
Distribution Semantics of Probabilistic Logic Programs

There are two major approaches to integrating probabilistic reasoning into logical representations: **distribution semantics** and **maximum entropy**.

- Is there a **sound interpretation of ASP**, in particular of **stable models** to any of the two approaches above?
- Under such interpretations, **what probabilistic problems can be addressed?** MARG? MLE? MAP? Decision?
- **What is the relation to other logic and uncertainty approaches?** Independent Choice Logic? Abduction? Stochastic Logic Programs? etc.

Maximum Entropy Summary

ME approaches annotate uncertainties only at the level of a logical theory. That is, they assume that the predicates in the BK are labelled as either true or false, but the label may be incorrect.

These approaches are not based on logic programming, but rather on first-order logic. Consequently, the underlying semantics are different: rather than consider proofs, **these approaches consider models or groundings of a theory**.

This difference primarily changes what uncertainties represent. For instance, Markov Logic Networks (MLN) represent programs as a set of weighted clauses. The weights in MLN do not correspond to probabilities of a formula being true but, intuitively, to a log odds between a possible world (an interpretation) where the clause is true and a world where the clause is false.

Distribution Semantics

DS approaches explicitly annotate uncertainties in BK. To allow such annotation, they extend Prolog with two primitives for stochastic execution: probabilistic facts and annotated disjunctions.

Probabilistic facts are the most basic stochastic primitive and they take the form of logical facts labelled with a probability p . **Each probabilistic fact represents a Boolean random variable that is true with probability p and false with probability $1 - p$.** *This is very close to facts in ASP. A “simple” syntax extension would be enough to capture probability annotations. **What about the semantics of such programs?***

Whereas probabilistic facts introduce non-deterministic behaviour on the level of facts, annotated disjunctions introduce non-determinism on the level of clauses. Annotated disjunctions allow for multiple literals in the head, where only one of the head literals can be true at a time.

Core Distribution Semantics

- Let F be a set of **grounded probabilistic facts** and $P : F \rightarrow [0, 1]$.

For example, F and P result from

0.9 :: edge(a, c) .

0.7 :: edge(c, b) .

0.6 :: edge(d, c) .

0.9 :: edge(d, b) .

- **Facts are assumed marginally independent:**

$$\forall a, b \in F, P(a \wedge b) = P(a)P(b).$$

- The **probability of** $S \subseteq F$ is

$$P_F(S) = \prod_{f \in S} P(f) \prod_{f \notin S} (1 - P(f)).$$

- Let R be a set of **definite clauses** defining further (new) predicates.

For example, R is

path(X, Y) :- edge(X, Y) .

path(X, Y) :- edge(X, Z), path(Z, Y) .

- Any combination $S \cup R$ has an **unique least Herbrand model**,

$$W = M(S \cup R).$$

- The set of ground facts S is often called an **explanation** of the world $W = M(S \cup R)$.
- The **probability of a possible world** W is

$$P(W) = \sum_{S \subseteq F: W=M(S \cup R)} P_F(S).$$

- The **probability of a ground proposition** p is (defined as) the probability that p has a proof:

$$P(p) = \sum_{S: S \cup R \vdash p} P_F(S) = \sum_{W: p \in W} P(W).$$

- A proposition may have many proofs in a single world $M(S \cup R)$. Without further guarantees, the probabilities of those proofs cannot be summed. The definition above avoids this problem. **** > For example, a proof of path(a, b) employs (only) the facts edge(a, c) and edge(c, b) i.e. these facts are an explana-

tion of path (a, b). Since these facts are (marginally) independent, **the probability of the proof is**

$$\begin{aligned} P(\text{path}(a, b)) &= P(\text{edge}(a, c) \wedge \text{edge}(c, b)) \\ &= P(\text{edge}(a, c)) \times P(\text{edge}(c, b)) \\ &= 0.9 \times 0.7 \\ &= 0.63. \end{aligned}$$

> This is the only proof of path (a, b) so $P(\text{path}(a, b)) = 0.63$. >> On the other hand, since path (d, b) has two explanations, edge (d, b) and edge (d, c), edge (c, b):

$$\begin{aligned} P(\text{path}(d, b)) &= P(\text{edge}(d, b) \vee (\text{edge}(d, c) \wedge \text{edge}(c, b))) \\ &= 0.9 + 0.6 \times 0.7 - 0.9 \times 0.6 \times 0.7 \\ &= 0.942. \end{aligned}$$

- With this **semantics of the probability of a possible world**, the probability of an arbitrary proposition is still hard to compute, because of the *disjunct-sum* problem.

Because computing the probability of a fact or goal under the distribution semantics is hard, systems such as Prism [4] and Probabilistic Horn Abduction (PHA) [8] impose additional restrictions that can be used to improve the efficiency of the inference procedure.

The key assumption is that the explanations for a goal are mutually exclusive, which overcomes the disjoint-sum problem. If the different explanations of a fact do not overlap, then its probability is simply the sum of the probabilities of its explanations. This directly follows from the inclusion-exclusion formulae as under the exclusive-explanation assumption the conjunctions (or intersections) are empty (*Statistical Relational Learning*, Luc De Raedt and Kristian Kersting, 2010)

This assumption/restriction is quite *ad-hoc* and overcoming it deserves further inquiry.

- Reading Fabio Gagliardi Cozman, Denis Deratani Mauá, *The joy of Probabilistic Answer Set Programming: Semantics - complexity, expressivity, inference* (2020) gave a big boost securing my initial intuition.
- The problem is in disjunctive clauses, such as the one in the following example.

```
% prob(a) = 0.7  
a ; -a.  
b ; c :- a.
```

- More specifically, in the example above, **the explanation a entails two possible worlds, ab and ac. How to assign probability of each one?**

Assigning Probabilities on “Multiple Worlds per Explanation” Scenarios

Clause Annotations

Assign a probability to each case in the head of the clause. For example, annotate $P(b|a) = 0.8$.

This case needs further study on the respective consequences, specially concerning the joint probability distribution.

- In particular, $P(b|a) = 0.8$ entails $P(\neg b|a) = 0.2$. But no stable model asserts $\neg b$.
- Also, unless assumed the contrary, the independence of b and c is unknown.

Learn from Observations

Leave the probabilities uniformly distributed and use observations to update their values.

Under this approach, how do observations affect the assigned probabilities?

- In particular, how to update the probabilities of the worlds $a \ b$ and $a \ c$ given observations such as a , b , ab , $a\neg b$, $\neg ab$ or abc ?
- **Idea.** Update uniformly the probability of the **worlds compatible with the observation**.
 - For example, the observation a is compatible with both worlds ab and ac ; $a\neg b$ is compatible with all worlds that contain a and do not contain b , etc.
 - **But** in this case, the effect of observing a is the same as observing ab , ac or even abc ; all these observations are compatible with the worlds ab and ac .

Leave One World Out

Use strict compatibility, add another world and update its probability on incompatible observations; The probability of this world measures the model's shortness.