Probabilistic Logics and Probabilistic Networks

Jan-Willem Romeijn, Philosophy, Groningen
Jon Williamson, Philosophy, Kent

ESSLLI 2008

Course Page:
http://www.kent.ac.uk/secl/philosophy/jw/2006/progicnet/ESSLLI.htm

Where are we?

Yesterday: an introduction into Progic. Plan for today:

Standard semantics: Kolmogorov probability, interval-valued probability, sets of probability.

Representation and interpretation: validity of the standard semantics.

Networks in the standard semantics: credal networks and credal sets, algorithms for credal sets, using nets in the standard semantics.

And where we are heading tomorrow:

Evidential probability: weak probabilistic logical sytems, partial entailment.

Probabilistic argumentation: logical and probabilistic information, representation and interpretations.

Fiducial probability: classical statistical inference, fiducial arguments, representation and interpretation in the Progicnet schema.

Networks for fiducial probability: functional models and networks, aiding fiducial inference with networks.

1 Standard semantics

The standard probabilistic semantics (or standard semantics for short) is the most basic semantics for probabilistic logic.

1.1 Kolmogorov probabilities

We define probability as a measure function over an algebra, a family of sets.

Definition 1.1 (Probability Space) *A probability space is a tuple* $(\Omega, \mathcal{F}, P)$*, where* Ω *is a sample space of elementary events,* $\mathcal{F}$ *is a* σ*-algebra of subsets of* Ω*, and* $P : \mathcal{F} \to [0,1]$ *is a probability measure satisfying the Kolmogorov axioms:*

P1. $P(E) \geq 0$*, for all* $E \in \mathcal{F}$*;*

P2. $P(\Omega) = 1$*;*

P3. $P(E_1 \cup E_2 \cup \cdots) = \sum_i P(E_i)$*, for any countable sequence* $E_1, E_2, \ldots$ *of pairwise disjoint events* $E_i \in \mathcal{F}$*.*

Probability over a language

We can connect algebras to languages as follows.

Definition 1.2 (Probability Structure) *A probability structure is a quadruple $M = (\Omega, \mathcal{F}, P, I)$, where $(\Omega, \mathcal{F}, P)$ is a probability space and I is an interpretation function associating each elementary event $\omega \in \Omega$ with a truth assignment on the propositional variables Φ in a language $\mathcal{L}$ such that $I(\omega, A) \in \{true, false\}$ for each $\omega \in \Omega$ and for every $A, B, C, \ldots \in \Phi$.*

Since P is defined on events rather than sentences, we need to link events within a probability structure M to formulas in Φ.

Proposition 1.3 *For arbitrary propositional formulas φ and ψ,*

1. $[\![\varphi \wedge \psi]\!]_M = [\![\varphi]\!]_M \cap [\![\psi]\!]_M$,
2. $[\![\varphi \vee \psi]\!]_M = [\![\varphi]\!]_M \cup [\![\psi]\!]_M$,
3. $[\![\neg\varphi]\!]_M = \Omega \setminus [\![\varphi]\!]_M$.

This is exactly the opposite of defining the probability space in terms of the language, as discussed yesterday. Both perspectives come down to the same thing. We use capital letters to denote both propositions within $\mathcal{L}$ and the corresponding events within $\mathcal{F}$.

Probability theory as a logic

The axioms of probability theory and the interpretation of the algebra as a language already provide a complete logic of sharp probability values.

- Fagen, Halpern, and Megiddo provide a proof theory for the standard semantics on a propositional language. Deciding satisfiability is NP-complete.
- There are obstacles to providing a proof theory for probability logics on more expressive languages, because a language that is expressive enough afford probabilistic reasoning about probability statements will extend beyond the complexity of first-order reasoning about real numbers and natural numbers.
- There is no need here to interpret probability as physical, epistemic, subjective, objective or what have you, in the same way as that there is no immediate need to interpret truth values in classical logic.

1.2 Interval-Valued probabilities

The same idea of probabilistic logic is at the basis of Bayesian inference, which originates in Ramsey and De Finetti, and is explicitly advocated by Howson, Morgan, Halpern, and many others.

Decision theory uses this logic in the background to represent uncertain belief, namely as willingness to bet. But what if someone wants to call the whole bet off? What if there is a difference between the proces at which you want to buy and sell a bet?

Unsharp probabilities
The sharp values are by no means a necessity for this type of probabilistic logic. Sets of probability emerge rather naturally.

> **Example 1.4** *If* $P(E) = 0.6$ *and* $P(F) = 0.7$*, and this is all that is known about* E *and* F*, we can derive that* $0.7 \leq P(E \cup F) \leq 1$*, and that* $0.3 \leq P(E \cap F) \leq 0.6$*.*

This example can be generalized to the following proposition.

Proposition 1.5 (Carpet shifting) *If $P(E)$ and $P(F)$ are defined in M, then:*

1. *$P(E \cap F)$ lies within the interval*

$$[\max(0, (P(E) + P(F)) - 1), \min(P(E), P(F))], \text{ and}$$

2. *$P(E \cup F)$ lies within the interval*

$$[\max(P(E), P(F)), \min(P(E) + P(F), 1)].$$

Interval-valued probability assignments follow rather naturally from the standard semantics using sharp assignments. We extend the use of interval-valued assignments to premises using inner and outer measures.

Example 1.6 *Suppose $F \notin \mathcal{F}$, our probability structure, but F is logically related to events in $\mathcal{F}$: F contains E and F is contained within G. Both E and G are within $\mathcal{F}$.*

- *When there is no measurable event contained in F that dominates E, then E is a kernel event for F.*
- *When every measurable event containing F dominates G, then G is a covering event for F.*

The measures of F's kernel and cover then yield non-trivial bounds on F with respect to M, since otherwise $P(F)$ would be undefined.

Inner and outer measures
We can formalise these kernels and covers by means of inner and outer measures.

> **Definition 1.7 (Inner and Outer Measure)** *Let $\mathcal{F}$ be a subalgebra of an algebra $\mathcal{F}'$, $P : \mathcal{F} \to [0, 1]$ a probability measure defined on the space $(\Omega, \mathcal{F}, P)$, and E an arbitrary set in $\mathcal{F}' \setminus \mathcal{F}$. Then define the* inner measure $\underline{P}$ induced by P *and the* outer measure $\overline{P}$ induced by P *as:*
>
> $\underline{P}(F) = \sup\{P(G) : E \subseteq F, E \in \mathcal{F}\}$ (inner measure of F)*;*
>
> $\overline{P}(F) = \inf\{P(G) : G \supseteq F, G \in \mathcal{F}\}$ (outer measure of F)*.*

Some properties of inner and outer measures:

- $\underline{P}(E \cup F) \geq \underline{P}(E) + \underline{P}(F)$, when E and F are disjoint (superadditivity)
- $\overline{P}(E \cup F) \leq \overline{P}(E) + \overline{P}(F)$, when E and F are disjoint (subadditivity)
- $\underline{P}(E) = 1 - \overline{P}(\overline{E})$
- $\underline{P}(E) = \overline{P}(E) = P(E)$, if $E \in \mathcal{F}$

1.3 Sets of probabilities

We turn to the relationship between inner- and outer measures, and sets of probabilities.

Theorem 1.8 (Horn and Tarski, 1948) *Suppose P is a measure on a (finitely additive) probability structure* $M = (\Omega, \mathcal{F}, P, I)$ *such that* $\mathcal{F} \subseteq \mathcal{F}'$. *Define* $\mathbb{P}$ *as the set of all extensions* P' *of P to* $\mathcal{F}'$. *Then for all* $E \in \mathcal{F}'$:

(i.) $\underline{P}(E) = \underline{\mathbb{P}}(E) \equiv \inf\{P'(E) : P' \in \mathbb{P}\}$, *and*

(ii.) $\overline{P}(E) = \overline{\mathbb{P}}(E) \equiv \sup\{P'(E) : P' \in \mathbb{P}\}$.

From interval values to sets
The Horn-Tarski result links the inner-measure of an event E to the lower probability $\underline{\mathbb{P}}(E)$ for a particular set of probability measures over $\mathcal{F}'$, namely those which correspond to P on all the events $E \in \mathcal{F}$. For the probability P' over events $F \in \mathcal{F}' \setminus \mathcal{F}$, there are no further restrictions.

- While every interval-valued probability can be associated with a set of probability distributions over a richer algebra, the converse does not hold.
- For a set of probability distributions to be representable as an interval-valued assignment, the set needs to be convex.

Convexity
Generally, a set X is called convex if and only if X is closed under all mappings that take two elements from X into a new element that lies in between them, for some suitable notion of in-betweenness.

> **Example 1.9** *Picture two points, x, y, that are members of a real line segment X defined on a linear space Z. The straight line segment from x to y is the set* $\{\lambda x + (1-\lambda)y : \lambda \in [0,1]\}$*. This equation specifies the convex hull of the set* $\{x, y\}$*, and the endpoints x and y are its extremal points.*

A *credal set* $\mathbb{K}$ is a convex set of probability functions, i.e. $P_1, P_2 \in \mathbb{K}$ implies $\lambda P_1 + (1-\lambda)P_2 \in \mathbb{K}$ for all $\lambda \in [0,1]$.

> **Theorem 1.10 (Walley 1991)** *If* $\mathbb{K}$ *is a convex set of probability functions, then there is an interval-valued probability P such that*

(iii.) $\underline{P}(E) = \underline{\mathbb{K}}(E) \equiv \inf\{P(E) : P \in \mathbb{K}\}$*; and*

(iv.) $\overline{P}(E) = \overline{\mathbb{K}}(E) \equiv \sup\{P(E) : P \in \mathbb{K}\}$.

2 Representation and interpretation

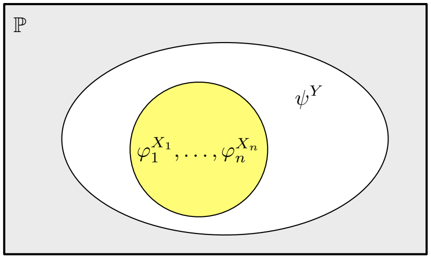
Recall the Progicnet schema,

$$\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \approx\!\!\!\!\models \psi^Y.$$

How can we relate the standard semantics to this semantical entailment? What does it represent, and how can it be interpreted?

2.1 Validity

Validity in the standard semantics comes down to the inclusion of the set of probability functions satisfying the premises, $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n}$, in the set of assignments satisfying the conclusion, ψ^Y.



Some remarks on what expressions can appear as premise:

- Normally premises are of the form a^X, presenting a direct restriction of the probability for a to X, that is, $P(a) \in X \subset [0,1]$, where X might also be a sharp probability value $P(a) = x$.
- Another possible premise is $(a|b)^x$, meaning that $P(a|b) = x$, so

$$(a|b)^x \iff \forall y \in (0,1] : b^y, (a \wedge b)^{xy}.$$

- Other premises concern independence relations between propositional variables, e.g., $A \perp\!\!\!\perp B$, meaning that $P(A,B) = P(A)P(B)$, so

$$\exists x, y \in [0,1] : a^x, b^y, (a \wedge b)^{xy}, (a \wedge \bar{b})^{x(1-y)}, (\bar{a} \wedge b)^{(1-x)y},$$

  or $A \perp\!\!\!\perp B|C$, which means that $P(A,B|C) = P(A|C)P(B|C)$.

2.2 Interpretation

In the interpretation of $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \models \psi^Y$, the premises are constraints to convex sets of probability assignments, and the conclusion is always a convex constraint, which is guaranteed to hold if all the constraints of the premises hold. Note that the problem is to find the *smallest* Y for which the equation holds.

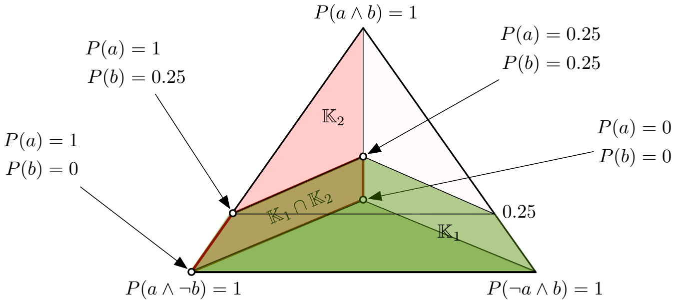


Figure 1: The set of all possible probability measures P, depicted as a tetrahedron, together with the credal sets obtained from the given probabilistic constraints.

Example 2.1 *Consider two premises* $(a \wedge b)^{[0,0.25]}$ *and* $(a \vee \neg b)^1$*. The set of all possible probability measures* P *can be depicted by a tetrahedron (3-simplex) with maximal probabilities for the state descriptions* $a \wedge b$*,* $a \wedge \neg b$*,* $\neg a \wedge b$*, and* $\neg a \wedge \neg b$ *at each of its four extremities. In this simplex we have depicted the convex sets* $\mathbb{K}_1$ *and* $\mathbb{K}_2$ *obtained from the constraints* $P(a \wedge b) \in [0, 0.25]$ *and* $P(a \vee \neg b) = 1$*. From the (convex) intersection* $\mathbb{K}_1 \cap \mathbb{K}_2$*, which includes all probability functions that satisfy both constraints, we see that* $Y = [0, 1]$ *attaches to the conclusion* a*, whereas* $Y = [0, 0.25]$ *attaches to* b*.*

3 Networks for the standard semantics

Recall the recipe of the Progicnet programme:

- **Representation**: Formulate a question of the form of the Progicnet schema.
- **Interpretation**: Decide upon appropriate semantics.
- **Network Construction**: Construct a probabilistic network to represent the models of the left-hand side of the Progicnet schema.
- **Inference**: Apply the common machinery of the network algorithm to answer the question posed in the first step.

The common machinery was introduced yesterday. We using probabilistic networks as a calculus for probabilistic logic: they play a role that is analogous to that of proof in classical logic.

3.1 Credal networks and credal sets

A Bayesian network consists of a directed acyclic graph (DAG) on a set of variables A_i, together with the probability functions $P(A_i|PAR_i)$ for each variable A_i conditional on its parents PAR_i in the graph.

Example 3.1 *An example of a Bayesian network is provided by the graph below together with corresponding conditional probability functions specified by:*

$$P(a_1) = 0.1, \qquad P(a_2|a_1) = 0.9, \qquad P(a_3|a_2) = 0.4,$$
$$P(a_2|\bar{a}_1) = 0.3, \qquad P(a_3|\bar{a}_2) = 0.2.$$

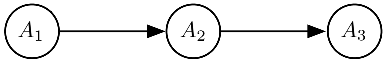


Figure 2: Directed acyclic graph in a Bayesian network.

A credal network is a generalised Bayesian network, representing a set of probability functions.

Example 3.2 *A credal network may have the local conditional distributions constrained by*

$$P(a_1) \in [0.1, 0.2], \qquad P(a_2|a_1) = 0.9, \qquad P(a_3|a_2) \in [0.4, 1],$$
$$P(a_2|\bar{a}_1) \in [0.3, 0.4], \qquad P(a_3|\bar{a}_2) = 0.2.$$

Extensions of the network

Further independence assumptions are made to determine this credal set from the credal network, i.e., from the graph and the bounds on probability assignments.

Natural Extension. No independence assumptions at all. The credal set contains all probability assignments that comply with the constraints of the given net.

Strong Extension. Restriction to the convex hull of the *extremal points* defined by the local conditional credal sets, and the conditional independencies determined from the graph via the Markov condition.

Complete Extension. Assume that *each* probability function in the joint credal set satisfies the Markov condition with respect to the given graph. In that case the credal network can be thought of as the set of Bayesian networks based on the given DAG whose conditional probabilities satisfy the constraints imposed by the conditional credal sets.

Not all probability assignments in the strong extension need to comply with the independence relations suggested by the graph. Only the extremal points necessarily satisfy these contraints, and it is only these extremal points that are employed in the derivations.

Strong extension of credal net
The notion of extension is tied up with the notion of convexity. Depending on the choice of the coordinates for the space of probability functions, the strong extension of a network may be identical to the complete extension.

> **Example 3.3** *Consider the credal network on variables A and B defined by the empty graph and local credal sets $P(a) \in [0.3, 0.7]$, $P(b) \in [0, 0.5]$. This credal network defines four extremal points, each representable by a Bayesian network on the empty graph, implying via the Markov condition that $A \perp\!\!\!\perp B$. The strong extension of this credal network is simply the convex hull of these four extremal points.*

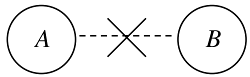


Figure 3: No dependence relation between variables A and B, represented by a graph in which nodes A and B are not connected.

Normal coordinates for the credal set from the example

To determine the credal set from the example, we may specify the possible distributions over the four atomic states $a \wedge b$, $a \wedge \neg b$, $\neg a \wedge b$, and $\neg a \wedge \neg b$, and take as coordinates $x_1 = P(a \wedge b)$, $x_2 = P(a \wedge \neg b)$, $x_3 = P(\neg a \wedge b)$, and $x_4 = P(\neg a \wedge \neg b)$.

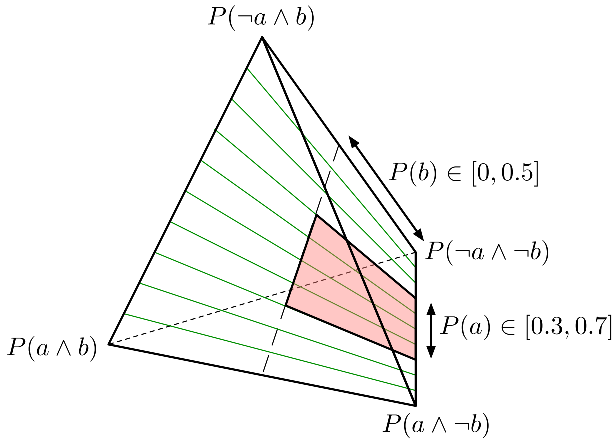


Figure 4: Atomic-state coordinates.

Alternative coordinates for credal sets

One can depict the full distribution over four possibilities using different coordinates. Represent the probability function by coordinates

$$\alpha = P(a) = 1 - P(\bar{a}),$$
$$\beta_1 = P(b|a) = 1 - P(\bar{b}|a),$$
$$\beta_0 = P(b|\bar{a}) = 1 - P(\bar{b}|\bar{a}).$$

This results in the cube on the right hand side of the figure below. Variables A and B are independent, $A \perp\!\!\!\perp B$, iff $\beta_0 = \beta_1$.

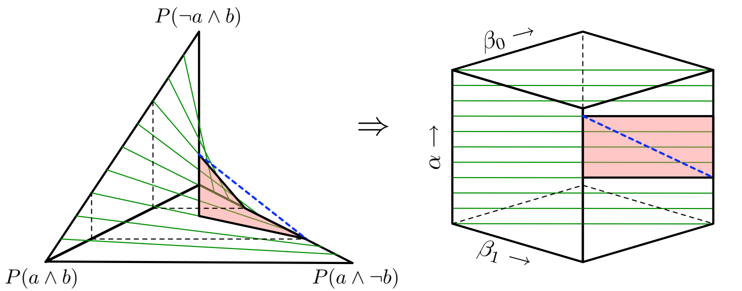


Figure 5: Two different coordinate systems.

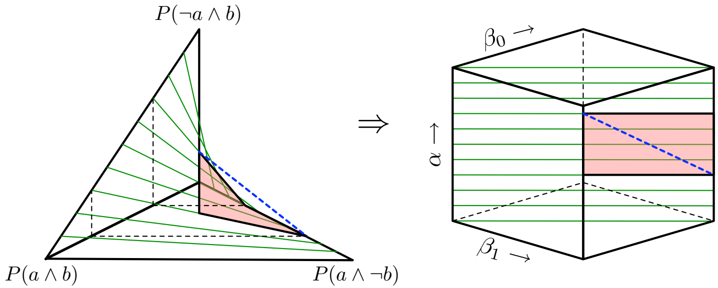


Figure 6: Two different coordinate systems.

With the new coordinates the convex hull of the extremal points lies on the diagonal surface: all probability functions in the strong extension satisfy the independence relation imposed by the graph. So the strong extension coincides with the complete extension!

Choosing the coordinate system

It is important when using a strong extension to specify the coordinate system.

- In probabilistic logic it is typical to use the former, atomic-state coordinate system (x_1, x_2, x_3, x_4). This ensures that if the extremal points all satisfy a linear constraint of the form $P(\varphi) \in X$ then so will every member of the strong extension.
- The network-structure coordinate system $(\alpha, \beta_0, \beta_1)$ is often typical in discussions of causality, where it is more important that the strong extension preserve independence relations than linear constraints.

Parameterised credal sets

It is important to note that there are credal sets that cannot be represented by a credal network of the above form.

> **Example 3.4** *Consider the credal set that consists of all probability functions on* $V = \{A, B\}$ *that satisfy the constraints* $P(ab) = 0.3$ *and* $P(b|\bar{a}) = 0$*. This implies* $P(a) \in [0.3, 1]$*, but the possible values of* $P(b|a)$ *depend on the particular value* x *of* $P(a)$ *in* $[0.3, 1]$*, since* $P(b|a) = 0.3/x$*. While it is true that* $P(b|a) \in [0.3, 1]$*, the value of* $P(b|a)$ *can not be chosen independently of that of* $P(a)$*.*

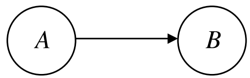


Figure 7: Full graph for A and B.

To represent the credal set of the above example we need to appeal to what we call a *parameterised credal network*. The graph is as before but now the conditional credal sets involve a parameter: $x := P(a) \in [0.3, 1]$, $P(b|a) = 0.3/x$, and $P(b|\bar{a}) = 0$.

3.2 Algorithms for credal nets

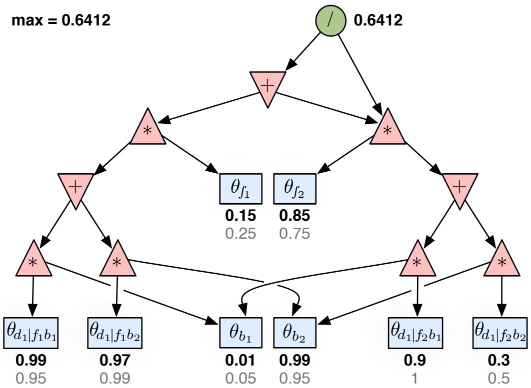
We briefly discuss two algorithms concerning credal networks: the hill-climbing algorithm and an algorithm for turning independence assumptions and bounds into credal networks with bounds on network coordinates.

Calculating intervals for complex sentences
For details on the network algorithm, see the slides of lecture 1 and the Progicnet paper. An overview of the algorithm:

- Despite the huge number of available algorithms for probabilistic networks, only a few are valuable candidates for solving problems of the form of the Progicnet schema.

- Given that the conclusion ψ on the right hand side of the Progicnet schema may be an arbitrary logical formula, the algorithm must be able to compute probabilities $P(\psi)$ or respective intervals $[\underline{P}(\psi), \overline{P}(\psi)]$ of arbitrarily complex queries.

- The general idea of Progicnet is to compile a probabilistic network, splitting up inference into an expensive *compilation phase* and a cheap *query answering phase*.



- The compilation step is to represent the topological structure of a probabilistic network by propositional sentences. The resulting propositional encoding of the network is then transformed into a generic graphical representation called *deterministic, decomposable negation normal form*.
- We restrict the infinitely large search space to a finite search space which involves all possible combinations of lower and upper bounds of the given intervals.
- To approximate lower and upper bounds when querying a credal network, we can apply the hill-climbing approach to the search space defined by the credal set's extremal points.

Determining a credal network

Say that we want to deal with an inference problem in the standard semantics, and that the premises include certain explicit probabilistic independencies. An overview of the algorithm:

Input: a set $V = \{A_1, \ldots, A_M\}$ of propositional variables, a set $\mathcal{I}$ of probabilistic independence relations of the form $A_i \perp\!\!\!\perp A_j$ or $A_i \perp\!\!\!\perp A_j | A_k$, and premises $\varphi_1^{X_1}, \ldots, \varphi_N^{X_N}$ involving those variables.

Construction of a graph: application of the adapted PC-algorithm of Pearl to find the smallest network $\mathcal{G}$ that satisfies the set of independence relations $\mathcal{I}$.

Derivation of constraints on network coordinates: deriving the restrictions to the conditional probability assignments in the network $\mathcal{G}$ by a search along the boundaries of the credal set.

Output: a graph $\mathcal{G}$ and a set of extremal points S in terms of the network coordinates, corresponding to the independence relations $\mathcal{I}$ and the premises $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n}$.

Finding the extremal points

This can be a laborious task: if in total there are $\dim(\Gamma) = N$ coordinates involved in the constraints, there are at most $N!2^N$ extremal points. But generally the number of points will be much smaller. To see what the algorithm comes down to in practice, consider the following example.

Example 3.5 *Let A_j for $j = 1, 2, 3$ be the propositional variables V, and let $I = \{A_2 \perp\!\!\!\perp A_3 \mid A_1\}$ be the set of independence relations. Pearl's adapted PC-algorithm gives us the graph shown below. The network coordinates are*

$$P(a_1) = \gamma_1,$$
$$P(a_2|a_1^i) = \gamma_{2|1}^i,$$
$$P(a_3|a_1^i) = \gamma_{3|1}^i.$$

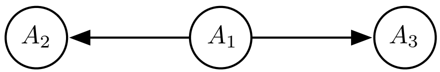


Figure 8: The graph expressing that $A_2 \perp\!\!\!\perp A_3 \mid A_1$.

Now imagine that we have the following constraints on the variables in the graph: $\varphi_1^{X_1} = ((a_1 \vee \neg a_2) \to a_3)^{[1/3, 2/3]}$ *and* $\varphi_2^{X_2} = (a_3)^{1/2}$*. We then apply the algorithm for the derivation of constraints on the coordinates* γ_k *as follows.*

- *We collect and order the rank numbers of the propositional variables appearing in the premises in two vectors* $s_1 = \langle 1, 2, 3 \rangle$ *and* $s_2 = \langle 1, 3 \rangle$*.*
- *We rewrite the expression* $\varphi_1 = (a_1 \vee \neg a_2) \to a_3$ *in a disjunctive normal form, and compute the probabilities in terms of the network coordinates,*

Disjunct	*Vector* e_1	*Probability* Γ_1^m
$\neg a_1 \wedge \neg a_2 \wedge \neg a_3$	$e_1(1) = \langle 0, 0, 0 \rangle$	$(1-\gamma_1)(1-\gamma_{2\|1}^0)(1-\gamma_{3\|1}^0)$
$a_1 \wedge \neg a_2 \wedge \neg a_3$	$e_1(2) = \langle 1, 0, 0 \rangle$	$\gamma_1(1-\gamma_{2\|1}^1)(1-\gamma_{3\|1}^1)$
$a_1 \wedge a_2 \wedge \neg a_3$	$e_1(3) = \langle 1, 1, 0 \rangle$	$\gamma_1\gamma_{2\|1}^1(1-\gamma_{3\|1}^1)$
$a_1 \wedge \neg a_2 \wedge a_3$	$e_1(4) = \langle 1, 0, 1 \rangle$	$\gamma_1(1-\gamma_{2\|1}^1)\gamma_{3\|1}^1$
$a_1 \wedge a_2 \wedge a_3$	$e_1(5) = \langle 1, 1, 1 \rangle$	$\gamma_1\gamma_{2\|1}^1\gamma_{3\|1}^1$
$(a_1 \vee \neg a_2) \to a_3$		$(1-\gamma_1)(1-\gamma_{2\|1}^0)(1-\gamma_{3\|1}^0) + \gamma_1$

and similarly for a_3*, whose probability is expressed as* $\Gamma_2^m = (1-\gamma_1)\gamma_{3|1}^0 + \gamma_1\gamma_{3|1}^1$*.*

- *Impose the constraints* $P(\varphi_i) = \sum_m \Gamma_i^m \in X_i$*, collecting them in the following system of linear equations:*

$$(1-\gamma_1)(1-\gamma^0_{2|1})(1-\gamma^0_{3|1}) + \gamma_1 \geq 1/3$$
$$(1-\gamma_1)(1-\gamma^0_{2|1})(1-\gamma^0_{3|1}) + \gamma_1 \leq 2/3$$
$$(1-\gamma_1)\gamma^0_{3|1} + \gamma_1\gamma^1_{3|1} = 1/2.$$

- *We compute the extremal points by finding solutions to this system of equations with maximal and minimal values for the respective coordinates.*
- *Once all extremal points have been found, the algorithm halts. We output the graph depicted above and a set of extremal points.*

The derivation of constraints on the coordinates of the network is computationally costly, especially for dense graphs. But remember that after the derivation of graph and constraints, further inference is relatively fast. The compilation of the network is an investment into future inferences.

3.3 Using nets in the standard semantics

In the standard semantics, the inferences run from credal sets, determined by constraints on the probability assignments, to other credal sets. How can networks be used in these inferences?

The poverty of standard semantics

- Networks can be used to speed up the inference process. But it is only when we assume the so-called strong or complete extension of a credal network that we can fully employ their computational and conceptual advantages.
- Taking the strong or complete extension of an incomplete graph amounts to making specific independence assumptions. The idea is to use the specifics of the domain of application, as captured by the semantics of the inference, to motivate these assumptions.
- The standard semantics does not have any specific domain of application, and so the additional independence assumptions cannot be motivated from semantic considerations.
- All the constraints on the probability assignments derive from explicit premises. The standard semantics on itself imposes no further constraints.

Dilation

The use of credal sets leads to a rather awkward result when it comes to updating the probability assignments to new information by conditionalisation. The probability interval may get, what is called, dilated.

> **Example 3.6** *Consider the credal set given by* $P(a) = 1/2$ *and* $P(b) = 1/2$*, but without an independence relation between A and B. In terms of the network coordinates* $\langle \alpha, \beta_0, \beta_1 \rangle$*, we can write* $P(a) = \alpha = 1/2$ *and* $P(b) = \alpha\beta_1 + (1-\alpha)\beta_0 = 1/2$*, so that the credal set is defined by* $\alpha = 1/2$ *and* $\beta_1 = 1 - \beta_0$*. This set is represented by the thick line in the figure below.*

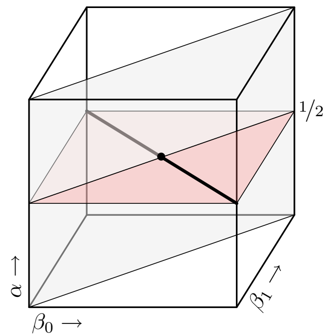


Figure 9: The credal set represented in the coordinate system $\langle \alpha, \beta_0, \beta_1 \rangle$, as the thick line in the square $\alpha = 1/2$. For each of the points in this credal set we have $P(b) = 1/2$, but both β_0 and β_1 can take any value in $[0, 1]$.

Imagine that we decide to observe whether a or $\bar{a}$. Whether we learn a or $\bar{a}$, the probability for b changes from to the whole interval after the update: the probability for b becomes $P(b|a) = \beta_1 \in [0, 1]$ or $P(b|\bar{a}) = \beta_0 \in [0, 1]$. The probability for b is dilated by the partition $\{a, \bar{a}\}$.

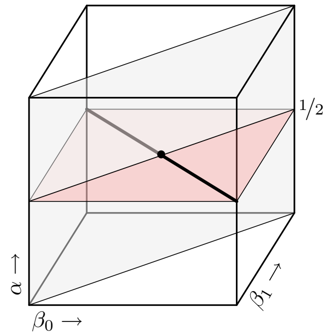


Figure 10: The credal set associated with independence between A and B is depicted as a shaded area, $\beta_0 = \beta_1$. By the independence assumption, the credal set becomes the singleton $\{\langle 1/2, 1/2, 1/2 \rangle\}$.

This awkward situation comes about because the restrictions on the probability assignment do not tell us anything about the relations between A and B. Dilation is avoided if we assume an independence relation between A and B. Learning whether a or $\bar{a}$ then leaves the probability of b invariant.