \neq a$

Substitution: $U \rightarrow a$

$PU(\mathcal{R}) = (\mathcal{R}_3) =$

$\langle (\text{likes}(a, V)) [0.05], V \neq a \rangle$

$\langle (\text{likes}(a, V) \mid \text{likes}(V, a)) [0.9], a \neq V \rangle$

$\langle (\text{likes}(U, a) \mid \text{likes}(a, U)) [0.9], U \neq a \rangle$

$\langle (\text{likes}(U, V) \mid \text{likes}(V, U)) [0.9], U \neq V, V \neq a, U \neq a \rangle$

Figure 1. The KREATOR protocol of the PU transformation steps from $\mathcal{R}_{MI} = \mathcal{R}_1$ to \mathcal{R}_2 and from \mathcal{R}_2 to $\mathcal{R}_3 = PU(\mathcal{R}_{MI})$ for \mathcal{R}_{MI} from Example 3.

Table 1. Conditionals occurring in \mathcal{R}_1 , \mathcal{R}_2 , and \mathcal{R}_3 given by the \mathcal{PU} transformation steps from $\mathcal{R}_{MI} = \mathcal{R}_1$ to \mathcal{R}_2 and from \mathcal{R}_2 to $\mathcal{R}_3 = \mathcal{PU}(\mathcal{R}_{MI})$ for \mathcal{R}_{MI} from Example 3 (cf. Figure 1) using the abbreviation $l(x, y)$ for $likes(x, y)$. Conditional R_1 in \mathcal{R}_1 is replaced by $R_{1.1}$ and $R_{1.2}$ in \mathcal{R}_2 , and conditional $R_{1.2}$ in \mathcal{R}_2 is replaced by $R_{1.2.1}$ and $R_{1.2.2}$ in \mathcal{R}_3 .

\mathcal{R}_1	\mathcal{R}_2	\mathcal{R}_3
	$R_{1.1}: \langle (l(U, a) l(a, U)) \rangle [0.9],$ $U \neq a$	
$R_1: \langle (l(U, V) l(V, U)) \rangle [0.9],$ $U \neq V$	$R_{1.2}: \langle (l(U, V) l(V, U)) \rangle [0.9],$ $U \neq V, V \neq a$	$R_{1.2.1}: \langle (l(a, V) l(V, a)) \rangle [0.9],$ $V \neq a$
		$R_{1.2.2}: \langle (l(U, V) l(V, U)) \rangle [0.9],$ $U \neq V, V \neq a, U \neq a$
$R_2: \langle (l(a, V) \top) \rangle [0.05],$ $V \neq a$		

Table 2. Maximum entropy probabilities of the ground instances of the conditionals in \mathcal{R}_1 , \mathcal{R}_2 , and \mathcal{R}_3 (cf. Table 1) under PCI-aggregation semantics; for PCI-grounding semantics, $ME_{\Delta}(\mathcal{R}_1)(g) = ME_{\Delta}(\mathcal{R}_2)(g) = ME_{\Delta}(\mathcal{R}_3)(g) = ME_{\otimes}(\mathcal{R}_3)(g)$ holds since the \mathcal{PU} transformation process does not change the maximum entropy model under grounding semantics and because \mathcal{R}_3 is parametrically uniform.

\mathcal{R}_1	\mathcal{R}_2	\mathcal{R}_3	ground instance g	$ME_{\otimes}(\mathcal{R}_1)(g)$	$ME_{\otimes}(\mathcal{R}_2)(g)$	$ME_{\otimes}(\mathcal{R}_3)(g)$
	$R_{1.1}$		$l(b, a) l(a, b)$	0.96674480	0.90000000	0.90000000
			$l(c, a) l(a, c)$	0.96674480	0.90000000	0.90000000
R_1		$R_{1.2.1}$	$l(a, b) l(b, a)$	0.46016768	0.45380549	0.89999999
	$R_{1.2}$		$l(a, c) l(c, a)$	0.46016768	0.45380549	0.89999999
		$R_{1.2.2}$	$l(b, c) l(c, b)$	0.96674480	0.96860780	0.90000000
			$l(c, b) l(b, c)$	0.96674480	0.96860780	0.90000000
R_2			$l(a, b)$	0.05000000	0.05000000	0.05000000
			$l(a, c)$	0.05000000	0.05000000	0.05000000

It is interesting to note that for the ground instances originating from R_1 there are two distinct probabilities under $ME_{\otimes}(\mathcal{R}_1)$, three probabilities under $ME_{\otimes}(\mathcal{R}_2)$, and as implied by Proposition 2 one probability under $ME_{\otimes}(\mathcal{R}_3)$. In all cases, PCI-aggregation semantics ensures that the distinct probabilities aggregate to the probability stated in the corresponding conditionals.

For the comparison of PCI-grounding and PCI-aggregation, it is also interesting to compare their ME behavior with respect to queries that are not instances of a conditional given in the knowledge base. For example, for $likes(b, c)$ we observe

$$ME_{\otimes}(\mathcal{R}_1)(likes(b, c)) = 0.64220609$$

$$ME_{\otimes}(\mathcal{R}_2)(likes(b, c)) = 0.64490162$$

$$ME_{\otimes}(\mathcal{R}_3)(likes(b, c)) = 0.58699481$$

and for $likes(b, a)$ we get

$$ME_{\otimes}(\mathcal{R}_1)(likes(b, a)) = 0.10504266$$

$$ME_{\otimes}(\mathcal{R}_2)(likes(b, a)) = 0.09916142$$

$$ME_{\otimes}(\mathcal{R}_3)(likes(b, a)) = 0.05000000$$

for PCI-aggregation semantics, while

$$ME_{\Delta}(\mathcal{R}_i)(likes(b, c)) = 0.58699481$$

$$ME_{\Delta}(\mathcal{R}_i)(likes(b, a)) = 0.05000000$$

holds for $i \in \{1, 2, 3\}$ under PCI-grounding semantics.

6. Conclusions and Further Work

In this paper, we considered maximum entropy based semantics for relational probabilistic conditionals. FO-PCL [16] employs a grounding semantics and uses instantiation restrictions for the free variables occurring in a conditional, requiring all admissible instances of a conditional to have the given probability. Aggregating semantics [15] defines probabilistic satisfaction by interpreting the intended probability of a conditional with free variables only as a guideline for the probabilities of its instances that aggregate to the conditional's given probability, while the actual probabilities for grounded instances may differ.

While the original definition of aggregation semantics [15] considered only conditionals without constraints representing instantiation restrictions, we developed the framework PCI extending aggregation semantics so that instantiation restrictions can also be taken into account, but without giving up the flexibility of aggregating over distinct probabilities. In comparison with [15], under PCI-aggregation semantics one can restrict the set of groundings of a conditional over which aggregating with respect to a conditional takes place by providing a corresponding constraint formula for the conditional. From a knowledge representation point of view, this can be useful in various situations, for instance when we talk about a particular relationship among individuals while already knowing that a specific individual like Clyde is an exception with respect to the given relationship.

Note that PCI captures both grounding semantics and aggregating semantics without instantiation restrictions as special cases. For the case that a knowledge base is parametrically uniform, PCI-grounding and PCI-aggregation semantics coincide when employing the maximum entropy principle, while for a knowledge base that is not parametrically uniform the two ME semantics induce different models in general. We illustrated the differences and common features of both semantics on a concrete knowledge base, using the KREATOR environment for computing the ME models and answering queries with respect to these distributions. We expect that observations of this kind will support the discussion of both formal and common sense properties of probabilistic first-order inference in general and inference according to the principle of maximum entropy in a first-order setting in particular.

Acknowledgments

We would like to thank the anonymous referees of this paper for their helpful comments and valuable suggestions.

Author Contributions

This paper is joint work by the authors.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Pearl, J. *Probabilistic Reasoning in Intelligent Systems*; Morgan Kaufmann: San Mateo, CA, USA, 1988.
2. Cowell, R.; Dawid, A.; Lauritzen, S.; Spiegelhalter, D. *Probabilistic Networks and Expert Systems*; Springer: New York, NY, USA; Berlin/Heidelberg, Germany, 1999.
3. Jaynes, E. *Papers on Probability, Statistics and Statistical Physics*; D. Reidel Publishing Company: Dordrecht, The Netherlands, 1983.
4. Shore, J.; Johnson, R. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theory* **1980**, *IT-26*, 26–37.
5. Paris, J. *The Uncertain Reasoner's Companion—A Mathematical Perspective*; Cambridge University Press: Cambridge, UK, 1994.
6. Paris, J.; Vencovska, A. In defence of the maximum entropy inference process. *Int. J. Approx. Reason.* **1997**, *17*, 77–103.
7. Paris, J. Common Sense and Maximum Entropy. *Synthese* **1999**, *117*, 75–93.
8. Kern-Isberner, G. *Conditionals in Nonmonotonic Reasoning and Belief Revision*; Lecture Notes in Artificial Intelligence LNAI 2087; Springer: Berlin/Heidelberg, Germany, 2001.
9. Paris, J.B. What You See Is What You Get. *Entropy* **2014**, *16*, 6186–6194.
10. Delgrande, J. On first-order conditional logics. *Artif. Intell.* **1998**, *105*, 105–137.

11. Fisseler, J. *Learning and Modeling with Probabilistic Conditional Logic; Dissertations in Artificial Intelligence*; IOS Press: Amsterdam, The Netherlands, 2010; Volume 328.
12. Halpern, J. *Reasoning about Uncertainty*; MIT Press: Cambridge, MA, USA, 2005.
13. Getoor, L.; Taskar, B. *Introduction to Statistical Relational Learning*; Getoor, L., Taskar, B., Eds.; MIT Press: Cambridge, MA, USA, 2007.
14. Kern-Isberner, G.; Beierle, C.; Finthammer, M.; Thimm, M. Comparing and Evaluating Approaches to Probabilistic Reasoning: Theory, Implementation, and Applications. In *Transactions on Large-Scale Data- and Knowledge-Centered Systems VI*; Hameurlain, A., Küng, J., Wagner, R., Liddle, S.W., Schewe, K.-D., Zhou, X., Eds.; Volume 7600, Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2012; pp. 31–75.
15. Kern-Isberner, G.; Thimm, M. Novel Semantical Approaches to Relational Probabilistic Conditionals. In *Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning, KR'2010, Toronto, ON, Canada, 9–13 May 2010*; Lin, F., Sattler, U., Truszczynski, M., Eds.; AAAI Press: Menlo Park, CA, USA, 2010; pp. 382–391.
16. Fisseler, J. First-order probabilistic conditional logic and maximum entropy. *Log. J. IGPL* **2012**, *20*, 796–830.
17. Finthammer, M.; Beierle, C. How to Exploit Parametric Uniformity for Maximum Entropy Reasoning in a Relational Probabilistic Logic. In *Logics in Artificial Intelligence, Proceedings of 13th European Conference, JELIA 2012, Toulouse, France, 26–28 September 2012*; Fariñas del Cerro, L.; Herzig, A.; Mengin, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7519, Lecture Notes in Computer Science; pp. 189–201.
18. Rödder, W.; Reucher, E.; Kulmann, F. Features of the Expert-System-Shell SPIRIT. *Log. J. IGPL* **2006**, *14*, 483–500.
19. Finthammer, M.; Beierle, C. Instantiation Restrictions for Relational Probabilistic Conditionals. In *Scalable Uncertainty Management, Proceedings the 6th International Conference on Scalable Uncertainty Management, Marburg, Germany, 17–19 September 2012*; Hüllermeier, E., Link, S., Fober, T., Seeger, B., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7520, Lecture Notes in Computer Science; pp. 598–605.
20. Beierle, C.; Krämer, A. Achieving Parametric Uniformity for Knowledge Bases in a Relational Probabilistic Conditional Logic with Maximum Entropy Semantics. *Ann. Math. Artif. Intell.* **2015**, *73*, 5–45.
21. Beierle, C.; Höhnerbach, M.; Marto, M. Implementation of a Transformation System for Relational Probabilistic Knowledge Bases Simplifying the Maximum Entropy Model Computation. In *Proceedings of the Twenty-Seventh International Florida Artificial Intelligence Research Society Conference (FLAIRS 2014), Pensacola Beach, FL, USA, 21–23 May 2014*; AAAI Press: Menlo Park, CA, USA, 2014; pp. 486–489.
22. Finthammer, M.; Thimm, M. An Integrated Development Environment for Probabilistic Relational Reasoning. *Log. J. IGPL* **2012**, *20*, 831–871.

23. Beierle, C.; Kuche, S.; Finthammer, M.; Kern-Isberner, G. A Software System for the Computation, Visualization, and Comparison of Conditional Structures for Relational Probabilistic Knowledge Bases. In Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2015), Hollywood, FL, USA, 18–20 May 2015; AAAI Press: Menlo Park, CA, USA, 2015; In press.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).