# Relational Probabilistic Conditional Reasoning at Maximum Entropy

Matthias Thimm[1], Gabriele Kern-Isberner[1], and Jens Fisseler[2]

[1] Department of Computer Science, Technische Universität Dortmund, Germany
[2] Faculty of Mathematics and Computer Science, FernUniversität in Hagen, Germany

**Abstract.** This paper presents and compares approaches for reasoning with relational probabilistic conditionals, i. e. probabilistic conditionals in a restricted first-order environment. It is well-known that conditionals play a crucial role for default reasoning, however, most formalisms are based on propositional conditionals, which restricts their expressivity. The formalisms discussed in this paper are relational extensions of a propositional conditional logic based on the principle of maximum entropy. We show how this powerful principle can be used in different ways to realize model-based inference relations for first-order probabilistic knowledge bases. We illustrate and compare the different approaches by applying them to several benchmark examples, and we evaluate each approach with respect to properties adopted from default reasoning. We also compare our approach to Bayesian logic programs (BLPs) from the field of statistical relational learning which focuses on the combination of probabilistic reasoning and relational knowledge representation as well.

## 1 Introduction

Conditional logic [9] is a popular choice for the representation of common sense knowledge and rules. A conditional $(B \mid A)$ expresses the relation "If $A$ then usually (mostly, likely, probably) $B$" between two formulas $A$ and $B$ of some underlying logic. In contrast to classical implication $A \Rightarrow B$ a conditional models *defeasible* belief and as such models of conditionals do not have to strictly obey these relations. Conditionals can be quantified yielding a probabilistic conditional logic. A quantified conditional $(B \mid A)[\alpha]$ can be interpreted as a constraint for probability distributions via $P \models (B \mid A)[\alpha]$ iff $P(B \mid A) = \alpha$. Usually, the underlying logic for representing $A$ and $B$ is propositional and as such the expressive power of probabilistic conditional logic is limited. In the past ten years the area of *statistical relational learning* (or *probabilistic inductive logic programming*) developed many approaches to extend traditional probabilistic models for reasoning like Bayes and Markov Networks [10] to relational (first-order) representations of knowledge. Among these are Bayesian Logic Programs [7], and Markov Logic Networks [11], to name only a few. Most of these approaches employ a grounding of relational probabilistic problems to propositional ones in order to benefit from reasoning techniques developed for propositional probabilistic reasoning. In this paper, we discuss relational extensions of probabilistic conditional logic.

*Example 1 (Common Cold).* Assume we want to model uncertain knowledge pertaining to the possible causes resp. the probability of catching a common cold. A simple

representation using uncertain *if-then*-rules can be given via

$$\begin{array}{lllll} R1: & \mathit{cold}(U) & & & [0.01] \\ R2: \textbf{if } \mathit{susceptible}(U) & & \textbf{then } \mathit{cold}(U) & [0.1] & \\ R3: \textbf{if } \mathit{contact}(U,V) & \textbf{and } \mathit{cold}(V) & \textbf{then } \mathit{cold}(U) & [0.6] & \end{array} \tag{1}$$

The uncertain rule R1 states that one normally does not have a common cold, i. e. only with a diminutive probability of $0.01$. Rule R2 denotes that a person catches a common cold with probability $0.1$ if this person is susceptible to it, and rule R3 represents the knowledge that person $U$, which is in contact with another person $V$ which has the common cold, also gets a common cold with probability $0.6$.

In contrast to the weights of formulas in Markov Logic Networks [11] the values of probabilistic conditionals have a probabilistic interpretation and as such exhibit a more intuitive way of representing uncertain knowledge. However, assigning probabilities to conditionals that contain variables may be ambiguous and their interpretation may be subjective or statistical [1]. This paper aims at investigating formal semantics and reasoning techniques for relational probabilistic logic. We built on approaches of previous works [6, 2] that rely on the principle of maximum entropy, a popular choice for model-based reasoning in propositional probabilistic conditional logic [3, 4]. By selecting the unique model of a set of probabilistic conditionals that maximizes entropy and as such represents the given knowledge in the most unbiased way, we obtain inference mechanisms that are optimal from an information-theoretical point of view, cf. [4]. In this paper, we investigate the performance of these approaches with respect to the system P properties for default reasoning [8]. Furthermore, we compare the behavior of our inference mechanims and illustrate that approaches for statistical relational learning are not apt for relational probabilistic default reasoning.

The rest of this paper is organized as follows. We continue by giving the syntax for our relational extension of probabilistic conditional logic and providing a brief introduction to Bayesian Logic Programs in Sec. 2. Afterwards in Sec. 3 we discuss common sense properties that should be fulfilled by reasonable inference relations. In Sec. 4 we propose and discuss three different approaches for defining semantics to relational probabilistic conditional logic and apply these for probabilistic reasoning in Sec. 5. Finally, we review related work in Sec. 6 and conclude in Sec. 7.

## 2 Relational Probabilistic Knowledge Representation

Let $\mathcal{L}$ be a propositional relational language, i. e. the fragment of a first-order language over a signature $\Sigma$ containing only predicates and constants. An *atom* is a predicate together with a list of terms, which may be constants or variables or a mixture of these. Formulas are built with atoms using the usual connectives disjunction, conjunction, and negation but without any quantifiers. If appropriate we abbreviate conjunctions $A \wedge B$ by $AB$. We denote variables with a beginning uppercase, constants with a beginning lowercase letter, and vectors of these with $\boldsymbol{X}$ resp. $\boldsymbol{a}$. A ground formula, i. e. a formula that does not contain any variables, is called a *sentence*. A possible world semantics is provided by Herbrand interpretations over the Herbrand universe $\mathcal{H}$ that contains all constants in $\Sigma$. Herbrand interpretations correspond to complete conjunctions of

ground literals from the Herbrand base. Let $\Omega$ be the set of all such possible worlds $\omega$. A possible world $\omega$ *satisfies* a ground atom $A$, denoted by $\omega \models A$, iff $A \in \omega$. Satisfaction of arbitrary sentences is defined in the usual way.

The conditional relational language $(\mathcal{L}|\mathcal{L})$ consists of all (qualitative) conditionals of the form $(B(\boldsymbol{c_B}, \boldsymbol{X})|A(\boldsymbol{c_A}, \boldsymbol{X}))$ with $A(\boldsymbol{c_A}, \boldsymbol{X}), B(\boldsymbol{c_B}, \boldsymbol{X})$ being formulas from $\mathcal{L}^{rel}$. In this a bit sloppy notation, the vectors $\boldsymbol{c_A}, \boldsymbol{c_B}$ contain all constants occurring in $A$ and $B$, and without loss of generality, we assume $\boldsymbol{X}$ to cover exactly all variables occurring both in $A$ and in $B$. For any $\phi = (B(\boldsymbol{c_B}, \boldsymbol{X})|A(\boldsymbol{c_A}, \boldsymbol{X})) \in (\mathcal{L}|\mathcal{L})$, let $\mathcal{H}^\phi$ be the set of all constant vectors $\boldsymbol{a}$ used for the proper groundings of the variables $\boldsymbol{X}$ occurring in $\phi$ from the Herbrand universe $\mathcal{H}$. The language $(\mathcal{L}|\mathcal{L})^p$ consists of all probabilistic conditionals of the form $\phi[\mu]$ with $\phi \in (\mathcal{L}|\mathcal{L})$ and $\mu \in [0, 1]$. Conditionals can not be nested, but $\mathcal{L}$ should be considered as a fragment of $(\mathcal{L}|\mathcal{L})$ by identifying relational propositional formulas $A(\boldsymbol{c_A}, \boldsymbol{X})$ with conditionals $(A(\boldsymbol{c_A}, \boldsymbol{X})|\top)$ with tautological antecedent. Conditionals that contain variables are called *open* conditionals while conditionals that contain no variables are called *ground* conditionals.

*Example 2.* We represent Ex. 1 using $(\mathcal{L}|\mathcal{L})^p$. Let $\mathcal{R}_{cold} = \{r_1, r_2, r_3, r_4, r_5\}$ be defined as

$r_1 = (cold(X))[0.01]$ $\quad$ $r_2 = (cold(X) \mid susc(X))[0.1]$
$r_3 = (cold(X) \mid contact(X, Y), cold(Y))[0.6]$ $\quad$ $r_4 = (contact(X, X))[0]$
$r_5 = (contact(X, Y) \mid contact(Y, X))[1]$

Conditionals $r_4$ resp. $r_5$ ensure that the relation induced by literals of the predicate $contact$ are irreflexive resp. symmetric.

Let $Int_{prob}(\Sigma)$ consist of all probability functions $P$ on $\Omega$, which assign to each possible world a (subjective) probability of it being the real world. Let $\models^\#$ be a semantic entailment relation between probability functions and probabilistic relational conditionals, specifying when $P \in Int_{prob}(\Sigma)$ is a *#-model* of $\phi[\mu] \in (\mathcal{L}|\mathcal{L})^p$: $P \in Mod^\#(\phi[\mu])$ iff $P \models^\# \phi[\mu]$. We will present several ways of instantiating the parametrical superscript $\#$ below. As usual, $\models^\#$ can be lifted to a classical (monotonic) entailment relation between formulas: $\mathcal{R} \models^\# \phi[\mu]$ iff $Mod^\#(\mathcal{R}) \subseteq Mod^\#(\phi[\mu])$ for $\mathcal{R} \subseteq (\mathcal{L}|\mathcal{L})^p$ and $\phi[\mu] \in (\mathcal{L}|\mathcal{L})^p$. If $\mathcal{S} \subseteq (\mathcal{L}|\mathcal{L})^p$ is another set of probabilistic relational conditionals, then $\mathcal{R} \models^\# \mathcal{S}$ iff $\mathcal{R} \models^\# \phi[\mu]$ for all $\phi[\mu] \in \mathcal{S}$, and $\mathcal{R} \equiv^\# \mathcal{S}$ iff $Mod^\#(\mathcal{R}) = Mod^\#(\mathcal{S})$.

We will compare our formalisms with a specific approach for statistical relational learning. Although these approaches were not developed for default reasoning current research on combining probability theory and relational knowledge representation focuses on this area. For example, Bayesian logic programming combines logic programming and Bayesian networks [7]. Due to space restriction and matters of presentation we only give a simplified definition for BLPs in the following. The basic structure for knowledge representation in Bayesian logic programs are *Bayesian clauses* which model probabilistic dependencies between atoms. Let $\mathbb{B}$ denote the set Boolean truth values $\mathbb{B} = \{\mathsf{true}, \mathsf{false}\}$. A *Bayesian clause* $c$ is an expression $(H \mid B_1, \ldots, B_n)$ with atoms $H, B_1, \ldots, B_n$. To each such clause, a *conditional probability distribution* $\mathsf{cpd}_c : \mathbb{B}^{n+1} \to [0, 1]$ is associated such that

$$\mathsf{cpd}_c(\mathsf{true}, x_1, \ldots, x_n) + \mathsf{cpd}_c(\mathsf{false}, x_1, \ldots, x_n) = 1 \quad \text{for all } (x_1, \ldots, x_n) \in \mathbb{B}^n \quad .$$

A function $\mathsf{cpd}_c$ for a Bayesian clause $c$ expresses the conditional probability distribution $P(\mathsf{head}(c) \mid \mathsf{body}(c))$ and thus partially describes an underlying probability distribution $P$.

In order to aggregate probabilities that arise from applications of different Bayesian clauses with the same head BLPs make use of *combining rules*. A combining rule $\mathsf{cr}_p$ for a predicate $p/n$ is a function $\mathsf{cr}_p$ that assigns to the conditional probability distributions of a set of Bayesian clauses a new conditional probability distribution that represents the *joint* probability distribution obtained from aggregating the given clauses. For example, given clauses $c_1 = (b(X) \mid a_1(X))$ and $c_2 = (b(X) \mid a_2(X))$ the result $f = \mathsf{cr}_b(\{\mathsf{cpd}_{c_1}, \mathsf{cpd}_{c_2}\})$ of the combining rule $\mathsf{cr}_b$ is a function $f : \mathbb{B}^3 \to [0,1]$ for the combined clause $(b(X) \mid a_1(X), a_2(X))$. Appropriate choices for such functions are *average* or *noisy-or*, cf. [7]. For example, noisy-or is defined as $no(p_1, p_2) = 1 - (1 - p_1)(1 - p_2)$.

A *Bayesian logic program* $B$ is a tuple $B = (C, D, R)$ with a (finite) set of Bayesian clauses $C = \{c_1, \ldots, c_n\}$, a set of conditional probability distributions (one for each clause in $C$) $D = \{\mathsf{cpd}_{c_1}, \ldots, \mathsf{cpd}_{c_n}\}$, and a set of combining functions (one for each Bayesian predicate appearing in $C$) $R = \{\mathsf{cr}_{p_1}, \ldots, \mathsf{cr}_{p_m}\}$. Semantics are given to Bayesian logic programs via transformation into propositional forms, i. e. into Bayesian networks [10]. Given a specific (finite) universe $U$ a Bayesian network $BN$ can be constructed by introducing a node for every ground Bayesian atom in $B$ and computing the corresponding (joint) conditional probability distributions. For a more detailed description of Bayesian Logic Programs we refer to [7].

## 3 Default Reasoning Properties – System $P^{prob}$

The classical probabilistic entailment relation $\models^{\#}$ specified in Sec. 2 will usually be quite weak, as is the case for propositional probabilistic logic. In this paper, we will focus on investigating non-monotonic inference relations $\mid\!\sim^{\#}$ by which relational probabilistic conditionals can be inferred plausibly from knowledge bases.

So, let $\mid\!\sim^{\#}$ describe a relation $\mathcal{R} \mid\!\sim^{\#} \phi[\mu]$ with $\mathcal{R} \subseteq (\mathcal{L}|\mathcal{L})^p$ and $\phi[\mu] \in (\mathcal{L}|\mathcal{L})^p$. We will present three different approaches for realizing $\mid\!\sim^{\#}$ in the following section, being based on different probabilistic entailment relations $\models^{\#}$ and the principle of maximum entropy, respectively. In order to be able to evaluate and compare these approaches, we will first set up a set of postulates applicable to such inference relations which are inspired by the system P properties from default reasoning ([8], see also [1]). Let $\mathcal{R}, \mathcal{R}_1, \mathcal{R}_2 \subseteq (\mathcal{L}|\mathcal{L})^p$ and $\phi[\mu], \psi[\nu] \in (\mathcal{L}|\mathcal{L})^p$.

**(Reflexivity)** For all $\phi[\mu] \in \mathcal{R}$, it holds that $\mathcal{R} \mid\!\sim^{\#} \phi[\mu]$.
**(Left Logical Equivalence)** If $\mathcal{R}_1 \equiv^{\#} \mathcal{R}_2$, then $\mathcal{R}_1 \mid\!\sim^{\#} \phi[\mu]$ iff $\mathcal{R}_2 \mid\!\sim^{\#} \phi[\mu]$.
**(Right Weakening)** If $\mathcal{R} \mid\!\sim^{\#} \phi[\mu]$ and $\phi[\mu] \models^{\#} \psi[\nu]$, then $\mathcal{R} \mid\!\sim^{\#} \psi[\nu]$.
**(Cumulativity)** If $\mathcal{R} \mid\!\sim^{\#} \phi[\mu]$, then $\mathcal{R} \mid\!\sim^{\#} \psi[\nu]$ iff $\mathcal{R} \cup \{\phi[\mu]\} \mid\!\sim^{\#} \psi[\nu]$.

Note that *cumulativity* subsumes both *cautious monotony* and *cut* [8].

Besides the common cold example (Ex. 2), we will illustrate the properties of our different semantical approaches using another benchmark example taken from [1].

*Example 3 (From [1]).* Consider the knowledge base $\mathcal{R}_{chirps} = \{r_1, r_2, r_3, r_4\}$ with

$r_1 = (chirps(X) \mid bird(X))[0.9] \quad r_2 = (chirps(X) \mid magpie(X), moody(X))[0.2]$
$r_3 = (bird(X) \mid magpie(X))[1] \quad r_4 = (magpie(tweety))[1]$

The knowledge represented in $\mathcal{R}_{chirps}$ concerns the default probabilities that a bird chirps ($r_1$) and that a moody magpie chirps ($r_2$). Knowing that every magpie is a bird ($r_3$) and given an actual magpie Tweety ($r_4$) the question at hand is to which probability Tweety chirps. As we have no knowledge whether Tweety is moody or not we cannot commit to any specific "reference class".

## 4 Semantics for Relational Probabilistic Conditional Logic

In this section we investigate different possibilities to define the semantic entailment relation $\models^{\#}$ and the nonmonotonic inference relation $\mathrel{|\!\sim}^{\#}$. In difference to the propositional case, assigning semantics to relational probabilistic conditionals is not straightforward. Nonetheless, we want to get some compatibility to the propositional case. Let $(B|A)[\mu]$ be a *ground* conditional, i. e. $(B|A)[\mu]$ contains no variables and therefore is of the form $(B(\boldsymbol{c_B}, \boldsymbol{X})|A(\boldsymbol{c_A}, \boldsymbol{X}))[\mu]$ with $\boldsymbol{X}$ being the empty vector. Then a probability distribution $P \in Int_{prob}(\Sigma)$ should be a $\#$-model of $(B|A)[\mu]$ iff it is a probabilistic model in the classical propositional sense: $P \in Mod^{\#}((B|A)[\mu])$ iff $P \models^{\#} (B|A)[\mu]$. As $P$ is defined on Herbrand interpretations the above is well-defined given that $P(A) = \sum_{\omega \in \Omega, \omega \models A} P(\omega)$ and $P(B|A) = P(AB)/P(A)$ for any sentences $A$ and $B$. But if a conditional $(B(\boldsymbol{c_B}, \boldsymbol{X})|A(\boldsymbol{c_A}, \boldsymbol{X}))[\mu]$ contains variables, the expression $\omega \models A(\boldsymbol{c_A}, \boldsymbol{X})$ for a possible world $\omega \in \Omega$ is not well-defined given our underlying Herbrand semantics and so is the relation $\models^{\#}$. In order to extend the semantical satisfaction of conditionals (see above) to conditionals that may contain variables, we investigate different strategies in the following subsections. Moreover, in order to obtain a non-monotonic inference relation $\mathrel{|\!\sim}^{\#}$ from $\models^{\#}$ we employ for each of the approaches the principle of maximum entropy [3, 4] to the corresponding set of $\#$-models. The entropy $H(P)$ of a probability distribution $P$ is defined as $H(P) = -\sum_{\omega \in \Omega} P(\omega) \log P(\omega)$. Given a set of conditionals $\mathcal{R}$ and a concrete semantical entailment relation $\models^{\#}$ we define the (usually, unique) probability distribution $ME^{\#}(\mathcal{R})$ with maximum entropy as follows

$$ME^{\#}(\mathcal{R}) = \operatorname*{argmax}_{P \models^{\#} \mathcal{R}} H(P). \tag{2}$$

Using $ME^{\#}(\mathcal{R})$ we define an inference relation $\mathrel{|\!\sim}^{\#}_{ME}$ as

$$\mathcal{R} \mathrel{|\!\sim}^{\#}_{ME} \phi[\mu] \qquad \text{iff} \qquad ME^{\#}(\mathcal{R}) \models^{\#} \phi[\mu], \tag{3}$$

for any conditional $\phi[\mu] \in (\mathcal{L}|\mathcal{L})^p$.

### 4.1 Grounding Semantics with Constraints

The first formalism uses a *grounding semantics* for relational probabilistic conditionals [2], similar to formalisms for statistical relational learning, see e. g. Markov Logic

Networks [11]. Within this formalism, any relational probabilistic conditional $\phi[\mu] = (B(\boldsymbol{c_B}, \boldsymbol{X})|A(\boldsymbol{c_A}, \boldsymbol{X}))[\mu] \in (\mathcal{L}|\mathcal{L})^p$ induces a set *gnd*$(\phi[\mu])$ of *ground instances*, which are obtained by substituting the free variables $\boldsymbol{X}$ by all combinations of constants in $\Sigma$. However, straightforward substitution easily yields inconsistent ground conditionals. Assume we are given the probabilistic relational conditional $(p(U, V)|p(V, U))[\mu]$. If both variables are substituted with the same constant, e. g. $c$, there exists no probability distribution $P$ satisfying $P(p(c, c)|p(c, c)) = \mu$, except if $\mu = 1.0$. To avoid such inconsistencies, the grounding semantics approach supplements conditionals with *constraint formulas*, which restrict the set of admissible combinations of constants when grounding. An *atomic constraint formula* is a *term equation* $t_1 = t_2$, with terms $t_1, t_2$, and predicate symbol $=$, denoting equality. A negated term equation is called a *term disequation* and is written as $t_1 \neq t_2$. Complex equational formulas are built using the usual logical connectives conjunction, disjunction and negation, but without quantifiers. Constraint formulas are interpreted by ground substitutions. A ground substitution $\sigma$ satisfies an equality constraint $t_1 = t_2$ iff $\sigma(t_1)$ and $\sigma(t_2)$ evaluate to the same constant. This satisfaction relation is canonically extended to complex constraint formulas.

For any $\langle\phi[\mu], C\rangle = \langle(B(\boldsymbol{c_B}, \boldsymbol{X})|A(\boldsymbol{c_A}, \boldsymbol{X}))[\mu], C\rangle$, i. e. a probabilistic relational conditional $\langle\phi[\mu], C\rangle$ with an associated constraint formula $C$, we assume that the variables occurring in $C$ are a subset of $\boldsymbol{X}$. Interpreting the (possible) elements of $\mathcal{H}^\phi$ as ground substitutions, $C$ restricts the set of ground instances of $\phi[\mu]$ by requiring that $\boldsymbol{a} \in \mathcal{H}^\phi$ satisfies $C$:

$$\begin{aligned} &\mathit{gnd}(\langle(B(\boldsymbol{c_B}, \boldsymbol{X})|A(\boldsymbol{c_A}, \boldsymbol{X}))[\mu], C\rangle) \\ &:= \left\{ (B(\boldsymbol{c_B}, \boldsymbol{a})|A(\boldsymbol{c_A}, \boldsymbol{a}))[\mu] \;\middle|\; \begin{array}{c} \boldsymbol{a} \in \mathcal{H}^{\langle(B(\boldsymbol{c_B}, \boldsymbol{X})|A(\boldsymbol{c_A}, \boldsymbol{X})), C\rangle}, \\ \boldsymbol{a} \text{ satisfies } C \end{array} \right\}. \end{aligned}$$

We can now define the semantic entailment relation $\models^{\mathit{gnd}}$ between a probability distribution $P \in \mathit{Int}_{\mathit{prob}}(\Sigma)$ and a probabilistic relational conditional $\langle\phi[\mu], C\rangle$:

$$P \models^{\mathit{gnd}} \langle\phi[\mu], C\rangle \qquad \text{iff} \qquad \forall \phi_{\mathit{gnd}}[\mu] \in \mathit{gnd}(\langle\phi[\mu], C\rangle) : P(\phi_{\mathit{gnd}}) = \mu.$$

That is, $P$ is a model of the probabilistic relational conditional $\langle\phi[\mu], C\rangle$ iff it is a model of all admissible ground instances of $\langle\phi[\mu], C\rangle$. As an additional condition, we require that all Herbrand interpretations which contain a ground atom that is not part of any ground instance of $\langle\phi[\mu], C\rangle$ have probability $0.0$. This is because the grounding semantics actually restricts the Herbrand universe to only contain ground atoms which are part of at least one ground instance of $\langle\phi[\mu], C\rangle$. Hence possible worlds containing ground atoms which are not part of any ground instance are considered impossible.

### 4.2 Averaging Semantics

While the previous approach relies on expressing relational conditionals in propositional terms and thus interprets open relational conditionals in a classical sense, in this and the next subsection we develop semantics using a non-classical interpretation. Both semantics have been previously introduced in [6]. Our first approach gives semantics to probabilistic conditionals by averaging conditional probabilities. The motivation for this semantics stems from the intuition that probabilistic rules such as

$(B(\boldsymbol{c_B}, \boldsymbol{X})|A(\boldsymbol{c_A}, \boldsymbol{X}))[\alpha]$, given an adequately large universe, should describe an expected value on the probability of $(B(\boldsymbol{c_B}, \boldsymbol{d_B})|A(\boldsymbol{c_A}, \boldsymbol{d_A}))[\alpha]$ for some randomly chosen $\boldsymbol{d_B}, \boldsymbol{d_A}$. Thus, given the actual probabilities of $(B(\boldsymbol{c_B}, \boldsymbol{d_B})|A(\boldsymbol{c_A}, \boldsymbol{d_A}))$ for each possible instantiation we expect the *average* of these probabilities should match $\alpha$. Hence, let $\models^{\varnothing}$ be the semantic entailment relation on probabilistic conditionals defined as $P \models^{\varnothing} (B(\boldsymbol{c_B}, \boldsymbol{X})|A(\boldsymbol{c_A}, \boldsymbol{X}))[\alpha]$ iff

$$\frac{\sum_{\boldsymbol{a} \in \mathcal{H}^{(B(\boldsymbol{c_B}, \boldsymbol{X})|A(\boldsymbol{c_A}, \boldsymbol{X}))}} P((B(\boldsymbol{c_B}, \boldsymbol{a})|A(\boldsymbol{c_A}, \boldsymbol{a}))[\alpha])}{|\mathcal{H}^{(B(\boldsymbol{c_B}, \boldsymbol{X})|A(\boldsymbol{c_A}, \boldsymbol{X}))}|} = \alpha \tag{4}$$

Intuitively spoken, a probability distribution $P$ $\varnothing$-satisfies a conditional $\phi[\mu]$ if the average of the individual instantiations of $\phi[\mu]$ is $\alpha$. As one can see, for a ground conditional $(B|A)[\alpha]$ the relation $\models^{\varnothing}$ coincides with the propositional case.

### 4.3 Aggregating Semantics

Our third semantical approach is inspired by statistical approaches. However, instead of counting objects, or tuples of objects, respectively, that make a formula true, we sum up the probabilities of the correspondingly instantiated formulas. In this way, both population-based and subjective belief aspects of probabilities can be combined. More precisely, we propose a mean value of subjective probabilities to interpret probabilistic rules.

To make the key idea of the approach clear, consider the relational probabilistic conditional $(B(\boldsymbol{c_B}, \boldsymbol{X})|A(\boldsymbol{c_A}, \boldsymbol{X}))[\alpha]$. If some first-order interpretation $\omega$ with a fixed domain is given, its statistical interpretation is provided by the relative frequency

$$\frac{|\{\boldsymbol{a} \mid \omega \models A(\boldsymbol{c_A}, \boldsymbol{a})B(\boldsymbol{c_B}, \boldsymbol{a})\}|}{|\{\boldsymbol{a} \mid \omega \models A(\boldsymbol{c_A}, \boldsymbol{a})\}|} = \alpha,$$

i. e. the number of tuples of individuals $\boldsymbol{a}$ is counted that satisfy the premise and the antecedent, in relation to the number of tuples that satisfy only the premise. Aggregating the information coming from all models of $A(\boldsymbol{c_A}, \boldsymbol{a})B(\boldsymbol{c_B}, \boldsymbol{a})$, resp. $A(\boldsymbol{c_A}, \boldsymbol{a})$, for each $\boldsymbol{a}$, gives rise to a subjective, population-based probability:

$$\frac{\sum_{\boldsymbol{a}} P(A(\boldsymbol{c_A}, \boldsymbol{a})B(\boldsymbol{c_B}, \boldsymbol{a}))}{\sum_{\boldsymbol{a}} P(A(\boldsymbol{c_A}, \boldsymbol{a}))} = \alpha,$$

If we allow $P$ to represent (subjective) beliefs, then the above equation expresses the average subjective belief that in any situation in which we observe individuals $\boldsymbol{a}$ satisfying $A(\boldsymbol{c_A}, \boldsymbol{a})$, we expect them to satisfy $B(\boldsymbol{c_B}, \boldsymbol{a})$ as well with probability $\alpha$. This switches the view from a frequentistic perspective to a possible worlds semantics.

So, the entailment relation $\models^{\odot}$ between functions from $Int_{prob}(\Sigma)$ and relational probabilistic conditionals is defined by $P \models^{\odot} (B(\boldsymbol{c_B}, \boldsymbol{X}) \mid A(\boldsymbol{c_A}, \boldsymbol{X}))[\alpha]$ iff

$$\frac{\sum\limits_{\boldsymbol{a} \in \mathcal{H}^{(B(\boldsymbol{c_B}, \boldsymbol{X})|A(\boldsymbol{c_A}, \boldsymbol{X}))}} P(A(\boldsymbol{c_A}, \boldsymbol{a})B(\boldsymbol{c_B}, \boldsymbol{a}))}{\sum\limits_{\boldsymbol{a} \in \mathcal{H}^{(B(\boldsymbol{c_B}, \boldsymbol{X})|A(\boldsymbol{c_A}, \boldsymbol{X}))}} P(A(\boldsymbol{c_A}, \boldsymbol{a}))} = \alpha\ . \tag{5}$$

As for $\models^{\varnothing}$, for a ground conditional, the operator $\models^{\odot}$ coincides with the usual propositional interpretation using conditional probabilities.

## 5 Relational Probabilistic Conditional Reasoning

In the following, we discuss the inference operators $\mathrel{|\!\sim}^{gnd}_{ME}$, $\mathrel{|\!\sim}^{\varnothing}_{ME}$, and $\mathrel{|\!\sim}^{\odot}_{ME}$ that derive from the application of the different semantics $\models^{gnd}$, $\models^{\varnothing}$, and $\models^{\odot}$, respectively. All three formalisms implement a model-based probabilistic inference, using the maximum entropy model of a knowledge base as its most appropriate model. This ensures that all inference relations defined in the previous sections comply with all basic demands for relational probabilistic reasoning, as the next proposition shows.

**Proposition 1.** *Let $\models^{\#}$ be any of the semantical entailment relations defined above. Then the inference relation $\mathrel{|\!\sim}^{\#}_{ME}$ satisfies (Reflexivity), (Left Logical Equivalence), (Right Weakening), and (Cumulativity).*

*Proof. We only show satisfaction of (Cumulativity). The proofs of the other properties are similar.*

**(Cumulativity)** *It holds $Mod^{\#}(\mathcal{R} \cup \{\phi[\mu]\}) \subseteq Mod^{\#}(\mathcal{R})$ and $ME(\mathcal{R}) \in Mod^{\#}(\mathcal{R} \cup \{\phi[\mu]\})$ as $\mathcal{R} \mathrel{|\!\sim}^{\#}_{ME} \phi[\mu]$. Suppose $ME(\mathcal{R} \cup \{\phi[\mu]\}) \neq ME(\mathcal{R})$, then $H(ME(\mathcal{R} \cup \{\phi[\mu]\})) > H(ME(\mathcal{R}))$ and $ME(\mathcal{R} \cup \{\phi[\mu]\})$ should be the ME-model of $\mathcal{R}$ as well because $ME(\mathcal{R} \cup \{\phi[\mu]\}) \in Mod^{\#}(\mathcal{R})$. Hence, $ME(\mathcal{R} \cup \{\phi[\mu]\}) = ME(\mathcal{R})$ and therefore $\mathcal{R} \mathrel{|\!\sim}^{\#}_{ME} \psi[\nu]$ iff $\mathcal{R} \cup \{\phi[\mu]\} \mathrel{|\!\sim}^{\#}_{ME} \psi[\nu]$ for any $\psi[\nu]$.* □

It is obvious that the satisfaction of the common sense properties discussed in Sec. 3 is mainly due to the principle of maximum entropy and independent of the actual used semantical entailment relation. This is not surprising as *ME*-inference is an optimal, model-based inference operation, and the semantical entailment relation is used in defining the properties themselves.

As for BLPs (dis-)satisfaction of these default reasoning properties is not so easy to see as the formalism of BLPs is much less based on classical logic. Consider again the postulate

**(Cumulativity)** If $\mathcal{R} \mathrel{|\!\sim}^{\#} \phi[\mu]$, then $\mathcal{R} \mathrel{|\!\sim}^{\#} \psi[\nu]$ iff $\mathcal{R} \cup \{\phi[\mu]\} \mathrel{|\!\sim}^{\#} \psi[\nu]$.

Let $B$ be a Bayesian Logic Program, $c$ a Bayesian clause, and $\mathsf{cpd}_c$ a conditional probability distribution for $c$. BLPs allow only to determine probabilities for some ground atom given some set of ground atoms as evidence. So the expression $B \mathrel{|\!\sim} (c, \mathsf{cpd}_c)$ is not well-defined for a clause $c$ that contains variables, cf. [13]. However, if $c$ contains no variables the probability of the head of $c$ for every truth assignment of the body atoms can be computed. But even for ground clauses (Cumulativity) is not satisfied for BLPs. Consider the following example.

*Example 4.* Let $B$ be a BLP consisting of the single clause $c = (B(X) \mid A(X))$ with $\mathsf{cpd}_c(\mathsf{true}, \mathsf{true}) = \mathsf{cpd}_c(\mathsf{true}, \mathsf{false}) = \mathsf{cpd}_c(\mathsf{false}, \mathsf{true}) = \mathsf{cpd}_c(\mathsf{false}, \mathsf{false}) = 0.5$. Let *noisy-or* be the combining rule for $B$. For some constant $a$ it follows clearly that for $c' = (B(a) \mid A(a))$ with $\mathsf{cpd}_{c'} = \mathsf{cpd}_c$ is holds that $B \mathrel{|\!\sim} (c', \mathsf{cpd}_{c'})$. However,

when determining the probability of $c'$ in $B \cup \{(c', \mathsf{cpd}_{c'})\}$ different results arise due to aggregating via noisy-or, e. g. the probability of $B(a)$ given that $A(a)$ holds computes to $1 - (1 - 0.5)(1 - 0.5) = 0.75$.

Similar problems arise when translating the other default reasoning postulates to the BLP framework. For example, (Reflexivity) is trivially dissatisfied if a BLP contains at least one clause with variables. These categorical problems are not surprising as BLPs were not developed for default reasoning per se. Further discussions on this topic can be found in [13].

In order to comprehend the differences between the individual approaches we go on by investigating their behavior in the benchmark examples introduced above.

*Example 5.* We investigate Ex. 2 that has also been discussed in the introduction. In order to investigate this example for the grounding semantics, we have to modify the rules slightly. This is because rule $r_3 : (cold(X) \mid contact(X, Y), cold(Y))[0.6]$ yields inconsistencies if the variables $X$ and $Y$ are substituted with the same constant, even if rule $r_4$ is part of the knowledge base. Hence we must complement this conditional with the constraint formula $X \neq Y$, thereby forbidding these substitutions:

$$r_3' : \langle (cold(X) \mid contact(X, Y), cold(Y))[0.6], X \neq Y \rangle.$$

One thing to notice in the formalization of the knowledge base in Ex. 2 is that we have not represented any specific knowledge on particular individuals. Due to this representation the inferences drawn from $\mathcal{R}_{cold}$ using any of the proposed semantics $\mid\!\sim^{gnd}_{ME}$, $\mid\!\sim^{\varnothing}_{ME}$, and $\mid\!\sim^{\odot}_{ME}$ are identical and as follows (assuming that the given signature contains three constants $\{a, b, c\}$):

$$\begin{aligned}
&\mathcal{R}_{cold} \mid\!\sim^{\#} (cold(a))[0.01] \\
&\mathcal{R}_{cold} \mid\!\sim^{\#} (cold(b) \mid susc(b))[0.1] \\
&\mathcal{R}_{cold} \mid\!\sim^{\#} (cold(c) \mid contact(c, a) \wedge cold(a))[0.6] \\
&\mathcal{R}_{cold} \mid\!\sim^{\#} (cold(c) \mid contact(c, a) \wedge cold(a) \wedge cold(b))[0.9]
\end{aligned}$$

where $\mid\!\sim^{\#}$ is one of $\{ \mid\!\sim^{gnd}_{ME}, \mid\!\sim^{\varnothing}_{ME}, \mid\!\sim^{\odot}_{ME} \}$. In order to understand why the inferences are identical consider the conditional $(cold(X) \mid susc(X))[0.1]$. For grounding semantics this conditional yields the ground instance $(cold(a) \mid susc(a))[0.1]$ (among others). Now consider the averaging semantics which basically demands that the average probability of $(cold(F) \mid susc(F))$ for $F \in \{a, b, c\}$ is $0.1$. As $\mathcal{R}_{cold}$ does not say anything about different conditions for $\{a, b, c\}$ the most rational thing to do is to treat all instantiations equally. This is also pursued by the maximum entropy inference procedure because any deviation from this uniform assignment would yield a higher entropy. Hence, in order to have an average probability of $0.1$ for all three instances the inference procedure exactly assigns a probability of $0.1$ to all three instances. A similar explanation applies to aggregating semantics.

If we add probabilistic facts like $(contact(a, b))[1]$ or $(cold(c))[1]$ to $\mathcal{R}_{cold}$ the situation changes and now different inferences can be drawn from the different semantics.

The scenario above cannot easily be modeled with BLPs. As BLPs rely on Bayesian networks one important requirement is *acyclicity* of the represented knowledge. In the above example, the probability of some $a$ catching a cold may depend on the probability

of $b$ catching a cold which itself may depend again on the probability of $a$ catching a cold. In order to represent the example properly with a BLP these cycles have to be broken, e. g. by assuming some order on the individuals and by inhibiting $contact$ to be symmetric. However, these changes would alter the modeled knowledge drastically. Note that our approach does not forbid cyclic dependencies as the status of conditionals is validated in a global way, taking every dependency into account.

*Example 6.* We now come to Ex. 3 and assume that our given signature contains three constants $\{tweety, huey, dewey\}$ which are also all assumed to be birds. We obtain the following inferences[3]:

$$\mathcal{R}_{chirps} \mathrel{|\!\sim}^{gnd}_{ME} (chirps(tweety))[0.90] \qquad \mathcal{R}_{chirps} \mathrel{|\!\sim}^{gnd}_{ME} (chirps(huey))[0.90]$$
$$\mathcal{R}_{chirps} \mathrel{|\!\sim}^{\varnothing}_{ME} (chirps(tweety))[0.86] \qquad \mathcal{R}_{chirps} \mathrel{|\!\sim}^{\varnothing}_{ME} (chirps(huey))[0.92]$$
$$\mathcal{R}_{chirps} \mathrel{|\!\sim}^{\odot}_{ME} (chirps(tweety))[0.86] \qquad \mathcal{R}_{chirps} \mathrel{|\!\sim}^{\odot}_{ME} (chirps(huey))[0.92]$$

Here, the grounding semantics yields the same probabilities for *Tweety* and *Huey* regarding $chirps$. As there is no knowledge on whether *Tweety* is moody conditional $r_1$ is responsible for yielding a probability of $0.9$ for $chirps(tweety)$. As for both the averaging and aggregating semantics we obtain identical results in this example. For both averaging and aggregating semantics *Tweety* is assumed to chirp, with a slightly lower probability than *Huey*. This complies with our intuition, as *Tweety* is known to be a magpie, and in case it is moody, its probability to chirp would decrease considerably. For *Huey*'s probability of chirping, we observe some compensating effect caused by the situation of *Tweety* being moody which is rarely the case ($0.12$ for both averaging and aggregating semantics).

Example 3 can be represented as a BLP $B$ as it contains no cyclic dependencies. For example, conditional $r_1$ can be represented as a Bayesian clause $c_1 = (chirps(X) \mid bird(X))$ with $\mathsf{cpd}_c(\mathsf{true}, \mathsf{true}) = 0.9$, $\mathsf{cpd}_c(\mathsf{false}, \mathsf{true}) = 0.9$, $\mathsf{cpd}_c(\mathsf{true}, \mathsf{false}) = 0.5$, and $\mathsf{cpd}_c(\mathsf{false}, \mathsf{false}) = 0.5$, the latter two probabilities being some default assumptions for the case when $X$ is no bird. However, inference in $B$ depends crucially on the combining rule chosen for $chirps$. If Tweety is moody both clauses deriving from the conditionals $r_1$ and $r_2$ are applicable and the resulting probabilities $0.9$ and $0.2$ have to be combined. In this scenario, the combining rule *noisy-and* defined via $na(p_1, p_2) = p_1 p_2$ would be an appropriate choice yielding a combined probability of $0.18$. If noisy-or would be chosen this results in a probability of $0.92$ which shows that combining rule have to be chosen very carefully. Another problem with the BLP representation is that $B$ is not able to compute any probability for $chirps(huey)$ as there is no evidence on whether $huey$ is a magpie or even a bird.

## 6 Related Work

There are several other approaches to defining a probabilistic semantics for a fragment of first-order logic, some of which also make use of the principle of maximum entropy.

[3] All probabilities are rounded off to two decimal places.

The research presented in [1, 3] aims at combining subjective and statistical probabilistic knowledge by deriving subjective probabilistic beliefs about a specific individual from statistical knowledge about sets of individuals, considering approximative probabilities and limits. Although this approach gives the same results as the principle of maximum entropy, the authors argue that the principle of maximum entropy cannot be applied on knowledge bases containing $n$-ary predicates, $n > 1$. Approaches allowing the representation of statistical probabilities suffer from these problems arising from the fact that the size of the universe constraints the representable probabilities. This is not the case for the semantics presented in this work, as they have no underlying frequentistic interpretation. Moreover, the application of the principle of maximum entropy to knowledge bases with arbitrary predicates seems to be unproblematic, but this has to be investigated more thoroughly in further work.

The grounding semantics in particular is similar to the probabilistic logic program semantics with entailment under maximum-entropy and the closed-world assumption as introduced in [5]. However, the maximum entropy inference relation $\mid\!\sim_{ME}^{gnd}$, which is defined for $\models^{gnd}$ via Equations (2) and (3), is independent of the query. That is, the maximum entropy model defined for a given set $\mathcal{R}$ via the grounding semantics is independent of the query, whereas the maximum entropy model defined via entailment under maximum-entropy and the closed-world assumption depends on the given query.

## 7 Conclusions and Discussion

We have introduced and evaluated three different semantics for relational probabilistic conditionals that differ with respect to their approaches to dealing with the conflicts and inconsistencies arising from the quantification of conditionals with precise probabilities. The aggregating as well as the averaging semantics try to deal with these conflicts by allowing some "exceptional" individuals to deviate from the overall behavior of a given population, while the grounding semantics utilizes constraints to restrict the set of individuals which may be used for generating the ground instances of the probabilistic conditionals.

We have shown that all semantics satisfy common sense properties inspired by similar properties for default reasoning and we have compared them on several example knowledge bases. We have also shown that approaches to statistical relational learning are inadequate for probabilistic default reasoning in relational settings. It turned out that all proposed semantics coincide on knowledge bases that do not model knowledge on exceptional individuals. In the presence of specific knowledge on individuals, however, the inferences drawn from the different approaches may vary significantly. While the grounding semantics seems to yield the most robust and predictable inferences it suffers from the additional demand to specify constraint formulas to inhibit an inconsistent grounding of the knowledge base. Nonetheless, inference based on grounding semantics benefits from research on propositional inference using maximum entropy and thus can be solved quite efficiently [12]. Both the averaging and aggregating semantics do not need constraint formulas but require a universe of sufficient size in order to compensate for exceptions explicitly represented. Furthermore, they allow for a smoother interpretation of conditionals and consider the interactions between the rep-

resented knowledge more deeply. While the averaging and aggregating semantics may differ only slightly, from a computational point of view, inference based on aggregating semantics is easier as Equation (5) describes a linear constraint whereas Equation (4) describes a non-linear constraint. From the point of view of modeling, the grounding semantics is most adequate for a population with well-defined homogeneous subpopulations whereas the average semantics provides probabilities that are means of the corresponding subjective probabilities, expressing that on the average, e.g., individuals show a certain behavior with the respective probability. Hence, they compute a statistics of subjective (conditional) probabilities. Finally, the aggregating semantics mimics the form of statistical probabilities but replaces frequencies by subjective estimations and allows for even more compensation effects. In particular, by assigning low probabilities to formulas involving abnormal individuals, the influence of such individuals on probabilities of general statements can be weakened.

Further work will comprise a more thorough evaluation of the formalisms presented here, as well as the development of appropriate (with respect to the underlying semantics) methods for learning probabilistic relational conditionals from data and efficient methods for inference.

**Acknowledgements.** The research reported here was partially supported by the Deutsche Forschungsgemeinschaft (grants KE 1413/2-1 and BE 1700/7-1).

## References

1. Bacchus, F., Grove, A.J., Halpern, J.Y., Koller, D.: From Statistical Knowledge Bases to Degrees of Belief. Artificial Intelligence 87(1-2), 75–143 (1996)
2. Fisseler, J.: Learning and Modeling with Probabilistic Conditional Logic, Dissertations in Artificial Intelligence, vol. 328. IOS Press (2010)
3. Grove, A.J., Halpern, J.Y., Koller, D.: Random Worlds and Maximum Entropy. Journal of Artificial Intelligence Research (JAIR) 2, 33–88 (1994)
4. Kern-Isberner, G.: Conditionals in Nonmonotonic Reasoning and Belief Revision. No. 2087 in LNCS, Springer (2001)
5. Kern-Isberner, G., Lukasiewicz, T.: Combining probabilistic logic programming with the power of maximum entropy. Artificial Intelligence 157, 139–202 (2004)
6. Kern-Isberner, G., Thimm, M.: Novel semantical approaches to relational probabilistic conditionals. In: Proceedings of the 12th Int. Conf. on Knowledge Representation (KR) (2010)
7. Kersting, K., De Raedt, L.: Bayesian logic programming: Theory and tool. In: Getoor, L., Taskar, B. (eds.) An Introduction to Statistical Relational Learning. MIT Press (2005)
8. Makinson, D.: General theory of cumulative inference. In: Non-monotonic Reasoning, pp. 1–18. Springer Lecture Notes on Artificial Intelligence 346, Berlin (1989)
9. Nute, D., Cross, C.: Conditional logic. In: Handbook of Philosophical Logic, vol. 4, pp. 1–98. Kluwer (2002)
10. Pearl, J.: Probabilistic Reasoning in intelligent Systems: Networks of plausible inference. Morgan Kaufmann (1998)
11. Richardson, M., Domingos, P.: Markov logic networks. Machine Learning 62(1-2), 107–136 (2006)
12. Rödder, W., Meyer, C.H.: Coherent Knowledge Processing at Maximum Entropy by SPIRIT. In: Proc. of the Twelfth Conf. on Uncertainty in Artificial Intelligence. pp. 470–476 (1996)
13. Thimm, M., Finthammer, M., Kern-Isberner, G., Beierle, C.: Comparing Approaches to Relational Probabilistic Reasoning: Theory and Implementation (submitted)