

# A Credal Extension of Independent Choice Logic

Alessandro Antonucci and Alessandro Facchini

Istituto Dalle Molle di Studi Sull'Intelligenza Artificiale (IDSIA)  
Lugano (Switzerland)  
{alessandro,alessandro.facchini}@idsia.ch

**Abstract.** We propose an extension of Poole's *independent choice logic* based on a relaxation of the underlying independence assumptions. A *credal* semantics involving multiple joint probability mass functions over the possible worlds is adopted. This represents a conservative approach to probabilistic logic programming achieved by considering *all* the mass functions consistent with the probabilistic facts. This allows to model tasks for which independence among some probabilistic choices cannot be assumed, and a specific dependence model cannot be assessed. Preliminary tests on an object ranking application show that, despite the loose underlying assumptions, informative inferences can be extracted.

**Keywords:** Probabilistic logic programming · Imprecise probabilities · PSAT · Independence.

## 1 Introduction

*Probabilistic logic programming* (PLP) is an emerging area in AI research aiming to develop reasoning tools for relational domains under uncertainty [9]. PLP can be either intended as a combination of logic programs with probabilistic statements, or a part of the wider current interest in probabilistic programming [4] specialised to highly structured probability spaces. After some early proposals such as Probabilistic Horn Abduction [26] and Independent Choice Logic [27], Distribution Semantics [28] and Probabilistic Datalog [14], PLP is currently subject of intense research as well as the theoretical basis for many real-world applications [8].

Most of these theories make assumptions such as the program acyclicity and the mutual independence of the probabilistic facts. This leads to the specification of a single probability mass function over the possible words (*least models*) associated to the logic program. In more recent times, several approaches are trying to go beyond some of these assumptions, but without giving away independence. In some of these cases [7,22], this leads to the adoption of a *credal* semantics, i.e., a PLP does not define a single mass function but a set of joint mass functions over the least models. In this paper we propose a different path consisting in a generalisation of Poole's independent choice logic which keeps the acyclicity condition on programs, but relaxes the independence assumptions. Such an extension can be intended as a credal, conservative, semantics considering *all* the joint mass functions over the least models being also consistent with the probabilistic facts. To introduce our motivations for such proposal, consider the following example.

*Example 1.* Two urns contain red, green and blue balls. Let  $\mu_1 = [0.60; 0.30; 0.10]$  and  $\mu_2 = [0.20; 0.35; 0.45]$  denote the normalised proportions of the colours in the urns.<sup>1</sup> Assuming that a ball is randomly drawn from each urn, the joint probability over the nine possible outcomes is the left matrix in Table 1. The matrix on the right corresponds to a different situation in which a ball is randomly drawn from the first urn, while the second ball is required to have a colour different from that of the first. To achieve that, an unbiased coin is flipped to decide which one of the two permitted colours should be picked. Both joint mass functions are consistent with the above marginal probabilities (i.e., the sums of the rows/columns are the values in  $\mu_1$  and  $\mu_2$ ).

|       |       |       |                       |       |       |       |
|-------|-------|-------|-----------------------|-------|-------|-------|
| ●     | ●     | ●     | ← ball <sub>1</sub> → | ●     | ●     | ●     |
| 0.120 | 0.210 | 0.270 | ●                     | 0.000 | 0.300 | 0.300 |
| 0.060 | 0.105 | 0.135 | ●                     | 0.150 | 0.000 | 0.150 |
| 0.020 | 0.035 | 0.045 | ●                     | 0.050 | 0.050 | 0.000 |
|       |       |       | ↑                     |       |       |       |
|       |       |       | ball <sub>2</sub>     |       |       |       |

**Table 1.** Two joint mass functions over two ternary variables sharing the same marginal probabilities. Dark grey cells corresponds to a query of interest (first ball non-red, second ball non-green).

This trivial example makes clear that marginal probabilities are in general not sufficient to uniquely determine a joint mass function unless additional assumptions, such as independence or an explicit dependence, are made. If no further assumptions can be made, the most conservative approach corresponds to consider *all* the joint mass functions consistent with the constraints and computing inferences with respect to the multiple probabilistic specification. This is clarified by the following example.

*Example 2.* With the same setup as in Example 1, we want to compute the probability of having the first ball not red and the second not green. Assuming independence, this query has probability  $(1 - \mu_1(r))(1 - \mu_2(g)) = 0.56$ . The same result can be obtained by summing the (consistent) dark grey values in the matrix on the left of Table 1. Under the coin-based dependence relation, the result becomes instead 0.65 (sum of the dark grey values in the matrix on the right). With no dependence/independence assumptions, the probability of the query can be only said to belong to the interval  $[0.5, 0.7]$ . These values are the maximum and the minimum obtained from a linear programming task involving the nine joint probabilities as optimisation variables and the consistency with the marginals (together with non-negativity and normalization) as linear constraints.

The feasible region of the linear programming task in the above example is a convex set of (joint) probability mass functions, i.e., a *credal set* [3,20]. Inferences based on credal sets are conservatively intended as the computation of the lower/upper bounds

<sup>1</sup> We use semicolons to separate the elements of an array and commas to separate the two bounds of an interval.

of the query with the mass function varying in the credal set. The interval estimate is also called *imprecise probability*. When coping with relational domains, adding independence statements leading to a unique, “precise”, specification might be not always a tenable assumption, and the imprecise-probabilistic technique we present in this paper can be regarded as the most conservative approach to the modeling of a condition of *ignorance* about the dependencies among the atoms in a program.

After reviewing some necessary background concepts in Section 2, our proposal is first introduced by an example in Section 3 and then formalised as a theory in Section 4. An inference algorithm is derived in Section 5, while the results of some preliminary experiments are reported in Section 6. Conclusions and outlooks are in Section 7.

## 2 Background

We review here some background information about logic programming and its probabilistic extension.

*Logic Programming.* A *term*  $t$  is defined as being either a constant or a variable. An *atom*  $r(t_1, \dots, t_k)$  is thence obtained by applying a relational symbol  $r$  to a sequence of terms  $t_1, \dots, t_k$ . We identify Boolean propositional variables with 0-ary predicates. An atom  $a$  or its negation  $\neg a$  is called a *literal*. A *clause* has form

$$a_0 \leftarrow a_1, \dots, a_m, \neg a_{m+1}, \dots, \neg a_n, \quad (1)$$

where  $a_i$  are atoms for each  $i = 0, \dots, n$ ,  $a_0$  is called the *head* of the clause, while the other atoms are the *body*. *Facts* are clauses with empty body. If a clause is not a fact is called a *rule*. A *ground atom* (literal, clause) is an atom (literal, clause) that does not contain any variable. The *grounding* of a clause is a clause obtained by uniformly replacing constants for the variables in the considered clause. A *logic program*  $\mathbf{P}$  is a finite set of clauses. The *Herbrand base* of a program is the set of all ground instances of atoms in the program. A *query*  $Q$  for a program  $\mathbf{P}$  is a set of ground literals whose positive part belong to the Herbrand base of  $\mathbf{P}$ . A logic program is *acyclic* if there is an assignment of a positive integer to each element of the Herbrand base such that for every grounding of a rule the number assigned to the head is greater then each number assigned to an element of the body. An *interpretation*  $\iota$  of a program  $\mathbf{P}$  is a function assigning a truth value to each member of the Herbrand base for  $\mathbf{P}$ . An interpretation  $\iota$  is said to be a *stable model* for  $\mathbf{P}$  if for every ground atom  $a$ ,  $a$  is true in  $\iota$  if and only if  $a$  is a fact in  $\mathbf{P}$  or it is the head of the grounding of a rule in  $\mathbf{P}$  whose base is true in  $\iota$ . A negation  $\neg a$  is true in  $\iota$  if and only if the atom  $a$  is not true in  $\iota$ .

The following classical result [2] is used in the rest of the paper.

**Proposition 1.** *An acyclic logic program has a unique stable model.*

*Probabilistic Logic Programming.* There are different ways of representing probabilistic information with logic programming. A possible approach is to annotate or extend clauses with probabilities (see, e.g., [21,24]). Another, nowadays very popular, approach consists in adding to a logic program *independent* probabilistic alternatives

representing mutually independent random events with a finite number of different outcomes, such as tossing a coin or rolling a die. Within this latter stream, various languages have been proposed such as PRISM [28], ICL [27], pD [15], LPAD [30] or ProbLog [10]. Despite the differences in their syntactical presentation, under Sato’s distribution semantics all these formalisms are in general comparable in their expressive power [29]. In this work we focus on Poole’s ICL. The choice of this formalism, rather than another among the languages listed above, is due to the fact that its explicit set-theoretical formulation of probabilistic choices will simplify the presentation of our approach. ICL syntax and semantics are introduced here below.

*Independent Choice Logic (ICL).* An ICL theory is a triple  $\langle \mathbf{P}, \mathbf{C}, \mu \rangle$  such that:

- $\mathbf{P}$  is an acyclic logic program.
- $\mathbf{C}$ , called the *choice space*, is a family of non-empty sets of ground atoms. The elements of  $\mathbf{C}$  are called *alternatives*, and the elements of each alternative *atomic choices*. Atomic choices from the same or different alternatives cannot unify with each other, nor with the head of any clause in  $\mathbf{P}$ .
- $\mu$  specifies a probability mass function over each alternative, i.e.,  $\mu : \bigcup_{C \in \mathbf{C}} C \rightarrow [0, 1]$  with  $\sum_{a \in C} \mu(a) = 1$  for each  $C \in \mathbf{C}$ .

The following example demonstrates the ICL syntax.

*Example 3.* Consider the setup of Example 1 under the assumption of independence between the colours of the two balls (Table 1 left). Let  $a_i^j$  denote the Boolean variable true if and only if the colour  $j$  has been drawn from the  $i$ -th urn. The example corresponds to an ICL theory over the ground atoms  $\{a_i^j\}_{i=1,2}^{j=r,g,b}$ . The logical program is empty, i.e.,  $\mathbf{P} = \emptyset$ , as the logical constraints over the atoms can be directly embedded in the choice space  $\mathbf{C}$ . Such a space contains two alternatives, i.e.,  $\mathbf{C} := \{C_1, C_2\}$ , with  $C_i = \{a_i^r, a_i^g, a_i^b\}$  for each  $i = 1, 2$ . Finally, the probabilities are directly obtained from the marginals, i.e.,  $\mu(C_i) = \mu_i$  for each  $i = 1, 2$ .

ICL semantics is defined in terms of *possible worlds*. A *total choice*  $\mathfrak{c}$  for the choice space  $\mathbf{C}$  is a choice function selecting exactly one atomic choice from each alternative in  $\mathbf{C}$ , and we associate to  $\mathfrak{c}$  a possible world  $\omega_{\mathfrak{c}}$ . For each total choice  $\mathfrak{c}$  the program  $\mathbf{P} \cup \{a \leftarrow a \in \text{image}(\mathfrak{c})\}$  is acyclic, and therefore by Proposition 1 it has a unique stable model  $\iota_{\mathfrak{c}}$ . We thence say that an atom  $a$  is true at  $\omega_{\mathfrak{c}}$ , written  $\omega_{\mathfrak{c}} \models a$ , if  $a$  is true in  $\iota_{\mathfrak{c}}$ , and states  $\omega_{\mathfrak{c}} \models \neg a$  if  $\omega_{\mathfrak{c}} \not\models a$ . Given a query  $Q$ , we write  $\omega \models Q$  whenever  $\omega \models q$ , for any member of  $q \in Q$ .

As alternatives are assumed to be independent, a ICL theory defines a unique mass function  $\mu'$  over the collection  $\Omega$  of all the possible worlds as follows:

$$\mu'(\omega_{\mathfrak{c}}) := \prod_{a \in \text{image}(\mathfrak{c})} \mu(a) \cdot \prod_{a \notin \text{image}(\mathfrak{c})} (1 - \mu(a)). \quad (2)$$

*Example 4.* The ICL theory in Example 3 has nine possible worlds corresponding to the elements of  $C_1 \times C_2$ . For each  $\omega$ , the associated probability assigned by  $\mu$  as in Equation (2) reproduces the value in Example 1. The query in Example 2 is  $Q := \{\neg a_1^r, \neg a_2^g\}$  and it is true only in four possible worlds. The corresponding success probability is  $\mu'(Q) = \sum_{\omega \models Q} \mu'(\omega) = 0.210 + 0.270 + 0.035 + 0.045$  as in Example 2.

### 3 A Motivating Example

Consider a possibility space determined by the fact that, on a working day, a person called Andrea is using her car or not, she is working late or not, and that in Milan, the city where she is living, it is raining or not. We denote by  $r$  the atomic fact that it rains in Milan, by  $c$  the fact that she uses the car, and by  $w$  the fact that she is working late. Hence, the elements of the possibility space we consider correspond to the eight possible worlds associated with  $r, c, w$  in Table 2, where  $\mathfrak{t}$  denotes true and  $\mathfrak{f}$  false.

|                            | $\omega_1$     | $\omega_2$     | $\omega_3$     | $\omega_4$     | $\omega_5$     | $\omega_6$     | $\omega_7$     | $\omega_8$     |
|----------------------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| $r$ (Rain in Milan)        | $\mathfrak{t}$ | $\mathfrak{t}$ | $\mathfrak{t}$ | $\mathfrak{t}$ | $\mathfrak{f}$ | $\mathfrak{f}$ | $\mathfrak{f}$ | $\mathfrak{f}$ |
| $c$ (Andrea using her car) | $\mathfrak{t}$ | $\mathfrak{t}$ | $\mathfrak{f}$ | $\mathfrak{f}$ | $\mathfrak{t}$ | $\mathfrak{t}$ | $\mathfrak{f}$ | $\mathfrak{f}$ |
| $w$ (Andrea working late)  | $\mathfrak{t}$ | $\mathfrak{f}$ | $\mathfrak{t}$ | $\mathfrak{f}$ | $\mathfrak{t}$ | $\mathfrak{f}$ | $\mathfrak{t}$ | $\mathfrak{f}$ |
| $\mu'(\omega_i)$           | 0.01           | 0.04           | 0.01           | 0.04           | 0.09           | 0.36           | 0.09           | 0.36           |
| $\mu^{(1)}(\omega_i)$      | 0.1            | 0.0            | 0.0            | 0.0            | 0.0            | 0.4            | 0.1            | 0.4            |
| $\mu^{(2)}(\omega_i)$      | 0.1            | 0.0            | 0.0            | 0.0            | 0.1            | 0.3            | 0.0            | 0.5            |
| $\mu^{(3)}(\omega_i)$      | 0.0            | 0.1            | 0.0            | 0.0            | 0.0            | 0.4            | 0.2            | 0.3            |
| $\mu^{(4)}(\omega_i)$      | 0.0            | 0.1            | 0.0            | 0.0            | 0.2            | 0.2            | 0.0            | 0.5            |
| $\mu^{(5)}(\omega_i)$      | 0.0            | 0.0            | 0.1            | 0.0            | 0.1            | 0.4            | 0.0            | 0.4            |
| $\mu^{(6)}(\omega_i)$      | 0.0            | 0.0            | 0.1            | 0.0            | 0.0            | 0.5            | 0.1            | 0.3            |
| $\mu^{(7)}(\omega_i)$      | 0.0            | 0.0            | 0.0            | 0.1            | 0.2            | 0.3            | 0.0            | 0.4            |
| $\mu^{(8)}(\omega_i)$      | 0.0            | 0.0            | 0.0            | 0.1            | 0.0            | 0.5            | 0.2            | 0.2            |

**Table 2.** A possibility space as a set of possible worlds, and nine mass functions over it.

Let us compute the probability that Andrea will hang out with friends under the following assumptions.

- The probability of rain in Milan is 0.1, that for Andrea using her car 0.5 and that for her working late 0.2.
- If Andrea is with her car or it is raining, she will visit her parents.
- If she is neither visiting her parents nor working late, then she will be hanging out with friends.

Let  $p$  denote the fact that Andrea visits her parents, and  $h$  the fact associated to her hanging out with friends. In a PLP dialect such as ICL, if we introduce symbols  $nr, nc$  and  $nw$  to denote the complementary atomic facts associated respectively to  $r, c$  and  $w$ , the example corresponds to a theory with:

- $\mathbf{C} = \{C_1, C_2, C_3\}$ , with alternatives  $C_1 = \{r, nr\}$ ,  $C_2 = \{c, nc\}$ ,  $C_3 = \{w, nw\}$ ;
- $\mu(r) = 0.1, \mu(nr) = 0.9, \mu(c) = 0.5, \mu(nc) = 0.5, \mu(w) = 0.2, \mu(nw) = 0.8$ ;
- and  $\mathbf{P} = \{p \leftarrow c, p \leftarrow r, h \leftarrow \neg p, nw\}$ .

Hence, for instance,  $\omega_i \models c$  if and only if  $\omega_i \not\models nc$  if and only if  $\omega_i \not\models \neg c$ , with  $i \in \{1, \dots, 8\}$ . It also holds that  $\omega_i \models p$ , for  $i \in \{1, \dots, 6\}$  and  $\omega_i \models \neg p$ , for

$i = 7, 8$ . Similarly  $\omega_8 \models h$  and  $\omega_i \models \neg h$  for  $i \neq 8$ . Since alternatives are assumed to be independent, a unique probability mass function  $\mu'$  (see Table 2) is defined over  $\Omega := \{\omega_1, \dots, \omega_8\}$  and, for instance, the success probability of the query  $h$  is  $\mu'(\omega_8) := 0.9 \cdot 0.5 \cdot 0.8 = 0.36$ . Yet, it is easy to identify realistic situations in which such independence assumptions are violated. E.g., Andrea using her car might be not independent of raining in Milan. While the modelling of a deterministic dependence (e.g.,  $c \leftarrow r$ ) or a probabilistic influence (e.g.,  $\mu(c \mid r) = 0.95$  and  $\mu(c \mid nr) = 0.45$ ) can be described in standard PLPs, a condition of complete ignorance about the relations between two or more atomic facts requires a generalisation of the semantics. For instance, if we do not make any assumption about the (in)dependence relations among the possible choices corresponding to the three variables in our example, the whole set of mass functions consistent with the marginals is the convex hull of the eight mass functions  $\{\mu^{(i)}\}_{i=1}^8$  in Table 2 and the probability of the query  $h$  can only be said to belong to the range  $[0.2, 0.5]$ . This interval shrinks to  $[0.32, 0.40]$  if no assumptions can be made about the independence between choices on elements of  $C_1$  and choices on elements of  $C_2$ , but both are assumed to be independent with respect to choices on  $C_3$ .

In the next section we formalise these ideas in a general framework, called *credal choice logic* (CCL), which provides an extension of Poole's independent choice logic in which the independence condition is relaxed.

## 4 Credal Choice Logic

*Syntax.* From a syntactical point of view, the idea pursued in this paper is to consider elements of the choice space not as independent alternatives but as *independent families*, each including possibly correlated alternatives. This is formally stated in the following definition.

**Definition 1.** A CCL theory  $\mathcal{T}$  is a triple  $\langle \mathbf{P}, \mathcal{C}, \mu \rangle$  where:

- $\mathbf{P}$  is an acyclic logic program.
- $\mathcal{C} = \{\mathbf{C}_1, \dots, \mathbf{C}_k\}$  is a family of choice spaces (i.e., a set of sets of non-empty sets of ground atoms). Alternatives and atomic choices are intended as in ICL. In particular we assume that:
  1. for each choice space  $\mathbf{C} \in \mathcal{C}$ , no atomic choice in  $\bigcup \mathbf{C}$  should unify with the head of any clause in  $\mathbf{P}$ ;
  2. for each pair  $\mathbf{C}, \mathbf{C}' \in \mathcal{C}$  with  $\mathbf{C} \neq \mathbf{C}'$ , the sets of atomic choices  $\bigcup \mathbf{C}$  and  $\bigcup \mathbf{C}'$  are disjoint.
- $\mu$  specifies a probability mass function over each alternative of each choice space, i.e.,  $\mu : \bigcup \bigcup \mathcal{C} \rightarrow [0, 1]$  with  $\sum_{a \in \mathbf{C}} \mu(a) = 1$ , for every  $\mathbf{C} \in \mathcal{C}$ , and every  $\mathbf{C} \in \mathcal{C}$ .

Contrary to ICL, in a CCL theory the alternatives in a given choice space are not assumed to be independent. Consequently there could be a choice space  $\mathbf{C}$  with alternatives  $C_1, C_2 \in \mathbf{C}$  such that  $C_1 \neq C_2$  but  $C_1 \cap C_2 \neq \emptyset$ . On the other hand, any ICL theory can be formalised as a CCL theory  $\langle \mathbf{P}, \mathcal{C}, \mu \rangle$  whose choice spaces  $\mathbf{C} \in \mathcal{C}$  are singletons.

*Example 5.* Dropping the independence assumption about the relation between Andrea going by car and raining in Milan in the example in the previous section transforms the corresponding ICL theory in a CCL theory  $\mathcal{T}_{\text{friends}} := \langle \mathbf{P}, \{\mathbf{C}_1, \mathbf{C}_2\}, \mu \rangle$ , where  $\mathbf{C}_1 = \{C_1, C_2\}$ ,  $\mathbf{C}_2 = \{C_3\}$ , while  $\mathbf{P}$  and  $\mu$  are defined as in the original example. The fact that alternatives  $C_1$  and  $C_2$  belong to the same choice space means that the choice among elements of  $C_1$  (namely between  $c$  and  $nc$ ) is not assumed to be independent from the choice among elements of  $C_2$  (namely between  $r$  and  $nr$ ). On the other hand the fact that the alternative  $C_3$  does not belong to the same choice space as  $C_1$  and  $C_2$  means that the choice between  $w$  and  $nw$  is independent of the two previous choices.

*Semantics.* As in ICL, the CCL semantics is defined in terms of possible worlds. A total choice for a family  $\mathcal{C}$  of choice spaces is a choice function  $\mathfrak{c}_i$  on  $\bigcup \mathcal{C}$  such that, for  $C$  and  $C'$  belonging to the same choice space  $\mathbf{C}$ , the following coherence condition is satisfied:

$$\mathfrak{c}_i(C) = \mathfrak{c}_i(C'), \quad (3)$$

whenever  $\mathfrak{c}_i(C) \in C \cap C'$ . Hence a total choice selects coherently (in the sense of Equation (3)) exactly one atomic choice from each alternative of every choice space of  $\mathcal{C}$ .

Now, consider the program  $\mathbf{P}(\mathfrak{c}) := \mathbf{P} \cup \{a \leftarrow \mid \varphi \in \text{image}(\mathfrak{c})\}$ . From the disjointness conditions,  $\mathbf{P}(\mathfrak{c})$  is necessarily acyclic and therefore by Proposition 1 it has a unique stable model  $I_{(\mathfrak{c})}$ . As for ICL, for each total choice  $\mathfrak{c}$  we define a corresponding possible world  $\omega_{\mathfrak{c}}$  and the associated notion of being true in it.

In the case of ICL, the probability for a possible world is given as in Equation (2) by the product of the probabilities of the atomic choices true in it and one minus the probabilities of the atomic choices false in it. The distribution semantics we defined for CCL acts similarly but on sets of probabilities.

To illustrate the idea, notice that a total choice  $\mathfrak{c}$  restricted to a choice space  $\mathbf{C}_i$  determines a partial selection  $\mathfrak{c}_i$ . Such partial selection can be identified with all total choices extending it, and thus with the collection  $E_{\mathfrak{c}_i}$  of possible world in which all elements in the image of  $\mathfrak{c}_i$  are true. Such collection represents the possible worlds an agent cannot tell apart if she is given only the partial information provided by  $\mathfrak{c}_i$ . Let  $\Omega_i$  be the set of all collections associated with  $\mathbf{C}_i$ . Assume  $\mu_i$  is a probability mass function over  $\Omega_i$ , and define  $\mu_i(a) := \sum_{\mathfrak{c}_i: a \in \text{image}(\mathfrak{c}_i)} \mu_i(E_{\mathfrak{c}_i})$ , for  $a \in \bigcup \mathbf{C}_i$ . We say that  $\mu_i$  agrees with  $\mu$  on  $\mathbf{C}_i$  if  $\mu_i(a) = \mu(a)$ , for every  $a \in \bigcup \mathbf{C}_i$ . Hence, call  $\mathcal{M}_i$  the set of all probability distributions over  $\Omega_i$  agreeing with  $\mu$  on  $\mathbf{C}_i$ .

Each possible world  $\omega_{\mathfrak{c}} \in \Omega$  can be identified with the unique sequence  $(E_{\mathfrak{c}_i} : i \in \{1, \dots, k\}) \in \times_{i \in \{1, \dots, k\}} \Omega_i$  such that  $\{\omega_{\mathfrak{c}}\} = \bigcap_{i \in \{1, \dots, k\}} E_{\mathfrak{c}_i}$ . For  $\mu_i \in \mathcal{M}_i$  with  $i \in \{1, \dots, k\}$ , let  $\mu' := \prod_{i \in \{1, \dots, k\}} \mu_i$  be the mass function on  $\Omega$  defined by:

$$\mu'(\omega_{\mathfrak{c}}) = \prod_{i \in \{1, \dots, k\}} \mu_i(E_{\mathfrak{c}_i}). \quad (4)$$

We can therefore associate with a CCL theory  $\mathcal{T}$  a joint credal set  $\mathcal{M}_{\mathcal{T}}$  obtained as the following set of factorising mass functions:

$$\left\{ \prod_{i \in \{1, \dots, k\}} \mu_i \mid \mu_i \in \mathcal{M}_i, i \in \{1, \dots, k\} \right\}. \quad (5)$$

It is immediate to verify the following result.

**Proposition 2.** *Let  $\mathcal{T}$  be a CCL theory.  $\mathcal{M}_{\mathcal{T}}$  is the largest closed convex set of probability mass functions  $\mu'$  on  $\Omega$  that agree with  $\mu$ , that is  $\mu'(a) = \mu(a)$  for every  $a \in \bigcup \mathcal{C}$ .*

In the imprecise-probability jargon, the condition in Equation 5 is called *strong independence* between the variables associated to the different choice spaces [3]. Accordingly, we call the credal set  $\mathcal{M}_{\mathcal{T}}$  *strong extension* of the marginals specified by  $\mu$ . In particular, we have that for  $\mu_i \in \mathcal{M}_i$  there is  $\mu' \in \mathcal{M}_{\mathcal{T}}$  such that:

$$\mu_i(E_{c_i}) = \sum_{\omega \in E_{c_i}} \mu'(\omega) = \mu'(\text{image}(c_i)). \quad (6)$$

Given a CCL theory  $\mathcal{T}$ , a question is therefore whether or not a probability mass function over the possible worlds consistent with the marginals exists. By Equation (5) it is enough to answer such question when  $\mathcal{C}$  contains a single choice space. But this can be shown by induction on the number of alternatives. Hence, the following holds.

**Proposition 3.** *The credal set  $\mathcal{M}_{\mathcal{T}}$  of a CCL theory  $\mathcal{T}$  is non-empty.*

*Example 6.* Consider the CCL theory  $\mathcal{T}_{\text{friends}} := \langle \mathbf{P}, \{\mathbf{C}_1, \mathbf{C}_2\}, \mu \rangle$  in Example 5. There are four partial choices  $c_1^j$  with respect to  $\mathbf{C}_1$  and thus four collections  $E_{c_1^j}$  of possible worlds and, similarly, two partial choices  $c_2^j$  with respect to  $\mathbf{C}_2$  and two collections  $E_{c_2^j}$  of possible worlds (see Table 3). We have that  $\mathcal{M}_1$  is the collection of all probability mass functions  $\mu_1 : \{E_{c_1^1}, \dots, E_{c_1^4}\} \rightarrow [0, 1]$  such that  $\mu(r) = \mu_1(E_{c_1^1}) + \mu_1(E_{c_1^2})$ ,  $\mu(nr) = \mu_1(E_{c_1^3}) + \mu_1(E_{c_1^4})$ ,  $\mu(c) = \mu_1(E_{c_1^1}) + \mu_1(E_{c_1^3})$  and  $\mu(nc) = \mu_1(E_{c_1^2}) + \mu_1(E_{c_1^4})$ . On the other hand,  $\mathcal{M}_2$  is given by taking  $\mu_2 : \{E_{c_2^1}, E_{c_2^2}\} \rightarrow [0, 1]$  defined simply as  $\mu_2(E_{c_2^1}) = \mu(w)$  and  $\mu_2(E_{c_2^2}) = \mu(nw)$ . Hence  $\mathcal{M}_{\mathcal{T}_{\text{friends}}} = \{\mu' = \mu_1 \cdot \mu_2 \mid \mu_1 \in \mathcal{M}_1, \mu_2 \in \mathcal{M}_2\}$ , where  $\mathcal{M}_1$  is the convex hull of  $\mu_1^{(1)} = [0.0; 0.1; 0.4; 0.5]$  and  $\mu_1^{(2)} = [0.1; 0.0; 0.5; 0.4]$ , while  $\mathcal{M}_2$  has a single element  $\mu_2 = [0.2; 0.8]$ .

| $(i, j)$              | (1, 1)                   | (1, 2)                   | (1, 3)                   | (1, 4)                   | (2, 1)                                       | (2, 2)                                       |
|-----------------------|--------------------------|--------------------------|--------------------------|--------------------------|--|--|
| $E_{c_i^j}$           | $\{\omega_1, \omega_2\}$ | $\{\omega_3, \omega_4\}$ | $\{\omega_5, \omega_6\}$ | $\{\omega_7, \omega_8\}$ | $\{\omega_1, \omega_3, \omega_5, \omega_7\}$ | $\{\omega_2, \omega_4, \omega_6, \omega_8\}$ |
| $\text{image}(c_i^j)$ | $\{r, c\}$               | $\{r, nc\}$              | $\{nr, c\}$              | $\{nr, nc\}$             | $\{w\}$                                      | $\{nw\}$                                     |

**Table 3.** Classes (of possible worlds) for  $\mathcal{M}_1$  and  $\mathcal{M}_2$ .



## 5 Inference

In the previous section, by relaxing the independence assumptions, we introduced CCL as a generalisation of ICL. Here we discuss how to compute inferences in CCL.

Given a CCL theory  $\mathcal{T}$ , the CCL semantics leads to the specification of a credal set  $\mathcal{M}_{\mathcal{T}}$  over  $\Omega$  called the *strong extension* of  $\mathcal{T}$ . Given a *query*  $Q$  (see definition in Section 2) for the program  $\mathbf{P}$  in  $\mathcal{T}$ , inference is intended as the computation of the lower  $\underline{\mu}(Q)$  and upper  $\bar{\mu}(Q)$  bounds of the success probability of the query with respect to the strong extension, i.e.,

$$\underline{\mu}(Q) := \min_{\mu(\Omega) \in \mathcal{M}_{\mathcal{T}}} \sum_{\omega \models Q} \mu(\omega), \quad (7)$$

and similarly, with the maximum replacing the minimum for the upper success probability  $\bar{\mu}(Q)$ . As  $\mathcal{M}_{\mathcal{T}}$  is non-empty (Proposition 3), Equation (7) always returns proper inferences. Moreover, being induced by a finite number of linear constraints over the joint probabilities (non-negativity, normalisation, and consistency with the marginal probabilistic facts), the convex set  $\mathcal{M}_{\mathcal{T}}$  has only a finite number of extreme points and Equation (7) can be regarded as a linear programming task, whose solution corresponds to an extreme point of  $\mathcal{M}_{\mathcal{T}}$ . E.g., the interval  $[0.2, 0.5]$  obtained in Section 3 for the query  $h$  under a complete relaxation of the independence assumptions can be regarded as a CCL inference over a strong extension corresponding to the convex hull of the mass functions  $\{\mu^{(i)}\}_{i=1}^8$  in Table 2. Such a brute-force approach cannot be used in general as the number of extreme points might be exponentially large with respect to the input size. In the rest of this section we show that even if, as expected, inference in CCL is NP-hard, these inferences can be reduced to a classical task for which dedicated solver and algorithms have been developed.

To do that, let us first denote as  $\text{CCL}_{\alpha}$  the decision task associated to an inference, i.e., given a  $\alpha \in [0, 1]$ , deciding whether or not  $\alpha \in [\underline{\mu}(Q), \bar{\mu}(Q)]$ . In case the theory contains a single choice space,  $\text{CCL}_{\alpha}$  can be reduced to *probabilistic satisfiability* (PSAT). PSAT is a generalisation of the classical *satisfiability* (SAT) decision problem, in which each element of a finite set of Boolean formulas is paired with a probabilistic assessment [25]. The task is to decide whether or not the assessments are consistent, i.e., to determine whether or not a joint mass function over the variable assigning the given probability to each formula exists. A PSAT instance is thus a set  $\{P(\varphi_i) = \alpha_i \mid i < n\}$  where each  $\varphi_i$  is a Boolean formula and  $\alpha_i$  is the associated probability. As the problem generalises SAT for propositional calculus it is NP-hard. The NP-completeness of the task has been proved in [17] and, unlike SAT, PSAT remains NP-complete even if each clause does not contains more than two literals.

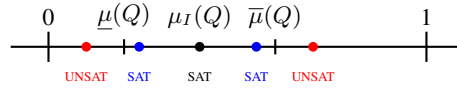
We first describe how PSAT can be used to solve  $\text{CCL}_{\alpha}$ . Let  $\langle \mathbf{P}, \mathcal{C} = \{\mathbf{C}\}, \mu \rangle$  be a CCL theory with a single choice space. First, we convert  $\mathbf{P}$  to an equivalent Boolean formula  $\varphi_{\mathbf{P}}$  (see, e.g., [19] for some conversions), and encode the choice space  $\mathbf{C}$  as a conjunction of XOR:

$$\varphi_{\mathbf{C}} := \bigwedge_{C \in \mathbf{C}} \bigoplus_{a \in C} a. \quad (8)$$

The class of models of  $\varphi_{\mathbf{C}} \wedge \varphi_{\mathbf{P}}$  corresponds to the stable models of the original CCL theory. The associated PSAT instance is therefore the set  $\Sigma_{\alpha} := \{P(\varphi_{\mathbf{C}} \wedge \varphi_{\mathbf{P}}) =$

$1\} \cup \{P(a) = \mu(a) \mid a \in \bigcup \mathbf{C}\} \cup \{P(\wedge Q) = \alpha\}$ . Clearly  $\alpha \in [\underline{\mu}(Q), \overline{\mu}(Q)]$  if and only if  $\Sigma_\alpha$  is satisfiable. With a small abuse of notation, we denote as  $\text{SAT}(\text{CCL}_\alpha)$  a Boolean variable true if and only if  $\Sigma_\alpha$  is satisfiable.

The computation of  $[\underline{\mu}(Q), \overline{\mu}(Q)]$  can be therefore achieved by iterating the above reduction for different values of  $\alpha$  according to a bracketing scheme that identifies both the bounds (Figure 1). The procedure is basically a bisection method that recognises UNSAT values of  $\alpha$  as outer approximations (red points in Figure 1) and SAT values as inner approximations (blue points in Figure 1). To solve PSAT instances, solvers such of those developed by [12] and [6] can be used. To achieve a precision  $\epsilon$  in the estimates, the number of calls of these solvers is  $O(\log \epsilon^{-1})$ .



**Fig. 1.** Inner and outer approximations of CCL inferences.

A key point for the bracketing algorithm is the computation of an inner (i.e., SAT) value  $\mu_I(Q)$  (e.g., black point in Figure 1). We achieve this by translating it into the inference task of computing in ProbLog the marginal probability of the query given some specific evidence. More precisely, we first translate  $\mathcal{T}$  into ProbLog by distinguishing different occurrences of the same atomic choices and adding, for each duplicated atomic choice, a clause of the form  $f \leftarrow a, a'$ , where  $f$  is some fixed new symbol. Reading  $f$  as denoting the false, the newly introduced clauses mimics the coherence conditions on total choices defined on CCL theories. Then, we modify the probabilistic assignment  $\mu$  to each atomic choice  $a$  in such a way that the marginal probability of  $a$  given the evidence  $f = \mathbf{f}$  (i.e.  $f$  is false) coincide with  $\mu(a)$ . Hence, the value  $\mu_I(Q)$  corresponds to the ProbLog computation of the marginal probability for the query  $Q$  given the evidence  $f = \mathbf{f}$ . Such task can be reduced to weighted model counting [11].

The above procedure should be adopted when coping with CCL theories having a single choice space as in the case of object ranking (see Section 6). For the general case, we suggest the following (outer) approximation scheme:

$$\underline{\mu}(Q) \geq \sum_{\omega_c \models Q} \underline{\mu}(\omega_c) = \sum_{\omega_c \models Q} \prod_{i=1}^k \underline{\mu}(\text{image}(c_i)), \quad (9)$$

where the inequality follows from Equation (7) by simply swapping the minimum and the sum, while the equality with the term on the right-hand side follows from the factorisation in Equation (4) and then by applying Equation (6).

*Example 7.* Consider the CCL theory  $\mathcal{T}_{\text{friends}}$  from Example 5. We calculate the lower success probability of query  $h$ . As the theory contains multiple choice spaces, we should

use Equation (9). However notice that  $\omega_i \models h$  exactly when  $i = 8$  (see Section 3) and therefore:

$$\underline{\mu}(\{h\}) = \underline{\mu}(\omega_8) = \underline{\mu}_1(\{nr, nc\}) \cdot \underline{\mu}_2(\{nw\}),$$

which gives the numerical value 0.32, as expected from Section 3. Analogously we get 0.40 for the corresponding upper success probability.

## 6 Empirical Analysis

In this section we report the results of a very first empirical analysis of the approach to CCL inference described in the previous section. Unlike most of the statistical models based on credal sets [3], CCL has no parameters directly affecting the *imprecision* level of the inferences (i.e., the difference between the upper and the lower probability of a query). A first important question is therefore whether or not, our relaxation of the standard independence assumptions in PLP, allows to obtain non-vacuous inferences from a query. Another point is whether or not available PSAT algorithms are able to solve instances induced by CCL theories.<sup>2</sup>

To achieve that we consider *object ranking*, that is the task of deriving a complete ranking over a set of  $n$  objects from a data set  $\mathcal{D}$  of complete rankings (if features are also considered, the term preference/label learning is used instead) [16]. We assume rankings not directly available: only the  $n^2$  *marginal* counts reporting how many times a certain object gets a certain rank are available. Note that, in general, complete information about the rankings cannot be recovered from these counts (Figure 2).

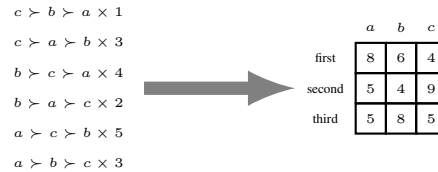


Fig. 2. From complete rankings to marginal counts.

Consider for instance a horse race. Call  $r_j$  the unary predicate denoting the property “ending the race in  $j$ -th position”, and  $h_i$  the constant associated to the  $i$ -th horse with  $i, j = 1, \dots, n$ . The ground atom  $r_j(h_i)$  means “horse  $i$  ended the race in  $j$ -th position”. Under our assumptions, we can learn from the data the (marginal) probabilities  $\alpha_{i,j} := \mu(r_j(h_i))$ . Object ranking can be based on a joint mass function over these  $n^2$  ground toms. This function should reproduce the marginal probabilities being also consistent with the obvious logical constraints (“one and only one horse ends the race in  $j$ -th

<sup>2</sup> Here we use the solver [12], freely available at <http://psat.sourceforge.net>.

position”, and “one and only one position is reached by the  $i$ -th horse at the end of the race”). These constraints are encoded in the following formula:

$$\varphi_{\text{ranking}} := \bigwedge_{j=1}^n \bigoplus_{i=1}^n r_j(h_i) \wedge \bigwedge_{i=1}^n \bigoplus_{j=1}^n r_j(h_i), \quad (10)$$

where  $\oplus$  is the exclusive disjunction (XOR). Algebraically, the task corresponds therefore to the specification of a joint mass function over  $n^2$  Boolean variables, i.e.,  $2^{n^2}$  probabilities. Only  $n!$  probabilities are non-zero as soon as we impose the constraints in Equation (10), while the  $n^2$  marginal probability induce an equal number of (linear) constraints. Thus, we might have non-unique specifications even with four objects only.

Consistently with what above, although complete rankings over  $n$  objects are available for training, we only learn the  $n^2$  marginal probabilities. This is done by smoothing the frequencies with a Laplace prior of equivalent size two. The corresponding CCL theory is defined as the program-free CCL theory  $\mathcal{T}_{\text{ranking}} := \langle \emptyset, \mathcal{C}, \mu \rangle$  with a single choice space, i.e.,  $\mathcal{C} = \{\{C_1, \dots, C_n\} \cup \{C'_1, \dots, C'_n\}\}$  with  $C_\ell = \{r_1(h_\ell), \dots, r_n(h_\ell)\}$  and  $C'_\ell = \{r_\ell(h_1), \dots, r_\ell(h_n)\}$ , and by stating  $\mu(r_j(h_i)) := \alpha_{i,j}$  for each  $i, j = 1, \dots, n$ . This eventually induces PSAT instances with the same probabilistic facts and logical constraints obtained by a CNF (*conjunctive normal form*) conversion of Equation (10).<sup>3</sup>

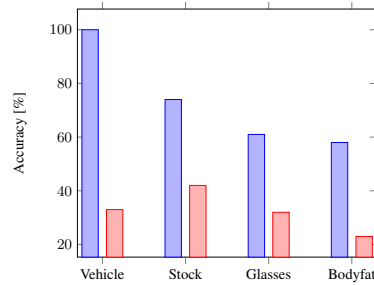
As a first benchmark, we use four classical UCI datasets (Vehicle, Stock, Glasses, and Bodyfat) with  $n = 4, 5, 6, 7$ . We evaluate pairwise preferences between pairs of objects. We write  $i' \succ i''$  to denote the fact that object  $i'$  has a higher ranking than object  $i''$ . To do that, we extend the theory  $\mathcal{T}_{\text{ranking}}$  by adding to the (initially empty) program  $\mathbf{P}$ , the following clauses:

$$\left\{ q \leftarrow r_{j'}(h_{i'}), r_{j''}(h_{i''}) \mid \begin{array}{l} j', j'' = 1, \dots, n \\ \text{s.t. } j' > j'' \end{array} \right\}, \quad (11)$$

where  $q$  is a new symbol. Hence, the query corresponding to the property  $i' \succ i''$  will be given by  $Q_{i' \succ i''} := \{q\}$ . For this query we compute both  $\underline{\mu}(Q_{i' \succ i''})$  and  $\overline{\mu}(Q_{i' \succ i''})$  and, on the basis of these values, we decide whether  $i' \succ i''$  or  $i'' \succ i'$ . Note that when coping with imprecise, interval-valued, inferences, a condition of indecision between the two options can be also observed.

This is the case if  $[\underline{\mu}(Q_{i' \succ i''}), \overline{\mu}(Q_{i' \succ i''})]$  overlaps the decision threshold .5, otherwise a clear preference is returned. This decision (denoted as CCL) is compared against the one based on the original data with the complete rankings regarded here as a ground truth and the one based on the marginal probabilities treated as independent (denoted as ICL). The ICL accuracy on the queries is evaluated separately on the pairs on which CCL is determinate and the ones on which CCL is indeterminate (i.e., indecision is returned). These results are in Figure 3. The separation between these two accuracies is clear: CCL becomes undecided on the tasks on which a less conservative approach would be less accurate, thus providing a more robust approach to the inferences.

<sup>3</sup> A XOR rewrites as a disjunction together with negations of pairwise conjunctions.



**Fig. 3.** ICL accuracy for preference queries on instances in which CCL is determinate (blue) and indeterminate (red).

## 7 Conclusions

In the last years, several works in the probabilistic logic tradition have proposed formalisms to explicitly deal with independence, see e.g. [1,5,18,23]. Our approach differs in the sense that, rather than attempting to combine probabilistic logics and probabilistic networks, is closer to logic programming. Indeed, we have introduced CCL, a conservative extension of Poole’s ICL in which both independent and non independent choices can be modelled. In the proposed setting, a theory specifies the (credal) set of all probability mass functions over least models compatible with the marginals on atomic choices. In guise of example, we applied CCL to object ranking and have shown how to infer the lower and upper probabilities of a query. In future work, on top of presenting more complex applications and deeper experiments of the proposed formalism, we plan to compare our work with related approaches such as for instance the one discussed by [13] in the context of probabilistic databases.

## References

1. Andersen, K.A., Hooker, J.N.: Bayesian logic. *Decision Support Systems* **11**(2), 191–210 (1994)
2. Apt, K.R., Bezem, M.: Acyclic programs. *New Generation Computing* **9**(3), 335–363 (1991)
3. Augustin, T., Coolen, F., de Cooman, G., Troffaes, M.: *Introduction to imprecise probabilities*. John Wiley & Sons (2014)
4. Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.A., Guo, J., Li, P., Riddell, A.: Stan: a probabilistic programming language. *Journal of Statistical Software* **20**, 1–37 (2016)
5. Cozman, F., de Campos, C., da Rocha, J.C.: Probabilistic logic with independence. *International Journal of Approximate Reasoning* **49**(1), 3–17 (2008)
6. Cozman, F., di Ianni, L.: Probabilistic satisfiability and coherence checking through integer programming. *International Journal of Approximate Reasoning* **58**, 57–70 (2015)
7. Cozman, F., Mauá, D.: On the semantics and complexity of probabilistic logic programs. *Journal of Artificial Intelligence Research* **60**, 221–262 (2017)
8. De Raedt, L.: Applications of probabilistic logic programming. In: *International Conference on Inductive Logic Programming* (2015)

9. De Raedt, L., Kimmig, A.: Probabilistic (logic) programming concepts. *Machine Learning* **100**(1), 5–47 (2015)
10. De Raedt, L., Kimmig, A., Toivonen, H.: Problog: a probabilistic Prolog and its application in link discovery. In: *International Joint Conference on Artificial Intelligence*. pp. 2462–2467 (2007)
11. Fierens, D., Van den Broeck, G., Renkens, J., Shterionov, D., Gutmann, B., Thon, I., Janssens, G., De Raedt, L.: Inference and learning in probabilistic logic programs using weighted Boolean formulas. *Theory and Practice of Logic Programming* **15**(3), 358–401 (2015)
12. Finger, M., De Bona, G.: Probabilistic satisfiability: logic-based algorithms and phase transition. In: *International Joint Conference on Artificial Intelligence*. pp. 528–533 (2011)
13. Flesca, S., Furfaro, F., Parisi, F.: Consistency checking and querying in probabilistic databases under integrity constraints. *Journal of Computer and System Sciences* **80**(7), 1448–1489 (2014)
14. Fuhr, N.: Probabilistic Datalog: a logic for powerful retrieval methods. In: *International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 282–290. ACM (1995)
15. Fuhr, N.: Probabilistic Datalog: implementing logical information retrieval for advanced applications. *Journal of the Association for Information Science and Technology* **51**(2), 95–110 (2000)
16. Fürnkranz, J., Hüllermeier, E.: Preference learning: an introduction. In: *Preference learning*, pp. 1–17. Springer (2010)
17. Georgakopoulos, G., Kavvadias, D., Papadimitriou, C.: Probabilistic satisfiability. *Journal of Complexity* **4**(1), 1–11 (1988)
18. Haenni, R., Romeijn, J.W., Wheeler, G., Williamson, J.: *Probabilistic logics and probabilistic networks*, vol. 350. Springer Science & Business Media (2010)
19. Janhunen, T.: Representing normal programs with clauses. In: *European Conference on Artificial Intelligence*. pp. 358–362. IOS Press (2004)
20. Levi, I.: *The enterprise of knowledge: an essay on knowledge, credal probability, and chance*. MIT press (1983)
21. Lukasiewicz, T.: Probabilistic logic programming. In: *European Conference on Artificial Intelligence*. pp. 388–392 (1998)
22. Lukasiewicz, T.: Probabilistic description logic programs. *International Journal of Approximate Reasoning* **45**(2), 288–307 (2007)
23. Michels, S., Hommersom, A., Lucas, P.J., Velikova, M.: A new probabilistic constraint logic programming language based on a generalised distribution semantics. *Artificial Intelligence* **228**, 1–44 (2015)
24. Ng, R., Subrahmanian, V.S.: Probabilistic logic programming. *Information and Computation* **101**(2), 150–201 (1992)
25. Nilsson, N.J.: Probabilistic logic. *Artificial Intelligence* **28**(1), 71–87 (1986)
26. Poole, D.: Probabilistic Horn abduction and Bayesian networks. *Artificial Intelligence* **64**(1), 81–129 (1993)
27. Poole, D.: The independent choice logic for modelling multiple agents under uncertainty. *Artificial Intelligence* **94**(1), 7–56 (1997)
28. Sato, T.: A statistical learning method for logic programs with distribution semantics. In: *International Conference on Logic Programming*. pp. 715–729 (1995)
29. Vennekens, J., Verbaeten, S.: *Logic Programs with Annotated Disjunctions*. Technical Report CW 368, K.U.Leuven (2003)
30. Vennekens, J., Verbaeten, S., Bruynooghe, M.: Logic programs with annotated disjunctions. In: *International Conference on Logic Programming*. pp. 431–445. Springer (2004)