Noname manuscript No.
(will be inserted by the editor)

Imprecise Probabilistic Query Answering using Measures of Ignorance and Degree of Satisfaction

Anbu Yue[1] · **Weiru Liu**[1] · **Anthony Hunter**[2]

the date of receipt and acceptance should be inserted later

Abstract In conditional probabilistic logic programming, given a query, the two most common forms for answering the query are either a probability interval or a precise probability obtained by using the maximum entropy principle. The former can be noninformative (e.g., interval $[0, 1]$) and the reliability of the latter is questionable when the priori knowledge is imprecise. To address this problem, in this paper, we propose some methods to quantitatively measure if a probability interval or a single probability is sufficient for answering a query. We first propose an approach to measuring the ignorance of a probabilistic logic program with respect to a query. The measure of ignorance (w.r.t. a query) reflects how reliable a precise probability for the query can be and a high value of ignorance suggests that a single probability is not suitable for the query. We then propose a method to measure the probability that the exact probability of a query falls in a given interval, e.g., a second order probability. We call it the degree of satisfaction. If the degree of satisfaction is high enough w.r.t. the query, then the given interval can be accepted as the answer to the query. We also prove our measures satisfy many properties and we use a case study to demonstrate the significance of the measures.

1 Introduction

Probabilistic knowledge presents in many real-world applications. Typical examples include medical expert systems, engineering experiments modelling and analysis, etc. Probabilistic logics have been intensively studied in the literature. One important element of many formal languages for representing probabilistic knowledge is the interval restriction for conditional probabilities, also called conditional constraints [Luk98]. Extensive work on probabilistic reasoning about propositional conditional constraints have been carried out (e.g., [DGdMP93], [FH94]).

Logic programming is a well established knowledge representation and reasoning formalism in artificial intelligence and deductive databases. The need for representing uncer-

[1] School of Electronics, Electrical Engineering and Computer Science,
Queen's University Belfast, Belfast BT7 1NN, UK. E-mail: {a.yue, w.liu}@qub.ac.uk
[2] Department of Computer Science, University College London,
Gower Street, London WC1E 6BT, UK. E-mail: a.hunter@cs.ucl.ac.uk

tainty in the logic programming framework is already reported by a great number of publications [Luk98, BGR04, BH07, RKT07, Fuh00, NS91, NS92, DD04] etc.

These probabilistic logic programs have been designed from different perspectives and have different syntactic forms and semantics. In *conditional probabilistic logic programming* [CPQC03, Luk01b], knowledge is represented by interval restrictions for probabilities on conditional events, which is in the form of $(\psi|\phi)[l, u]$. A probabilistic conditional event $(\psi|\phi)[l, u]$ is interpreted as *given ϕ, the probability of ψ falls in the interval* $[l, u]$ where $[l, u] \subseteq [0, 1]$.

In *Causal Probabilistic Logic Programming* [BGR04, BH07], a rule $Pr(\psi|_c\phi) = y$ is interpreted as *"if ϕ happens, this fact will cause the probability of ψ being y"*. In *Success Probabilistic Logic Programming* [RKT07, Fuh00], a rule $\psi \leftarrow \phi$ is associated with a probability p_r, which represents the probability that this rule is true. In [NS91, NS92, DD04, Saa07, BSS09], a probabilistic rule is of the form $\psi[l_1, u_1] \leftarrow \phi[l_2, u_2]$, which means that *"if the probability of ϕ is in the interval $[l_2, u_2]$ then the probability of ψ is in between l_1 and u_1"*. In this logic, we may get a set of intervals for a query, whose actual probability can fall in any of the intervals.

In the field of clinical trials, statistical data summarizing trials results provide some indications about the relationship between medical treatments and their effects. If some rules are derived from these data, the probability of such a rule should not be interpreted as the probability of a rule-head being true causing the rule-tail being true, rather, it is the probability of the effects had the treatment been carried out. For example, rule $(mortality|drug\_name)[l, u]$ cannot be interpreted as *"using this drug causes death with probability in [l,u]"*, rather, it says that with the treatment of this drug, the probability of the mortality (of the patient) is in this interval (so that the cause(s) of the death is something else, not this drug). It is clear that clinical trials data are usually in the form of conditional probability, not in the form of $\psi[l_1, u_1] \leftarrow \phi[l_2, u_2]$. In addition, the effects of a treatment do not fall into different probability intervals. Therefore, based on our application background in clinical trials, we focus only on conditional probabilistic logic programming in this paper.

Conditional probabilistic logic programming is a framework to represent and reason with imprecise (conditional) probabilistic knowledge. An agent's knowledge is represented by a *probabilistic logic program* (PLP) which is a set of (conditional) logical formulas with probability intervals. The impreciseness of an agent's knowledge is explicitly represented by assigning a probability interval to every logical formula (representing a conditional event) indicating that the probability of a formula shall be in the given interval.

Given a PLP and a query against the PLP, traditionally, a probability interval is returned as the answer. This interval implies that the true probability of the query shall be within the given interval. However, when this interval is too wide, it provides no useful information. For instance, if a PLP contains knowledge

$$\{(fly(X)|bird(X)[0.98, 1], (bird(X)|magpie(X))[1, 1]\}$$

then the answer to the query *Can a magpie fly?* (i.e., $?(fly(t)|magpie(t))$) is a trivial bound $[0, 1]$. One way to enhance the reasoning power of a PLP is to apply the maximum entropy principle [KIL04]. Based on this principle, a single probability distribution is selected and it is assumed to be the most acceptable one for a query among all possible probability distributions. As a consequence, a precise probability is given for a query even when the agent's original knowledge is imprecise. In the above example, by applying the maximum entropy principle, 0.98 is returned as the answer for the query. Intuitively, accepting a precise probability from (a prior) imprecise knowledge can be risky. When an agent's knowledge is rich enough then a single probability could be reliable, however, when an agent's knowledge is (very) imprecise, an interval is more appropriate than a single probability.

Therefore, in probabilistic logic programming as well as other condition probabilistic logics, there is a question that has not been fully investigated, that is, how useful a probabilistic logic program (PLP) is to answering a given query? This question's importance is twofold: first, it helps to analyze if a PLP is adequate to answer a query and second, if a PLP is sufficiently relevant to a query, then shall a single probability be obtained or shall a probability interval be more suitable? If it is an interval that is more suitable, then how can we get a more meaningful interval (which is satisfactory to a certain extent), rather than a loose bound?

To answer the above questions, in this paper, we propose two concepts, *the measure of ignorance* and *the measure of the degree of satisfaction*, w.r.t. a PLP and a query. The former analyzes the impreciseness of the PLP w.r.t. a query, and the latter measures which (tighter) interval is sufficiently informative to answer a query.

The main contributions of this paper are as follows. First, we formally analyze conditional probabilistic logic programs and the maximum entropy principle. Although the assumption of applying the maximum entropy principle is intuitive and widely accepted, such an assumption introduces some new unsupported knowledge, and thus, we need to know to what extent an answer given under the maximum entropy is reliable. Second, we propose a general framework which formally defines the measure of ignorance and the measure of the degree of satisfaction, and the postulates for these two measures. We also provide several consequence relations based on the degree of satisfaction. Third, by using the divergence of probabilistic distribution, we instantiate our framework, and show that the measure of ignorance and the measure of the degree of satisfaction have many desirable properties and provide much useful information about a PLP w.r.t. a query. Fourth, we prove that our framework is an extension of both reasoning with probabilistic logic programs and reasoning under the maximum entropy principle. Finally, we prove that these measures can be viewed as a second-order probability. More specifically, a high level of ignorance means a high probability about the given PLP (an agents knowledge) is towards total absence of knowledge. The degree of satisfaction is the second-order probability about the actual probability for a conditional event given in the query falls in the given interval (provided in the query).

This paper is organized as follows. After a brief reviewing of probabilistic logic programming in Section 2, we formally analyze probabilistic logic programming and the maximum entropy principle, and provide our general framework in Section 3. The instantiation of the framework is given in Section 4. In Section 5, we use examples to illustrate that our framework can provide additional information when assessing query results. In Section 6, we present algorithms used in the implementation of the framework and discuss our experiments with statistical data in the field of clinical trials. After comparing with related works in Section 7, we conclude this paper in Section 8.

2 Probabilistic Logic Programming

2.1 Syntax and semantics

We briefly review conditional probabilistic logic programming here [Luk98,Luk01b]. We use Φ to denote a finite set of predicate symbols and constant symbols, $\mathcal{V}$ to denote a set of *object variables*, and $\mathcal{B}$ to denote a set of *bound constants* which describe the bound of probabilities where bound constants are in [0,1]. We use $a, b, \ldots$ to denote constants from Φ and $X, Y \ldots$ to denote object variables from $\mathcal{V}$. An *object term* t is a constant from Φ or an object variable from $\mathcal{V}$. An *atom* is of the form $p(t_1, \ldots, t_k)$, where p is a predicate symbol and $t_1, \ldots, t_k$ are object terms. We use Greek letters $\phi, \varphi, \psi, \ldots$ to denote *events* (or *formulas*) which are obtained from atoms by logic connectives $\wedge, \vee, \neg$ as usual. A *conditional event* is of the form $(\psi|\phi)$ where ψ and ϕ are events, and ϕ is called the *antecedent* and ψ is called the *consequent*. A *probabilistic formula*, denoted as $(\psi|\phi)[l, u]$, means that the probability of conditional event $\psi|\phi$ is between l and u, where l, u are bound constants. A set of probabilistic formulas is called a *conditional probabilistic logic program (PLP)*, a PLP is denoted as P in the rest of the paper.

A *ground* term (resp. event, conditional event, probabilistic formula, or PLP) is a term (resp. event, conditional event, probabilistic formula or PLP) that does not contain any object variables in $\mathcal{V}$.

All the constants in Φ form the Herbrand universe, denoted as HU_Φ, and the Herbrand base, denoted as HB_Φ, is the finite nonempty set of all events constructed from the predicate symbols in Φ and constants in HU_Φ. A subset I of HB_Φ is called a *possible world* and $\mathcal{I}_\Phi$ is used to denote the set of all possible worlds over Φ. A function σ that maps each object variable to a constant is called an *assignment*. It is extended to object terms by $\sigma(c) = c$ for all constant symbols from Φ. An event ϕ satisfied by I under σ, denoted by $I \models_\sigma \phi$, is defined inductively as:

$I \models_\sigma p(t_1, \ldots, t_n)$ iff $p(\sigma(t_1), \ldots, \sigma(t_n)) \in I$;
$I \models_\sigma \phi_1 \wedge \phi_2$ iff $I \models_\sigma \phi_1$ and $I \models_\sigma \phi_2$;
$I \models_\sigma \phi_1 \vee \phi_2$ iff $I \models_\sigma \phi_1$ or $I \models_\sigma \phi_2$;
$I \models_\sigma \neg\phi$ iff $I \not\models_\sigma \phi$

A possible world I *satisfies* or is a *model* of ϕ, denoted by $I \models \phi$, iff $I \models_\sigma \phi$ for all assignment σ. A possible world I *satisfies* or is a *model* of a set of formulas $\mathcal{F}$, denoted by $I \models \mathcal{F}$, iff for all assignment σ and for all $\phi \in \mathcal{F}, I \models_\sigma \phi$. An event ϕ is a *logical consequence* of $\mathcal{F}$, denoted as $\mathcal{F} \models \phi$, iff all models of $\mathcal{F}$ satisfy ϕ.

In this paper, we use $\top$ to represent the (ground) tautology, and we have that $I \models \top$ for all I and all assignments σ; we use $\bot$ to denote $\neg\top$.

If Pr is a function (or distribution) on $\mathcal{I}_\Phi$ (i.e., as $\mathcal{I}_\Phi$ is finite, Pr is a mapping from $\mathcal{I}_\Phi$ to the unit interval [0,1] such that $\sum_{I \in \mathcal{I}_\Phi} Pr(I) = 1$), then Pr is called a *probabilistic interpretation*. For an assignment σ, the probability assigned to an event ϕ by Pr, is denoted as $Pr_\sigma(\phi)$ where $Pr_\sigma(\phi) = \sum_{I \in \mathcal{I}_\Phi, I \models_\sigma \phi} Pr(I)$. When ϕ is ground, we simply write it as $Pr(\phi)$. When $Pr_\sigma(\phi) > 0$, the conditional probability, $Pr_\sigma(\psi|\phi)$, is defined as $Pr_\sigma(\psi|\phi) = Pr_\sigma(\psi \wedge \phi)/Pr_\sigma(\phi)$. When $Pr_\sigma(\phi) = 0$, $Pr_\sigma(\psi|\phi)$ is undefined. Also, when $(\psi|\phi)$ is ground, we simply write it as $Pr(\psi|\phi)$.

A probabilistic interpretation Pr satisfies or is a *probabilistic model* of a probabilistic formula $(\psi|\phi)[l, u]$ under assignment σ, denoted by $Pr \models_\sigma (\psi|\phi)[l, u]$, iff $u \geq Pr_\sigma(\psi|\phi) \geq l$ or $Pr_\sigma(\phi) = 0$. A probabilistic interpretation Pr satisfies or is a *probabilistic model* of a probabilistic formula $(\psi|\phi)[l, u]$, denoted by $Pr \models (\psi|\phi)[l, u]$, iff Pr satisfies $(\psi|\phi)[l, u]$ under all assignments. A probabilistic interpretation Pr satisfies or is a *probabilistic model* of a PLP P iff for all assignment σ and for all $(\psi|\phi)[l, u] \in P, Pr \models_\sigma (\psi|\phi)[l, u]$. A probabilistic formula $(\psi|\phi)[l, u]$ is a *consequence* of the PLP P, denoted by $P \models (\psi|\phi)[l, u]$, iff all probabilistic models of P satisfy $(\psi|\phi)[l, u]$. A probabilistic formula $(\psi|\phi)[l, u]$ is a *tight consequence* of P, denoted by $P \models_{tight} (\psi|\phi)[l, u]$, iff $P \models (\psi|\phi)[l, u]$, $P \not\models (\psi|\phi)[l, u']$, $P \not\models (\psi|\phi)[l', u]$ for all $l' < l$ and $u' > u$ ($l', u' \in [0, 1]$). It is worth noting that if $P \models (\phi|\top)[0, 0]$ then $P \models (\psi|\phi)[1, 0]$ where $[1, 0]$ stands for the empty set.

A query is of the form $?(\psi|\phi)$ or $?(\psi|\phi)[l,u]$, where ψ and ϕ are ground events and l,u are real numbers. For query $?(\psi|\phi)$, under tight consequence reasoning, a bound $[l,u]$ is given by the answer, such that $P \models_{tight} (\psi|\phi)[l,u]$. For query $?(\psi|\phi)[l,u]$, the bound $[l,u]$ is given by the user. It is returned as *True* (or *Yes*) if $P \models (\psi|\phi)[l,u]$ and *False* (or *No*) if $P \not\models (\psi|\phi)[l,u]$.

Using probabilistic logic programs, we can represent imprecise probabilistic knowledge.

Example 1 Let P be a PLP that contains only one constant $tweety$ and

$$P = \left\{ \begin{array}{l} (fly(X)|bird(X))[0.98,1], \\ (bird(X)|magpie(X))[1,1] \end{array} \right\}$$

Table 1 Probabilistic models for P

$Index$	$HerbrandModel$	$Probability$
I_1	$\emptyset$	x_1
I_2	$\{magpie(tweety)\}$	ϵ_1
I_3	$\{bird(tweety)\}$	x_2
I_4	$\{bird(tweety), magpie(tweety)\}$	x_3
I_5	$\{fly(tweety)\}$	x_4
I_6	$\{fly(tweety), magpie(tweety)\}$	ϵ_2
I_7	$\{fly(tweety), bird(tweety)\}$	x_5
I_8	$\{fly(tweety), bird(tweety), magpie(tweety)\}$	x_6

where
$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + \epsilon_1 + \epsilon_2 = 1$,
$\frac{x_5+x_6}{x_2+x_3+x_5+x_6} \in [0.98,1]$ or $x_2 + x_3 + x_5 + x_6 = 0$,
$\frac{x_3+x_6}{x_3+x_6+\epsilon_1+\epsilon_2} = 1$ or $x_3 + x_6 + \epsilon_1 + \epsilon_2 = 0$, thus $\epsilon_1 = \epsilon_2 = 0$

Intuitively, we expect that the probability of $(fly(tweety)|magpie(tweety))$ to be high. Table 1 gives all the possible probabilistic models for P, which satisfy all the constraints listed in the table. There are eight Herbrand models ($\mathcal{I}_{\Phi} = \{I_0, \ldots, I_8\}$) of P. The third column gives the probability for each of them. So,

$$Pr(fly(tweety)|magpie(tweety))) = \frac{x_6}{x_6 + x_3}.$$

It is easy to see that there are many probability distributions that satisfy all the constraints induced from the PLP. Let $x_1 = x_4 = x_6 = 0, x_5 = 0.98, x_2 = 0, x_3 = 0.02$, we can get a probabilistic model Pr_1 for P and $Pr_1(fly(tweety)|magpie(tweety)) = 0$. Let $x_1 = x_4 = x_2 = x_3 = 0, x_5 = x_6 = 0.5$, we get another probabilistic model Pr_2 for P and $Pr_2(fly(tweety)|magpie(tweety)) = 1$. Therefore,

$$P \models_{tight} (fly(tweety)|magpie(tweety))[0,1]$$

that means we have no idea about whether a magpie can fly.

$$\begin{array}{l}
\sum_{I\in\mathcal{I}_\Phi} y_I = 1\\
\sum_{I\in\mathcal{I}_\Phi, I\models\neg\psi\wedge\phi} -ly_I + \sum_{I\in\mathcal{I}_\Phi, I\models\psi\wedge\phi}(1-l)y_I \geq 0, \text{ (for all } (\psi|\phi)[l,u]\in P, l>0)\\
\sum_{I\in\mathcal{I}_\Phi, I\models\neg\psi\wedge\phi} uy_I + \sum_{I\in\mathcal{I}_\Phi, I\models\psi\wedge\phi}(u-1)y_I \geq 0, \text{ (for all } (\psi|\phi)[l,u]\in P, u<1)\\
y_I \geq 0 \text{ (for all } I\in\mathcal{I}_\Phi)
\end{array}$$

Fig. 1 Linear constraints $LC(P, \mathcal{I}_\Phi)$

2.2 Maximum entropy principle

One possible method to enhance the reasoning power of probabilistic logic programs is reasoning by the distribution with maximum entropy [KIL04, KIR04]. The principle of maximum entropy is a well known technique to represent probabilistic knowledge. Entropy quantifies the indeterminateness inherent to a distribution Pr by $H(Pr) = -\sum_{I\in\mathcal{I}_\Phi} Pr(I) log Pr(I)$. Given a logic program P, the *principle of maximum entropy model* (or *me-model*), denoted by $me[P]$, is defined as:

$$H(me[P]) = \max H(Pr) = \max_{Pr\models P} -\sum_{I\in\mathcal{I}_\Phi} Pr(I)\log Pr(I) \qquad (1)$$

$me[P]$ is the unique probabilistic interpretation Pr that is a probabilistic model of P and that has the greatest entropy among all the probabilistic models of P.

In Example 1, let Pr be the probability distribution with maximum entropy that satisfies the PLP P, then $Pr(fly(tweety)|magpie(tweety)) = 0.98$.

Let P be a ground PLP, we say that $(\psi|\phi)[l,u]$ is a *me-consequence* of P, denoted by $P \models^{me} (\psi|\phi)[l,u]$, iff

- P is unsatisfiable, or
- $me[P] \models (\psi|\phi)[l,u]$.

we say that $(\psi|\phi)[l,u]$ is a *tight me-consequence* of P, denoted by $P \models^{me}_{tight} (\psi|\phi)[l,u]$, iff

- P is unsatisfiable, $l = 1$, $u = 0$, or
- $P \models \bot \leftarrow \phi$, $l = 1$, $u = 0$, or
- $me[P](\phi) > 0$ and $me[P](\psi|\phi) = l = u$.

Applying the principle of maximum entropy solves the problem of inferring noninformative probabilistic intervals. For instance, we have that
$P \models^{me}_{tight} (fly(tweety)|magpie(tweety))[0.98, 0.98]$, where P is as given in Example 1.

2.3 Implementation of reasoning with PLP

A logic program can be treated as a set of inequality constraints $LC(P, \mathcal{I}_\Phi)$ shown in Fig. 1. A solution $y^\star_I, I \in \mathcal{I}_\Phi$ that satisfies $LC(P, \mathcal{I}_\Phi)$ gives a probabilistic model of P, that is $Pr(I) = y^\star_I$.

Theorem 1 ([KIL04]) *Let P be a PLP. Then*

1. *For every model Pr of P, there exists a solution $(y_I)_{I\in\mathcal{I}_\Phi}$ of the system of linear constraints $LC(P, \mathcal{I}_\Phi)$, such that $Pr(I) = y_I$ for all $I \in \mathcal{I}_\Phi$.*
2. *For every solution of $LC(P, \mathcal{I}_\Phi)$, there exists a model Pr of P such that $y_I = Pr(I)$ for all $I \in \mathcal{I}_\Phi$.*

The next theorem shows that the reasoning problems can be reduced to the optimization problem subject to the linear constraints $LC(P, \mathcal{I}_{\Phi})$:

Theorem 2 ([KIL04]) *Let P be a PLP and $P \not\models \bot \leftarrow \phi$. Then $P \models_{tight} (\psi|\phi)[l, u])$, where l (respectively, u) is the minimal (respectively, maximal) value of the following linear program over the variables $y_I (I \in \mathcal{I}_{\Phi})$:*

$$\textit{Minimize (respectively, maximize)} \quad \frac{\sum_{I \in \mathcal{I}_{\Phi}, I \models \psi \wedge \phi} y_I}{\sum_{I \in \mathcal{I}_{\Phi}, I \models \phi} y_I}$$

$$\textit{Subject to } LC(P, \mathcal{I}_{\Phi}) \textit{ and } \sum_{I \in \mathcal{I}_{\Phi}, I \models \phi} y_I > 0$$

Proposition 1 ([KIL04]) *Let P be a PLP, and let ψ be a ground event. Then, $P \models^{me} (\psi|\top)[0, 0]$ iff $P \models_{tight} (\psi|\top)[0, 0]$*

Theorem 3 ([KIL04]) *Let P be a PLP. Then $P \models^{me}_{tight} (\psi|\phi)[l, u])$, where $l = u = me[P](\psi|\phi)$, and $me[P]$ is computed by solving the following entropy maximization problem over the variables $y_I (I \in \mathcal{I}_{\Phi})$:*

$$\textit{Maximize} \quad -\sum_{I \in \mathcal{I}_{\Phi}} y_I \log y_I \quad \textit{subject to} \quad LC(P, \mathcal{I}_{\Phi}) \tag{2}$$

Example 2 Considers the PLP given in Example 1. The constraints $LC(P, _{\Phi})$ induced by P is the following:

$$\begin{array}{l}
\sum_{1 \leq i \leq 8} y_{I_i} = 1 \\
0.02 * y_{I_7} + 0.02 * y_{I_8} - 0.98 * y_{I_3} - 0.98 * y_{I_4} \geq 0 \\
y_{I_3} + y_{I_4} \geq 0 \\
-y_{I_2} - y_{I_3} \geq 0 \\
y_{I_2} + y_{I_3} \geq 0 \\
y_{I_i} \geq 0, 1 \leq i \leq 8
\end{array}$$

3 Reasoning with Imprecise Probability

3.1 A formal analysis of PLP

In information theory, information entropy is a measure of the uncertainty associated with a random variable. Entropy quantifies information in a piece of data. Informally speaking, $-\log p(X = x_i)$ means the degree of surprise when one observes that the random variable turns out to be x_i. In other words, $-\log p(X = x_i)$ reflects the information one receives from the observation. The entropy is an expectation of the information one may receive from a random domain by observing random events. Inspired by this, we define a knowledge entropy, which reflects how much an agent knows the truth value of ψ given ϕ prior any observations. Informally, the more surprised an agent is by the observation, the more knowledge it learns from the observation, and thus, the less prior knowledge its has about ψ given ϕ before observing ψ given ϕ.

Definition 1 Let P be a PLP, and $(\psi|\phi)$ be a conditional event. Suppose that Pr is a probabilistic model for P, then the *knowledge entropy* of inferring ψ from ϕ under Pr is defined by:

$$K_{Pr}(\psi|\phi) = 1 + Pr(\psi|\phi)\log Pr(\psi|\phi) + Pr(\neg\psi|\phi)\log Pr(\neg\psi|\phi)$$

It is obvious that $K_{Pr}(\psi|\phi) = K_{Pr}(\neg\psi|\phi)$ and $K_{Pr}(\psi|\phi) \in [0,1]$. Trivially, we have $K_{Pr}(\phi|\phi) = 1$ and $K_{Pr}(\neg\phi|\phi) = 1$, since from Pr, an agent can exactly know the truth value of ψ and its negation given ϕ.

By extending the above definition, we can define a knowledge measurement for a PLP.

Definition 2 Let P be a PLP, and $(\psi|\phi)$ be a conditional event. Suppose that Pr is a probabilistic model for P and $Pr(\phi) > 0$, then the *knowledge measurement* $K_P(\psi|\phi)$ is defined by:

$$\begin{aligned} minK_P(\psi|\phi) &= \min_{Pr \models P} K_{Pr}(\psi|\phi), \\ maxK_P(\psi|\phi) &= \max_{Pr \models P} K_{Pr}(\psi|\phi), \\ K_P(\psi|\phi) &= [minK_P(\psi|\phi), maxK_P(\psi|\phi)]. \end{aligned}$$

The measurement $K_P(\psi|\phi)$ is used to characterize the usefulness of a priori knowledge contained in PLP P for inferring ψ when knowing or observing ϕ. When ψ or $\neg\psi$ can be inferred from ϕ under P, P contains all the necessary knowledge of inferring ψ given ϕ, and we have that $minK_P(\psi|\phi) = 1$. When knowledge in P excludes the possibility that the probability of ψ (or $\neg\psi$) may be 1 given ϕ, i.e., $P \cup \{(\psi|\phi)[1,1]\}$ (or $P \cup \{(\psi|\phi)[0,0]\}$) is unsatisfiable, then the knowledge contained in P cannot fully support ψ given ϕ, so $maxK_P(\psi|\phi) < 1$. Specifically, if the conclusion that ψ is more (or less) likely to be true than $\neg\psi$ (, i.e. the probability of ψ given ϕ is bigger (or smaller) than $\neg\psi$ given ϕ,), then $maxK_P(\psi|\phi) > 0$.

We can define a partial order $\preceq$ over the set $\{[x,y]|x,y \in [0,1]\}$ as $[a,b] \preceq [c,d]$ iff $a \leq c, b \geq d$, and $[a,b] \prec [c,d]$ iff $[a,b] \preceq [c,d]$ and $a < c$ or $b > d$. We say that a PLP P is more precise than P' w.r.t. $\psi|\phi$, if $K_P(\psi|\phi) \preceq K_{P'}(\psi|\phi)$, denoted as $P \preceq^k_{(\psi|\phi)} P'$.

If $minK_P(\psi|\phi) \neq maxK_P(\psi|\phi)$, then the knowledge contained in P is not sufficient to decide the probability of ψ given ϕ, that is, the knowledge contained in P about inferring ψ given ϕ is imprecise. In order to infer the actual probability of ψ given ϕ under P, we need additional knowledge.

Proposition 2 *Let P and P' be two PLPs. If $P \models P'$ then $P \preceq^k_{(\psi|\phi)} P'$ for any conditional event $(\psi|\phi)$.*

Proof Suppose that $P \models_{tight} (\psi|\phi)[l,u]$ and $P' \models_{tight} (\psi|\phi)[l',u']$. If $P \models P'$, then $[l,u] \subseteq [l',u']$. So, $minK_P(\psi|\phi) \leq minK_{P'}(\psi|\phi)$ and $maxK_P(\psi|\phi) \geq maxK_{P'}(\psi|\phi)$, therefore, $P \preceq^k_{(\psi|\phi)} P'$. □

This proposition suggests that the consequence relation $\models$ considers all the statements in a PLP while the knowledge measurement focuses only on the knowledge about ψ given ϕ. $maxK_P(\psi|\phi)$ reflects the best knowledge consistent with P that can help to infer the truth value of ψ given ϕ. Since $P \models P'$, P contains more knowledge than P', thus $minK_P(\psi|\phi) \geq minK_{P'}(\psi|\phi)$. It is also possible that this knowledge excludes some other information that is useful to infer the truth value of ψ given ϕ, thus $maxK_P(\psi|\phi) \leq maxK_{P'}(\psi|\phi)$.

In the view of knowledge entropy, reasoning under the maximum entropy principle implicitly introduces some extra knowledge to enhance the reasoning power of PLP.

Example 3 Consider again the PLP in Example 1. There exists a probabilistic model Pr_1 for P such that

$$Pr_1(fly(tweety)|magpie(tweety)) = 0.5, K_{Pr_1}(fly(tweety)|magpie(tweety)) = 0$$

On the other hand, there exist probabilistic models Pr and Pr' for P such that

$$\begin{array}{l} Pr(fly(tweety)|magpie(tweety)) = 1 \\ Pr'(fly(tweety)|magpie(tweety)) = 0 \end{array}$$

and

$$K_{Pr}(fly(tweety)|magpie(tweety)) = K_{Pr'}(fly(tweety)|magpie(tweety)) = 1$$

So, $K_P(fly(tweety)|magpie(tweety)) = [0, 1]$. With the maximum entropy, we get that

$$me[P](fly(tweety)|magpie(tweety)) = 0.98$$

and thus

$$1 > K_{me[P]}(fly(tweety)|magpie(tweety)) > 0$$

This suggests that some knowledge is introduced.

For example, $me[P](magpie(tweety)|bird(tweety)) = 0.5$, which is not supported by P. The rationale behind the maximum entropy principle is to represent given probabilistic information as faithfully as possible, by maximizing admissible indeterminateness. Specific to this example, with the maximum entropy principle, an assumption that magpies are distinguishable from typical birds is introduced (this assumption cannot be represented in PLP), and actually, such an assumption enriches the knowledge contained in P. □

From the above example, we know that reasoning under the maximum entropy cannot be taken for granted as reasoning based on minimal knowledge, but actually based on some implicit knowledge. We should be aware that although the assumption seems intuitive, it may be wrong.

Example 4 Let $P_1 = (headUp(X)|toss(X))[0.5, 0.5], P_2 = (headUp(X)|toss(X))[0, 1]$ be PLPs. Here, P_1 says that tossing a fair coin may result in head-up with probability 0.5, however, in P_2, we do not know whether the coin is fair.

In this example, the knowledge in P_1 is richer than that in P_2 since from P_1 we know the coin is fair. Using the maximum entropy principle, we get that

$$P_1 \models^{me} (headUp(coin)|toss(coin))[0.5, 0.5],$$

$$P_2 \models^{me} (headUp(coin)|toss(coin))[0.5, 0.5].$$

This result suggests that the difference between P_1 and P_2 is omitted under the maximum entropy reasoning. By calculating the knowledge entropy of P_1 and P_2, we know that $K_{P_1}(headUp(coin)|toss(coin)) = [0, 0]$ and $K_{P_2}(headUp(coin)|toss(coin)) = [0, 1]$. Thus we know that P_1 is more precise than P_2.

Obviously, the conclusion $(headUp(coin)|toss(coin))[0.5, 0.5]$ is more acceptable under P_1 than under P_2. □

From the above examples, we know that, accepting a conclusion obtained by reasoning under the maximum entropy principle may imply that we are willing to introduce extra knowledge into a given PLP and there is no guarantee that this extra knowledge is always correct.

In the next subsection, we provide a general framework for analyzing and reasoning with imprecise PLPs. For this purpose, we provide two concepts: ignorance and degree of satisfaction. The ignorance reflects the richness of the knowledge contained in a PLP and the degree of satisfaction of a query $?(\psi|\phi)[l, u]$ reflects the possibility that the actual probability of $(\psi|\phi)$ falls in the given bound $[l, u]$.

3.2 General framework for measuring imprecise knowledge

Intuitively, the knowledge measurement $K_P(\psi|\phi)$ indicates to some extent the ignorance about the conditional event $(\psi|\phi)$ when using knowledge contained in P. But unfortunately, such interval can not sufficiently reflect the ignorance about $(\psi|\phi)$. This is not surprising, since $K_P(\psi|\phi)$ is determined only by the tight probability bound of the conditional event $(\psi|\phi)$, other knowledge is not considered in $K_P(\psi|\phi)$.

Example 5 Let P be a PLP:

$$P = \left\{ \begin{array}{l} (fly(X)|bird(X))[0.9, 1], (bird(X)|magpie(X))[1, 1] \\ (sickMagpie(X)|magpie(X))[0, 0.1], (magpie(X)|sickMagpie(X))[1, 1] \end{array} \right\}$$

From P, we can infer that
$P \models_{tight} (fly(t)|magpie(t))[0, 1]$,
$P \models_{tight} (fly(t)|sickMagpie(t))[0, 1]$, $P \models^{me}_{tight} (fly(t)|magpie(t))[0.9, 0.9]$, and
$P \models^{me}_{tight} (fly(t)|sickMagpie(t))[0.9, 0.9]$.

Here, we have $K_P(fly(t)|sickMagpie(t)) = K_P(fly(t)|magpie(t))$. However, since the proportion of sick magpies in birds is smaller than the proportion of magpies in birds, the knowledge that birds can fly should be more cautiously applied to sick magpies than magpies in general. In other words, the statement that more than 90% of birds can fly is more about magpies than sick magpies. Therefore, to accept that 90% of magpies can fly is more rational than to accept 90% of sick magpies can fly. However, these analyses can not be obtained directly from comparing the bounds inferred from P.

We define an ignorance measurement to characterize the knowledge incompleteness restricted to a given (conditional) event:

Definition 3 (Ignorance) Let $\mathcal{PL}$ be the set of all PLPs and $\mathcal{E}$ be a set of all conditional events. Function $\mathsf{IG} : \mathcal{PL} \times \mathcal{E} \mapsto [0, 1]$ is called a *measure* [1] *of ignorance*, iff for any PLP P and conditional event $(\psi|\phi)$ it satisfies the following postulates

[1] In mathematical analysis, a measure m is a function, such that $m : 2^S \mapsto [0, \infty]$ and

1. $m(E_1) \geq 0$ for any $E \subseteq S$,
2. $m(\emptyset) = 0$,
3. if $E_1, E_2, E_3, \ldots$ is a countable sequence of pairwise disjoint subsets of S, the measure of the union of all the E_i's is equal to the sum of the measures of each E_i, that is, $m(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} m(E_i)$.

[Boundedness] $\mathsf{IG}(P, \psi|\phi) \in [0, 1]$.
[Preciseness] $\mathsf{IG}(P, \psi|\phi) = 0$ iff $P \models_{tight} (\psi|\phi)[u, u]$ or $P \models \bot \leftarrow \phi$.
[Total Ignorance] $\mathsf{IG}(\emptyset, \psi|\phi) = 1$, if $\not\models_{cl} \phi \rightarrow \psi$ and $\not\models_{cl} \phi \rightarrow \neg\psi$.
[Soundness] If $\mathsf{IG}(P, \psi|\phi) = 1$ then $P \models (\psi|\phi)[0, 1]$.
[Irrelevance] If P and another PLP P' do not contain common syntaxes, i.e.$\Phi \cap \Phi' = \emptyset$, then $\mathsf{IG}(P, \psi|\phi) = \mathsf{IG}(P \cup P', \psi|\phi)$.

For simplicity, we use $\mathsf{IG}_P(\psi|\phi)$ to denote $\mathsf{IG}(P, \psi|\phi)$ for a given PLP P and a conditional event $(\psi|\phi)$. Value $\mathsf{IG}_P(\psi|\phi)$ defines the *ignorance* about $(\psi|\phi)$ from P.

If $P = \emptyset$, only tautologies can be inferred from P. Therefore, from any PLP P, $\mathsf{IG}_P(\psi|\phi) \leq \mathsf{IG}_\emptyset(\psi|\phi)$, which means that an empty PLP has the biggest ignorance value for any conditional event. When $\mathsf{IG}_P(\psi|\phi) = 0$, event $(\psi|\phi)$ can be inferred precisely from P, since a single precise probability for $(\psi|\phi)$ can be obtained from P. The ignorance measurement focuses on the knowledge about $(\psi|\phi)$ contained in P, which means that irrelevant knowledge does not provide a better understanding of this conditional event.

Proposition 3 *Let P be a PLP and $(\psi|\phi)$ be a conditional event. If $\mathsf{IG}_P(\psi|\phi) = 1$, then $K_P(\psi|\phi) = [0, 1]$. If $\mathsf{IG}_P(\psi|\phi) = 0$, then $K_P(\psi|\phi) = [a, a]$, for some $a \in [0, 1]$.*

Proof If $\mathsf{IG}_P(\psi|\phi) = 0$ then $P \models_{tight} (\psi|\phi)[0, 1]$, and thus $K_P(\psi|\phi) = [0, 1]$. If $\mathsf{IG}_P(\psi|\phi) = 1$ then $P \models_{tight} (\psi|\phi)[p, p]$ for some $p \in [0, 1]$. So, for any $Pr \models P$, $Pr(\psi|\phi) = p$. By the definition of $K_P(\psi|\phi)$, we have that $K_P(\psi|\phi) = [a, a]$, where $a = 1 + a * \log a + (1 - a) * log(1 - a)$. It is easy to check that $a \in [0, 1]$. □

When querying a PLP, the tight consequence reasoning gives a bound as the answer which can be too cautious and not very informative; however, me-consequence gives a precise probability for a query, and it is too risky to simply accept it. We argue that, sometimes we do not need to know a precise probability which is not reliable enough; we may want to know whether the probability falls in a given bound with high enough possibility. In other words, we may want to strike a balance between a less informative bound which is true for sure and an intuitively precise probability which is not reliable enough. Consider the ignorance of a conditional event, when the ignorance is 0, then accepting the precise probability given by the maximum entropy principle is guaranteed to be right. But how about the situation that the ignorance of a conditional event is very small? It suggests that the knowledge contained in the PLP is rich enough to infer an informative bound, but the knowledge is not rich enough to infer a precise probability. In order to extract an informative and reliable interval, we first measure the degree of satisfaction for a query (with bound). If a degree is high enough, then regarding the query as true is reliable, since it is very possible that the actual probability falls in the given interval.

Definition 4 **(Degree of Satisfaction)** Let $\mathcal{PL}$ be the set of all PLPs and $\mathcal{F}$ be a set of all probabilistic formulas. Function $\mathsf{SAT} : \mathcal{PL} \times \mathcal{F} \mapsto [0, 1]$ is called a measure of *degree of satisfaction* iff for any PLP P and ground probabilistic formula $\mu = (\psi|\phi)[l, u]$, it satisfies the following postulates:

[Reflexivity] $\mathsf{SAT}(P, \mu) = 1$, iff $P \models \mu$.
[Rationality] $\mathsf{SAT}(P, \mu) = 0$ if $P \cup \{\mu\}$ is unsatisfiable.
[Monotonicity]
$\mathsf{SAT}(P, \mu) \geq \mathsf{SAT}(P, (\psi|\phi)[l', u'])$, if $[l', u'] \subseteq [l, u]$.
$\mathsf{SAT}(P, \mu) > \mathsf{SAT}(P, (\psi|\phi)[l', u'])$, if $[l', u'] \subset [l, u]$
and $\mathsf{SAT}(P, (\psi|\phi)[l', u']) < 1$.

[Cautious Monotonicity] Let $P' = P \cup \{(\psi|\phi)[l', u']\}$, where $P \models^{me} (\psi|\phi)[l', u']$.
If $1 \geq \mathsf{SAT}(P, \mu) \geq 0$ then $\mathsf{SAT}(P', \mu) \geq \mathsf{SAT}(P, \mu)$,

For simplicity, we use $\mathsf{SAT}_P(\mu)$ to denote $\mathsf{SAT}(P, \mu)$.

The reflexivity property says that every consequence is totally satisfied. Rationality says that 0 is given as the degree of satisfaction of an unsatisfiable probabilistic formula. Monotonicity says that if we expect a more precise interval for a query, then the chance that the exact probability of the query is *not* in the interval is greater. Cautious monotonicity says that, if P and P' are equivalent except for the bound of $(\psi|\phi)$, and P' contains more knowledge about $(\psi|\phi)$, then the degree of satisfaction of μ under P' should be bigger than that of μ under P.

Proposition 4 *Function* SAT *is* consistent with the maximum entropy principle*, that is, it satisfies the following conditions for any PLP P and any conditional event $(\psi|\phi)$ with $P \not\models \bot \leftarrow \phi$ and $[l, u] \in [0, 1]$*

$$\mathsf{SAT}_P((\psi|\phi)[l, u]) \begin{cases} = 0 & \textit{if } P \models^{me} (\psi|\phi)[l', l'], \textit{ and } l' \notin [l, u] \\ > 0 & \textit{if } P \models^{me} (\psi|\phi)[l', l'], \textit{ and } l' \in [l, u] \end{cases}$$

Proof Assume that $P \models_{tight} (\psi|\phi)[l_0, u_0]$ and $P \models^{me}_{tight} (\psi|\phi)[l', l']$. For simplicity, we denote $\mu = (\psi|\phi)[l, u]$.

1) If $l' \in [l, u]$.

1.1) If $l_0 = u_0 = l'$. According to Postulate Reflexivity, $\mathsf{SAT}_P(\mu) = 1$.

1.2) If $l_0 < l'$ or $u_0 > l'$.

$\mathsf{SAT}_P(\mu) \geq \mathsf{SAT}_P((\psi|\phi)[l_1, u_1]) > \mathsf{SAT}_P((\psi|\phi)[l', u']) \geq 0$ where $l_1 = \max(l, l_0), u_1 = \min(u, u_0)$.

Therefore, in this case $\mathsf{SAT}_P(\mu) > 0$

2) If $l' \notin [l, u]$.

Let $P' = P \cup \{(\psi|\phi)[l', l']\}$.

According to Postulate Rationality, $\mathsf{SAT}_{P'}(\mu) = 0$. According to Cautious Monotonicity, $\mathsf{SAT}_{P'}(\mu) \geq \mathsf{SAT}_P(\mu)$.

Therefore, in this case $\mathsf{SAT}_P(\mu) = 0$. □

For a query $?(\psi|\phi)[l, u]$, when $\mathsf{SAT}_P((\psi|\phi)[l, u]) < 1$ it means that the exact probability of $(\psi|\phi)$ in $[l, u]$ could be wrong based on the knowledge in P.

In our framework, given a PLP P, a conditional event $(\psi|\phi)$, and a probabilistic formula $(\psi|\phi)[l, u]$, the ignorance value $\mathsf{IG}_P(\psi|\phi)$ and the degree of satisfaction $\mathsf{SAT}_P(\mu)$ reveal different aspects of the impreciseness of the knowledge in P w.r.t. $(\psi|\phi)$ and μ. The former says how much this P can tell about $(\psi|\phi)$ and the latter says to what degree a user can be satisfied with the bound $[l, u]$ attached to $(\psi|\phi)$.

Proposition 5 *Let P be a PLP and $(\psi|\phi)$ be a conditional event. If* $\mathsf{IG}_P(\psi|\phi) = 0$ *then* $\mathsf{SAT}_P((\psi|\phi)[l, l]) = 1$ *for some* $l \in [0, 1]$.

Proof From the Preciseness property of ignorance measure IG, and $\mathsf{IG}_P(\psi|\phi) = 0$, we have that $P \models_{tight} (\psi|\phi)[l, l]$ for some $l \in [0, 1]$. From the Reflexivity property of degree of satisfaction, we have that $\mathsf{SAT}_P((\psi|\phi)[l, l]) = 1$. □

The above proposition says that when the knowledge contained in P totally ignores the conditional event $(\psi|\phi)$, then all the knowledge contained in P is irrelevant to the query $?(\psi|\phi)[l, l]$.

Definition 5 Let $\mathsf{SAT}_P(\mu)$ be the degree of satisfaction for a PLP P and $\mu = (\psi|\phi)[l, u]$ be a probabilistic formula. We define two consequence relations as

- $P \models^{SAT \geq w} \mu$ iff $\mathsf{SAT}_P(\mu) \geq w$,
- $P \models^{SAT \geq w}_{tight} \mu$ iff $P \models^{SAT \geq w} \mu$ and $P \not\models^{SAT \geq w} (\psi|\phi)[l', u']$ for every $[l', u'] \subset [l, u]$.

Proposition 6 *Let $\mathsf{SAT}_P(\mu)$ be the degree of satisfaction for a PLP P and a probabilistic formula $\mu = (\psi|\phi)[l, u]$, then*

$$\begin{array}{l} P \models \mu \textit{ iff } P \models^{SAT=1} \mu \\ P \models_{tight} \mu \textit{ iff } P \models^{SAT=1}_{tight} \mu \\ P \models^{me}_{tight} \mu \textit{ iff } \lim_{\epsilon \to 0+} P \models^{SAT \geq \epsilon}_{tight} \mu \end{array}$$

Proof From the Reflexivity property of SAT and the definition of $\models^{SAT=1}$, we have that $P \models \mu$ iff $P \models^{SAT=1} \mu$ and from the definition of $\models^{SAT=1}_{tight}$, we have that $P \models_{tight} \mu$ iff $P \models^{SAT=1}_{tight} \mu$.

Assume that $pme = me[P](\psi|\phi)$.

If $\mathsf{SAT}_P(\mu)$ is also consistent with the maximum entropy principle, then $\models^{SAT \geq \epsilon}_{tight} (\psi|\phi)[l', u']$ for some $[l', u'] \subseteq [0, 1]$ and $pme \in [l', u']$. As ϵ becomes smaller, the bound $[l', u']$ gets tighter. Since SAT satisfies the Monotonicity property, the bound is approaching to the bound $[pme, pme]$. Therefore, $P \models^{me}_{tight} \mu$ iff $\lim_{\epsilon \to 0+} P \models^{SAT \geq \epsilon}_{tight} \mu$. □

In this proposition, we use $SAT = 1$ instead of $SAT \geq 1$, since the degree of satisfaction can not be greater than 1.

The above proposition says that our framework is a generalization of PLP under its original semantics as well as under the maximum entropy principle. That is, the classical consequence relations $\models$ and $\models_{tight}$ are too cautious - they are equivalent to requiring the degree of satisfaction of μ w.r.t P to be 1, which means that the true probability of $(\psi|\phi)$ must fall in the bound $[l, u]$. On the other hand, reasoning under the maximum entropy principle ($\models^{me}_{tight}$) is credulous – it excludes all the other possible probability distributions except for the most possible one.

Given a query $?(\psi|\phi)[l, u]$ against a PLP P, the degree of satisfaction $\mathsf{SAT}_P(\mu)$ tells the probability that $Pr(\psi|\phi) \in [l, u]$. For a query $?(\psi|\phi)$, the bound $[l, u]$ returned by $P \models_{tight} (\psi|\phi)[l, u]$ may be noninformative as discussed above. In our framework, we provide three possible routes to generate a more informative interval $[l', u']$ with $\mathsf{SAT}_P((\psi|\phi)[l', u']) \geq a$, where a is threshold given by the user. First, a user may want to know the highest acceptable lower bound, so l is increased to a smallest value l' s.t. $\mathsf{SAT}_P((\psi|\phi)[l', u]) \geq a$ holds. Second, a user may want to know the lowest upper bound, so u is decreased to be u' until $\mathsf{SAT}_P((\psi|\phi)[l, u']) \geq a$ is true. Third, a user may want to create an interval $[l', u']$ around $me[P]$, the precise probability given by the maximum entropy principle, where $\mathsf{SAT}_P((\psi|\phi)[l', u']) \geq a$ holds. To formalize these three scenarios, we define three consequence relations $\models^{SAT \geq a}_{maxLow}$, $\models^{SAT \geq a}_{minUp}$ and $\models^{SAT \geq a}_{arroundMe}$ for them respectively as

- $P \models^{SAT \geq a}_{maxLow} (\psi|\phi)[l', u]$ iff $P \models^{SAT \geq a}_{tight} (\psi|\phi)[l', u]$ where $P \models_{tight} (\psi|\phi)[l, u]$
- $P \models^{SAT \geq a}_{minUp} (\psi|\phi)[l, u']$ iff $P \models^{SAT \geq a}_{tight} (\psi|\phi)[l, u']$, where $P \models_{tight} (\psi|\phi)[l, u]$
- $P \models^{SAT \geq a}_{arroundMe} (\psi|\phi)[l', u']$ iff $P \models^{SAT \geq a}_{tight} (\psi|\phi)[l', u']$ where $P \models_{tight} (\psi|\phi)[l, u]$, and $\exists b \geq 0, P \models^{me}_{tight} (\psi|\phi)[m, m]$, $l' = \max\{l, m - b\}$, $u' = \min\{u, m + b\}$

Example 6 Let $P = \{(fly(t)|bird(t))[0.90, 1],\ (bird(t)|magpie(t))[1, 1]\}$ be a PLP. From P, we can only infer that
$P \models_{tight} (fly(t)|magpie(t))[0, 1]$, and $P \models^{me}_{tight} (fly(t)|magpie(t))[0.9, 0.9]$.

As discussed above, the bound $[0,1]$ is meaningless and there is not enough knowledge to infer that exactly 90% magpies can fly. In reality, taking $[0.9,0.9]$ as the answer for this query is too risky, and there is no need to get a precise probability for the query. A more informative interval $[l,u]$ than $[0,1]$ would be required. Assume that a user is happy when there is a 80% (i.e. $a=0.8$) chance that the actual probability of the query is in $[l,u]$, then we are able to use the above three consequence relations to get the following

$$P \models^{SAT\geq 0.8}_{maxLow} (fly(t)|magpie(t))[0.7,1]$$
$$P \models^{SAT\geq 0.8}_{minUp} (fly(t)|magpie(t))[0,0.96]$$
$$P \models^{SAT\geq 0.8}_{arroundMe} (fly(t)|magpie(t))[0.7,1]$$

From the highest lower bound 0.7, a user can assume that a magpie very likely can fly. The user should not think that all magpies can fly either, since the lowest upper bound 0.96 is less than 1. The bound $[0.7,1]$ gives an estimate for the probability that a magpie can fly.

4 Instantiation of Our Framework

4.1 Quasi-distance

How to measure the *distance* between probability distributions is a major topic in probability theory and information theory. One of the most common measures for comparing probability distributions is the KL-divergence:

Definition 6 Let Pr and Pr' be two probability distributions over the same set $\mathcal{I}_{\Phi}$. The KL-divergence between Pr and Pr' is defined as:

$$KL(Pr\|Pr') = -\sum_{I\in\mathcal{I}_{\Phi}} Pr(I)\log\frac{Pr'(I)}{Pr(I)}$$

KL-divergence is also called *relative entropy*.

It should be noted that $KL(Pr,Pr')$ is undefined if $Pr'(I)=0$ and $Pr(I)\neq 0$. This means Pr has to be absolutely continuous w.r.t. Pr' for $KL(Pr\|Pr')$ to be defined.

Let $\mathbf{Pr}_{\Phi}$ be the set of all probability distributions on the set of interpretations $\mathcal{I}_{\Phi}$. Let $\mathbf{Pr}_1$ and $\mathbf{Pr}_2$ be two subsets of $\mathbf{Pr}_{\Phi}$, $\mathbf{Pr}_1$ and $\mathbf{Pr}_2$ are *separated* if each is disjoint from the other's *closure* [2]. A subset $\mathbf{Pr}$ of $\mathbf{Pr}_{\Phi}$ is called *inseparable* if it cannot be partitioned into two separated subsets. For example, the intervals $[0,0.3]$, $[0.4,1]$ are separated and each of them is inseparable in the set of real numbers $\mathcal{R}$. Emptyset $\emptyset$ is defined as inseparable. Obviously, any subset $\mathbf{Pr}$ can be partitioned into a set of inseparable sets. Formally, there exists $\mathbf{Pr}_1,\mathbf{Pr}_2,\ldots$, such that every $\mathbf{Pr}_i$ is inseparable, $\mathbf{Pr}_i\cap\mathbf{Pr}_j=\emptyset$ $(i\neq j)$, and $\mathbf{Pr}=\bigcup_i\mathbf{Pr}_i$.

Definition 7 Let $(\psi|\phi)$ be a conditional event and $\mathbf{Pr}$ be a subset of $\mathbf{Pr}_{\Phi}$. Suppose that $\mathbf{Pr}$ is inseparable, and $\mathbf{Pr}\models_{tight}(\psi|\phi)[l,u]$ [3]. We define $\delta^{ub}: 2^{\mathbf{Pr}_{\Phi}}\times\mathcal{F}\rightarrow[0,1]$ and

[2] The closure of a set S is the smallest *closed* set containing S.

[3] It is defined that $\mathbf{Pr}\models_{tight}(\psi|\phi)[l,u]$ iff $l=\min_{Pr\in\mathbf{Pr}}Pr(\psi|\phi)$ and $u=\max_{Pr\in\mathbf{Pr}}Pr(\psi|\phi)$. In order to complete the definition, if $\forall Pr\in\mathbf{Pr}$, $Pr(\phi)=0$ then it is defined that $\mathbf{Pr}\models_{tight}(\psi|\phi)[1,0]$.

$\delta^{lb} : 2^{\mathbf{Pr}_\Phi} \times \mathcal{F} \to [0,1]$ as

$$\delta^{ub}(\mathbf{Pr}, (\psi|\phi)) = \min_{\substack{Pr \in \mathbf{Pr} \\ Pr \models (\psi|\phi)[u,u]}} KL(Pr||Pr_{unif})$$

$$\delta^{lb}(\mathbf{Pr}, (\psi|\phi)) = \min_{\substack{Pr \in \mathbf{Pr} \\ Pr \models (\psi|\phi)[l,l]}} KL(Pr||Pr_{unif})$$

where Pr_{unif} is the uniform distribution on $\mathcal{I}_\Phi$.

For simplicity, we use $\delta^{ub}_{\mathbf{Pr}}(\psi|\phi)$ to denote $\delta^{ub}(\mathbf{Pr}, (\psi|\phi))$ and use $\delta^{lb}_{\mathbf{Pr}}(\psi|\phi)$ to denote $\delta^{lb}(\mathbf{Pr}, (\psi|\phi))$.

Value $\delta^{ub}_{\mathbf{Pr}}(\psi|\phi)$ (resp. $\delta^{lb}_{\mathbf{Pr}}(\psi|\phi)$) measures how much additional information needs to be added to the uniform distribution in order to infer the upper (resp. lower) bound of the conditional event $(\psi|\phi)$ given subset $\mathbf{Pr}$.

Definition 8 Let $\mathbf{Pr}$ be an inseparable subset of $\mathbf{Pr}_\Phi$ and $(\psi|\phi)$ be a conditional event defined on Φ. Let $\mathbf{Pr}_{IS}$ contain all the inseparable subsets of $\mathbf{Pr}_\Phi$. We define $\vartheta_{\psi|\phi} : \mathbf{Pr}_{IS} \mapsto [0,1]$ as $\vartheta_{\psi|\phi}(\mathbf{Pr}) = \text{sign}(p_{me} - u) * \delta^{ub}_{\mathbf{Pr}}(\psi|\phi) - \text{sign}(p_{me} - l) * \delta^{lb}_{\mathbf{Pr}}(\psi|\phi)$, where $p_{me} = Pr_{unif}(\psi|\phi)$. Here, $\text{sign}(x) = 1$ if $x \geq 0$ and $\text{sign}(x) = -1$ otherwise.

Let σ_Φ denote the smallest collection such that σ_Φ contains all the inseparable subsets of $\mathbf{Pr}_\Phi$ and it is closed under complement and countable unions of its members. Therefore, $\langle \mathbf{Pr}_\Phi, \sigma_\Phi \rangle$ is a *measurable space* over the set $\mathbf{Pr}_\Phi$. Obviously, $\mathbf{Pr}_\Phi \in \sigma_\Phi$, and if $\mathbf{Pr} = \{Pr \mid Pr \models P\}$ for any PLP P, then $\mathbf{Pr} \in \sigma_\Phi$.

We extend function $\vartheta_{\psi|\phi}$ to any subset of σ_Φ.

Definition 9 Let $\mathbf{Pr}$ be a subset of σ_Φ and $(\psi|\phi)$ be a conditional event defined on Φ. Define $\vartheta_{\psi|\phi} : \sigma_\Phi \mapsto [0,1]$ as $\vartheta_{\psi|\phi}(\mathbf{Pr}) = \sum_{\mathbf{Pr_i} \in \mathfrak{P}} \vartheta_{(\psi|\phi)} \mathbf{Pr_i}$ where $\mathfrak{P}$ is a partition of $\mathbf{Pr}$ such that each element of $\mathfrak{P}$ is inseparable.

Informally, value $\vartheta_{(\psi|\phi)}(\mathbf{Pr})$ measures how *wide* the probability distributions in $\mathbf{Pr}$ are when inferring ψ given ϕ. For example, when all the distributions in $\mathbf{Pr}$ assign the same probability for the conditional event $(\psi|\phi)$, then the set $\mathbf{Pr}$ is acting like a single distribution when inferring ψ given ϕ, and in this case, $\mathbf{Pr}$ has *width* 0 for inferring ψ given ϕ.

From the definition, we know that function $\vartheta_{(\psi|\phi)}$ is a *measure*. Since it is a measure, we can define a probability distribution based on it, and we show that this probability distribution can be used as an instantiation of ignorance in the next subsection.

It is worth noting that since the set of all probabilistic models of a PLP is a convex set and thus is inseparable, we can use $\vartheta_{(\psi|\phi)}$ to measure the probabilistic models of a PLP. We will discuss this further in the next subsection.

4.2 Instantiation of ignorance

Definition 10 Let P be a PLP and $(\psi|\phi)$ be a conditional event. Then a KL-divergence based ignorance denoted as $\mathsf{IG}^{KL}_P(\psi|\phi)$ is defined as

$$\mathsf{IG}^{KL}_P(\psi|\phi) = \vartheta_{(\psi|\phi)}(\mathbf{Pr})/\vartheta_{(\psi|\phi)}(\mathbf{Pr}_\Phi)$$

where $\mathbf{Pr} = \{Pr \mid Pr \models P\}$.

Since $\vartheta_{\psi|\phi}$ is a measure, IG_P^{KL} is a uniform probability distribution. Thus, $\mathsf{IG}_P^{KL}(\psi|\phi)$ is the probability that a randomly selected probability distribution from set $\mathbf{Pr}_{\Phi}$ assigns $\psi|\phi$ a probability value that is in the interval $[l,u]$, where $P \models_{tight} (\psi|\phi)[l,u]$. If this probability is close to 1, then reasoning with P is similar to reasoning with an empty PLP; when it is close to 0, it indicates that a tighter bound for $(\psi|\phi)$ can be inferred from P.

In the above definition, $\mathsf{IG}_P^{KL}(\psi|\phi)$ is undefined when $P \models \bot \leftarrow \phi$, since $\vartheta_{(\psi|\phi)}(\cdot)$ is undefined. So, we extend IG_P^{KL} as $\mathsf{IG}_P^{KL}(\psi|\phi) = 0$ if $P \models \bot \leftarrow \phi$.

Proposition 7 *The measure IG^{KL} satisfies the properties given in Definition 3.*

Proof Since $\mathsf{IG}_P^{KL}(\cdot)$ is a second-order probability function, $\mathsf{IG}_P^{KL}(\psi|\phi) \in [0,1]$.

If $P \models \bot \leftarrow \phi$, then $\mathsf{IG}_P^{KL}(\psi|\phi) = 0$ by definition. Assume that $P \not\models \bot \leftarrow \phi$. If $\mathsf{IG}_P^{KL}(\psi|\phi) = 0$, then $\vartheta_{(\psi|\phi)}(\mathbf{Pr}) = 0$, and so, $P \models (\psi|\phi)[l,l]$ for some $l \in [0,1]$. Therefore, IG^{KL} satisfies the property Preciseness.

From the definition of IG^{KL}, it is easy to see that $\mathsf{IG}_{\emptyset}^{KL}(\psi|\phi) = 1$.

On the other hand, if $\mathsf{IG}_P^{KL}(\psi|\phi) = 1$ then $P \models (\psi|\phi)[l,u]$ iff $\emptyset \models (\psi|\phi)[l,u]$. Since $\emptyset \models (\psi|\phi)[0,1]$, $P \models (\psi|\phi)[0,1]$.

Suppose that PLP P and another PLP P' do not contain common syntaxes, i.e.$\Phi \cap \Phi' = \emptyset$. Let $Pr' \models P \cup P'$. Since $\Phi \cap \Phi' = \emptyset$, we can construct another probability distribution Pr as $Pr(I) = \sum_{I \subseteq I'} Pr'(I')$ for all $I \subseteq \mathcal{I}_{\Phi}$. Obviously, $Pr \models P$. By Definitions 7 and 8, two probability distributions that satisfy P' are chosen to calculate $\delta^{lb}(\mathbf{Pr},(\psi|\phi))$, $\delta^{ub}(\mathbf{Pr},(\psi|\phi))$, $\vartheta_{\psi|\phi}(\mathbf{Pr})$, and thus $\mathsf{IG}_{P' \cup P}^{KL}(\psi|\phi)$. From them, two probability distributions are constructed and they satisfy P, and give the conditional event $(\psi|\phi)$ the lower bound and upper bound respectively. With these two distributions, we can calculate $\mathsf{IG}_P^{KL}(\psi|\phi)$, as well as $\mathsf{IG}_{P' \cup P}^{KL}(\psi|\phi)$. □

Proposition 8 *Let P be a PLP and $(\psi|\phi)$ be a conditional event. Suppose that $P \models_{tight} (\psi|\phi)[l,u]$ and $pm = me[P](\psi|\phi)$. Then $IG_P^{KL}(\psi|\phi) = IG_{P_1}^{KL}(\psi|\phi) + IG_{P_2}^{KL}(\psi|\phi)$, where $P_1 = P \cup \{(\psi|\phi)[pm,u]\}, P_2 = P \cup \{(\psi|\phi)[l,pm]\}$.*

Proof Let $\mathbf{Pr} = \{Pr \mid Pr \models P\}$, $\mathbf{Pr}_1 = \{Pr \models P_1\}$, and $\mathbf{Pr}_2 = \{Pr \models P_2\}$. Then $\mathbf{Pr} = \mathbf{Pr}_1 \cup \mathbf{Pr}_2$. Since $\vartheta_{(\psi|\phi)}$ is a measure, $\vartheta_{(\psi|\phi)}(\mathbf{Pr}) = \vartheta_{(\psi|\phi)}(\mathbf{Pr}_1) + \vartheta_{(\psi|\phi)}(\mathbf{Pr}_2)$. Thus, $IG_P^{KL}(\psi|\phi) = IG_{P_1}^{KL}(\psi|\phi) + IG_{P_2}^{KL}(\psi|\phi)$. □

This proposition says that the ignorance of a PLP about a conditional event is the sum of the ignorance of lacking knowledge supporting probability distributions above and below the maximum entropy probability. The ignorance can also be calculated according to the maximum entropy as below.

Proposition 9 *Let P be a PLP and $(\psi|\phi)$ be a conditional event. Suppose that $P \models_{tight} (\psi|\phi)[l,u]$, $\mathbf{Pr} = \{Pr \mid Pr \models P\}$, and $\emptyset \models_{tight}^{me} (\psi|\phi)[p_{me},p_{me}]$, then $\vartheta_{(\psi|\phi)}(\mathbf{Pr}) = \mathrm{sign}(u - p_{me}) * \max_{Pr \models P^u} H(Pr) - \mathrm{sign}(l - p_{me}) * \max_{Pr \models P^l} H(Pr)$ where $P^u = P \cup \{(\psi|\phi)[u,u]\}$ and $P^l = P \cup \{(\psi|\phi)[l,l]\}$.*

Table 2 KL Ignorance values for Example 7

Conditional Event	Ignorance value
$(haveLegs(tweety)\|penguin(tweety))$	0.0065
$(haveLegs(robin)\|bird(robin))$	0.0027
$(fly(robin)\|bird(robin) \wedge red(robin))$	0.0545
$(fly(robin)\|bird(robin))$	0.1555
$(fly(tweety)\|penguin(tweety))$	0.0068

Proof Assume that $P_1 = P \cup \{(\psi|\phi)[p_{me}, u]\}, P_2 = P \cup \{(\psi|\phi)[l, p_{me}]\}$. Suppose that $\mathbf{Pr}_1 = \{Pr \mid Pr \models P_1\}$ and $\mathbf{Pr}_2 = \{Pr \mid Pr \models P_2\}$. Then we have

$$\begin{aligned}
\vartheta_{(\psi|\phi)}(\mathbf{Pr}_1) &= (u - sign(p_{me}))\delta^{ub}(\mathbf{Pr}_1, (\psi|\phi)) - (sign(p_{me} - p_{me})\delta^{lb}(\mathbf{Pr}_1, (\psi|\phi)) \\
&= (u - sign(p_{me}))\delta^{ub}(\mathbf{Pr}_1, (\psi|\phi)) \\
&= (u - sign(p_{me})) \min_{\substack{Pr \in \mathbf{Pr}_1 \\ Pr \models (\psi|\phi)[u,u]}} KL(Pr||Pr_{unif}) \\
&= (u - sign(p_{me})) \max_{Pr \models P^u} H(Pr) \\
\vartheta_{(\psi|\phi)}(\mathbf{Pr}_2) &= (p_{me} - sign(p_{me})\delta^{ub}(\mathbf{Pr}_1, (\psi|\phi)) - (sign(p_{me}) - l)\delta^{lb}(\mathbf{Pr}_1, (\psi|\phi)) \\
&= (sign(p_{me}) - l)\delta^{lb}(\mathbf{Pr}_2, (\psi|\phi))) \\
&= (sign(p_{me}) - l) \min_{\substack{Pr \in \mathbf{Pr}_2 \\ Pr \models (\psi|\phi)[l,l]}} KL(Pr||Pr_{unif}) \\
&= (sign(p_{me}) - l) \max_{Pr \models P^l} H(Pr)
\end{aligned}$$

So, $\vartheta_{(\psi|\phi)}(\mathbf{Pr}) = \vartheta_{(\psi|\phi)}(\mathbf{Pr}_1) + \vartheta_{(\psi|\phi)}(\mathbf{Pr}_2) = (u - sign(p_{me})) \max_{Pr \models P^u} H(Pr) + (sign(p_{me}) - l) \max_{Pr \models P^l} H(Pr)$. □

Example 7 Let PLP P be

$$P = \left\{ \begin{array}{l} (fly(X)|bird(X))[0.9, 0.98], \\ (bird(X)|penguin(X))[1, 1], \\ (fly(X)|penguin(X))[0, 0.05], \\ (haveLegs(X)|bird(X))[0.98, 1] \end{array} \right\}$$

Consider the following conditional events:

$$\begin{aligned}
E_1 &= (haveLegs(tweety)|penguin(tweety)) \\
E_2 &= (haveLegs(robin)|bird(robin)) \\
E_3 &= (fly(robin)|bird(robin) \wedge red(robin)) \\
E_4 &= (fly(robin)|bird(robin)) \\
E_5 &= (fly(tweety)|penguin(tweety))
\end{aligned}$$

By calculating the KL-ignorance value for conditional events $E_1, \ldots, E_5$, we have that $\mathsf{IG}_P^{KL}(E_1) = 0.0065$, $\mathsf{IG}_P^{KL}(E_2) = 0.0027$, etc., as shown in Table 2. From the table, we can see that $\mathsf{IG}_P^{KL}(E_1) > \mathsf{IG}_P^{KL}(E_2)$, due to the ignorance about whether the special birds, penguins, are typical birds or not (with respect to the property of having legs).

Comparing these ignorance values, we also see that $\mathsf{IG}_P^{KL}(E_3) < \mathsf{IG}_P^{KL}(E_4)$. It looks counter-intuitive, since being red is irrelevant to being able to fly or not. But such irrelevance is based on our common knowledge and not represented in P. Based on P, we actually do not know whether being red is irrelevant to the ability of flying. To be more precise, for example, if the predicate red is replaced by $abnormal$, then whether an abnormal bird can fly is unclear. Consider that the difference between red and $abnormal$ comes from our common knowledge and not from the PLP P, it is natural that the ignorance value for E_3 is smaller than that for E_4.

4.3 Instantiation of satisfaction function

Given a PLP P, a set of probability distributions can be induced such that $\mathbf{Pr} = \{Pr \mid Pr \models P\}$ and a unique probability distribution $me[P]$ in the set that has the maximum entropy can be determined. In $\mathbf{Pr}$, some distributions are likely to be the actual probability distribution. However, due to the lack of information, we do not know which one is the actual probability distribution. Based on the maximum entropy principle, $me[P]$ is assumed to be the most likely one, and the probability $me[P](\psi|\phi)$ is assumed to be the most likely probability for the event $(\psi|\phi)$. Intuitively, the probability value that is closer to $me[P](\psi|\phi)$ is more likely to be the actual probability of $(\psi|\phi)$. Based on this, an interval that contains values closer to $me[P](\psi|\phi)$ are more likely to contain the actual probability of $(\psi|\phi)$. Of course, a loose interval is always more likely to contain the actual probability of $(\psi|\phi)$ than a tighter interval.

From the KL-divergence, we can define how close a value is to $me[P]$ as:

$$\begin{aligned}
\nu_{P,(\psi|\phi)}^{pos}(v) &= \min_{Pr\models P, Pr(\psi|\phi)=v} KL(Pr||me),\\
&\quad \text{where } v \geq me[P]\\
\nu_{P,(\psi|\phi)}^{neg}(v) &= \min_{Pr\models P, Pr(\psi|\phi)=v} KL(Pr||me),\\
&\quad \text{where } v \leq me[P]\\
dis_{P,(\psi|\phi)}^{pos}(u,v) &= |\nu_{P,(\psi|\phi)}^{pos}(u) - \nu_{P,(\psi|\phi)}^{pos}(v)|\\
dis_{P,(\psi|\phi)}^{neg}(u,v) &= |\nu_{P,(\psi|\phi)}^{neg}(u) - \nu_{P,(\psi|\phi)}^{neg}(v)|
\end{aligned}$$

Let dis be $dis_{P,(\psi|\phi)}^{pos}$ (resp. $dis_{P,(\psi|\phi)}^{neg}$). It is easy to see that dis is a distance function on $\mathcal{R}^{[pme,u]}$ (resp. $\mathcal{R}^{[l,pme]}$), where $P \models_{tight} (\psi|\phi)[l,u]$, $pme = me[P](\psi|\phi)$ and $\mathcal{R}^{[a,b]} = \{x \mid x \in [a,b], x \in \mathcal{R}\}$, i.e. dis satisfies the following:

- $dis(u,v) \geq 0$
- $dis(u,v) = 0$ iff $u = v$
- $dis(u,v) = dis(v,u)$
- $dis(u,v) \leq dis(u,x) + dis(x,v)$

Again, from the distance functions $dis_{P,(\psi|\phi)}^{pos}$ and $dis_{P,(\psi|\phi)}^{neg}$, a probability distribution can be defined. So, by KL-divergence, the possible probabilities of a conditional event $(\psi|\phi)$ are measurable. Assume that every probability is equally probable, then the (second order) probability that the actual (first order) probability of $(\psi|\phi)$ falls in an interval $[a,b]$ is the *length* of $[a,b]$ divided by the *length* of $[l,u]$, where $P \models_{tight} (\psi|\phi)[l,u]$, according to the distance function $dis^{pos}_{P,(\psi|\phi)}$ and $dis^{neg}_{P,(\psi|\phi)}$. Formally, we define the degree of satisfaction as this second order probability.

Definition 11 Let P be a PLP and $(\psi|\phi)$ be a conditional event. Suppose that $P \models_{tight} (\psi|\phi)[l,u]$ and $P \models^{me}_{tight} (\psi|\phi)[p_{me}, p_{me}]$, then we have that:

$$\mathsf{SAT}^{KL}_P((\psi|\phi)[a,b]) = \\ \begin{cases} 0.5 * (\frac{dis^{pos}_{P,(\psi|\phi)}(p_{me},\min(u,b))}{dis^{pos}_{P,(\psi|\phi)}(p_{me},u)} + \frac{dis^{neg}_{P,(\psi|\phi)}(p_{me},\max(a,l))}{dis^{neg}_{P,(\psi|\phi)}(p_{me},l)}), & \text{if } p_{me} \in [a,b] \\ 0, & \text{otherwise} \end{cases}$$

Proposition 10 *Let P be a PLP, then the function SAT^{KL}_P defined in Definition 11 satisfies all the postulates in Definition 4.*

Proof Suppose that $\mu = (\psi|\phi)[l,u]$ is a probabilistic formula.

- **Reflexivity** If $P \models_{tight} \mu$, then
  $\mathsf{SAT}_P(\mu) = 0.5 * (\frac{dis^{pos}_{P,(\psi|\phi)}(p_{me},\min(u,u))}{dis^{pos}_{P,(\psi|\phi)}(p_{me},u)} + \frac{dis^{neg}_{P,(\psi|\phi)}(p_{me},\max(l,l))}{dis^{neg}_{P,(\psi|\phi)}(p_{me},l)}) = 1$
  On the other hand, assume that $P \models_{tight} (\psi|\phi)[l',u']$.
  If $\mathsf{SAT}_P(\mu) = 1$ then $dis^{pos}_{P,(\psi|\phi)}(p_{me}, \min(u,u')) = dis^{pos}_{P,(\psi|\phi)}(p_{me}, u)$ and $dis^{pos}_{P,(\psi|\phi)}(p_{me}, \max(l,l')) = dis^{pos}_{P,(\psi|\phi)}(p_{me}, l)$. Then $u' = u$ and $l' = l$, so $P \models (\psi|\phi)[l,u]$.
- **Rationality** If $P \cup \{\mu\}$ is unsatisfiable, and $P \models_{tight} (\psi|\phi)[l',u']$, then $[l',u'] \cap [l,u] = \emptyset$. Suppose that $P \models^{me}_{tight} (\psi|\phi)[p_{me}, p_{me}]$, then $p_{me} \in [l',u']$ and $p_{me} \notin [l,u]$.
  Therefore $\mathsf{SAT}_P(\mu) = 0$.
- **Monotonicity** Suppose that $[l',u'] \subseteq [l,u]$, and $P \models_{tight} (\psi|\phi)[l_0,u_0]$,
  then $dis^{pos}_{P,(\psi|\phi)}(p_{me}, \min(u',u_0)) \leq dis^{pos}_{P,(\psi|\phi)}(p_{me}, \min(u,u_0))$ and
  $dis^{pos}_{P,(\psi|\phi)}(p_{me}, \max(l',l_0)) \leq dis^{pos}_{P,(\psi|\phi)}(p_{me}, \max(l,l_0))$.
  Therefore, $\mathsf{SAT}(P,\mu) \geq \mathsf{SAT}(P,(\psi|\phi)[l',u'])$.
  Similarly, If $[l',u'] \subset [l,u]$ then $\mathsf{SAT}(P,\mu) > \mathsf{SAT}(P,(\psi|\phi)[l',u'])$.
- **Cautious Monotonicity** Assume that $P \models_{tight} (\psi|\phi)[l_0,u_0]$, $P' \models_{tight} (\psi|\phi)[l_1,u_1]$.
  Let $P' = P \cup \{(\psi|\phi)[l',u']$ for some $[l',u']$ s.t. $P \models^{me} (\psi|\phi)[l',u']$. It is easy to verify that $me[P] \models P'$. Suppose that $\mathbf{Pr}' = \{Pr \mid Pr \models P'\}$ and $\mathbf{Pr} = \{Pr \mid Pr \models P\}$, then $\mathbf{Pr}' \subseteq \mathbf{Pr}$. Since $H(me[P]) > H(Pr)$ for all $Pr \in \mathbf{Pr}$, $H(me[P]) > H(Pr)$ for all $Pr \in \mathbf{Pr}'$, thus $me[P] = me[P']$. By the definition of dis^{pos} and dis^{neg}, we have that $dis^{pos}_{P,(\psi|\phi)}(p_{me}, a) = dis^{pos}_{P',(\psi|\phi)}(p_{me}, a)$ and $dis^{neg}_{P,(\psi|\phi)}(p_{me}, b)) = dis^{neg}_{P',(\psi|\phi)}(p_{me}, b))$ for any a, b s.t. $[a,b] \subseteq [l_1,u_1]$.
  If $p_{me} \notin [l,u]$ then $\mathsf{SAT}^{KL}_P(\mu) = \mathsf{SAT}^{KL}_{P'}(\mu) = 0$. Now, assume that $p_{me} \in [l,u]$, so,
  $\mathsf{SAT}^{KL}_P((\psi|\phi)[l,u]) = 0.5 * (\frac{dis^{pos}_{P,(\psi|\phi)}(p_{me},\min(u,u_0))}{dis^{pos}_{P,(\psi|\phi)}(p_{me},u_0)} + \frac{dis^{neg}_{P,(\psi|\phi)}(p_{me},\max(l,l_0))}{dis^{neg}_{P,(\psi|\phi)}(p_{me},l_0)})$
  $\mathsf{SAT}^{KL}_{P'}((\psi|\phi)[l,u]) = 0.5 * (\frac{dis^{pos}_{P',(\psi|\phi)}(p_{me},\min(u,u_1))}{dis^{pos}_{P',(\psi|\phi)}(p_{me},u_1)} + \frac{dis^{neg}_{P',(\psi|\phi)}(p_{me},\max(l,l_1))}{dis^{neg}_{P',(\psi|\phi)}(p_{me},l_1)})$
  We have $dis^{pos}_{P',(\psi|\phi)}(p_{me}, u_1) = dis^{pos}_{P,(\psi|\phi)}(p_{me}, u_1) \leq dis^{pos}_{P,(\psi|\phi)}(p_{me}, u_0)$,
  $dis^{neg}_{P',(\psi|\phi)}(p_{me}, l_1) = dis^{neg}_{P,(\psi|\phi)}(p_{me}, l_1) \leq dis^{neg}_{P,(\psi|\phi)}(p_{me}, l_0)$.
  Since $P' \models P$, we have that $u_1 \leq u_0$ and $l_1 \geq l_0$.
  1) If $u \leq u_1$, then $\min(u,u_0) = \min(u,u_1) = u$.
  Since $dis^{pos}_{P,(\psi|\phi)}(p_{me}, u) = dis^{pos}_{P',(\psi|\phi)}(p_{me}, u)$, we have that
  $\frac{dis^{pos}_{P',(\psi|\phi)}(p_{me},\min(u,u_1))}{dis^{pos}_{P',(\psi|\phi)}(p_{me},u_1)} \geq \frac{dis^{pos}_{P,(\psi|\phi)}(p_{me},\min(u,u_0))}{dis^{pos}_{P,(\psi|\phi)}(p_{me},u_0)}$

2) If $u_1 \leq u \leq u_0$, then $\frac{dis^{pos}_{P',(\psi|\phi)}(p_{me}, \min(u,u_1))}{dis^{pos}_{P',(\psi|\phi)}(p_{me}, u_1)} = 1$ and $\frac{dis^{pos}_{P,(\psi|\phi)}(p_{me}, \min(u,u_0))}{dis^{pos}_{P,(\psi|\phi)}(p_{me}, u_0)} \leq 1$.

3) If $u \geq u_0$, then $\frac{dis^{pos}_{P',(\psi|\phi)}(p_{me}, \min(u,u_1))}{dis^{pos}_{P',(\psi|\phi)}(p_{me}, u_1)} = 1$ and $\frac{dis^{pos}_{P,(\psi|\phi)}(p_{me}, \min(u,u_0))}{dis^{pos}_{P,(\psi|\phi)}(p_{me}, u_0)} = 1$.

Concluding the above, we have that $\frac{dis^{pos}_{P',(\psi|\phi)}(p_{me}, \min(u,u_1))}{dis^{pos}_{P',(\psi|\phi)}(p_{me}, u_1)} \geq \frac{dis^{pos}_{P,(\psi|\phi)}(p_{me}, \min(u,u_0))}{dis^{pos}_{P,(\psi|\phi)}(p_{me}, u_0)}$.

By applying the same analysis on l, we know $\mathsf{SAT}^{KL}_{P'}((\psi|\phi)[l,u]) \geq \mathsf{SAT}^{KL}_{P}((\psi|\phi)[l,u])$

- The fact that SAT^{KL} satisfies the conditions in Definition 4 follows from the definition of SAT^{KL}.

□

Example 8 Let P_1 and P_2 be as given in Example 4. Then

$$\begin{aligned}
&\mathsf{SAT}^{KL}_{P_1}((headUp(coin)|toss(coin))[0, 0.5]) \\
&= \mathsf{SAT}^{KL}_{P_1}((headUp(coin)|toss(coin))[0.5, 1]) \\
&= 0.5.
\end{aligned}$$

So, in P_1, we do not know whether head-ip is more probable than tail-up. However
$\mathsf{SAT}^{KL}_{P_2}((headUp(coin)|toss(coin))[0.5, 0.5]) = 1$,
we are sure that the probability of head-up and tail-up is the same.

4.4 Logic properties

The postulates Right Weakening (RW), Reflexivity (Ref), Left Logical Equivalence (LLE), Cautious Cut, Cautious Monotonicity (CM), and Or proposed by Kraus et al. [KLM90], are commonly regarded as being particularly desirable for any reasonable notion of nonmonotonic entailment $\Vdash$.

In [KIL04], these postulates are reformulated for probabilistic reasoning:

RW. If $(\phi|\top)[l,u] \Vdash (\psi|\top)[l',u']$, and $P \Vdash (\phi|\phi')[l,u]$,
then $P \Vdash (\psi|\phi')[l',u']$.

Ref. $P \Vdash (\phi|\phi)[1,1]$.

LLE. If $\phi \equiv \phi'$ is logically valid, then $P \Vdash (\psi|\phi)[l,u]$ iff $P \Vdash (\psi|\phi')[l,u]$.

Cautious Cut. If $P \Vdash (\phi'|\phi)[1,1]$ and $P \Vdash (\psi|\phi \wedge \phi')[l,u]$, then $P \Vdash (\psi|\phi)[l,u]$.

CM. If $P \Vdash (\phi'|\phi)[1,1]$ and $P \Vdash (\psi|\phi)[l,u]$, then $P \Vdash (\psi|\phi \wedge \phi')[l,u]$.

Theorem 4 ([KIL04]) *The consequence relations $\models$ and $\models^{me}$ satisfy RW. Ref. LLE. Cut. CM. for all PLPs, all ground events ψ, ϕ, ϕ' and all $l, u, l', u' \in [0,1]$.*

In Section 3, we have shown that the consequence relations $\models^{SAT=1}$ and $\models^{SAT>0}$ coincide with $\models$ and $\models^{me}$ respectively. Below, we will show that $\models^{SAT\geq a}$ with $a > 0$ also have some nice properties.

Proposition 11 *Let P be a PLP, ψ, ϕ, ϕ' be ground events, and $l, u, l', u' \in [0,1]$. Suppose that the consequence relation $\models^{SAT\geq a}$ with $a > 0$ is defined from SAT^{KL}, then $\models^{SAT\geq a}$satisfies*

Ref. $P \models^{SAT\geq a} (\phi|\phi)[1,1]$.

LLE. *If* $\phi \equiv \phi'$ *is logically valid, then* $P \models^{SAT\geq a} (\psi|\phi)[l,u]$
iff $P \models^{SAT\geq a} (\psi|\phi')[l,u]$.

Cautious Cut. *If* $P \models (\phi'|\phi)[1,1]$ *and* $P \models^{SAT\geq a} (\psi|\phi \wedge \phi')[l,u]$,
then $P \models^{SAT\geq a} (\psi|\phi)[l,u]$.

CM. *If* $P \models (\phi'|\phi)[1,1]$ *and* $P \models^{SAT\geq a} (\psi|\phi)[l,u]$,
then $P \models^{SAT\geq a} (\psi|\phi \wedge \phi')[l,u]$.

Proof Property Ref directly follows the Reflexivity property of SAT, since $P \models_{thight} (\phi|\phi)[1,1]$. Property LLE follows the fact that our definition of SAT^{KL} is syntax independent. Property Cautious Cut and property CM hold for the same reason as LLE holds, since $\phi \wedge \phi' \equiv \phi$ under P when $P \models (\phi'|\phi)[1,1]$. □

The consequence relation $\models^{SAT\geq a}$ does not satisfy the reformulated RW, i.e., from the fact that $(\phi|\top)[l,u] \models^{SAT\geq a} (\psi|\top)[l',u']$ and $P \models^{SAT\geq a} (\phi|\phi')[l,u]$, we cannot draw the conclusion that $P \models^{SAT\geq a} (\psi|\phi')[l,u]$ in general.

These properties indicate that the consequence relation $\models^{SAT\geq a}$ is plausible for nonmonotonic reasoning.

5 Examples

In this section, we illustrate the usefulness of our framework with some examples.

Example 9 Let P be a PLP as given in Example 11. In our framework, we calculate the KL-ignorance and KL-satisfaction for our queries. We have $\mathsf{IG}^{KL}_{(fly(t)|magpie(t))}(P) = 0.11$ and $\mathsf{IG}^{KL}_{(fly(t)|sickMagpie(t))}(P) = 0.0283$. This indicates that P is more useful to infer the proportion of magpies that can fly than to infer the proportion of sick magpies that can fly. We also have that $\mathsf{SAT}^{KL}_{P}((fly(t)|magpie(t))[0.8,1]) = 0.58$, $\mathsf{SAT}^{KL}_{P}((fly(t)|sickMagpie(t))[0.8,1]) = 0.53$. By comparing these KL degrees of satisfaction, we know that magpies are more likely to fly than sick magpies.

Example 10 Let P be as given in Example 7. Consider a query

$$?(haveLegs(tweety)|penguin(tweety))[l,u]$$

with different values of l and u, we can calculate its KL degree of satisfaction as shown in Table 3. In the table, we can see that

$$\mathsf{SAT}^{KL}_{P}((haveLegs(tweety)|penguin(tweety))[0.1,1]) = 0.994,$$

$$\mathsf{SAT}^{KL}_{P}((haveLegs(tweety)|penguin(tweety))[0.2,1]) = 0.987, etc$$

So the degree of satisfaction of the query decreases as the bound becomes tighter.

It is worth noting that $P \models^{me}_{tight} (haveLegs(tweety)|penguin(tweety))[0.98,0.98]$. However, the degree of satisfaction of $(haveLegs(tweety)|penguin(tweety))[0.97,0.99]$ is only 0.081, which indicates that the probability 0.98 is not fully acceptable. This is because we do not know whether penguins are typical birds (with respect to the property of having legs).

On the contrary, we can accept that $(haveLegs(tweety)|penguin(tweety))[0.8,1]$, which means that we can also infer that $tweety$ has legs given that it is a penguin. Although we are

Table 3 KL degree of a satisfaction of query $?(haveLegs(tweety)|penguin(tweety))[l,u]$ with different values of l and u

Query bound [l,u]	degree of satisfaction SAT_P^{KL}
$[0.0, 1]$	1
$[0.1, 1]$	0.994
$[0.2, 1]$	0.987
$[0.3, 1]$	0.977
$[0.4, 1]$	0.965
$[0.5, 1]$	0.948
$[0.6, 1]$	0.925
$[0.7, 1]$	0.890
$[0.8, 1]$	0.829
$[0.9, 1]$	0.687
$[0.97, 0.99]$	0.081

When querying $?(haveLegs(tweety)|penguin(tweety))$, we also get that
$\mathsf{IG}_P^{KL}(haveLegs(tweety)|penguin(tweety)) = 0.0065$
$P \models_{tight} (haveLegs(tweety)|penguin(tweety))[0,1]$
$P \models_{tight}^{me} (haveLegs(tweety)|penguin(tweety))[0.98, 0.98]$

not entirely sure about this, it is more reliable than believing $97\% - 99\%$ magpies can fly (compare 0.829 to 0.081).

In this example, a user wants to know the lower bound for the probability that a penguin has legs. From the the non-informative interval $[0,1]$, the user can only know that it is possible that the probability that a penguin has leg is 0 (say, a penguin is possibly a special kind of bird), which is useless. By $\models_{maxLow}^{SAT \geq a}$, the user can infer a non-trivial lower bound for the query. Given $a = 0.8$, it is inferred that more than 80% penguin may have legs.

When adding knowledge $(haveLegs(X)|penguin(X))[0.5,1]$ into P, we get P' and $\mathsf{SAT}_{P'}^{KL}((haveLegs(tweety)|penguin(tweety))[0.8,1]) = 0.867$, which is bigger than 0.829. This fact indicates that new pieces of knowledge that support *a penguin has legs* are obtained.

□

Consider the inheritance problem. Intuitively, we expect that a subclass can inherit its superclass's attributes. But if we permit inheritance with exception, then a special subclass may lack the attributes that its superclass has. The more specific a subclass is, the more possible that it lacks the attributes that its superclass has.

Example 11 Let P be a PLP:

$$P = \left\{ \begin{array}{l} (fly(X)|bird(X))[0.9,1] \\ (bird(X)|magpie(X))[1,1] \\ (magpie(X)|sickmagpie(X))[1,1] \end{array} \right\}$$

When querying $?(fly(t)|magpie(t))$ and $?(fly(t)|sickmagpie(t))$, we get that
$P \models_{tight} (fly(t)|magpie(t))[0,1]$, $P \models_{tight}^{me} (fly(t)|magpie(t))[0.9, 0.9]$,
$P \models_{tight} (fly(t)|sickmagpie(t))[0,1]$, and $P \models_{tight}^{me} (fly(t)|sickMagpie(t))[0.9,0.9]$. Thus we can not differ magpies and sick magpies in their ability of flying, although sick magpies are more special than magpies, and therefore they are less likely to be able to fly than magpies. In contrast, in our framework, $\mathsf{SAT}_P^{KL}((fly(t)|magpie(t))[0.8,1]) = 0.58$, and $\mathsf{SAT}_P^{KL}((fly(t)|sickmagpie(t))[0.8,1]) = 0.53$. By comparing these KL degree of satisfaction, we know that magpies are more likely than sick magpies to fly. □

Example 12 (Route planning) [KIL04]. Assume that John wants to pick up Mary after she stops working. To do so, he must drive from his home to her office. Now, John has the following knowledge at hand: Given a road (ro) from R to S, the probability that he can reach (re) S from R without running into a traffic jam is greater than 0.7. Given a road in the south (so) of the town, this probability is even greater than 0.9. A friend just called him and gave him advice (ad) about some roads without any significant traffic. Clearly, if he can reach S from T and T from R, both without running into a traffic jam, then he can also reach S from R without running into a traffic jam. Furthermore, John has some concrete knowledge about the roads, the roads in the south of the town, and the roads that his friend was talking about. For example, he knows that there is a road from his home (h) to the university (u), from the university to the airport (a), and from the airport to Mary's office (o). Moreover, John believes that his friend was talking about the road from the university to the airport with a probability between 0.8 and 0.9 (he is not completely sure about it, though). The above and some other probabilistic knowledge is expressed by the following PLP P:

$$P = \left\{ \begin{array}{l} ro(h,u)[1,1], \\ ro(u,a)[1,1], \\ ro(a,o)[1,1], \\ ad(h,u)[1,1], \\ ad(u,a)[0.8,0.9], \\ so(a,o)[1,1], \\ (re(R,S)|ro(R,S))[0.7,1], \\ (re(R,S)|ro(R,S) \wedge so(R,S))[0.9,1], \\ (re(R,S)|ro(R,S) \wedge ad(R,S))[1,1], \\ (re(R,S)|re(R,T) \wedge re(T,S))[1,1] \end{array} \right\}$$

John wants to know the probability of him running into a traffic jam, which can be expressed by query: $Q_0 = ?(re(h,o)|\top)$.

In [KIL04], Q_0 can be answered by $P \models_{tight} (re(h,o)|\top)[0.7,1]$, and $P \models^{me}_{tight} (re(h,o)|\top)[0.93,0.93]$. John can either accept a noninformative bound $[0.7,1]$ or accept an unreliable precise probability 0.93, and no further reasoning can be done.

Using our method, we get that $\mathsf{IG}^{KL}_P(re(h,o)|\top) = 0.066$. The ignorance value $\mathsf{IG}^{KL}_P(re(h,o)|\top)$ indicates that the knowledge is reliable about $(re(h,o)|\top)$. However, the actual probability of $(re(h,o)|\top)$ may be still different from 0.93, since $\mathsf{IG}^{KL}_P(re(h,o)|\top) > 0$.

John is wondering whether he can reach Mary's office from his home, such that the probability of him running into a traffic jam is smaller than 0.10. This can be expressed by the following probabilistic query: $Q_1 = ?(re(h,o)|\top)[0.90,1]$. John is also wondering whether the probability of him running into a traffic jam is smaller than 0.10, if his friend was really talking about the road from the university to the airport. This can be expressed as a probabilistic query: $Q_2 = ?(re(h,o)|ad(u,a))[0.90,1]$.

In [KIL04], in the traditional probabilistic logic programming both Q_1 and Q_2 are given the answer "No"; by applying the maximum entropy principle Q_1 is given the answer "No" and Q_2 is given the answer "Yes". For Q_1 John will accept the answer "No", however, for Q_2, John may be confused and does not know which answer he should trust.

Using our method, we calculate the degrees of satisfaction of these two queries. For Q_1, $\mathsf{SAT}^{KL}_P(Q_1) = 0$, which means the bound $[0.9,1]$ does not contain the probability given by the maximum entropy principle, and thus John has no confidence that he can reach Mary's office on time. For Q_2, $\mathsf{SAT}^{KL}_P(Q_2) = 0.724$, the relative high value "0.724" can help John to decide whether he should set off to pick up Mary.

Table 4 Degrees of satisfaction for queries Q_1 and Q_2

Bound	$(re(h,o)\vert\top)$	Bound	$(re(h,o)\vert ad(u,a))$
$[0,1]$	1	$[0,1]$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[0.70,1]$	1	$[0.88,\ 1]$	1
$[0.75,1]$	0.785	$[0.897,1]$	0.75
$[0.80,1]$	0.658	$[0.922,1]$	0.60
$[0.86,1]$	0.500	$[0.94,\ 1]$	0.50
$[0.90,1]$	0.000		

Using our method, John gets an estimate of the probability that he can reach Mary's office from his home without running into a traffic jam. If it is a special day for him and Mary, he hopes that his estimate is more accurate, otherwise, he can tolerate a less accurate estimate. Formally, he needs to decide the threshold a for $\models^{SAT\geq a}_{maxLow}$. For example, for Q_2, he may set $a = 0.6$ for a normal day, and $a_I = 0.75$ for an important day. Therefore, he can infer that $P \models^{SAT\geq 0.6}_{maxLow}$ $(re(h,o)|ad(u,a))[0.922,1]$ and $P \models^{SAT\geq 0.75}_{maxLow}$ $(re(h,o)|ad(u,a))[0.897,1]$. If it is an ordinary day and the lowest probability is bigger than 0.90, then he can set off. On an important day, he will need to investigate more about the traffic (to decrease the ignorance of $(re(h,o)|ad(u,a))$) or he has to revise his plan, since $0.897 < 0.9$.

On the another hand, we also analyze the usefulness of the advice from his friend. By analyzing his friend's knowledge, we have $\mathsf{IG}^{KL}_P(re(h,o)|ad(u,a)) = 0.0184$. This means that his friend's advice is indeed useful, since this ignorance value is significantly smaller than $\mathsf{IG}^{KL}_P(re(h,o)|\top)$. So, John needs to call his friend to make sure that his friend is really talking about the road from the university to the airport.

The degrees of satisfaction for various intervals are given in Table 4. From the table, we can see that the degree of satisfaction decreases as the interval becomes tighter. This means that the second order probability that the actual probability of $(\psi|\phi)$ falls in $[l,u]$ is getting smaller. □

6 Implementation of Our Framework and A Case Study

6.1 Implementation

To efficiently return a query result given a PLP, we implemented the algorithms proposed in [KIL04,Luk01b] for reasoning with PLPs. Using these algorithms, a PLP can be translated into a linear or nonlinear optimization problem. We implemented these algorithms in Java and solved the underlying optimization problem using a component in Matlab.

In addition, we also implemented the calculation of ignorance and degree of satisfaction with the algorithms given below, KLIgnorance (Algorithm 1) and KLSatisfaction (Algorithm 3). These two algorithms rely on the algorithms provided in [KIL04,Luk01b] as well as the software Matlab to optimize a PLP.

In terms of complexity, our algorithms call Algorithm tight_0_concequence (Fig. 2.) and Algorithm tight_me_concequence (Fig. 3.) 1 or 2 times. It is stated in [KIL04] that the complexity for Algorithm tight_0_concequence (Fig. 2.) is FP^{NP}-complete and Algorithm tight_me_concequence (Fig. 3.) *fall outside the range of such standard complexity analysis (where the upper complexity bound is based on the existence of a polynomial-size probabilis-*

tic interpretation that involves only rational numbers), since the me-model of a probabilistic logic program P may involve irrational numbers.[KIL04]. The same difficulties exist for analyzing the complexity of our algorithms. Furthermore, our algorithms also rely on the computation of nonlinear optimization problems subject to linear constraints as that for Algorithm tight_me_concequence (Fig. 3.). Therefore, the complexity of our algorithms falls in the same level as that of Algorithm tight_me_concequence (Fig. 3.), which is intractable [KIL04].

6.2 A case study on breast cancer clinical trials

Usually clinical trials provide a huge amount of statistical data. From these statistical data, we can compare the efficiency of drugs or therapies for different groups of patients. In order to make use of these data, we need to represent the statistical knowledge formally, and to provide analyzing tools for using such knowledge to answer queries related to individuals (maybe with some facts about the individuals). For this purpose, we use PLP as the formal representation language. As we discussed in Section 1, PLP is chosen because of its expressive power for imprecise probabilistic knowledge and its reasoning efficiency. Also, statistical data can be regarded as probabilistic data that is guaranteed by the law of large numbers in the field of probability theory, so using PLP to model statistical data drawn from trials is theoretically valid.

6.2.1 Observation vs. a prior facts

In PLPs, we use ground formulas to state a prior facts from statistics, i.e., something that must be true (statistically) is regarded as a fact. These facts are treated differently from observations about individuals. Observing an event (such as the test result of a particular test) does not infer that the event would happen for sure. So, observations cannot be represented as formulas of the form $(\psi(a)|\top)[1,1]$ in a PLP, doing so implies that we know $\psi(a)$ as being true even before it is observed. In other words, taking $\psi(a)$ as a probabilistic event, we cannot predict if $\psi(a)$ is true or false before we observed it. In our framework, all observations are stored in a separate database (named OBS) rather than in a PLP containing statistical knowledge. When querying $(\psi|\phi)[l,u]$ on PLP P, this observation database OBS is automatically called, so querying $(\psi|\phi)[l,u]$ is equivalent to querying $(\psi|\phi \wedge \bigwedge OBS)[l,u]$ on P.

6.2.2 Background knowledge

From a clinical trial, only statistical data are explicitly provided. This knowledge alone is not sufficient for reasoning, some background knowledge is also necessary. In order to process data in a trial, the background knowledge needed can be categorized into three groups

- additional statistical knowledge (to trials data), which is explicitly represented typically by a table, such as a statistics about the death rate in a particular age group;
- meta knowledge for a trial, such as the principle for choosing the participants, which is represented explicitly or implicitly in a trial report;
- background knowledge related to the trial, which may be omitted in a trial report and is shared by many trials, such as age distribution, natural death rate, a prior estimation of a disease, etc..

Algorithm 1 KLIgnorance

```
Input: PLP $P$ and a ground query $Q = ?(\psi|\phi)$
Output: ignorance value for $Q$
if $P$ is unsatisfiable then
    return $1$;
end if
if $P \models_{tight} (\phi|\top)[0,0]$ then
    return $1$;
end if
Compute the tight bounds $[l,u]$ for $(\psi|\phi)$ by Algorithm Tight_0_Consequence in Fig. 2.
Compute the simplified PLP $D$ index sets $R$ and associate numbers $a_r$ and optimal solution $y_r^\star$ ($r \in R$)
    by Algorithm Tight_me_Consequence in Fig. 3.
Compute the optimal value $ig_{neg}$ of the optimization problem:
    $ig_{neg} = \max \left( -\sum_{r \in R} y_r^l (\log y_r^l - \log a_r) \right)$
    subject to: $y_r^l$ satisfies $LC(\top, D^l, R)$, where $D^l = D \cup \{(\psi|\phi)[l,l]\}$
Compute the optimal value $ig_{pos}$ of the optimization problem:
    $ig_{pos} = \max \left( -\sum_{r \in R} y_r^u (\log y_r^u - \log a_r) \right)$
    subject to: $y_r^u$ satisfies $LC(\top, D^u, R)$, where $D^u = D \cup \{(\psi|\phi)[u,u]\}$.
Compute optimal solution $y'_r$ ($r \in R$) for $P' = \emptyset$ by Algorithm Tight_me_Consequence in Fig. 3.
    $p_{me} := me[P'](\psi|\phi)$.
Compute the optimal value $ig'_{neg}$ of the optimization problem:
    $ig'_{neg} = \max \left( -\sum_{r \in R} y_r^l (\log y_r^l - \log a_r) \right)$
    subject to: $y_r^l$ satisfies $LC(\top, D_0^l, R)$, where $D_0^l = \{(\psi|\phi)[l,l]\}$
Compute the optimal value $ig'_{pos}$ of the optimization problem:
    $ig'_{pos} = \max \left( -\sum_{r \in R} y_r^u (\log y_r^u - \log a_r) \right)$
    subject to: $y_r^u$ satisfies $LC(\top, D_0^u, R)$, where $D_0^u = \{(\psi|\phi)[u,u]\}$.
if $p_{me} < u$ then
    $s_1 := 1$
else
    $s_1 := -1$;
end if
if $p_{me} > l$ then
    $s_2 := 1$;
else
    $s_2 := -1$;
end if
$ig := (s_1 * ig_{pos} + s_2 * ig_{neg}) / (ig'_{pos} + ig'_{neg})$
return $ig$
```

Algorithm 2 KLDivergence

```
Input: PLP $P$, $me[P]$, a conditional event $(\psi|\phi)$, and a probability value $v$.
Output: $kl = \min_{Pr \models P, Pr(\psi|\phi)=v} KL(Pr||me[P])$
The $me[P]$ is obtained from Algorithm 1, and is represented as $y^{me}$.
Compute the tight bounds $[l', u']$ for $(\psi|\phi)$ by Algorithm Tight_0_Consequence in Fig. 2.
if $v \notin [l', u']$ then
    return ERROR
end if
Compute the optimal value $kl$ of the optimization problem:
    $kl = \min \left( \sum_{r \in R} y_r \log y_r - \sum_{r \in R} y^r \log y^{me} \right)$
subject to: $y_r$ satisfy $LC(\top, D^V, R)$, where $D^V = D \cup \{(\psi|\phi)[v, v]\}$.
return kl
```

Algorithm 3 KLSatisfaction

```
Input: PLP $P$ and a ground query $Q = ?(\psi|\phi)[l, u]$
Output: KL degree of satisfaction for $Q$
if $P \models_{tight} (\phi|\top)[0, 0]$ then
    return 1;
end if
if $l \geq u$ then
    return 0;
end if
Compute the tight bounds $[l', u']$ for $(\psi|\phi)$ by Algorithm Tight_0_Consequence in Fig. 2.
if $l < l'$ then
    $l := l'$
end if
if $u > u'$ then
    $u := u'$
end if
Compute $s_p = \nu^{pos}_{P,(\psi|\phi)}(u')$ by Algorithm 2.
Compute $s_n = \nu^{neg}_{P,(\psi|\phi)}(u')$ by Algorithm 2.
Compute $s'_p = \nu^{pos}_{P,(\psi|\phi)}(l)$ by Algorithm 2.
Compute $s'_n = \nu^{neg}_{P,(\psi|\phi)}(u)$ by Algorithm 2.
$sat := 0.5 * (s'_p/s_p + s'_n/s_n)$
return $sat$
```

6.2.3 Analysis of trials data for breast cancer

In this section, we model and query the meta-analysis results of early breast cancer trials[4]. This meta analysis of original individual trials aims to examine the effects of various treatments with early breast cancer. Here we consider the mortality of patients who have had the treatment of radiotherapy after breast conserving surgery (BCS) in node-negative disease. The statistical data we use are from Webfigure 6a. in the web-site of EBCTCG [5]. For instance from this table, we can get the statistics of 10-year mortality of 50-year-old breast cancer patients who have had the treatment of radiotherapy after breast conserving surgery in node-negative disease, as well as the 10-year mortality of the patients whose ER (Estro-

[4] Early Breast Cancer Trialists' Collaborative Group (EBCTCG)

[5] http://www.ctsu.ox.ac.uk/ ebctcg/local2000/annex.pdf

Table 5 Sub-table of Webfigure 6a. in web-site of EBCTG , Breast Cancer Mortality

Entry Age		
Subgroup	mortality of BCS+RT	mortality of BCS only
Age <50	18.3%	21.4%
Age 50-59	15.0%	19.7%
Age 60-69	16.6%	18.5%
Age 70+	-%	-%
ER Status (ER-poor vs ER-positive)		
Subgroup	mortality of BCS+RT	mortality of BCS only
ER-poor	22.3%	27.2%
ER-positive	16.9%	19.6%
ER-unknown	17.2%	18.3%
Total	17.4%	20.3%

gen Receptor) values are positive. But we do not have statistical data about the effects of ER value on 10-year mortality of 50-year-old patients. That is, we do not know the mortality of 50-year-old patients whose ER values are positive.

The knowledge (or data) in Webfigure 6a can be formally represented by a PLP P below with 13 rules. Question like *what is the mortality of a 50-year-old patient who has had the treatment of radiotherapy after breast conserving surgery and whose ER value is positive* can be formalized as $Q = ?(mort(name, Y10) \mid bcsRT(name, Y1) \wedge hasBC(name, Y1) \wedge er(name, Y1, positive) \wedge age(name, 50s) \wedge tenYear(Y1, Y10))$, where $name$ should be replaced by an individual's name whose 10-year mortality is our interest.

1 $(age(X,Y) \wedge age(X,Z) \wedge Y \neq Z \mid \top)[0,0]$
2 $(bcsRT(X,Y) \wedge bcsOnly(X,Y) \mid \top)[0,0]$
3 $(er(X,Y1,poor) \wedge er(X,Y1,positive) \mid \top)[0,0]$
4 $(er(X,Y1,poor) \wedge er(X,Y1,unknown))[0,0]$
5 $(er(X,Y1,positive) \wedge er(X,Y1,unknown))[0,0]$
6 $(mort(X,Y10) \mid bcsRT(X,Y1) \wedge hasBC(X,Y1)$
$\wedge tenYear(Y1,Y10))[0.174, 0.174]$
7 $(mort(X,Y10) \mid bcsRT(X,Y1) \wedge hasBC(X,Y1)$
$\wedge age(X,yt50) \wedge tenYear(Y1,Y10))[0.183, 0.183]$
8 $(mort(X,Y10) \mid bcsRT(X,Y1) \wedge hasBC(X,Y1)$
$\wedge age(X,50s) \wedge tenYear(Y1,Y10))[0.150, 0.150]$
9 $(mort(X,Y10) \mid bcsRT(X,Y1) \wedge hasBC(X,Y1)$
$\wedge age(X,60s) \wedge tenYear(Y1,Y10))[0.166, 0.166]$
10 $(mort(X,Y10) \mid bcsRT(X,Y1) \wedge hasBC(X,Y1)$
$\wedge er(X,Y1,poor) \wedge tenYear(Y1,Y10))[0.223, 0.223]$
11 $(mort(X,Y10) \mid bcsRT(X,Y1) \wedge hasBC(X,Y1)$
$\wedge er(X,Y1,positive) \wedge tenYear(Y1,Y10))[0.169, 0.169]$
12 $(mort(X,Y10) \mid bcsRT(X,Y1) \wedge hasBC(X,Y1)$
$\wedge er(X,Y1,unknown) \wedge tenYear(Y1,Y10))[0.172, 0.172]$

In this PLP, there are four constants related to attribute Age: $yt50$ and $ot70$ standing for *younger than 50* and *older than 70* respectively, $50s$ and $60s$ standing for *age in 50s* and *age in 60s* respectively. Rules 2, 3, 4, and 5 are from the meta-knowledge from the trials, which states the methods for dividing sub-groups on ER values. $bcsRT(X,Y)$ means that patient X has RT treatment after breast conserving surgery in year Y. $bcsOnly(X,Y)$ means X has breast conserving surgery in year Y only. $er(X,Y,Z)$ states that the test result of ER status for X in year Y is Z. $hasBC(X, Y)$ means that X has breast cancer in year Y. $mort(X, Y)$ means X died for breast cancer in year Y. $tenYear(Y, Z)$ states that year Z is ten years after year Y.

Rule 1 also comes from the background knowledge, it says that a person cannot have two different ages (or age groups). Condition $Y \neq Z$ in Rule 1 can be replaced by any two values Y and Z which cannot be held simultaneously. Rule 1 is in fact equivalent to a set of rules such as $(age(X, yt50) \wedge age(X, 60s)|\top)[0, 0]$, by replacing Y and Z with $Y = 50s$ and $Z = 60s$. The remaining rules come directly from the statistical data listed in Table 5, which is a sub-table of Webfigure 6a[6]. More precisely, Rules 6-12 correspond to the second column in the table.

Assume that we have a patient named $Mary$ who is 50 years old and is diagnosed as having breast cancer with positive result of ER test. A doctor decides to give her the treatment of radiotherapy (RT) after breast conserving surgery. From the statistical data, we know the 10-year mortality of breast cancer patients in their 50s who have had RT treatment after BCS and the 10-yeay mortality of breast cancer patients with a positive ER value. So, what can we tell about Mary's 10-year mortality after BCS and RT, given that her ER value is positive? Formally, this is to answer the query $Q = ?(mort(Mary, Y10) \mid bcsRT(Mary, Y1) \wedge hasBC(Mary, Y1) \wedge er(Mary, Y1, positive) \wedge age(Mary, 50s) \wedge tenYear(Y1, Y10))$.

Let us denote the conditional event in this query Q as E to simplify the notation (i.e., $Q = ?E$). With the given PLP, we have that $P \models_{tight} E[0, 1]$ and $P \models_{tight}^{me} E[0.1456, 0.1456]$. That is, we get a non-informative interval $[0, 1]$ and a precise probability 0.1456 as two possible answers to this query.

Note that from statistics, the 10-year mortality in the subgroup of patients in their 50s with BCS+RT is 15.0% and in the subgroup of ER-positive (ER value is positive) patients with BCS+RT is 16.9%. Since the probability of query Q given by maximum entropy (14.56%) is less than both 15.0% and 16.9%, this value seems reasonable since both subgroups for *Age in 50-59* and *ER-positive* have a lower 10-year mortality rate than other subgroups. So both factors together could further reduce the mortality rate. This value is also backed up by the ignorance value of E under P which is 0.017 and this small value implies that the knowledge in P is rich enough to answer Q with a single probability. In other words, the probability 14.56% given by maximum entropy is reasonable. However, on the other hand, the ignorance is bigger than 0, thus there is still a small chance that probability 14.56% could be wrong. In this case, we want to find out the interval where the true probability could lie and how satisfied we are with this interval. To do so, we need to measure the degree of satisfaction which is the second order of probability about the probability of a query being in a given interval. The second order probabilities for $Pr(Q) \in [l, u]$, where l, u have different values, are listed in Table 6. In this table, we take the maximum entropy probability, p_{me}, as a middle point to create various sized intervals $[l, u]$ with $l = p_{me} - \Delta, u = p_{me} + \Delta$ where ϵ is the base value for increase/decrease and $\Delta = k\epsilon$ indicates how many times (k times) more/less of ϵ we want to increase/decrease p_{me}. In this case, we set $\epsilon = 0.005$, and create the first interval $[0.1406, 0.1506]$ that contains $p_{me} = 0.1456$.

If a doctor wants to follow the treatment plan BCS+RT given that $Mary$ is in her 50s and her ER value is positive, the doctor could look up Table 6 to see how reasonable this treatment plan is. For instance, we have $p(Pr(Q) \in [0.1356, 0.1556]) = 0.00085$ which can be explained as: the probability that *the probability of $Mary$ being dead in 10-years time after BCS+RT treatment falling in between* 0.1356 *and* 0.1556 is 0.00085. In other words, it is very unlikely that $Mary$'s 10-year mortality is between 13.56% and 15.56%. If in the

[6] see http://www.ctsu.ox.ac.uk/ ebctcg/local2000/annex.pdf

Table 6 Probability bound for $Pr(Q)$ and the degree of satisfaction (2nd order probability) of $Pr(Q) \in [l, u]$

	Probability Bound $[l, u]$ for $Pr(Q)$	Degree of satisfaction
$\Delta = k\epsilon, \epsilon = 0.005$	$l = p_{me} - \Delta, u = p_{me} + \Delta$	$p(Pr(Q) \in [l, u])$
$\Delta = 1\epsilon = 0.005$	$[0.1406, 0.1506]$	0.00059
$\Delta = 2k\epsilon, k = 1$	$[0.1356, 0.1556]$	0.00085
$k = 2$	$[0.1256, 0.1656]$	0.00934
$k = 3$	$[0.1156, 0.1756]$	0.02102
$k = 4$	$[0.1056, 0.1856]$	0.03650
$k = 5$	$[0.0956, 0.1956]$	0.05642
$k = 6$	$[0.0856, 0.2056]$	0.07965
$k = 7$	$[0.0756, 0.2156]$	0.10620
$k = 8$	$[0.0656, 0.2256]$	0.13717
$\Delta = 4k\epsilon, k = 5$	$[0.0456, 0.2456]$	0.20465
$k = 6$	$[0.0256, 0.2656]$	0.28097
$k = 7$	$[0.0056, 0.2856]$	0.36836
$k = 8$	$[0, 0.3056]$	0.42478
$\Delta = (20k + 12)\epsilon, k = 2$	$[0, 0.4056]$	0.59181
$k = 3$	$[0, 0.5056]$	0.71571
$k = 4$	$[0, 0.6056]$	0.80642
$k = 5$	$[0, 0.7056]$	0.87390
$k = 6$	$[0, 0.8056]$	0.92588
$k = 7$	$[0, 0.9056]$	0.96571
$k = 8$	$[0, 1]$	1

table, there is an entry with a smaller value of u (the 10-year mortality is not beyond u) and a reasonably large value of degree of satisfaction, then the doctor could decide that this plan is worth following. On the other hand, if there does not exist an entry in the table that shows a high probability of a low 10-year mortality probability for $Mary$, then this treatment plan is questionable. In this particular case, we have an entry $([0, 0.3056], 0.42478)$ which shows that $p(Pr(Q) \in [0, 0.3056]) = 0.42478$, and it means that there is almost a 43% probability that $Mary$'s 10-year mortality probability is below 30%. If her doctor is happy enough with this estimation, then the treatment BCS+RT can go ahead.

7 Related Work

Since the early 1990s, there have been considerable research efforts on integrating logical programming with probability theory. These probabilistic logic programs have been studied from different perspectives and have different syntactic forms and semantics, including *conditional probabilistic logic programming* [CPQC03, Luk98, Luk01b], *Causal Probabilistic Logic Programming* [BGR04, BH07, Saa07], *Success Probabilistic Logic Programming* [RKT07, Fuh00], and some others [DD04].

In causal probabilistic logic programming [BGR04, BH07], a rule $Pr(\psi|_c\phi) = y$ is interpreted as *"if ϕ happens, this fact will cause the probability of ψ being y"*. A causal probability statement implicitly represents a set of conditional independence assumptions: given its cause ϕ, an effect ψ is probabilistically independent of all factors except the (direct or indirect) effects of ϕ (see [BGR04] for details). Formally, if $Pr(\psi|_c\phi_1) = y_1 \in P$ and $Pr(\psi|_c\phi_2) = y_2 \in P$ where $y_1 \neq y_2$, then no possible world satisfies both ϕ_1 and ϕ_2.

In [RKT07, Fuh00], a real number attached to a rule represents the probability that this rule is alliable (or satisfiable). In other words, a PLP in this view represents a set of (classical) logic programs, and the probability of each member is decided by all the probabilities of the rules. Then for any query, the answer is the probability of choosing a classical logic program from the set that can successfully infer the query. In this formalization, we can only query about the probability of ψ and cannot query about the probability of $(\psi|\phi)$, since $(\psi|\phi)$ is meaningless in classical logic programs.

In [NS91, NS92, DD04, Saa07, BSS09], probabilities are attached to literals, such as: $b[0.6, 0.7] \leftarrow a[0.2, 0.3]$, which means that if the probability of a is in between 0.2 and 0.3, then the probability of b is in between 0.6 and 0.7. Intuitively, the interpretation of rules is more close to casuality than conditioning. As a consequence, if we have another rule: $b[0.2, 0.3] \leftarrow c[0.5, 0.6]$, then $Pr(a) \in [0.2, 0.3]$ and $Pr(c) \in [0.5, 0.6]$ cannot be both true simultaneously,

In this paper, we have focused on the framework of conditional probabilistic logic programming for representing conditional events, because this framework is more suitable for modelling our applications, such as clinical trials information or dialog knowledge.

Because of its weakness in reasoning, subclasses cannot inherit the properties of its superclass in the basic semantics of PLP. In [Luk01a, Luk05, Luk07], Lukasiewicz provided another method to enhance the reasoning power mainly on the issue of inheritance. In this setting, *logic entailment strength* λ is introduced. With strength 1, subclasses can completely inherit the attributes of its superclass; with strength 0 subclasses cannot inherit the attributes of its superclass; with a strength between 0 and 1, subclasses can partially inherit the attributes of its superclass. Value strength appears to be similar to the degree of satisfaction in our framework, but they are totally different. First, λ is not a measurement for a query, but is given by a user to control the reasoning procedure, in other words, we cannot know beforehand the strength in order to infer a conclusion. Second, even if we can use a strength as a measurement, i.e. even if we can obtain the required strength to infer an expected conclusion, it is not an instance of degree of satisfaction, because the cautious monotonicity postulate in Definition 4 is not satisfied. Given a PLP P, assume that we can infer both $(\psi|\phi)[l_1, u_1]$ by strength $\lambda = \lambda_1$ and $(\psi|\phi)[l_2, u_2]$ by strength $\lambda = \lambda_2$. Now assume that $(\psi|\phi)[l_1, u_1]$ is added to P, however, in order to infer $(\psi|\phi)[l_2, u_2]$, we still need to have the strength $\lambda = \lambda_2$ being given. That is, adding additional information to P does not avoid requiring the strength λ_2 if $(\psi|\phi)[l_2, u_2]$ is to be inferred. In contrast, if we have $(\psi|\phi)[l_1, u_1]$ added in the PLP, then the degree of satisfaction of $(\psi|\phi)[l_2, u_2]$ will increase. Consider Example 11, with strength $\lambda = 0.5$, we can infer that $(fly(t)|magpie(t))[0.8, 1]$, $(fly(t)|sickMagpie(t))[0.8, 1]$. However, these two conclusions have different degrees of satisfaction.

In [RKI03, Röd03], the authors provided a second order uncertainty to measure the reliability of accepting the precise probability obtained by applying the maximum entropy principle as the answer to a query in propositional probabilistic logic. The second order uncertainty for $(\psi|\phi)$ and PLP P is defined as $(-\log l - \log u)$ where $P \models_{tight} (\psi|\phi)[l, u]$. Similarly, we provided an ignorance function to measure the usefulness of a PLP for answering a query. If a precise probability for a query is inferred from a PLP P then P contains full information about the query, and therefore accepting the probability is totally reliable. More precisely, their second order uncertainty is directly computed from the probability interval of the query inferred from P. In contrast, our ignorance is computed from the PLP, which provides more information than an interval. Therefore, our measure of ignorance is more accurate in reflecting the knowledge in a PLP. Consider Example 11 again, the second order uncertainty of $(fly(t)|magpie(t))$ and $(fly(t)|sickMagpie(t))$ are the same. However the degree of satisfaction for the two queries are different.

In [BSS09], the authors defined a higher-order probability distribution over the probabilities that a query can be inferred from, given a probabilistic logic program Π. Informally speaking, the higher order probability that a query Q is entailed by Π with probability in $[a, b]$ is interpreted as the ratio between the number of probabilistic models of Π that give Q a probability value within the interval $[a, b]$ and the total number of probabilistic models of Π. As stated in the paper, there are no assumptions about the dependencies or correlations between the events represented in the probabilistic logic program, including the maximum entropy principle, that is, their method is based the assumption of *ignorance*. In contrast, our method tries to balance between the assumption of the maximum entropy principle with the ignorance of the knowledge contained in a PLP. Therefore, our approach can be seen as a step forward towards addressing the problem of ignorance.

8 Conclusion

To be able to accurately answer a query is critical in many intelligent systems. When the underlying knowledge is uncertain, e.g., probabilistic, this problem is more evident. So, what is the probability of an answer is indeed for a given query, when the knowledge used itself is uncertain? One way is to attach an interval to the answer indicating that the probability of the answer is in the interval, another is to generate a single precise probability. The maximum entropy principle is widely used for this latter purpose.

Although the maximum entropy principle is intuitive and widely accepted in information theory, it is too risky to simply apply it to answer a query when the knowledge used is not certain. In order to tell how much we can *trust* a result for a query given a PLP with imprecise knowledge, we proposed a framework to measure both ignorance and the degree of satisfaction of an answer to a query under a given PLP. Using the consequence relations provided in this paper, we can get an informative and reliable interval as the answer for a query or alternatively we know how much we can trust a single probability. The proofs that our framework is an extension of both traditional conditional probabilistic logic programming and the maximum entropy principle (in terms of consequence relations) show that our framework is theoretically sound.

We demonstrated our framework with some examples from the literature and from other research projects we are involved in. The results show that providing degrees of satisfaction and ignorance are useful for making decisions, when there seem to be several choices and a system does not have other information to suggest which result to choose.

Acknowledgements This work is funded by the EPSRC projects with reference numbers: EP/D070864/1 and EP/D074282/1.

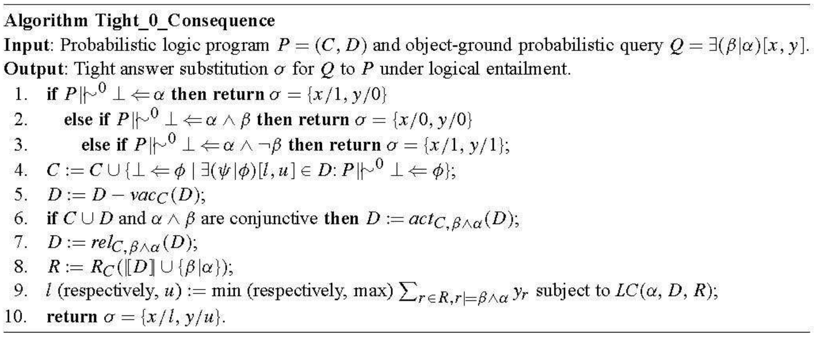
A Appendix: Algorithms

In this section, we provide a brief description of the algorithms proposed in [KIL04] which are used in our algorithms. We consider only ground PLPs here.

At first, a PLP is divided into two sets C and D. The first set contains conditional events with the $[1, 1]$ interval and the other purely probabilistic intervals, that is, $C = \{(\psi|\phi)[1, 1] \in P\} \cup \{(\psi|\phi)[0, 0] \in P\}$ and $D = P \setminus C$. In this section, we denote that $P = (C, D)$.

The idea of these algorithms is to generate equivalent classes of possible worlds and the possible worlds in each equivalent class are indifferentiable under the knowledge contained by the PLP.

Fig. 2 Algorithm Tight_0_Consequence, from [KIL04]



```
**Algorithm Tight_0_Consequence**
**Input**: Probabilistic logic program $P = (C, D)$ and object-ground probabilistic query $Q = \exists(\beta|\alpha)[x, y]$.
**Output**: Tight answer substitution $\sigma$ for $Q$ to $P$ under logical entailment.
if $P \|\!\!\sim^0 \bot \Leftarrow \alpha$ then return $\sigma = \{x/1, y/0\}$
    else if $P \|\!\!\sim^0 \bot \Leftarrow \alpha \wedge \beta$ then return $\sigma = \{x/0, y/0\}$
        else if $P \|\!\!\sim^0 \bot \Leftarrow \alpha \wedge \neg\beta$ then return $\sigma = \{x/1, y/1\}$;
$C := C \cup \{\bot \Leftarrow \phi \mid \exists(\psi|\phi)[l, u] \in D\colon P \|\!\!\sim^0 \bot \Leftarrow \phi\}$;
$D := D - vac_C(D)$;
if $C \cup D$ and $\alpha \wedge \beta$ are conjunctive then $D := act_{C,\beta\wedge\alpha}(D)$;
$D := rel_{C,\beta\wedge\alpha}(D)$;
$R := R_C([\![D]\!] \cup \{\beta|\alpha\})$;
$l$ (respectively, $u$) := min (respectively, max) $\sum_{r\in R, r\models\beta\wedge\alpha} y_r$ subject to $LC(\alpha, D, R)$;
return $\sigma = \{x/l, y/u\}$.
```

Given a PLP $P = (C, D)$ and a nonempty set of conditional events E, the set $R_C(E)$ is defined as the set of all mappings r that assign each conditional events $(\psi|\phi) \in E$ an element of $\{\psi \wedge \phi, \neg\psi \wedge \phi, \neg\phi\}$ such that $C \cup \{r(\psi|\phi)|(\psi|\phi) \in E\}$ is satisfiable.

Define $S_C(E)$ as the set of all possible worlds that satisfy C, and this set can be partitioned into subsets w.r.t. $R_C(E)$, such that $S_C(E) = \{S_r \mid r \in R_C(E)\}$ where $S_r = \{I \in I_\Phi \mid I \models C, \forall(\psi|\phi) \in E, I \models r(\psi|\phi)\}$.

An important result in [KIL04] is that, reasoning with a PLP P can be reduced to calculating a probability distribution over a set $S_C(E)$.

Let $P = (C, D)$ be a PLP. A probabilistic formula $(\psi|\phi)[l, u] \in D$ is called *vacuous* under C iff

- $C \models (\phi|\top)[0, 0]$, or
- $C \models (\psi \wedge \phi|\top)[0, 0]$ and $l = 0$, or
- $C \models (\psi|\phi)[1, 1]$ and $u = 1$

We use $vacc_C(D)$ to denote the set of all vacuous probabilistic formulas in D under C.

Let $P = (C, D)$ be a PLP, and let α be an event. Set $At(\alpha)$ is used to denote the set of all atoms that occur in α. Denote the Herbrand base defined on the set of predicates and constants that occur in P and α as $HB_{P,\alpha}$. The *decomposition* of $HB_{P,\alpha}$ w.r.t. P and α is a partition $\{H_1, \ldots, H_k\}$ of $HB_{P,\alpha}$ such that

- each probabilistic formula in P is defined over some H_i with $i \in \{1, \ldots, k\}$, and
- α is defined over some H_i with $i \in \{1, \ldots, k\}$, and
- $k \geq 1$ is maximal.

For $i \in \{1, \ldots, k\}$, we define D_i as the set of all probabilistic formulas from D that are defined over H_i. The *relevant subset* of D w.r.t. C and α, denoted by $rel_{C,\alpha}(D)$, is defined as the set D_i with minimal index $i \in \{1, \ldots, k\}$ such that $At(\alpha) \subseteq D_i$.

In order to reason under the maximum entropy principle, it needs to calculate the size of each element of the set $S_C(D \cup \{(\psi|\phi)\}$, where $?(\psi|\phi)$ is the query. Let $a_r = |S_r|$ where $r \in R_C(D \cup \{(\psi|\phi)\})$. For simplicity, we denote $R_C(D \cup \{(\psi|\phi)\}$ by R. Assume that $P \not\models (\phi|\top)[0, 0]$, then $P \models^{me}_{tight} (\psi|\phi)[d, d]$ where $d = (\sum_{r\in R, r\models\psi\wedge\phi} y^\star_r)/(\sum_{r\in R, r\models\phi} y^\star_r)$, and $y^\star_r$ $(r \in R)$ is the optimal solution of the following optimization problem over the variables y_r $(r \in R)$:

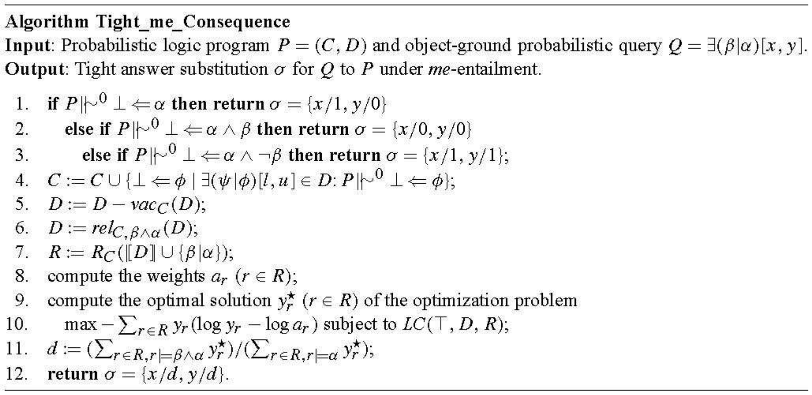
$$\max - \sum_{r\in R} y_r(\log y_r - \log a_r).$$

References

[BGR04] Chitta Baral, Michael Gelfond, and J. Nelson Rushton. Probabilistic reasoning with answer sets. In *LPNMR*, pages 21–33, 2004.

[BH07] Chitta Baral and Matt Hunsaker. Using the probabilistic logic programming language p-log for causal and counterfactual reasoning and non-naive conditioning. In *IJCAI*, pages 243–249, 2007.

Fig. 3 Algorithm Tight_me_Consequence, from [KIL04]



```
Algorithm Tight_me_Consequence
Input: Probabilistic logic program $P = (C, D)$ and object-ground probabilistic query $Q = \exists(\beta|\alpha)[x, y]$.
Output: Tight answer substitution $\sigma$ for $Q$ to $P$ under me-entailment.

if $P \Vdash\!\sim^0 \bot \Leftarrow \alpha$ then return $\sigma = \{x/1, y/0\}$
  else if $P \Vdash\!\sim^0 \bot \Leftarrow \alpha \wedge \beta$ then return $\sigma = \{x/0, y/0\}$
    else if $P \Vdash\!\sim^0 \bot \Leftarrow \alpha \wedge \neg\beta$ then return $\sigma = \{x/1, y/1\}$;
$C := C \cup \{\bot \Leftarrow \phi \mid \exists(\psi|\phi)[l, u] \in D\colon P \Vdash\!\sim^0 \bot \Leftarrow \phi\}$;
$D := D - vac_C(D)$;
$D := rel_{C,\beta\wedge\alpha}(D)$;
$R := R_C([\![D]\!] \cup \{\beta|\alpha\})$;
compute the weights $a_r$ $(r \in R)$;
compute the optimal solution $y_r^\star$ $(r \in R)$ of the optimization problem
  max $-\sum_{r\in R} y_r(\log y_r - \log a_r)$ subject to $LC(\top, D, R)$;
$d := (\sum_{r\in R, r\models\beta\wedge\alpha} y_r^\star)/(\sum_{r\in R, r\models\alpha} y_r^\star)$;
return $\sigma = \{x/d, y/d\}$.
```

[BSS09] Matthias Broecheler, Gerardo I. Simari, and V. S. Subrahmanian. Using histograms to better answer queries to probabilistic logic programs. In *Proceedings of the 25th International Conference on Logic Programming*, pages 40–54, Berlin, Heidelberg, 2009. Springer-Verlag.

[CPQC03] Vítor Santos Costa, David Page, Maleeha Qazi, and James Cussens. CLP($\mathcal{BN}$): Constraint logic programming for probabilistic knowledge. In *UAI*, pages 517–524, 2003.

[DD04] Alex Dekhtyar and Michael I. Dekhtyar. Possible worlds semantics for probabilistic logic programs. In *ICLP*, pages 137–148, 2004.

[DGdMP93] Didier Dubois, Lluis Godo, Ramon López de Mántaras, and Henri Prade. Qualitative reasoning with imprecise probabilities. *J. Intell. Inf. Syst.*, 2(4):319–363, 1993.

[FH94] Alan M. Frisch and Peter Haddawy. Anytime deduction for probabilistic logic. *Artif. Intell.*, 69(1-2):93–122, 1994.

[Fuh00] Norbert Fuhr. Probabilistic datalog: Implementing logical information retrieval for advanced applications. *JASIS*, 51(2):95–110, 2000.

[KIL04] G. Kern-Isberner and T. Lukasiewicz. Combining probabilistic logic programming with the power of maximum entropy. *Artificial Intelligence*, 157(1-2):139–202, 2004.

[KIR04] G. Kern-Isberner and W. Rödder. Belief revision and information fusion on optimum entropy. *International Journal of Intelligent Systems*, 19(9):837–857, 2004.

[KLM90] Sarit Kraus, Daniel J. Lehmann, and Menachem Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artif. Intell.*, 44(1-2):167–207, 1990.

[Luk98] Thomas Lukasiewicz. Probabilistic logic programming. In *ECAI*, pages 388–392, 1998.

[Luk01a] Thomas Lukasiewicz. Probabilistic logic programming under inheritance with overriding. In *UAI*, pages 329–336, 2001.

[Luk01b] Thomas Lukasiewicz. Probabilistic logic programming with conditional constraints. *ACM Trans. Comput. Log.*, 2(3):289–339, 2001.

[Luk05] Thomas Lukasiewicz. Weak nonmonotonic probabilistic logics. *Artif. Intell.*, 168(1-2):119–161, 2005.

[Luk07] Thomas Lukasiewicz. Nonmonotonic probabilistic logics under variable-strength inheritance with overriding: Complexity, algorithms, and implementation. *Int. J. Approx. Reasoning*, 44(3):301–321, 2007.

[NS91] Raymond Ng and V.S. Subrahmanian. A semantical framework for supporting subjective and conditional probabilities in deductive databases. *Journal of Automated Reasoning*, 10:565–580, 1991.

[NS92] Raymond Ng and V. S. Subrahmanian. Probabilistic logic programming. *Inf. Comput.*, 101(2):150–201, 1992.

[RKI03] Wilhelm Rödder and Gabriele Kern-Isberner. From information to probability: An axiomatic approach - inference isinformation processing. *Int. J. Intell. Syst.*, 18(4):383–403, 2003.

[RKT07] Luc De Raedt, Angelika Kimmig, and Hannu Toivonen. Problog: A probabilistic prolog and its application in link discovery. In *IJCAI*, pages 2462–2467, 2007.

[Röd03] Wilhelm Rödder. On the measurability of knowledge acquisition, query processing. *Int. J. Approx. Reasoning*, 33(2):203–218, 2003.

[Saa07] Emad Saad. Qualitative and quantitative reasoning in hybrid probabilistic logic programs. In *ISIPTA'07 - Fifth International Symposium On Imprecise Probability: Theories And Applications*, 2007.