Zugzwang
Logic and Artificial Intelligence
Why this title?

Francisco Coelho
fc@uevora.pt
Universidade de Évora and
NOVA**LINCS**

Bruno Dinis
bruno.dinis@uevora.pt
Universidade de Évora

May 2, 2023

Abstract

(*rewrite*) A major limitation of logical representations in real world applications is the implicit assumption that the background knowledge is perfect. This assumption is problematic if data is noisy, which is often the case. Here we aim to explore how answer set program specifications with probabilistic facts can lead to characterizations of probability functions *Why is this important? Is this what 'others in sota' are trying do to?* on the specification's domain.

1 Introduction and Motivation

(*Define and/or give references to all necessary concepts used in the paper*)

(*state of the art; references*) Answer set program (ASP) is a logic programming paradigm based on the stable model (SM) semantics of normal (logic) program (NP) that can be implemented using the latest advances in SAT solving technology. Unlike ProLog, ASP is a truly declarative language that supports language constructs such as disjunction in the head of a clause, choice rules, and hard and weak constraints.

(*references*) The distribution semantics (DS) is a key approach to extend logical representations with probabilistic reasoning. Probabilistic facts (PFs) are the most basic stochastic DS primitive and they take the form of logical facts, a, labelled with a probability, p, such as $p :: a$; Each PF represents

a boolean random variable that is true with probability p and false with probability $\overline{p} = 1 - p$. A (consistent) combination of the PFs defines a total choice (TC) $c = \{p :: a, \ldots\}$ such that

$$\mathrm{P}(C = c) = \prod_{a \in c} p \prod_{a \notin c} \overline{p}. \tag{1}$$

Our goal is to extend this probability, from TCs, to cover the *specification* domain. We use the term "specification" as set of rules and facts, plain and probabilistic, to decouple it from any computational semantics, implied, at least implicitly, by the term "program". We can foresee at least two key applications of this extended probability:

1. Support probabilistic reasoning/tasks on the specification domain.

2. Also, given a dataset and a divergence measure, the specification can be scored (by the divergence w.r.t. the *empiric* distribution of the dataset), and weighted or sorted amongst other specifications. These are key ingredients in algorithms searching, for example, optimal specifications of a dataset.

Our idea to extend probabilities starts with the stance that a specification describes an *observable system* and that observed events must be related with the SMs of that specification. From here, probabilities must be extended from total choices to SMs and then from SMs to any event.

Extending probability from TCs to SMs faces a critical problem, illustrated by the example in section 2, concerning situations where multiple SMs, ab and ac, result from a given TC, a, but there is not enough information to assign a single probability to each SM. We propose to address this issue by using algebraic variables to describe that lack of information and then estimate the value of those variables from empirical data.

In a related work, [5], epistemic uncertainty (or model uncertainty) is considered as a lack of knowledge about the underlying model. This lack of knowledge can be mitigated via further observations. This seems to presuppose a Bayesian approach to imperfect knowledge in the sense that having further observations allows to improve/correct the model. Indeed, the approach in the paper uses Beta distributions in order to be able to learn the full distribution. This approach seems to be specially fitted to being able to tell when some probability lies beneath some given value. (*Our approach seems to be similar in spirit. If so, we should mention this in the introduction.*)

(*Discuss the least informed strategy and the corolary that stable models should be conditionally independent on the total choice.*)

(*Give an outline of the paper.*)

2 A simple but fruitful example

(*Write an introduction to the section*)

Example 1. *Consider the following specification*

$$\begin{aligned} 0.3 :: a,\\ b \vee c \leftarrow a. \end{aligned} \tag{2}$$

This specification has three stable models, $\overline{a}, ab$ *and* ac *(see Figure 1). While it is straightforward to set* $P(\overline{a}) = 0.7$*, there is* no further information *to assign values to* $P(ab)$ *and* $P(ac)$*. Assuming that the stable model (SM) are (probabilistically) independent, we can use a parameter* ~~λ~~*/*θ *such that*

$$\begin{aligned} P(ab) &= 0.3\theta,\\ P(ac) &= 0.3(1-\theta). \end{aligned}$$

While uncertainty is inherent to the specification it can be mitigated with the help of a dataset: the parameter θ can be estimated from a empirical distribution (*or we can have a distribution of* θ).

In summary, if an ASP specification is intended to describe some observable system then:

1. Observations can be used to estimate the value of the parameters (such as θ above and others entailed from further clauses).
2. (*What about the case where we already know a distribution of* θ*?*)
3. With a probability set for the stable models, we want to extend it to all the events of the ~~specification~~/domain.
4. This extended probability can then be related to the *empirical distribution*, using a probability divergence, such as Kullback-Leibler; and the divergence value used as a *performance* measure of the specification with respect to the observations.
5. If that specification is only but one of many possible candidates then that performance measure can be used, *e.g.* as fitness, by algorithms searching (optimal) specifications of a dataset of observations.

   (*Expand this:*) If observations are not consistent with the models of the specification, then the specification is wrong and must be changed.

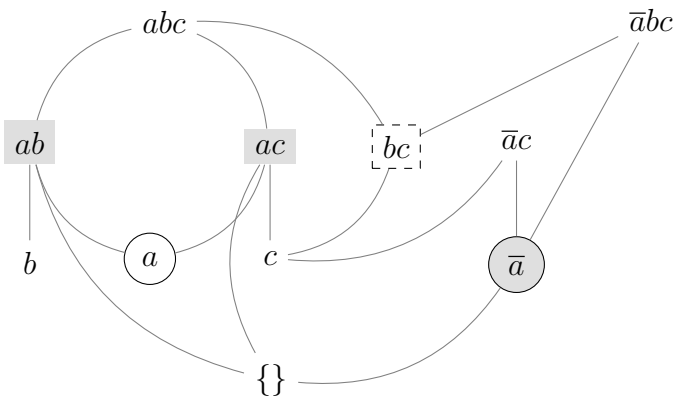

Figure 1: Events related to the stable models of example 1. The circle nodes are the TCs and the shaded nodes are the SMs.

Currently, we are addressing the problem of extending a probability function (possibly using parameters such as θ), defined on the SMs of a specification, to all the events of that specification. Of course, this extension must satisfy the Kolmogorov axioms of probability so that probabilistic reasoning is consistent with the ASP specification.

Conditional independence of stable worlds asserts a least informed strategy that we discussed in the introduction and make explicit here:

Assumption 1. *Stable models are conditionally independent, given their total choices.*

The stable models ab, ac from example 1 result from the clause $b \vee c \leftarrow a$ and the total choice a. These formulas alone impose no relation between b and c (given a), so none should be assumed. Dependence relations are further discussed in section 3.1.

3 Extending Probabilities

(*Somewhere, we need to shift the language from extending* probabilities *to extending* measures)

The diagram in fig. 1 illustrates the problem of extending probabilities from total choice nodes to stable models and then to general events in a *node-wise* process. This quickly leads to coherence problems concerning probability, with no clear systematic approach — Instead, weight extension can be based in the relation an event has with the stable models.

Given an ASP specification Introduce also the sets mentioned below *how?* , we consider the *atoms* $a \in \mathcal{A}$ and *literals*, $z \in \mathcal{L}$, *events* $e \in \mathcal{E} \iff e \subseteq \mathcal{L}$ and *worlds* $w \in \mathcal{W}$ (consistent events), *total choices* $c \in \mathcal{C} \iff c = a \vee \neg a$ and *stable models* $s \in \mathcal{S} \subset \mathcal{W}$.

Our path starts with a perspective of stable models as playing a role similar to *prime* factors. The stable models of a specification are the irreducible events entailed from that specification and any event must be ~~interpreted~~/considered under its relation with the stable models.

(*Introduce a structure with worlds, events, and stable models*) *seems irrelevant* This focus on the SMs leads to the following definition:

Definition 1. *A* stable structure *is a pair* (A, S) *where* A *is a set of atoms* can be extracted from S. *and* S *is a set of consistent events over* A.

(*expand this text to explain how the stable models form the basis of the equivalence relation*).

Definition 2. *The* stable core (SC) *of the event* $e \in \mathcal{E}$ *is*

$$[\![e]\!] := \{s \in \mathcal{S} \mid e \subseteq s \vee s \subseteq e\} \tag{3}$$

We now define an equivalence relation, $\sim$, so that two events are related if they are either both inconsistent or both consistent with the same stable core.

Definition 3. *For a given specification, let* $u, v \in \mathcal{E}$. *The equivalence relation* $\sim$ *is defined by*

$$u \sim v : \iff u, v \notin \mathcal{W} \vee \left(u, v \in \mathcal{W} \wedge [\![u]\!] = [\![v]\!]\right). \tag{4}$$

Observe that the minimality of stable models implies that, in eq. (3), either e is a stable model or one of $e \subseteq s, s \subseteq e$ is never true. This relation defines a partition of the events space, where each class holds a unique relation with the stable models. In particular, we denote each class by:

$$[e]_\sim = \begin{cases} \bot := \mathcal{E} \setminus \mathcal{W} & \text{if } e \notin \mathcal{E} \setminus \mathcal{W}, \\ \{u \in \mathcal{W} \mid [\![u]\!] = [\![e]\!]\} & \text{if } e \in \mathcal{W}, \end{cases} \tag{5}$$

The stable core defines a *canonical* representative of each class:

Theorem 1. *Let* $e \in \mathcal{E}$ *and* $[\![e]\!] = \{s_1, \ldots, s_n\} \subseteq \mathcal{S}$. *Then*

$$[e]_\sim = [s_1 \cup \cdots \cup s_n]_\sim. \tag{6}$$

We simplify the notation with $[s_1, \ldots, s_n]_\sim := [s_1 \cup \cdots \cup s_n]_\sim$. (This only works for consistent $s_1, \ldots, s_n$: $\{\{\}\} = [\overline{a}, ab, ac]_\sim \neq [a\overline{a}bc]_\sim = \bot$.)

Proof. (*tbd*) □

The subsets of the stable models, together with $\bot$, form a set of representatives. Consider again Example 1. As previously mentioned, the stable models are $\mathcal{S} = \overline{a}, ab, ac$ so the quotient set of this relation is $[\mathcal{E}]_\sim$:

$$\left\{\bot, \emptyset, [\overline{a}]_\sim, [ab]_\sim, [ac]_\sim, [\overline{a}, ab]_\sim, [\overline{a}, ac]_\sim, [ab, ac]_\sim, [\overline{a}, ab, ac]_\sim\right\} \tag{7}$$

For example,

$$\begin{array}{llll} [\{\}]_\sim = [\overline{a}, ab, ac]_\sim, & [a]_\sim = [ab, ac]_\sim, & [b]_\sim = [ab]_\sim, & [\overline{b}]_\sim = \emptyset, \\ [a\overline{c}]_\sim = \emptyset, & [ab]_\sim = \emptyset, & [b\overline{b}]_\sim = \bot, & [\overline{a}b]_\sim = [\overline{a}]_\sim, \\ [\overline{bc}]_\sim = \emptyset, & [abc]_\sim = [ab, ac]_\sim, & [a\overline{b}c]_\sim = [ac]_\sim, & [\overline{a}bc]_\sim = [\overline{a}]_\sim, \end{array}$$

- Since all events within an equivalence class are in relation with a specific set of stable models, *weights, including probability, should be constant within classes*:
$$\forall u \in [e]_\sim \left(\mathrm{P}(u) = \mathrm{P}(e)\right).$$
- So, instead of dealing with $64 = 2^6$ events, we need only to handle $9 = 2^3 + 1$ classes, well defined in terms of combinations of the stable models.

(*Check adaptation*) Our path to set a probability measure on $\mathcal{E}$ has two phases:

- Extending the probabilities, *as weights*, of the total choices to events.
- Normalization of the weights.

The "extension" phase, traced by equations (1) and (8 — 13), starts with the weight (probability) of total choices, $\mu(c) = \mathrm{P}(C = c)$, expands it to stable models, $\mu(s)$, and then, within the equivalence relation from Equation (4), to (general) events, $\mu(e)$, including (consistent) worlds.

Total Choices. Using (1), this case is given by

$$\mu(c) = \mathrm{P}(C = c) = \prod_{a \in c} p \prod_{a \notin c} \overline{p} \tag{8}$$

Stable Models. Each total choice c, together with the rules and the other facts of a specification, defines a set of stable models associated with that choice, that we denote by S_c.

Given a stable model $s \in \mathcal{S}$, a total choice c, and variables/values $\theta_{s,c} \in [0,1]$,

$$\mu(s,c) := \begin{cases} \theta_{s,c} & \text{if } s \in S_c \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

such that $\sum_{s \in S_c} \theta_{s,c} = 1$.

Classes. Each class is either the inconsistent class, $\bot$, or is represented by some set of stable models.

- **Inconsistent Class.** The inconsistent class contains events that are logically inconsistent. Since these events should never be observed:

$$\mu(\bot, c) := 0. \tag{10}$$

- **Independent Class.** A world that neither contains nor is contained in a stable model describes a case that, according to the specification, should never be observed. So the respective weight is set to zero:

$$\mu\big(\emptyset, c\big) := 0. \tag{11}$$

- **Other Classes.** The extension must be constant within a class, its value should result from the elements in the stable core, and respect the assumption 1:

$$\mu\big(\llbracket s_1, \ldots, s_n \rrbracket, c\big) := \prod_k \mu(s_k, c)\,. \tag{12}$$

Events. Each (general) event e is in the class defined by its stable core, $\llbracket e \rrbracket$. So, we set:

$$\mu(e,c) := \mu\big(\llbracket e \rrbracket, c\big)\,. \tag{13}$$

Equation (9) expresses the *specification's* lack of knowledge about the weight assignment, when a single total choice entails more than one stable model. In this case, how to distribute the respective weights? Our ~~answer~~/proposal to /address this problem consists in assigning an unknown weight, $\theta_{s,c}$, conditional depending??? on the total choice, c, to each stable model s. This approach allows the expression of an unknown quantity and future estimation, given observed data.

Equation (12) results from conditional independence of stable models.

3.1 Dependence

Our basic assertion about dependence relations between atoms of the underlying system is that they can be *explicitly expressed in the specification*. And, in that case, they should be.

For example, a dependence relation between b and c can be expressed by $b \leftarrow c \wedge d$, where d is an atomic choice that explicitly expresses the dependence between b and c. One would get, for example, a specification such as

$$0.3 :: a, b \vee c \leftarrow a, 0.2 :: d, b \leftarrow c \wedge d.$$

with stable models $\overline{ad}, \overline{a}d, a\overline{d}b, a\overline{d}c, adb$.

The interesting case is the subtree of the total choice ad. Notice that no stable model s contains adc because (i) adb is a stable model and (ii) if $adc \subset s$ then $b \in s$ so $adb \subset s$.

Following equations (**??**) and (**??**) What are these equations? this entails

$$\begin{cases} \mathrm{P}(W = adc \mid C = ad) = 0, \\ \mathrm{P}(W = adb \mid C = ad) = 1 \end{cases}$$

which concentrates all probability mass from the total choice ad in the adb branch, including the node $W = adbc$. This leads to the following cases:

x	$\mathrm{P}(W = x \mid C = ad)$
ad	1
adb	1
adc	0
$adbc$	1

so, for $C = ad$,

$$\begin{aligned} \mathrm{P}(W = b) &= \frac{2}{4} \\ \mathrm{P}(W = c) &= \frac{1}{4} \\ \mathrm{P}(W = bc) &= \frac{1}{4} \\ &\neq \mathrm{P}(W = b)\,\mathrm{P}(W = c) \end{aligned}$$

i.e. the events $W = b$ and $W = c$ are dependent and that dependence results directly from the segment $0.2 :: d, b \leftarrow c \wedge d$ in the specification.

Why does this not contradict Assumption 1?

Todo

Prove the four world cases (done), support the product (done) and sum (tbd) options, with the independence assumptions.

4 Developed Example

We continue with the specification from Equation (2).

Step 1: Total Choices. The total choices, and respective stable models, are

Total Choice (c)	$\mathrm{P}(C = c)$	Stable Models (s)
a	0.3	ab and ac.
$\overline{a} = \neg a$	$\overline{0.3} = 0.7$	$\overline{a}$.

Step 2: Stable Models. Suppose now that

Stable Models (s)	Total Choice (c)	$\mathrm{P}(S = c \mid C = c)$
$\overline{a}$	1.0	$\overline{a}$.
ab	0.8	a.
ac	$0.2 = \overline{0.8}$	a.

Step 3: Worlds. Following equations **??** — **??** we get:

Occ. (o)	S.M. (s)	Relation	T.C. (c)	$\mathrm{P}(W = w)$
$\emptyset$	all	contained	a, $\overline{a}$	1.0
a	ab, ac	contained	a	$0.8 \times 0.3 + 0.2 \times 0.3 = 0.3$
b	ab	contained	a	$0.8 \times 0.3 = 0.24$
c	ac	contained	a	$0.2 \times 0.3 = 0.06$
$\overline{a}$	$\overline{a}$	stable model	$\overline{a}$	$1.0 \times 0.3 = 0.3$
$\overline{b}$	none	independent	none	0.0
$\overline{c}$	none	...		
ab	ab	stable model	a	0.24
ac	ac	stable model	a	0.06
$a\overline{b}$	none	...		
$a\overline{c}$	none	...		
$\overline{a}b$	$\overline{a}$	contains	$\overline{a}$	1.0
$\overline{a}c$	$\overline{a}$	...		
$\overline{a}\overline{b}$	$\overline{a}$	...		
$\overline{a}\overline{c}$	$\overline{a}$	...		
abc	ab, ac	contains	a	$0.8 \times 0.2 = 0.016$

5 Final Remarks

(*develop this section.*)

- The measure of the inconsistent events doesn't need to be set to 0 and, maybe, in some cases, it shouldn't.

- The physical system might have *latent* variables, possibly also represented in the specification. These variables are never observed, so observations should be concentrated *somewhere else*.

Acknowledgements

This work is supported by NOVA**LINCS** (UIDB/04516/2020) with the financial support of FCT.IP.

References

[1] Fabio Gagliardi Cozman and Denis Deratani Mauá. "The joy of probabilistic answer set programming: semantics, complexity, expressivity, inference". In: *International Journal of Approximate Reasoning* 125 (2020), pp. 218–239.

[2] Andrew Cropper et al. "Inductive logic programming at 30". In: *Machine Learning* 111.1 (2022), pp. 147–172.

[3] Martin Gebser et al. "Answer set solving in practice". In: *Synthesis lectures on artificial intelligence and machine learning* 6.3 (2012), pp. 1–238.

[4] Fabrizio Riguzzi. *Foundations of probabilistic logic programming: Languages, semantics, inference and learning*. CRC Press, 2022.

[5] Victor Verreet et al. "Inference and learning with model uncertainty in probabilistic logic programs". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 36. 9. 2022, pp. 10060–10069.