Probability-possibility transformations: Application to credal networks

Salem Benferhat, Amélie Levray, and Karim Tabia

Univ Lille Nord de France, F-59000 Lille, France
UArtois, CRIL UMR CNRS 8188, F-62300 Lens, France
{benferhat, levray, tabia}@cril.fr
http://www.cril.fr

Abstract. This paper deals with belief graphical models and probability-possibility transformations. It first analyzes some properties of transforming a credal network into a possibilistic one. In particular, we are interested in satisfying some properties of probability-possibility transformations like dominance and order preservation. The second part of the paper deals with using probability-possibility transformations in order to perform MAP inference in credal networks. This problem is known for its high computational complexity in comparison with MAP inference in Bayesian and possibilistic networks. The paper provides preliminary experimental results comparing our approach with both exact and approximate inference in credal networks.

Keywords: Credal networks, possibilistic networks, probability-possibility transformations

1 Introduction

Belief graphical models such as Bayesian [4], credal [3] and possibilistic networks [2] are powerful tools for representing and reasoning with uncertain information. Bayesian networks allow to compactly encode a probability distribution thanks to the conditional independence relationships existing between the variables. Credal networks, based on the theory of credal sets, generalize Bayesian networks in order to allow some flexibility regarding the model parameters. Indeed, credal networks are often seen as probabilistic graphical models with relaxed parameters. They are for instance used in robustness analysis and for encoding incomplete and ill-known knowledge and reasoning with the knowledge of groups of experts. Possibilistic networks are the counterparts of Bayesian networks based on possibility theory [7, 17].
Many uncertainty frameworks exist, some of which are generalizations of some other ones. For instance, imprecise probability theory [10, 15] is a generalization of probability theory while possibility theory [7, 17] is an alternative non additive uncertainty theory particularly suited for handling incomplete, qualitative and partial information. In order to cast the information encoded within one setting into another uncertainty framework, transformations are used. They are mechanical transformations satisfying some desirable properties like consistency and order preservation. Many works are done

for instance for transforming probability measures into possibilistic ones [9, 17]. However, in the context of belief graphical models and knowledge bases, only few works addressed some related issues [1, 13]. Transformations can be useful in various contexts such as i) using the existing tools (e.g. algorithms and software) developed in one setting rather than developing everything from scratch for the other setting or ii) exploiting information provided in different uncertainty languages as it is often the case in some multiple expert applications. In this paper, we are mainly interested in probability-possibility transformations for computational complexity purposes. More precisely, in this preliminary work, our objective is to exploit probability-possibility transformations to efficiently perform MAP inference in credal networks where this task is NP^{PP}-hard in the general case [12]. The main contributions of the paper are:

– Proposing and analyzing a transformation allowing to turn a credal network into a possibilistic network.
– Proposing a kind of approximate approach for MAP inference in credal networks by means of probability-possibility transformations.
– Providing preliminary experimental studies showing that MAP inference could efficiently be carried out using our approach with a high accuracy rate.

2 A brief refresher on credal and possibilistic networks

This section briefly presents the belief graphical models dealt with in this paper.

2.1 Bayesian networks

Bayesian networks ($\mathcal{BN}$) are well-known probabilistic graphical models [4]. They allow to compactly represent a probability distribution over a set of variables of interest. A $\mathcal{BN}$ is specified by:

– *A graphical component* with nodes and edges forming a directed acyclic graph (DAG). Each node represents a variable A_i of the modeled problem and the edges encode independence relationships among variables.
– *A quantitative component*, where each variable A_i is associated with a local probability distribution $p(A_i|par(A_i))$ for A_i in the context of its parents $par(A_i)$.

The joint probability distribution encoded by a Bayesian network is computed using the following chain rule:

$$P(A_1, .., A_n) = \prod_{i=1}^{n} P(A_i|par(A_i)) \quad (1)$$

2.2 Credal networks

Credal networks are probabilistic graphical models based on imprecise probabilities. Imprecise probability theory [10, 15] generalizes probability theory to encode imprecise and ill-known information. A key notion in this theory is the one of credal set.

Definition 1 (Credal set). *A credal set is a convex set of probability distributions.*

Intuitively, if K is a convex set of probability measures, then $mixing$ any two distributions p_1 and p_2 from K will result in a distribution p belonging to K. Mixing here means linearly combining a set of distributions p_1 ... p_k as follows: p=$\sum_{i=1}^{k}(a_i * p_i)$ where $\sum_{i=1}^{k} a_i$=1. A credal set is often interpreted as a set of imprecise beliefs in the sense that the true uncertainty model (probability measure) is in this set but there is no way to determine it exactly due to lack of knowledge. In order to characterize a credal set, one can use a (finite[1]) set of extreme points (edges of the polytope representing the credal set), probability intervals or linear constraints.

Interval-based probability distributions (IPD for short) are a very natural and common way to specify imprecise and ill-known information. In an IPD IP, every interpretation $\omega_i \in \Omega$ is associated with a probability interval $IP(\omega_i)$=$[l_i, u_i]$ where l_i (resp. u_i) denotes the lower (resp. upper) bound of the probability of ω_i.

Definition 2 (Interval-based probability distribution). *Let Ω be the set of possible worlds. An interval-based probability distribution IP is a function that maps every interpretation $\omega_i \in \Omega$ to a closed interval $[l_i, u_i] \subseteq [0, 1]$.*

An IPD should satisfy the following constraints in order to ensure that the underlying credal set is not empty and every lower/upper probability bound is reachable.

$$\sum_{\omega_i \in \Omega} l_i \leq 1 \leq \sum_{\omega_i \in \Omega} u_i$$

$$\forall \omega_i \in \Omega, l_i + \sum_{\omega_{j \neq i} \in \Omega} u_j \geq 1 \; and \; u_i + \sum_{\omega_{j \neq i} \in \Omega} l_j \leq 1$$

In order to give a formal semantics for IPDs, let us first define the concept of compatible probability distribution.

Definition 3 (Compatible probability distribution). *A probability distribution p over Ω is said compatible with IP iff $\forall \omega_i \in \Omega$, $p(\omega_i) \in IP(\omega_i)$.*

Note that while a standard probability distribution p induces a complete order over the set of possible worlds Ω, an IPD IP may induce a partial order since some interpretations may be incomparable in case of overlapping intervals. In this paper, a credal set K_i associated with a variable A_i having an interval-based probability distribution IP denotes the closed convex set of (standard) probability distributions p that are compatible with IP. Let us now introduce probabilistic graphical models based on credal sets, called credal networks [3, 12].

Definition 4 (Credal network). *A credal network $\mathcal{CN}$=<G,K> is a probabilistic graphical model where*

- *G=<V, E> is a directed acyclic graph (DAG) encoding conditional independence relationships where V=$\{A_1, A_2, .., A_n\}$ is the set of variables of interest (D_i denotes the domain of variable A_i) and E is the set of edges of G.*
- *K=$\{K_1, K_2, .., K_n\}$ is a collection of local credal sets, each K_i is associated with the variable A_i in the context of its parents $par(A_i)$.*

[1] It is important to note that the number of extreme points can reach $N!$ where N is the number of interpretations [16].

Such credal networks are called separately specified credal networks as the only constraints on probabilities are specified in local tables for each variable in the context of its parents. Note that in practice, in local tables, one can either specify a set of extreme points characterizing the credal set as in JavaBayes[2] software or directly local IPDs.
A credal network $\mathcal{CN}$ can be seen as a set of Bayesian networks $\mathcal{BN}$s, each encoding a joint probability distribution. In this paper, we deal only with discrete variables and the semantics associated with a $\mathcal{CN}$ is a set of compatible $\mathcal{BN}$s, defined as follows:

Definition 5 (Compatible Bayesian network). *Let $\mathcal{CN}$=<G,K> be a credal network and $\mathcal{BN}$=<G,CPT> be a Bayesian networks. $\mathcal{BN}$ is said compatible with $\mathcal{CN}$ iff*

1. *$\mathcal{BN}$ and $\mathcal{CN}$ have exactly the same structure (hence they encode the same conditional independence relations).*
2. *For each variable A_i, $\forall a_i \in D_i$, $p_{\mathcal{BN}}(a_i|par(a_i)) \in K_i(a_i|par(a_i))$.*

According to this semantics, a credal network $\mathcal{CN}$ encodes a set of joint probability distributions, called extensions and denoted $ext(\mathcal{CN})$, where each joint distribution $p \in ext(\mathcal{CN})$ is encoded by a compatible Bayesian network. Given an extension $ext(\mathcal{CN})$, one can compute a joint IPD (interval-based joint probability distribution) as follows:

$$\underline{P}(a_1a_2..a_n) = \min_{p \in ext(\mathcal{CN})} (p(a_1a_2..a_n)) \quad (2)$$

$$\overline{P}(a_1a_2..a_n) = \max_{p \in ext(\mathcal{CN})} (p(a_1a_2..a_n)) \quad (3)$$

In Equations 2 and 3, $p(a_1a_2..a_n)$ is computed with the well-known chain rule in Bayesian networks (see Equation 1). Note that the vertices of $ext(\mathcal{CN})$ can be obtained by considering only the set of vertices of the local credal sets K_i associated with the variables [3]. As for marginalization and conditioning, they are defined as follows:
Let $K(A_1..A_n)$ be a credal set over the set of variables A=$\{A_1..A_n\}$. Let X and Y be two disjoint subsets of A such that $X \cup Y$=A. Then,

$$K(X) = CH(\{\sum_Y p(X,Y) \; with \; p(X,Y) \in K(A_1..A_n)\}) \quad (4)$$

where CH is the convex hull operator. As for conditioning, let e be an evidence, then

$$K(A_1..A_n|e) = CH(\{p(A_1..A_n|e) \; with \; p(A_1..A_n) \in K(A_1..A_n) \; and \; p(e) > 0\}) \quad (5)$$

2.3 Possibilistic networks

A possibilistic network $\mathcal{PN}$=<G,Θ> is specified by:

i) *A graphical component* G consisting of a directed acyclic graph (DAG) where vertices represent the variables and edges represent direct *dependence* relationships between variables.
ii) *A numerical component* Θ allowing to weight the uncertainty relative to each variable using local possibility tables. The possibilistic component consists in a set of local possibility tables θ_i=$\pi(A_i|par(A_i))$ for each variable A_i in the context of its parents $par(A_i)$ in the network $\mathcal{PN}$.

[2] http://www.cs.cmu.edu/~javabayes/Home/

Note that all the local possibility distributions θ_i must be normalized, namely $\forall i$=1..n, for each parent context $par(a_i)$, $\max_{a_i \in D_i}(\pi(a_i \mid par(a_i))$=1.

In the possibilistic setting, the joint possibility distribution is factorized using the following possibilistic counterpart of the chain rule:

$$\pi(a_1, a_2, .., a_n) = \otimes_{i=1}^{n}(\pi(a_i|par(a_i))). \quad (6)$$

where $\otimes$ denotes the product or the min-based operator depending on the quantitative or the qualitative interpretation of the possibilistic scale [7].

3 Probability-possibility transformations

3.1 Form probability distributions to possibilistic ones

Many probability-possibility transformations exist [6, 9, 17]. Most of the works address desirable properties and propose some transformations that satisfy such properties. Among these transformations, the optimal transformation (OT) [6] defines a consistency condition requiring that

1. the obtained possibility distribution π dominates the original probability distribution p (namely, $\phi \subseteq \Omega$, $P(\phi) \leq \Pi(\phi)$).
2. the obtained possibility distribution π preserves the order of elementary worlds encoded in p (namely, $\forall(\omega_i, \omega_j) \in \Omega^2$, $p(\omega_i) > p(\omega_j) \Rightarrow \pi(\omega_i) > \pi(\omega_j)$ and $p(\omega_i)$=$p(\omega_j)$ $\Rightarrow \pi(\omega_i)$=$\pi(\omega_j)$).

The optimal transformation (OT) transforms p into π as follows:

$$\pi_i = \sum_{j/p_j \leq p_i} p_j, \quad (7)$$

where π_i (resp. p_i) denotes $\pi(\omega_i)$ (resp. $p(\omega_i)$). The transformation of Equation 7 guarantees that the obtained possibility distribution π is the most specific[3] (hence most informative) one that is consistent and preserving the order of interpretations.

The author in [14] addressed the commutativity of transformations with respect to some operations but the aim was to show that the obtained distributions are not identical. Some of these issues were also dealt with in the context of fuzzy interval analysis [8]. In [1], we dealt with some issues about probability-possibility transformations especially those regarding reasoning tasks and graphical models. In particular, we showed that:

- there is no transformation that can preserve the order of arbitrary events through some reasoning operations like marginalization.
- for the independence of events and variables, we showed that there is no transformation that preserves the independence relations,
- when the uncertain information is encoded by means of graphical models, we showed that no transformation can preserve the order of interpretations and events.

In this paper, we deal with some of these issues in the context of credal networks.

[3] Let π' and π'' be two possibility distributions, π' is more specific than π'' iff $\forall \omega_i \in \Omega$, $\pi'(\omega_i) \leq \pi''(\omega_i)$

3.2 From interval-based probability distributions to possibilistic ones

When transforming uncertain information expressed by means of probability intervals to a possibility distribution, there is to the best of our knowledge only one work [11] where the authors learn possibility distributions from empirical data by transforming confidence intervals to possibility distributions. The starting point of this transformation is to consider an IPD as a means of encoding a partial order $\mathcal{M}$ over Ω. Indeed, contrary to precise probability distributions which encode complete order relations over Ω, interval-based ones encode partial orders in the form $\omega_i <_{IP} \omega_j$ in case where $u_i < l_j$. Let $\mathcal{M}$ be the partial order encoded by an IPD IP and let $\mathcal{C}$ be the set of linear extensions (complete orders) that are compatible with the partial order $\mathcal{M}$. The transformation proposed in [11] proceeds as follows:

– For every linear extension $\mathcal{C}_l \in \mathcal{C}$ and for each $\omega_i \in \Omega$, compute:

$$\pi^{C_l}(\omega_i) = \max_{p_1..p_n} (\sum_{p_j \leq p_i} p_i) \tag{8}$$

subject to the following constraints (in order to explore only compatible probability distributions satisfying the current linear extension C_l):

$$\begin{cases} p_i \in [l_i, u_i] \\ \sum_{i=1..n} p_i = 1 \\ p_1..p_n \; satisfy \; the \; linear \; extension \; C_l \end{cases}$$

– Build the distribution π that dominates all the distributions π^{C_l} as follows: $\forall \omega_i \in \Omega$,

$$\pi(\omega_i) = \max_{\mathcal{C}_l \in \mathcal{C}} (\pi^{C_l}(\omega_i)) \tag{9}$$

The motivation of using Equation 9 is to guarantee that the obtained possibility distribution π dominates the probability intervals IP. This transformation tries on one hand to preserve the order of interpretations induced by IP and the dominance principle requiring that $\forall \phi \subseteq \Omega$, $P(\phi) \leq \Pi(\phi)$ on the other hand.

There are two main drawbacks with the transformation of Equations 8 and 9:

– The first issue is about the computational complexity of such transformation. Applied directly, this latter can consider in the worst case $N!$ linear extensions where N is the number of possible worlds. The authors proposed in [11] an algorithm allowing to achieve some improvements during this transformation but it is still very costly when one considers variables having domains exceeding a dozen values, which is common in many applications.
– The second concern lies in the fact that this transformation does not guarantee that the obtained distribution is optimal is terms of specificity. Indeed, it was shown in [5] that the transformation of Equation 9 results in a loss of information as it is not the most specific one dominating the considered IPD. The authors in [5] suggest that any upper generalized R-cumulative distribution $\overline{F}$ built from one linear extension $\mathcal{C}_l \in \mathcal{C}$ can be viewed as a possibility distribution and it also dominates all the probability distributions that are compatible with the IPD. Let $\mathcal{C}_l$ be a linear

extension compatible with the partial order $\mathcal{M}$ induced by an IPD. Let ϕ_1, $\phi_2..\phi_n$ be subsets of Ω such that ϕ_i=$\{\omega_j|\omega_j \leq_{\mathcal{C}_l} \omega_i\}$. The upper cumulative distribution $\overline{F}$ built from one linear extension $\mathcal{C}_l$ is as follows (see [5] for more details):

$$\overline{F}(\phi_i) = \min(\sum_{\omega_j \in \phi_i} u_j, 1 - \sum_{\omega_j \notin \phi_i} l_j) \tag{10}$$

The obtained cumulative distribution $\overline{F}$ is a possibility distribution dominating IP and it is such that $\overline{P}(\phi_i)$=$\Pi(A_i)$.

Regarding the commutativity of transformations with respect to change operations like marginalization and conditioning used to answer MAP queries, since probability distributions are special cases of IPDs, it can be expected that for the commutativity issue, the transformations exhibit the same properties. This is the focus of the next section.

4 Commutativity of interval-based probability-possibility transformations with respect to marginalization and conditioning

This section checks whether the interval-based probability-possibility transformations are commutative with respect to two major change operations that are marginalization and conditioning. Namely, the question dealt with here is: Given an IPD IP, do we get exactly the same results when i) we first transform IP into a possibility distribution π then apply the change operation in the possibilistic setting and when ii) we first apply the change operation in the interval-based setting then transform the result into a possibility distribution. Proposition 1 provides the answer for marginalization :

Proposition 1. *Let TR be an interval-based probability-possibility transformation*[4]. *Then there exists an IPD IP, two events $\phi \subseteq \Omega$, $\psi \subseteq \Omega$ with $\phi \neq \psi$, and π= $TR(IP)$ such that $\overline{P}(\phi) < \underline{P}(\psi)$ but $\Pi(\phi) > \Pi(\psi)$.*

Proposition 1 asserts that no interval-based probability-possibility transformation can guarantee the preservation of the order of events as shown in the following example.

Example 1. Let IP be an IPD of Table 1 where Ω=$\{\omega_1, \omega_2, \omega_3, \omega_4\}$ and π= $TR(IP)$. In this example, α_1, α_2 and α_3 are possibility degrees such that $1 > \alpha_1 > \alpha_2 > \alpha_3$ in order to satisfy the preference preservation principle. Now, let ϕ and ψ be two events such that ϕ=$\{\omega_1\}$ and ψ=$\{\omega_2, \omega_3\}$. We have $\Pi(\phi)$=1$>\Pi(\psi)$=$\max(\alpha_1, \alpha_2)$ while $\overline{P}(\phi)$=.4$<\underline{P}(\psi) = .6$.

---

[4] In the rest of this paper, TR denotes an interval-based probability-possibility transformation satisfying the following principles:

– *Dominance:* The possibility distribution π obtained from the IPD IP by TR dominates every probability distribution p compatible with IP, namely $\forall \phi \subseteq \Omega$, $\pi(\phi) \geq p(\phi)$.

– *Order preservation:* Given two interpretations $\omega_i \in \Omega$ and $\omega_j \in \Omega$, $\pi(\omega_i) < \pi(\omega_j)$ iff $\overline{p}(\omega_i) < \underline{p}(\omega_j)$.

ω_i	$IP(\omega_i)$	$\pi(\omega_i)$
ω_1	[.36, .4]	1
ω_2	[.35,.35]	α_1
ω_3	[.25,.25]	α_2
ω_4	[0, .04]	α_3

Table 1. Example showing the loss of the order of events.

As shown in Example 1, the strict order of events is not preserved by TR because of the different behavior of the additivity axiom in the probabilistic setting and the maxitivity axiom of the possibilistic setting used by the marginalization operation.
As a consequence of Proposition 1, we have the following Lemma:

Lemma 1. *Let TR be an interval-based probability-possibility transformation. Then there exists an IPD IP over Ω=$\{\omega_1, \omega_2, .., \omega_n\}$ and a partition Ω'=$\{W_1, W_2...W_k\}$ of Ω with k<n. Let π=$TR(IP)$, IP' is obtained by marginalizing IP on Ω' according to Equation 4 and π' is obtained by marginalizing π on Ω' in the possibilistic setting. Then there may exist an event W_i∈Ω' such that*

$$\pi(W_i) \neq \pi'(W_i).$$

Proof (Proof sketch). The proof follows from Proposition 1 since if the order of events is not preserved then the underlying marginalized distributions must be different.

Let us now check the commutativity issue with respect to conditioning. For standard probability distributions, we have the following finding [1]:

Proposition 2. *Let p be a probability distribution over Ω and let $\phi\subseteq\Omega$ be an evidence. Let TR be a probability-possibility transformation, p' be a probability distribution obtained by conditioning p by ϕ, $\pi'' = TR(p')$ and π' is the possibility distribution obtained by conditioning π=$TR(p)$ by ϕ. Then, $\forall\omega_i, \omega_j \in \Omega$,*

$$\pi'(\omega_i)<\pi'(\omega_j) \text{ iff } \pi''(\omega_i)<\pi''(\omega_j).$$

Note that Proposition 2 is valid in both the product and the min-based possibilistic settings and it states that the order of interpretations is not affected by the order of applying the transformation and the conditioning operation. For IPDs, the following proposition states that the partial order encoded by IP after conditioning is preserved in the (complete) order induced by π after conditioning on the same evidence.

Proposition 3. *Let IP be an IPD over Ω and let $\phi\subseteq\Omega$ be an evidence. Let TR be an interval-based probability-possibility transformation, IP'=$IP(.|\phi)$ be a posterior probability distribution obtained by conditioning IP by ϕ, π''=$TR(IP')$ and π'=$\pi(.|\phi)$ is the possibility distribution obtained by conditioning π=$TR(IP)$ by ϕ. Then,*

$$\forall\omega_i, \omega_j \in \Omega,\ \pi'(\omega_i)<\pi'(\omega_j) \text{ iff } \pi''(\omega_i)<\pi''(\omega_j).$$

Proof (Proof sketch). The idea of the proof is that since conditioning in both the probabilistic and possibilistic settings consists in discarding the worlds that are not models of the evidence ϕ (by assigning them a 0 probability/possibility degree) then renormalizing the obtained distribution. Hence, the order of interpretations that are models of ϕ is not affected by the order of application of transformation/conditioning operations.

Let us now see how one can use probability-possibility transformations to perform some inference queries in credal networks.

5 A probability-possibility transformation based approach for inference in credal networks

In [13] a natural transformation of Bayesian networks into possibilistic networks is proposed using the existing probability-possibility transformations such as OT.

5.1 From credal networks to possibilistic networks

A straightforward way to transform a credal network into a possibilistic network is as follows:

Definition 6 (Credal-possibilistic network transformation). *Let $\mathcal{CN}$ be a credal network, $\mathcal{PN}_{\mathcal{CN}}$ is a possibilistic network obtained from $\mathcal{CN}$ and defined by:*

- *A graphical component G which is the same graph as the credal network hence $\mathcal{PN}_{\mathcal{CN}}$ encodes the same independence relations as $\mathcal{CN}$.*
- *A collection of local possibility tables π_i obtained by transforming local credal sets K_i with TR, a transformation from interval-based probability distribution into possibilistic ones.*

The advantage of transforming a graphical model using Definition 6 is to preserve the independence relationships while transforming only local tables.

Example 2. Let $\mathcal{CN}$ be the credal network of Figure 1 over two binary variables A and B. Using the transformation of Equation 9, the credal network $\mathcal{CN}$ of Figure 1 will be transformed to the possibilistic network $\mathcal{PN}$ of Figure 2.

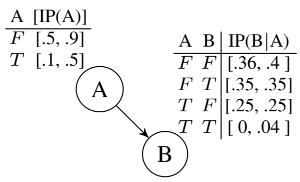


Fig. 1. Example of a credal network $\mathcal{CN}$.

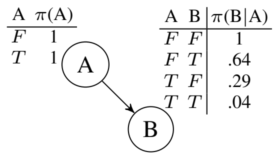


Fig. 2. The possibilistic network $\mathcal{PN}_{CN}$ obtained from the credal network $\mathcal{CN}$ of Fig. 1.

In the following, we address two main questions: i) Does the distribution $\pi_{\mathcal{PN}}$ dominate $IP_{\mathcal{CN}}$ (the joint interval-based distribution encoded by $\mathcal{CN}$)? and ii) Is the partial order of interpretations induced by $IP_{\mathcal{CN}}$ preserved by the transformation TR?
Regarding the first question, the two following propositions provide the answer. For elementary worlds $\omega_i \in \Omega$, Proposition 4 ensures that the computed possibility distribution dominates the corresponding probability degrees in case where the credal network $\mathcal{CN}$ is a Bayesian network (namely, all the intervals in $\mathcal{CN}$ are singletons).

Proposition 4. *Let TR be a probability-possibility transformation. Let $\mathcal{BN}$ be a standard Bayesian network and let $p_{\mathcal{BN}}$ be the underlying joint probability distribution encoded by $\mathcal{BN}$. Let $\mathcal{PN}$ be a possibilistic network such that $\mathcal{PN}$=$TR(\mathcal{BN})$ and $\pi_{\mathcal{PN}}$ be the joint possibility distribution encoded by $\mathcal{PN}$. Then $\forall \omega_i \in \Omega$,*

$$\pi_{\mathcal{PN}}(\omega_i){\geq}P_{\mathcal{BN}}(\omega_i).$$

Proof (Proof sketch).
Let ω_i=$a_1 a_2 ... a_n$ be an instantiation of the network variables A_1, $A_2 ... A_n$. We have in the product-based possibilistic setting, for every variable value a_i in its parents context $par(a_i)$, $p_{\mathcal{BN}}(a_i|par(a_i)){\leq}\pi_{\mathcal{PN}}(a_i|par(a_i))$, guaranteed by the transformation TR. Then $\prod_{i=1}^{n}(p_{\mathcal{BN}}(a_i|par(a_i)){\leq}\prod_{i=1}^{n}(\pi_{\mathcal{PN}}(a_i|par(a_i))$. The proof follows similarly for $\min$-based possibilistic networks.

Now, regarding arbitrary events $\phi{\subseteq}\Omega$, the issue is still open. If we use the optimal transformation OT, the following proposition states that the obtained joint possibility distribution does not guarantee to dominate the joint probability distribution.

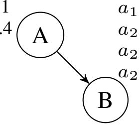
Proposition 5. *Let OT be the optimal probability-possibility transformation. There may exist a standard Bayesian network $\mathcal{BN}$ encoding a joint probability distribution denoted $p_{\mathcal{BN}}$. Let $\mathcal{PN}$ be a possibilistic network such that $\mathcal{PN}$= $OT(\mathcal{BN})$ and $\pi_{\mathcal{PN}}$ be the joint possibility distribution encoded by $\mathcal{PN}$. Then there may exist an event $\phi{\subseteq}\Omega$ such that*

$$\Pi_{\mathcal{PN}}(\phi){\ngeq}P_{\mathcal{BN}}(\phi)$$

The following counter-example shows that $\Pi_{\mathcal{PN}}(\phi){\geq}P_{\mathcal{BN}}(\phi)$ is not guaranteed when using the optimal transformation OT.

Example 3. Let $\mathcal{BN}$ be the Bayesian network of Figure 3 over two variables A and B having the domains D_A=$\{a_1, a_2\}$ and D_B=$\{b_1, b_2, b_3\}$ respectively.

A	p(A)	$\pi(A)$
a_1	.6	1
a_2	.4	.4



A	B	p(B|A)	$\pi(B\|A)$
a_1	b_1	.6	1
a_1	b_2	.3	.4
a_1	b_3	.1	.1
a_2	b_1	.5	1
a_2	b_2	.3	.5
a_2	b_3	.2	.2

Fig. 3. Example of a Bayesian network $\mathcal{BN}$ and the possibilistic network $\mathcal{PN}$ obtained from $\mathcal{BN}$ using the optimal transformation OT.

A	B	p(A,B)	$\pi(A, B)$
a_1	b_1	.36	1
a_1	b_2	.18	.4
a_1	b_3	.06	.1
a_2	b_1	.2	.4
a_2	b_2	.12	.2
a_2	b_3	.08	.08

Fig. 4. Joint probability and possibility distributions encoded by the networks $\mathcal{BN}$ and $\mathcal{PN}$ of Fig. 3.

The joint distributions encoded by the networks $\mathcal{BN}$ and $\mathcal{PN}$ are given in Figure 4. From Figure 4, one can compute $P(b_3)$=.06+.08=.14$>\Pi(b_3)$=$\max(.1, .08)$=.1.

Example 3 clearly shows that the transformation OT does not guarantee that when transforming a Bayesian network to a possibilistic network, the underlying joint possibility distribution dominates the corresponding probability distribution.
Now, how about the order of interpretations encoded by a credal network when it is transformed into a possibilistic network? The following proposition answers this question. Recall that the objective here is to check if the order of interpretations induced by $IP_{\mathcal{CN}}$ (the joint IPD encoded by the credal network $\mathcal{CN}$) is preserved in the obtained joint possibility distribution $\pi_{\mathcal{PN}}$ encoded by the possibilistic network $\mathcal{PN}$.

Proposition 6. *Let TR be a transformation from credal networks to possibilistic ones. Then there exists a credal network $\mathcal{CN}$ and two interpretations $\omega_i \in \Omega$ and $\omega_j \in \Omega$ such that*

$$\overline{p}_{CN}(\omega_i) < \underline{p}_{CN}(\omega_j) \text{ but } \pi_{PN}(\omega_i) \geq \pi_{PN}(\omega_j).$$

where $\underline{p}_{CN}$ and $\overline{p}_{CN}$ denote lower and upper bounds induced by $\mathcal{CN}$ and π_{PN} denotes the joint possibility distribution induced by $\mathcal{PN}$ using the transforamtion of Definition 6. The following gives a counter-example.

Example 4. Let $\mathcal{CN}$ be the credal network of Figure 5 over two disconnected variables A and B. Note that the IPD $IP(A)$ in $\mathcal{CN}$ is a permutation[5] of the IPD of B. Hence, the transformation of $IP(A)$ and $IP(B)$ by TR gives $\pi(A)$ and $\pi(B)$ where $\pi(B)$ is also a permutation of $\pi(A)$. In this example, since TR is assumed to preserve the partial order of interpretations, we have $1 > \alpha_1 > \alpha_2 > \alpha_3$. The probability and possibility degrees of interpretations $a_1 b_3$ and $a_2 b_2$ are

A	$IP(A)$	$\pi(A)$
a_1	[.36, .4]	1
a_2	[.26, .26]	α_1
a_3	[.24, .24]	α_2
a_4	[.1, .14]	α_3

B	$IP(B)$	$\pi(B)$
b_1	[.1, .14]	α_3
b_2	[.26, .26]	α_1
b_3	[.24, .24]	α_2
b_4	[.36, .4]	1

Fig. 5. Counter-example for Proposition 6.

$\underline{p}(a_1 b_3)$=0.36∗0.24=0.0864 and $\overline{p}(a_2 b_2)$=0.26∗0.26=0.0676. Clearly, $\underline{p}(a_1 b_3) > \overline{p}(a_2 b_2)$. Now, $\pi(a_1 b_3)$=min$(\alpha_2, 1)$ and $\pi(a_2 b_2)$=min(α_1, α_1) then, $\pi(a_1 b_3) < \pi(a_2 b_2)$. It is clear that the relative order of interpretations is reversed whatever is the used transformation in the ordinal setting. In the same way, in the product-based possibilistic setting, the relative order of interpretations can not be preserved by any transformation.

Up to now, the findings of this paper are rather negative but transformations from credal networks into possibilistic ones can be very helpful for certain types of queries in credal networks as it is shown in the following sections.

5.2 MAP inference in credal networks through credal-possibilistic network transformation

Inference in probabilistic graphical models generally consists in computing the probability of an event. In credal networks, this equivalently comes down to computing lower or upper probabilities of an event of interest. Let A=$\{A_1, A_2 ... A_n\}$ be the set of variables of the probabilistic model. Let $\mathcal{O} \subseteq A$ be the set of observed variables and let o be an instantiation of observation variables $\mathcal{O}$. Let also $\mathcal{Q} \subseteq A$ be the set of query variables and let q be instantiation of the query variables. There are three main query types when reasoning with belief graphical models:

- Computing the probability of an event q of interest (Pr) given an evidence o.
- Computing the most plausible explanation (MPE). Given an observation o of some variables, the objective is to compute the most probable instantiation q of all the remaining (unobserved) variables $\mathcal{Q}$. Note that here $\mathcal{O} \cup \mathcal{Q}$=$A$ and $\mathcal{Q} \cap \mathcal{O}$=$\emptyset$.

[5] The permutation property of probability-possibility transformations is discussed in [14].

- Computing the maximum a posteriori (MAP). Given some observations o of the values of some variables $\mathcal{O}$, the objective is to compute the most probable instantiation q of the query variables $\mathcal{Q}$. In MAP queries, $\mathcal{Q} \cap \mathcal{O} = \emptyset$. Note that MPE queries are special cases of MAP ones.

In credal networks, the inference problem equivalently comes down to compute either the lower or the upper bound of an event of interest. As for MPE and MAP queries, there are different criteria to choose the *most probable* instantiation of query variables given the observations. The commonly used criterion in credal networks is the one of *interval-dominance* and refers to non-dominated instantiations of query variables.

Definition 7 (Interval-dominance). *An instantiation q_i of query variables $\mathcal{Q}$ dominates another instantiation q_j iff $\underline{P}(q_i|o) > \overline{P}(q_j|o)$ where o is an instantiation of observation variables $\mathcal{O}$.*

The following table summarizes complexity results of inference in Bayesian and credal networks [12].

	Query	Polytree	Bounded treewidth	Multiply-connected
	Pr	Polynomial	Polynomial	PP-Complete
Bayesian	MPE	Polynomial	Polynomial	NP-Complete
Networks	MAP	NP-Complete	NP-Complete	NP^{PP}-Complete
	Pr	NP-Complete	NP-Complete	NP^{PP}-Complete
Credal	MPE	Polynomial	Polynomial	NP-Complete
Networks	MAP	Σ_2^P-Complete	Σ_2^P-Complete	NP^{PP}-Hard

It is obvious that even in polytrees, MAP inference is a hard task. In practice, the size of networks and the set of extreme points of local credal sets is often large. This motivates approximate inference approaches where the goal is to provide bounds of the real bounds of probabilities. In this work, we provide a kind of approximate inference approach for MAP inference in $\mathcal{CN}$s by transforming the credal network $\mathcal{CN}$ into a possibilistic network $\mathcal{PN}$ used to answer the queries. Note that the complexity of inference in possibilistic networks is similar to inference in Bayesian networks.

6 Experimental studies

The objective of this section is to empirically evaluate the accuracy of performing MAP inference in credal networks by transforming them into possibilistic networks. In order to evaluate our approach, we carried out a set of experimentations on the well-known and publicly available credal networks benchmark [6]. This latter contains a set of credal networks with different topologies and parameters in *.bif* format. Table 2 gives some details on the networks used in our experimentations. Table 2 shows that the number of variables in the used networks varies from 6 up to 37. As for variable domains, their sizes vary between 2 and 8. In this preliminary study, we are interested only in MAP queries where given some observed variables, the task is to find the most probable values of some other non observed variables, called query variables. In this experimentation, we report results where the number of observed variables and the observed values

[6] `http://ipg.idsia.ch/software/`

Networks	Topology	# Nodes	max domain size
$Alarm$	Multiply-connected	37	4
$Insurance$	Multiply-connected	27	5
$Poly$	Polytree	10, 20, 30	4
$Multi$	Multiply-connected	6, 10, 20	8

Table 2. Credal networks used in the experimentations.

are randomly chosen. The queries concern only one variable chosen randomly. Now, in order to compare the results of MAP inference in credal networks and their possibilistic counterpart, each query Q is submitted to a credal network $\mathcal{CN}$ then to the corresponding possibilistic network $\mathcal{PN}$ obtained from $\mathcal{CN}$. The results are compared through the accuracy criterion defined as follows:

$$accuracy(Q_1, Q_2...Q_n) = \frac{1}{n} \sum_{i:1..n} \frac{|CN_{MAP}(Q_i) \cap PN_{MAP}(Q_i)|}{|CN_{MAP}(Q_i) \cup PN_{MAP}(Q_i)|}, \quad (11)$$

where $CN_{MAP}(Q_i)$ (resp. $PN_{MAP}(Q_i)$) denotes the results of the query Q_i submitted to the network $\mathcal{CN}$ (resp. $\mathcal{PN}$). This criterion evaluates the coincidence between the results of $\mathcal{CN}$ to the MAP queries and the ones of $\mathcal{PN}$.
In Table 3, we provide the accuracy (see Equation 11) of MAP inference achieved through our credal-possibilistic network transformation approach with respect to the results of credal networks. More precisely, the column $Exact$ vs $Appr$ provides the accuracy of an approximate inference algorithm in credal networks achieved with the GL2U software[7] on each network category. The column $Exact$ vs CD (resp. $Exact$ vs MD) provides the accuracy of possibilistic networks obtained by our credal-possibilistic network transformation where local tables are transformed using the cumulative distribution of Equation 10 (resp. Masson and Denoeux's transformation [11] considering all the linear extensions). Note also that we evaluated only product-based possibilistic networks and the experiments are performed on a few dozen requests on a laptop.
The results of Table 3 clearly show that, on one hand, the credal-possibilistic network

Networks	$Exact$ vs $Appr$	$Exact$ vs CD	$Exact$ vs MD
$Alarm$	75%	100%	*timeout*
$Insurance$	50%	100%	*timeout*
$Poly$	83%	100%	*timeout*
$Multi$	90%	38%	48%

Table 3. Credal networks used in the experimentations.

transformation based approach can ensure a high accuracy rate and, on the other hand, the results are often better than those obtained with an approximate approach.

7 Conclusions

This paper dealt with probability-possibility transformations in the context of credal networks. We first analyzed some issues related to the commutativity of transformations with respect to marginalization and conditioning, two main change operations used for MAP inference. We then proposed an approach allowing to perform MAP

[7] `http://people.idsia.ch/~sun/gl2u-ii.htm`

inference in credal networks with a lower computational costs. Finally, we provided experimental studies showing the efficiency of the proposed approach in terms of accuracy. Future works will deal with extensive experimental studies as well as using our credal-possibilistic network transformation based approach for achieving classification with credal networks in real applications.

Acknowledgements

This work is done with the support of a CNRS funded project PEPS FaSciDo 2015 - MAPPOS.

References

1. S. Benferhat, A. Levray, and K. Tabia. On the analysis of probability-possibility transformations: Changing operations and graphical models. In *ECSQARU 2015, Compiegne, France, July 15-17*, 2015.
2. C. Borgelt and R. Kruse. Learning possibilistic graphical models from data. *Fuzzy Systems, IEEE Transactions on*, 11(2):159–172, Apr 2003.
3. F. G. Cozman. Credal networks. *Artificial Intelligence*, 120(2):199 – 233, 2000.
4. A. Darwiche. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, 2009.
5. S. Destercke, D. Dubois, and E. Chojnacki. Transforming probability intervals into other uncertainty models. In *EUSFLAT 2007 proceedings*, volume 2, pages 367–373, Ostrava, Czech Republic, 2007. Universitas Ostraviensis.
6. D. Dubois, L. Foulloy, G. Mauris, and H. Prade. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10(4):273–297, 2004.
7. D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, 1988.
8. D. Dubois and H. Prade. Random sets and fuzzy interval analysis. *Fuzzy Sets Syst.*, 42(1):87–101, July 1991.
9. G. J. Klir and J. F. Geer. Information-preserving probability- possibility transformations: Recent developments. In R. Lowen and M. Roubens, editors, *Fuzzy Logic*, pages 417–428. Kluwer Academic Publishers, Dordrecht, 1993.
10. I. Levi. *The enterprise of knowledge : an essay on knowledge, credal probability, and chance / Isaac Levi*. MIT Press Cambridge, Mass, 1980.
11. M.-H. Masson and T. Denoeux. Inferring a possibility distribution from empirical data. *Fuzzy Sets and Systems*, 157(3):319–340, 2006.
12. D. Mauá, C. P. de Campos, A. Benavoli, and A. Antonucci. Probabilistic inference in credal networks: New complexity results. *J. Artif. Intell. Res. (JAIR)*, 50:603–637, 2014.
13. Y.Ben Slimen, R. Ayachi, and N. Ben Amor. Probability-possibility transformation: - application to bayesian and possibilistic networks. In *WILF13 2013, Genoa, Italy, Nov. 19-22, 2013.*, pages 122–130.
14. T. Sudkamp. On probability-possibility transformations. *Fuzzy Sets and Systems*, pages 73–81, 1992.
15. P. Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24(23):125 – 148, 2000.
16. A. Wallner. Extreme points of coherent probabilities in finite spaces. *Int. J. Approx. Reasoning*, 44(3):339–357, March 2007.
17. L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 100:9–34, 1999.