Credal Sets Approximation by Lower Probabilities: Application to Credal Networks

Alessandro Antonucci[a] and Fabio Cuzzolin[b]

(a) Istituto Dalle Molle di Studi sull'Intelligenza Artificiale
Galleria 2, Via Cantonale, Manno-Lugano (Switzerland)
alessandro@idsia.ch

(b) Department of Computing, Oxford Brookes University
Wheatley Campus, Wheatley, Oxford (United Kingdom)
fabio.cuzzolin@brookes.ac.uk

Abstract. *Credal sets* are closed convex sets of probability mass functions. The *lower probabilities* specified by a credal set for each element of the power set can be used as constraints defining a second credal set. This simple procedure produces an outer approximation, with a bounded number of extreme points, for general credal sets. The approximation is optimal in the sense that no other lower probabilities can specify smaller supersets of the original credal set. Notably, in order to be computed, the approximation does not need the extreme points of the credal set, but only its lower probabilities. This makes the approximation particularly suited for *credal networks*, which are a generalization of Bayesian networks based on credal sets. Although most of the algorithms for credal networks updating only return lower posterior probabilities, the suggested approximation can be used to evaluate (as an outer approximation of) the *posterior credal set*. This makes it possible to adopt more sophisticated decision making criteria, without having to replace existing algorithms. The quality of the approximation is investigated by numerical tests.

Key words: Imprecise probability, lower probabilities, credal sets, credal networks, interval dominance, maximality.

1 Introduction

Consider the problem of modelling a condition of uncertainty about the state of a categorical variable. In the Bayesian framework, this problem is faced by assessing the probability of each possible outcome, thus specifying a (single) probability mass functions. Yet, there are situations where the assessment of a precise probabilistic value for each outcome can be difficult. In such cases, multiple assessments (e.g., through intervals) can be considered, leading to the specification of sets of, instead of single, probability mass functions. These sets, which are required to be convex by compelling rationality criteria related to the behavioural interpretation of probability, are called *credal sets* (CS, [1]) and represent a very general class of uncertainty models [2].

A (partial) characterization of a CS can be obtained by considering its *lower probabilities*, i.e., the infima over all the probability mass functions in the CS, of the probabilities assigned to the elements of the power set.[1] These bounds correspond to a number of constraints satisfied by the original CS. Yet, this is only a partial characterization, as the set of probability mass functions consistent with these constraints is in general a proper superset of the original CS.

This simple procedure is intended in this paper as an outer approximation for CSs. The approximation is proved to be optimal, in the sense that no other outer approximation based on lower probabilities can specify more informative CSs. Furthermore, the number of extreme points[2] of the approximating CS is bounded by the factorial of the number of possible values of the variable.

Notably, in order to achieve this approximation, only the lower probabilities of the CS are needed, while an explicit enumeration of its extreme points is not necessary. This makes the approximation particularly suited for *credal networks* [3], which are a generalization of Bayesian networks based on CSs. In fact, most of the algorithms for credal networks updating only return the lower posterior probabilities, and not the posterior CS. We show that the outer approximation of this posterior CS, as returned by the transformation we consider, can be computed by means of these standard algorithms. This makes it possible to adopt more refined criteria for making decisions on a credal network, without the need to replace existing algorithms. Although the outer approximation can eventually lead to over-cautious decisions, we show by extensive numerical simulations that this happens only in a minority of cases.

The paper is organized as follows. First, we review some background information about CSs (Section 2.1), lower probabilities (Section 2.2), and credal networks (Section 2.3). Then, in Section 3, we provide a characterization of the class of CSs associated to lower probabilities and we detail the transformation to obtain an outer approximation of a CS by means of its lower probabilities. Also some theoretical results characterizing the proposed technique are reported. The transformation is applied to credal networks in Section 4. Numerical tests to investigate the quality of the approximation are in Section 5. Conclusions and future outlooks are finally in Section 6, while the proofs of the theorems can be found in the Appendix.

2 Background

2.1 Credal Sets

Let X denote a generic variable, taking values in a finite set $\mathcal{X} := \{x^{(1)}, \ldots, x^{(n)}\}$. A probability mass function over X, which is a nonnegative real map over $\mathcal{X}$ normalized to one, will be denoted by $P(X)$. A *credal set* (CS) over X, which is a convex set of probability mass functions over X, will be be denoted by $K(X)$.

[1] The power set of a variable is made by all the subsets of its set of possible values.

[2] A point in a convex set is *extreme* if it cannot be obtained as a convex combination of other points in this set.

The *extreme points* of $K(X)$ (see Footnote 2) are denoted as $\mathrm{ext}[K(X)]$. Here we only consider CSs with a finite number of extreme points, i.e., such that $|\mathrm{ext}[K(X)]| < +\infty$.[3] Geometrically, a CS is therefore a polytope in the probability simplex, and can be equivalently specified through an explicit enumeration of its extreme points (V-representation) and a finite set of linear constraints (H-representation). Unlike the V-representation, which is clearly uniquely defined, different H-representations can specify the same CS. The notation $\overline{K}(X)$ is used for the vacuous CS, i.e., the (convex) set of all the probability mass functions over X. It is easy to note that $|\mathrm{ext}[\overline{K}(X)]| = |\mathcal{X}|$.

2.2 Lower Probabilities

A conjugate pair of *lower/upper probability* operators [1] is defined as a pair $(\underline{P}, \overline{P})$ of nonnegative real maps over the power set $2^{\mathcal{X}}$, such that: (i) $\underline{P}(\emptyset) = 0$; (ii) the operators are respectively super- and sub-additive, i.e.,

$$\underline{P}(A \cup B) \geq \underline{P}(A) + \underline{P}(B)$$

$$\overline{P}(A \cup B) \leq \overline{P}(A) + \overline{P}(B),$$

for each $A, B \in 2^{\mathcal{X}}$; (iii) the following conjugacy relation holds for each $A \in 2^{\mathcal{X}}$

$$\overline{P}(A) = 1 - \underline{P}(\mathcal{X} \setminus A). \tag{1}$$

According to (1), the operator $\overline{P}$ is completely determined by its conjugate $\underline{P}$ (and vice versa). In this paper, we only consider lower probability operators.

2.3 Credal Networks

Consider a collection of categorical variables $X_1, \ldots, X_v$.[4] Let these variables be in one-to-one correspondence with the nodes of a directed acyclic graph, and assume that this graph depicts conditional independence relations among the variables according to the Markov condition. In the Bayesian framework, this implies the following factorization for the joint probability $P(x_1, \ldots, x_v) = \prod_{i=1}^{v} P(x_i|\mathrm{pa}(X_i))$, where $\mathrm{Pa}(X_i)$ is the joint variable made of the *parents* of X_i according to the graph. This implicitly defines a *Bayesian network* over X.

In order to define a *credal network* [3] over the same variables and the same graph, we simply leave each conditional probability mass function $P(X_i|\mathrm{pa}(X_i))$ free to vary in a conditional CS $K(X_i|\mathrm{pa}(X_i))$. This defines a set of joint probability mass functions, whose convexification is a joint CS $K(X_1, \ldots, X_v)$, to be called the *strong extension* of the credal network. Each extreme point of the strong extension factorizes as the joint probability mass function of a Bayesian network, and its conditional probability mass functions are extreme points of the conditional CSs $K(X_i|\mathrm{pa}(X_i))$.

[3] The notation $|S|$ is used for the cardinality of the set S.

[4] The background information in this section is particularly brief for sake of space. More insights on credal networks can be found in [3] or [4].

A typical inference problem to be addressed in a credal network is *updating*, i.e., given some evidence x_E about a set of variables X_E, evaluate the lower probability $\underline{P}(x|x_E)$ (with respect to the strong extension) for each possible value $x \in \mathcal{X}$ of a variable of interest X. Despite its hardness [5], various algorithms for this problem has been developed ([6] for a survey and [7,8] for recent advances).

3 Credal Sets Associated to Lower Probabilities

Given a lower probability operator $\underline{P}$, let us consider the CS of its consistent probability mass functions, i.e.,

$$K_{\underline{P}}(X) := \left\{ P(X) \in \overline{K}(X) \left| \sum_{x \in A} P(x) \geq \underline{P}(A), \forall A \in 2^{\mathcal{X}} \right. \right\}. \tag{2}$$

The following result (conjectured by Weichselberger and proved by Wallner in [9]) provides a characterization of the the maximum number of extreme points of the CS in (2):

$$|\text{ext}[K_{\underline{P}}(X)]| \leq |\mathcal{X}|!, \tag{3}$$

this being true for each lower probability operator $\underline{P}$ defined as in Section 2.2. Note that, in the case of belief functions, stronger results on the form of the vertices can be proven [10].

Example 1. *Given a ternary variable X, consider the following CS*

$$K(X) = \text{CH} \left\{ \begin{bmatrix} .90 \\ .05 \\ .05 \end{bmatrix}, \begin{bmatrix} .10 \\ .40 \\ .50 \end{bmatrix}, \begin{bmatrix} .20 \\ .20 \\ .60 \end{bmatrix}, \begin{bmatrix} .20 \\ .70 \\ .10 \end{bmatrix}, \begin{bmatrix} .80 \\ .05 \\ .15 \end{bmatrix}, \begin{bmatrix} .45 \\ .25 \\ .30 \end{bmatrix}, \begin{bmatrix} .05 \\ .80 \\ .15 \end{bmatrix} \right\},$$

where CH *denotes the convex hull operator, while probability mass functions are denoted as vertical arrays. Standard techniques (e.g., [11]) can be used to verify that none of these seven probability mass functions is a convex combination of the remaining six, and hence* $|\text{ext}[K(X)]| = 7$.

Example 1 violates (3). This simply proves that not any CS can be obtained from a lower probability operator as in (2). The class of CSs associated with lower probability operators should be therefore regarded as a special class of CSs.[5] The idea of this paper is that this class is sufficiently large to provide a reasonable approximation of general CSs (at least from the point of view of decision making based on them, see Section 4).

In fact, the lower probabilities of a CS define a lower probability operator, which can be employed indeed to define a (new) CS. The whole procedure is formalized as follows.

[5] The only exception is the case of CSs over binary variables. If general CSs can have an arbitrary number of extreme points, a CS over a binary variable has at most two extreme points. In fact, for binary variables, any CS can be associated with a lower probability operator.

Transformation 1. *Given a CS $K(X)$, consider its extreme points* $\mathrm{ext}[K(X)]$ *(i.e., its unique, V-representation). Then, for each $A \in 2^{\mathcal{X}}$, compute the* lower probability*:*[6]

$$\underline{P}_K(A) := \min_{P(X)\in\mathrm{ext}[K(X)]} \sum_{x\in A} P(x). \tag{4}$$

Finally, consider the CS $\tilde{K}(X)$ associated as in (2) with the lower probability operator in (4).[7]

It is straightforward to verify that (4) specifies a lower probability operator as in Section 2.2. Thus, given a generic non-empty CS $K(X)$, Transformation 1 always returns a non-empty CS $\tilde{K}(X)$. The following is an example of the application of this transformation.

Example 2. *The application of Transformation 1 to the CS $K(X)$ in Example 1 returns a CS $\tilde{K}(X)$ whose H-representation, according to (2), is:*

$$\begin{cases} P(\{x^{(1)}\}) \leq .05 \\ P(\{x^{(2)}\}) \leq .05 \\ P(\{x^{(3)}\}) \leq .05 \\ P(\{x^{(1)}\} \cup \{x^{(2)}\}) \leq .40 \\ P(\{x^{(1)}\} \cup \{x^{(3)}\}) \leq .20 \\ P(\{x^{(2)}\} \cup \{x^{(3)}\}) \leq .10. \end{cases}$$

from which (see [11]) the following V-representation follows:

$$\tilde{K}(X) = \mathrm{CH}\left\{ \begin{bmatrix} .05 \\ .35 \\ .60 \end{bmatrix}, \begin{bmatrix} .05 \\ .80 \\ .15 \end{bmatrix}, \begin{bmatrix} .15 \\ .80 \\ .05 \end{bmatrix}, \begin{bmatrix} .35 \\ .05 \\ .60 \end{bmatrix}, \begin{bmatrix} .90 \\ .05 \\ .05 \end{bmatrix} \right\}.$$

Thus, as expected, (3) is now satisfied. Fig. 1 depicts the polytopes associated with $K(X)$ and $\tilde{K}(X)$ on the same probability simplex.

A characterization of this transformation is provided by the following result.

Theorem 1. *Consider a CS $K(X)$. Let $\underline{P}_K$ denote the corresponding lower probability operator as in (4), and $\tilde{K}(X)$ the output of Transformation 1. Then:*

(i) $K(X) \subseteq \tilde{K}(X)$*;*
(ii) $K(X) = \tilde{K}(X)$ *if and only if a lower probability operator $\underline{P}'$ such that $K_{\underline{P}'}(X) = K(X)$ exists;*
(iii) A lower probability operator $\underline{P}' \neq \underline{P}_K$ such that $K(X) \subseteq K_{\underline{P}'}(X) \subseteq \tilde{K}(X)$ cannot exist.

It is worth noting that, for its application, Transformation 1 does not need the extreme points of the CS, but only its lower probabilities. This feature suggests a possible application to credal networks, which is detailed in the next section.

[6] The minimum in (4) is the same we obtain by minimizing over the whole CS [12].

[7] We prefer to avoid the somehow heavy notation $K_{\underline{P}_K}(X)$.

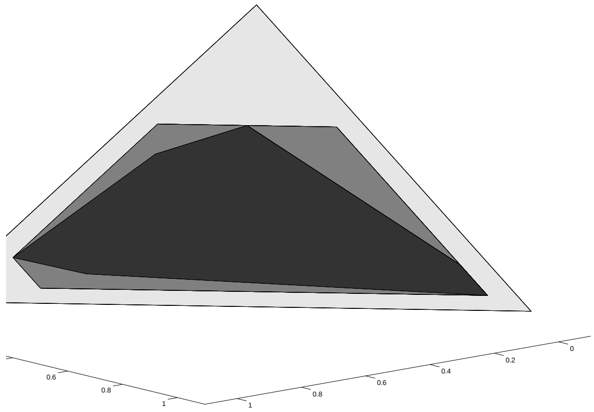

Fig. 1. The CS $K(X)$ as in Example 1 (dark gray), and the output $\tilde{K}(X)$ of Transformation 1 as in Example 2 (medium gray) on the same simplex (light gray).

4 Computing Posterior Credal Sets in a Credal Network

As noted in Section 2.3, a typical inference task to be discussed on a credal network is *updating* knowledge about a variable of interest, after the observation of some evidence x_E about X_E. This is generally intended as the computation of the posterior probability $\underline{P}(x|x_E)$ (with respect to the strong extension) for each $x \in \mathcal{X}$, and this is what most of the updating algorithms for credal networks do. Yet, in the imprecise-probabilistic framework, the proper model of the posterior knowledge about X should be better identified with the posterior CS $K(X|x_E)$. In order to estimate this CS, Transformation 1 can be used to obtain an outer approximation (thus, in a sense, an over-cautious model) of $\tilde{K}(X|x_E)$.

To this aim, the lower probabilities $\underline{P}(A|x_E)$ for each $A \in 2^{\mathcal{X}}$ are needed, while standard updating algorithms only return the lower probabilities for the singletons. To overcome this limitation, let us introduce the notion of *coarsening*.

Given a variable X and an element of its power set $A \in 2^{\mathcal{X}}$, the coarsening of X based on A, is a variable X_A such that $\mathcal{X}_A := \{A\} \cup \mathcal{X} \setminus A$. In other words, the coarsening of a variable shrinks the set of possible values by clustering all the elements of A into a single value (denoted as {A}). The coarsening over A of a probability mass function $P(X)$ is indeed defined as a probability mass function $P_A(X_A)$ such that $P_A(\{A\}) := \sum_{x \in A} P(x)$ and $P_A(x) := P(x)$ for each $x \in \mathcal{X} \setminus A$. Finally, the coarsening over A of a CS $K(X)$ is a CS $K_A(X_A)$ obtained as the convex hull of the coarsening of the extreme points of $K(X)$, i.e,

$$K_A(X_A) := \mathrm{CH}\{P_A(X_A)\}_{P(X) \in \mathrm{ext}[K(X)]}.$$

The following result holds.

Theorem 2. *Consider a credal network over the variables* $(X, X_1, \ldots, X_v)$ *and a subset* $A \in 2^{\mathcal{X}}$*, where* X *corresponds to a node with no children. Obtain a second credal network over the variables* $(X_A, X_1, \ldots, X_v)$ *with same graph and same conditional CSs, except those associated with* X_A*, which are the coarsening of those originally associated to* X*, i.e.,* $K(X_A|\mathrm{pa}(X_A)) := K_A(X_A|\mathrm{pa}(X))$*. Then:*

$$\underline{P}(A|x_E) = \underline{P}_A(\{A\}|x_E), \tag{5}$$

where $\underline{P}$ *and* $\underline{P}_A$ *denote inferences on the first and on the second network.*

Note that, in the statement of Theorem 2, the variable of interest is assumed to correspond to a node without children. If X has a children, say Y, the only problem is how to define the conditional CS $K(Y|\{A\})$ in terms of $\{K(Y|x)\}_{x\in A}$. The problem can be easily solved if X corresponds to a node without parents (see the extension of Theorem 2 in the appendix), while in more general cases, further inferences on the network should be computed.

According to Theorem 2, we can therefore regard the lower probabilities for non-singletons in a credal network as lower probabilities of singletons in a "coarsened" network. Thus, $\tilde{K}(X|x_E)$ can be obtained through standard updating algorithms, by simply iterating the computation in (5) for each $A \in 2^{\mathcal{X}}$.

This result is important in order to make decisions based on the posterior state of X after the observation of x_E. In fact, as only lower posterior probabilities of the singletons are typically available, decision are based on the *interval dominance* criterion, i.e., we reject the states of X whose upper probability is smaller than the lower probability of some other state. The set of unrejected states is therefore:

$$\mathcal{X}^{\mathrm{ID}}_{\underline{P}} := \left\{x \in \mathcal{X} \,\mathrm{s.t.}\, \nexists x' \in \mathcal{X} \,|\, \underline{P}(x') > \overline{P}(x)\right\},$$

and is regarded as the set of optimal states according to this criterion.[8]

Other, more informative, decision criteria (see [13]) cannot be adopted unless the posterior CS is available. As an example, if decisions are based on the *maximality* criterion, a state is rejected if, for each point (or equivalently extreme point) of the CS, there is another state with higher probability, i.e.,

$$\mathcal{X}^{\mathrm{MAX}}_{K} := \{x \in \mathcal{X} \,\mathrm{s.t.}\, \nexists x' \in \mathcal{X}|P(x') > P(x), \forall P(X) \in K(X)\}\,. \tag{6}$$

This clearly requires that the (extreme) points of the posterior are CS available. Yet, by exploiting the result in Theorem 2, we can use Transformation 1 to compute the outer approximation $\tilde{K}(X|x_E)$ of $K(X|x_E)$, and make decisions on the basis of the maximality criterion (or any other criterion) with the CS $\tilde{K}(X|x_E)$. As the CS we work with is an outer approximation of the *true* CS, this can eventually lead to over-cautious decisions, i.e., we can include in the set of optimal decisions, some states which in fact are not. Nevertheless, the numerical simulations in the next section show that this tends to happen only in a minority of cases.

[8] We can similarly proceed if a linear utility function has been defined.

5 Numerical Tests

Different techniques can be adopted to evaluate the quality of the outer approximation associated with Transformation 1. As an example, a geometrical approach would consist in comparing the area of the polytopes associate with the two CSs. Yet, as a CS is basically of model of uncertain knowledge to be used to make decisions, it seems more reasonable to compare *decisions* based on the two CSs.

In order to do that, we consider randomly generated CSs over variables with an increasing number of possible values and extreme points, and we compare the number of optimal states according to maximality in the original CS and in its outer approximation. As an obvious consequence of (i) in Theorem 1 and (6), we have that $\mathcal{X}_K^{\mathrm{MAX}} \subseteq \mathcal{X}_{\tilde{K}}^{\mathrm{MAX}}$. Thus, the difference between the two sets can be simply characterized by the difference between the cardinality of the corresponding sets of optimal states. This is shown in the following table.

$\lvert\mathcal{X}\rvert$	$\lvert\mathrm{ext}[K(X)]\rvert$	$\lvert\mathcal{X}_{\tilde{K}}^{\mathrm{MAX}}\rvert - \lvert\mathcal{X}_K^{\mathrm{MAX}}\rvert$
3	3	0.235
4	4	0.317
5	5	0.353
6	6	0.359
7	7	0.255

Table 1. Numerical evaluation of the quality of the approximation associated with Transformation 1. The third column reports the average of the difference between the number of states recognized as optimal by using the outer approximation and those associated with the original CS. For each row 10000 randomly generated CSs over a variable with $\lvert\mathcal{X}\rvert$ states and $\lvert\mathrm{ext}[K(X)]\rvert$ extreme points have been generated.

As a comment, we note that, on average, the approximation introduces a non-optimal state once every three or four CSs. These values might be regarded as a reasonable approximation, especially for variables with many states.

6 Conclusions

An outer approximation for CSs based on lower probabilities, together with some theoretical and numerical characterizations, has been presented. The approximation is particularly suited for applications to decision making on credal networks, and makes it possible to adopt more sophisticated decision criteria, without the need of newer inference algorithms. As a future work, we want to investigate possible analytical characterizations of the quality of the approximation, and identify an inner approximation with similar features.

Acknowledgments

The work reported in this paper has benefitted from discussions with Sébastien Destercke and Enrique Miranda.

Appendix

Proof (Theorem 1). As an obvious consequence of (4), for each $P(X) \in K(X)$,

$$\sum_{x \in A} P(x) \geq \underline{P}_K(A).$$

Thus, according to (2), we have that $P(X) \in \tilde{K}(X)$, and hence (i). Now we prove (ii). The *if* implication is trivial, as we simply have $\underline{P}' = \underline{P}_K$. To prove the *only if* implication, note that the existence of $\underline{P}'$ implies that $\underline{P}_K = \underline{P}'$, and hence $\tilde{K}(X) = K_{\underline{P}'}(X) = K(X)$. In order to prove (iii), let us follow an *ad absurdum* scheme. Accordingly, let $\underline{P}' \neq \underline{P}_K$ be the lower probability operator such that $K(X) \subseteq K_{\underline{P}'}(X) \subseteq \tilde{K}(X)$. Thus, for each $A \in 2^{\mathcal{X}}$:

$$\underline{P}'(A) = \min_{P'(X) \in K_{\underline{P}'}(X)} \sum_{x \in A} P'(X) \leq \min_{P(X) \in \tilde{K}(X)} \sum_{x \in A} P(X) = \min_{P(X) \in K(X)} \sum_{x \in A} P(X),$$

where the inequality holds because of the set inclusion and the last equality is because of the definition of Transformation 1. But, as $K_{\underline{P}}(X) \supseteq K(X)$, the inequality should be an equality, this contradicting the assumption $\underline{P}' \neq \underline{P}_K$.
□

Proof (Theorem 2). Let us first assume $A := \{x^{(1)}\} \cup \{x^{(2)}\}$ and the network Bayesian. Consider the joint states $(x^{(1)}, x_1, \ldots, x_v)$ and $(x^{(2)}, x_1, \ldots, x_v)$ in the original network, and the joint state $(\{A\}, x_1, \ldots, x_v)$ in the "coarsened" network. By exploiting the factorization described in Section 2.3, we have:

$$P(\{x^{(1)}\} \cup \{x^{(2)}\}, x_1, \ldots) = P(x^{(1)}, x_1, \ldots) + P(x^{(2)}, x_1, \ldots) = P_A(\{A\}, x_1, \ldots),$$

from which the thesis follows by simple marginalization and application of Bayes' rule. The same result holds for credal networks, because of the notion of strong extension and by simply observing that coarsening for CSs consist in the *Bayesian* coarsening of each extreme point.

Extension (of Theorem 2) to the case where X correspond to a node without parents. Let Y be a children of X, and set $A = \{x^{(1)}\} \cup \{x^{(2)}\}$. In the Bayesian case:

$$P(y|\{x^{(1)}\} \cup \{x^{(2)}\}) = \frac{P(y, \{x^{(1)}\} \cup \{x^{(2)}\})}{P(\{x^{(1)}\} \cup \{x^{(2)}\})} = \frac{P(y|x^{(1)})P(x^{(1)}) + P(y|x^{(2)})P(x^{(2)})}{P(x^{(1)}) + P(x^{(2)})}.$$

Thus, in the credal case:

$$K(Y|x^{(1)}\cup x^{(2)}) = \mathrm{CH}\left\{ \frac{P_i(Y|x^{(1)})P_k(x^{(1)}) + P_j(Y|x^{(2)})P_k(x^{(2)})}{P_k(x^{(1)}) + P_k(x^{(2)})} \right\}_{\substack{P_i(Y|x^{(1)}) \in K(Y|x^{(1)}) \\ P_j(Y|x^{(2)}) \in K(Y|x^{(2)}) \\ P_k(X) \in K(X)}} .$$

We similarly proceed if X has more than a single children and if A has cardinality greater than two.

References

1. Levi, I.: The Enterprise of Knowledge. MIT Press, London (1980)
2. Walley, P.: Measures of uncertainty in expert systems. Artificial Intelligence **83**(1) (1996) 1–58
3. Cozman, F.G.: Credal networks. Artificial Intelligence **120** (2000) 199–233
4. Antonucci, A., Zaffalon, M.: Decision-theoretic specification of credal networks: A unified language for uncertain modeling with sets of Bayesian networks. International Journal of Approximate Reasoning **49**(2) (2008) 345–361
5. de Campos, C.P., Cozman, F.G.: The inferential complexity of Bayesian and credal networks. In: Proceedings of the International Joint Conference on Artificial Intelligence, Edinburgh (2005) 1313–1318
6. Cozman, F.G.: Graphical models for imprecise probabilities. International Journal of Approximate Reasoning **39**(2-3) (2005) 167–184
7. de Campos, C.P., Cozman, F.G.: Inference in credal networks through integer programming. In de Cooman, G., Vejnarová, J., Zaffalon, M., eds.: ISIPTA '07 – Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications, Prague, Action M Agency (2007) 145–154
8. Antonucci, A., Sun, Y., de Campos, C.P., Zaffalon, M.: Generalized loopy 2U: a new algorithm for approximate inference in credal networks. International Journal of Approximate Reasoning (2010) (to appear).
9. Wallner, A.: Maximal number of vertices of polytopes defined by f-probabilities. In Cozman, F.G., Nau, R., Seidenfeld, T., eds.: ISIPTA 2005 – Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications, SIPTA (2005) 126–139
10. Cuzzolin, F.: On the credal structure of consistent probabilities. In: Logics in Artificial Intelligence - Lecture Notes in Computer Science. Volume 5293/2008. Springer Verlag, Heidelberg (2008) 126–139
11. Avis, D., Fukuda, K.: A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra. Discrete and Computational Geometry **8** (1992) 295–313
12. Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London (1991)
13. Troffaes, M.C.M.: Decision making under uncertainty using imprecise probabilities. International Journal of Approximate Reasoning **45**(1) (2007) 17–29