
Towards conservative inference in credal networks using belief functions: the case of credal chains

Marco Sangalli¹

Thomas Krak¹

Cassio de Campos¹

¹Uncertainty in AI, Eindhoven University of Technology, The Netherlands

ABSTRACT

This paper explores belief inference in credal networks using Dempster-Shafer theory. By building on previous work, we propose a novel framework for propagating uncertainty through a subclass of credal networks, namely chains. The proposed approach efficiently yields conservative intervals through belief and plausibility functions, combining computational speed with robust uncertainty representation. Key contributions include formalizing belief-based inference methods and comparing belief-based inference against classical sensitivity analysis. Numerical results highlight the advantages and limitations of applying belief inference within this framework, providing insights into its practical utility for chains and for credal networks in general.

Keywords. credal networks, belief functions, Dempster-Shafer theory, approximate inference

1. INTRODUCTION

Bayesian networks (BNs), as probabilistic graphical models, provide a powerful framework for modelling probabilistic relationships among variables. Credal networks (CNs) generalize BNs by allowing for partially specified probabilistic information. It is well-known that inference in both BNs and CNs is in general computationally hard [2, 9].

We consider the problem of computing inferences in CNs under complete independence [3, 9]. There are, broadly speaking, two ways by which we might approach this problem. We might either try to identify specific inference problems and/or graph topologies that circumvent the hardness results; or, we can try to develop approximate methods. The first path is arguably difficult, since NP-hardness was already established for marginal inference on relatively simple graph structures like poly-trees [4] and even trees with evidence [8]. Here we consider the second approach. In particular, we investigate the merit in leveraging the existing machinery from the theory of belief functions, also known as Dempster-

Shafer (D-S) theory [6, 12], to approximately compute inferences for a given CN under complete independence. At first glance, this has several upsides. For one, we can fully use the existing computational and algorithmic tools from that theory, which has recently been investigated specifically in the context of Bayesian networks generalized to belief functions [13], where the computational complexity of BN inferences is largely reduced. Moreover, a recent result [10] essentially implies that the global model that we obtain from this is guaranteed to be a conservative bound on the CN in which we are actually interested (even though this result does not hold for marginals, as we will see later).

We work with the subcase of CNs that have the structure of a chain. We require that local models in the CN can be translated (possibly by conservative approximation) to belief functions. These choices are enough to unveil a number of facts about the idea of using D-S theory in inferences for CNs. We investigate different ways in which local models can be made into belief functions. In particular, we are interested in the quality of this approach; i.e. we would like to know (and reduce) how much information is lost with respect to the exact inference in the CN chains in which the local uncertainty models are specified using probability intervals.

Since exact inference in such CNs is computationally tractable, we are able to directly compare these methods from D-S theory to the gold standard of the underlying model of interest and to reach conclusions about the quality of the approximations in chains and beyond. Despite optimizations that generate tighter bounds than the direct application of D-S theory, we conclude that—outside specific settings—such belief-based inference often generates too wide intervals to be a practically suitable approximation for the CN inferences of interest.

2. PRELIMINARIES

This section introduces the basic concepts of credal networks [1], Dempster-Shafer theory [5, 12] of belief functions, and other required concepts for this paper.

2.1. Bayesian and credal networks. A Bayesian network (BN) is a probabilistic graphical model where a

directed acyclic graph (DAG) encodes probabilistic relationships between random variables. Each node represents a variable, and edges indicate conditional dependencies. Credal networks (CNs) extend BNs by addressing uncertainty in probability specifications. Instead of fixed point estimates, CNs utilize sets of probability distributions (credal sets), allowing representation of multiple compatible probabilistic models. This approach provides a more comprehensive framework for reasoning under incomplete or ambiguous information, capturing broader uncertainty in probabilistic modelling through credal sets rather than single distributions.

A basic example of a CN is one with the DAG structure $A \rightarrow B$, where $A : \Omega \rightarrow E_A$ and $B : \Omega \rightarrow E_B$ are two discrete random variables, $E_A := \{a_1, \dots, a_n\}$, $E_B := \{b_1, \dots, b_m\}$, and whose credal sets are specified using collections of intervals:

$$\underline{p}_i^A \leq \mathbb{P}(A = a_i) \leq \bar{p}_i^A, \quad \underline{p}_j^i \leq \mathbb{P}(B = b_j | A = a_i) \leq \bar{p}_j^i.$$

Throughout, we consider CNs under *complete independence* [3, 9], which essentially means that the joint model that we consider can be expressed as the set of all BNs whose local models— $\mathbb{P}(A)$ and $\mathbb{P}(B|A)$ in this example—are compatible with these credal sets.

Solving an inference problem for such a CN amounts to computing (tight) lower and upper bounds on some probabilistic quantity with respect to all the BNs in this set. For instance, in this example we may want to compute lower and upper bounds for $\mathbb{P}(B = b_j)$, which are given by

$$\begin{aligned} \underline{\mathbb{P}}(B = b_j) &:= \min_{\mathbb{P} \in \text{CN}} \sum_{i=1}^n \mathbb{P}(B = b_j | A = a_i) \mathbb{P}(A = a_i), \text{ and} \\ \bar{\mathbb{P}}(B = b_j) &:= \max_{\mathbb{P} \in \text{CN}} \sum_{i=1}^n \mathbb{P}(B = b_j | A = a_i) \mathbb{P}(A = a_i). \end{aligned}$$

In a chain $X_1 \rightarrow \dots \rightarrow X_k$ we may do something similar to infer bounds on $\mathbb{P}(X_k = \tilde{x})$, using

$$\begin{aligned} \underline{\mathbb{P}}(X_k = \tilde{x}) &:= \min_{\mathbb{P} \in \text{CN}} \sum_{x_1, \dots, x_{k-1}} \mathbb{P}(x_1) \mathbb{P}(x_2 | x_1) \dots \mathbb{P}(\tilde{x} | x_{k-1}), \\ \bar{\mathbb{P}}(X_k = \tilde{x}) &:= \max_{\mathbb{P} \in \text{CN}} \sum_{x_1, \dots, x_{k-1}} \mathbb{P}(x_1) \mathbb{P}(x_2 | x_1) \dots \mathbb{P}(\tilde{x} | x_{k-1}), \end{aligned}$$

where the minimization and maximization are understood to be taken with respect to all the combinations of the local models specified in the CN.

2.2. Mass functions and belief functions. Let $(E, 2^E)$ be a discrete measurable space, where 2^E represents the power set of E . A mass function [7, 13] (also called basic probability assignment) m on E is a function $m : 2^E \rightarrow [0, 1]$ such that

- $m(\emptyset) = 0$

- $\sum_{V \in 2^E} m(V) = 1.$

Each set V such that $m(V) > 0$ is called a *focal set* of m . If the only focal set of m is E , the mass function is said to be *vacuous*. Meanwhile, if the only set with mass is not the whole space, m is said to be *deterministic*. If m gives masses only to singletons, it is said to be *Bayesian*; there exists a 1-1 correspondence between Bayesian basic probability assignments and discrete probability distributions.

Given a mass function m on E , its associated belief function $Bel_m : 2^E \rightarrow [0, 1]$ is defined as

$$Bel_m(W) := \sum_{V \subseteq W} m(V) \text{ for all } W \in 2^E. \quad (1)$$

The plausibility function $Pl_m : 2^E \rightarrow [0, 1]$ associated to the mass function m is defined as

$$Pl_m(W) := \sum_{V \cap W \neq \emptyset} m(V) \text{ for all } W \in 2^E. \quad (2)$$

These functions satisfy the conjugacy relation

$$Pl_m(W) = 1 - Bel_m(E \setminus W) \text{ for all } W \in 2^E.$$

Operations with mass functions. There are three important operations from D-S theory that we will use in this work: marginalization, vacuous extension, and Dempster's rule of combination.

Let $(E, 2^E)$ and $(F, 2^F)$ be two measurable spaces and consider $Y \subseteq E \times F$. We define the projection of Y on E as

$$Y^{\downarrow E} := \{e \in E : (e, f) \in Y\}.$$

Let m be a mass function on $E \times F$. The *marginalization* of m to E is a mass function $m^{\downarrow E}$ on E defined as [13]

$$m^{\downarrow E}(V) := \sum_{\substack{Y \subseteq E \times F \\ Y^{\downarrow E} = V}} m(Y) \text{ for all } V \in 2^E. \quad (3)$$

Let now m be a mass function on E . The *vacuous extension* $m^{\uparrow E \times F}$ of m to $E \times F$ is defined as

$$m^{\uparrow E \times F}(V \times F) := m(V) \text{ for all } V \in 2^E. \quad (4)$$

Finally, let m_1 and m_2 be two mass functions on $(E, 2^E)$. Dempster's rule of combination, denoted by \otimes , is an operation between two mass functions that yields a new mass function on the same space. It is defined as

$$m_1 \otimes m_2(V) := \frac{1}{1 - k} \sum_{\substack{V_1, V_2 \in 2^E \\ V_1 \cap V_2 = V}} m_1(V_1) m_2(V_2) \quad (5)$$

where $k \in [0, 1]$ is the *degree of conflict*, defined as

$$k := \sum_{\substack{V_1, V_2 \in 2^E \\ V_1 \cap V_2 = \emptyset}} m_1(V_1) m_2(V_2). \quad (6)$$

If $k = 0$ we say that there is no conflict between m_1 and m_2 ; conversely if $k = 1$ they are in total conflict and they cannot be combined. We shall also note that Dempster's rule of combination is both associative and commutative and, by vacuous extending, can combine mass functions defined on different state spaces.

2.3. Coherence of probability intervals. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $X : \Omega \rightarrow (E_X, 2^{E_X})$ be a discrete random variable taking values in $E_X := \{x_1, \dots, x_n\}$. Let $(\underline{p}^X, \overline{p}^X) := \{(\underline{p}_i^X, \overline{p}_i^X)\}_{i=1}^n$ be a probability interval for the random variable X , namely

$$\underline{p}_i^X \leq \mathbb{P}(X = x_i) \leq \overline{p}_i^X, \quad \text{for all } i = 1, \dots, n.$$

We say that $(\underline{p}^X, \overline{p}^X)$ is a *coherent* probability interval [7, 11] if it satisfies the following conditions:

[Coh1] $\underline{S}_X := \sum_{i=1}^n \underline{p}_i^X \leq 1$ and $\overline{S}_X := \sum_{i=1}^n \overline{p}_i^X \geq 1$;

[Coh2] $\underline{p}_i^X \geq 1 - \sum_{h \neq i} \overline{p}_h^X$, for all $i = 1, \dots, n$;

[Coh3] $\overline{p}_i^X \leq 1 - \sum_{h \neq i} \underline{p}_h^X$, for all $i = 1, \dots, n$.

Given a mass function m_X on E_X , we can generate a coherent probability interval $(\underline{p}^X, \overline{p}^X)$ for the random variable $X : \Omega \rightarrow (E_X, 2^{E_X})$ using the belief and plausibility functions:

$$\underline{p}_i^X = \text{Bel}_{m_X}(\{x_i\}), \text{ and } \overline{p}_i^X = \text{Pl}_{m_X}(\{x_i\}),$$

for all $i = 1, \dots, n$. It immediately follows from the definitions of belief and plausibility that $\text{Bel}_{m_X}(\{x_i\}) \leq \text{Pl}_{m_X}(\{x_i\})$ and the three conditions of coherence hold.

So, from a given mass function, we can always generate a coherent probability interval that agrees with it. However, the converse is not true; there are coherent probability intervals for which it is not possible to find a corresponding mass function [11]. In particular, one *can* always take the natural extension of a coherent probability interval, to obtain a coherent lower prevision [7]; see e.g. (11) further on. However, this lower prevision need not be a belief function, and hence need not have an associated (non-negative) mass function [7, 11]. Moreover, even if such a mass function does exist, it need not be unique; for a given coherent probability interval, there may in general be multiple mass functions whose associated belief and plausibility functions on the singletons, agree with this interval. This issue will be revisited in more detail in Section 3.1.

3. BELIEF INFERENCE IN CHAINS

In this section we expand the framework for belief inference previously introduced by Shenoy [13]. Let $A \rightarrow B$ be a credal network, where A and B are random variables of cardinality n and m respectively. Let us call

$E_A = \{a_1, \dots, a_n\}$ and $E_B = \{b_1, \dots, b_m\}$ the state spaces of A and B respectively. Suppose we are given a mass function m_A on E_A , and n mass functions $m_{B|a_i}$ on E_B , that correspond each to a probability interval for $\mathbb{P}(B|A = a_i)$, i.e.

$$\underline{p}_j^i := \text{Bel}_{m_{B|a_i}}(b_j) \leq \mathbb{P}(b_j|a_i) \leq \text{Pl}_{m_{B|a_i}}(b_j) =: \overline{p}_j^i \quad (7)$$

where we simplify the notation $\mathbb{P}(B = b_j|A = a_i) = \mathbb{P}(b_j|a_i)$. We can extend each mass function $m_{B|a_i}$ to a mass function $m_{B|a_i}$ on $E_A \times E_B$, using Smets' conditional embedding [13, 14], by defining, for all $V \in 2^{E_B}$, the focal sets

$$m_{B|a_i}(a_i \times V \cup (E_A \setminus \{a_i\} \times E_B)) = m_{B|a_i}(V).$$

The mass functions built in this way are the only ones on $E_A \times E_B$ that have the following properties:

- $(m_{B|a_i} \otimes m_{A=a_i}^{\uparrow E_A \times E_B})^{\downarrow E_B} = m_{B|a_i}$ for all $i = 1, \dots, n$, where $m_{A=a_i}$ is the deterministic mass function on E_A with $m_{A=a_i}(\{a_i\}) = 1$;
- Each $m_{B|a_i}^{\downarrow E_A}$ is vacuous on E_A ;
- $m_{B|A} := \bigotimes_{i=1}^n m_{B|a_i}$ is a mass function on $E_A \times E_B$, in which all of the pairwise combinations have no conflict, and $m_{B|A}^{\downarrow E_A}$ is vacuous on E_A .

In summary, we can express the local uncertainty models of this CN using m_A and $m_{B|A}$. The associated global model $m_{A,B}$ is then simply obtained by defining $m_{A,B} = m_A^{\uparrow E_A \times E_B} \otimes m_{B|A}$. It can be easily checked that the mass functions that are combined to obtain $m_{A,B}$ have no conflict, since the intersection of any two focal sets is always non-empty.

From this global model, we can derive any inferences of interest. In this work, we generally focus on marginal inference, e.g. the marginal uncertainty on B , expressed as $m_B = m_{A,B}^{\downarrow B}$. Using the mass function m_B , we may generate a probability interval on B , $(\underline{p}^B, \overline{p}^B)$, using Bel_{m_B} and Pl_{m_B} . Our first result is a closed-form expression for the values of this probability interval.

Theorem 3.1. *Let m_A , $m_{B|A}$, $m_{A,B}$, and m_B be as defined above. Then for all $j = 1, \dots, m$ it holds that*

$$\underline{p}_j^B := \text{Bel}_{m_B}(\{b_j\}) = \sum_{V \in 2^{E_A}} \left(m_A(V) \prod_{i: a_i \in V} \underline{p}_j^i \right), \text{ and} \quad (8)$$

$$\overline{p}_j^B := \text{Pl}_{m_B}(\{b_j\}) = 1 - \sum_{V \in 2^{E_A}} \left(m_A(V) \prod_{i: a_i \in V} (1 - \overline{p}_j^i) \right), \quad (9)$$

with \underline{p}_j^i and \overline{p}_j^i as in Equation (7).

This result is a particular case of Equation (5.1) in [15] and the proof can be found in the Appendix.

One may now wonder, how the interval $(\underline{p}^B, \bar{p}^B)$ obtained using Theorem 3.1, relates to the values $\underline{\mathbb{P}}(B)$ and $\bar{\mathbb{P}}(B)$ computed from a CN under complete independence, as in Section 2.1. At this stage we need to make the observation that for the CN in Section 2.1, our uncertainty model for A was a probability interval $(\underline{p}^A, \bar{p}^A)$. Conversely, our discussion here started from an (arbitrary) mass function m_A and, as mentioned in Section 2.3, there need not be a one-to-one correspondence between the two.

As the following example shows, it is sometimes possible to choose a mass function m_A that agrees with a given $(\underline{p}^A, \bar{p}^A)$, such that the values $(\underline{p}^B, \bar{p}^B)$ are *not* guaranteed bounds on $\underline{\mathbb{P}}(B)$ and $\bar{\mathbb{P}}(B)$.

Example 3.1. Consider a CN $A \rightarrow B$, where the random variable A has state space $E_A = \{a_1, \dots, a_4\}$ and satisfies $0.2 \leq \mathbb{P}(A = a_i) \leq 0.3$ for all $i = 1, \dots, 4$. Consider the mass function m_A on E_A that assigns mass 0.2 to singletons, $m(\{a_1, a_2\}) = m(\{a_3, a_4\}) = 0.1$, and zero mass elsewhere. It is easily verified that indeed $\text{Bel}_{m_A}(\{a_i\}) = 0.2$ and $\text{Pl}_{m_A}(\{a_i\}) = 0.3$ for all $i = 1, \dots, 4$.

Fix any $j \in \{1, \dots, |E_B|\}$, and suppose the collection of lower bounds for $\mathbb{P}(b_j|a_i)$ is $(\underline{p}_j^i)_{i=1}^4 = (0.2, 0.4, 0.8, 0.9)$.

If we compute \underline{p}_j^B using (8) we obtain a value of 0.54, whereas it is easy to check that

$$\underline{\mathbb{P}}(B = b_j) = \min_{\substack{\underline{p}_i^A \leq \mathbb{P}(a_i) \leq \bar{p}_i^A \\ \sum_i \mathbb{P}(a_i) = 1}} \sum_{i=1}^n \mathbb{P}(a_i) \underline{p}_j^i = 0.52.$$

This provides an example where the credal lower bound for $\mathbb{P}(B = b_j)$ is lower than \underline{p}_j^B . \diamond

We next focus on finding mass functions which ensure that intervals generated by belief inference, provide a conservative approximation of credal intervals.

3.1. Standard good mass and good probability intervals. We say that a coherent probability interval $(\underline{p}^A, \bar{p}^A)$ on $A : \Omega \rightarrow E_A$ is *good* if

$$\Delta_A := \bar{S}_A + (n-2)\underline{S}_A - (n-1) \geq 0, \quad (10)$$

where $n = |E_A|$, $\bar{S}_A := \sum_{i=1}^n \bar{p}_i^A$ and $\underline{S}_A := \sum_{i=1}^n \underline{p}_i^A$. If a coherent probability interval $(\underline{p}^A, \bar{p}^A)$ for the random variable A is *good*, then there always exists a mass function m_A with at most $2n+1$ focal sets, such that the associated belief and plausibility of singletons match this interval. One such mass function m_A that satisfies this property is characterized by

$$m_A(\{a_i\}) = \underline{p}_i^A$$

$$m_A(E_A \setminus \{a_i\}) = 1 - \bar{p}_i^A - \sum_{h \neq i} \underline{p}_h^A$$

$$m_A(E_A) = \Delta_A$$

for all $i = 1, \dots, n$. We shall call this mass function on E_A the *standard good mass* (SGM) associated to $(\underline{p}^A, \bar{p}^A)$.

Relation between Möbius inverse and standard good mass. It is well-known that there is a one-to-one correspondence between belief (resp. plausibility) functions, and mass functions. Indeed, Equation (1) (resp. (2)) gives the connection in one direction. To go from a belief function Bel to a mass function m , one uses the Möbius inverse

$$m(V) = \sum_{W \subseteq V} (-1)^{|V \setminus W|} \text{Bel}(W).$$

If we start from a coherent probability interval $(\underline{p}^A, \bar{p}^A)$ and want to use this machinery, we therefore first have to represent it as a belief function. It is also well-known that this is not always possible; see e.g. [11] for a characterization of when it is. Following this work, the basic idea is to take the natural extension of this interval,

$$\begin{aligned} \text{Bel}(V) &= \max \left\{ \sum_{i: a_i \in V} \underline{p}_i^A, 1 - \sum_{i: a_i \notin V} \bar{p}_i^A \right\}, \\ \text{Pl}(V) &= \min \left\{ \sum_{i: a_i \in V} \bar{p}_i^A, 1 - \sum_{i: a_i \notin V} \underline{p}_i^A \right\}. \end{aligned} \quad (11)$$

If (and only if) this extension is indeed a belief (resp. plausibility) function, we can then use the Möbius inverse to find an associated mass function.

In the specific case that the interval $(\underline{p}^A, \bar{p}^A)$ is *good*, we have the following result.

Theorem 3.2. *If $(\underline{p}^A, \bar{p}^A)$ is a good coherent probability interval for the random variable $A : \Omega \rightarrow E_A = \{a_1, \dots, a_n\}$, then the mass function m_A^M obtained using (11) and the Möbius inverse, coincides with the standard good mass associated to $(\underline{p}^A, \bar{p}^A)$.*

The proof of this result can be found in the Appendix.

3.2. Credal bound for standard good masses. In this section we present one of the main results of this paper, namely the fact that when the local probability intervals of a CN can be represented using SGMs, then the marginal inference $(\underline{p}^B, \bar{p}^B)$ computed using (8) and (9) is always a conservative approximation of the credal bounds $\underline{\mathbb{P}}(B)$ and $\bar{\mathbb{P}}(B)$.

Theorem 3.3. *If $(\underline{p}^A, \bar{p}^A)$ is a good coherent probability interval and m_A its associated standard good mass, then for all $j = 1, \dots, m$ it holds that*

$$\underline{p}_j^B \leq \min_c \sum_{i=1}^n \mathbb{P}(B = b_j | A = a_i) \mathbb{P}(A = a_i) \quad (12)$$

$$\bar{p}_j^B \geq \max_{\mathcal{C}} \sum_{i=1}^n \mathbb{P}(B = b_j | A = a_i) \mathbb{P}(A = a_i) \quad (13)$$

where $\mathcal{C} = \{\underline{p}_j^i \leq \mathbb{P}(B = b_j | A = a_i) \leq \bar{p}_j^i, \underline{p}_i^A \leq \mathbb{P}(A = a_i) \leq \bar{p}_i^A, \sum_i \mathbb{P}(A = a_i) = 1\}$.

Proof. We start by proving (12). Fix $j \in \{1, \dots, m\}$. Without loss of generality let us order $\{\underline{p}_j^i\}_{i=1}^n$ increasingly:

$$\underline{p}_j^1 \leq \dots \leq \underline{p}_j^n$$

We can rewrite the minimum in (12) as:

$$\min_{\mathcal{C}} \sum_{i=1}^n \mathbb{P}(b_j | a_i) \mathbb{P}(a_i) = \min_{\substack{\underline{p}_i^A \leq x_i \leq \bar{p}_i^A \\ \sum_{i=1}^n x_i = 1}} \sum_{i=1}^n \underline{p}_j^i x_i.$$

Since by [Coh1] it holds that

$$\sum_{i=1}^n \underline{p}_i^A \leq 1 \leq \sum_{i=1}^n \bar{p}_i^A,$$

there exists a least index \tilde{i} such that

$$\bar{p}_1^A + \dots + \bar{p}_{\tilde{i}-1}^A + x_{\tilde{i}} + \underline{p}_{\tilde{i}+1}^A + \dots + \underline{p}_n^A = 1$$

where $x_{\tilde{i}} \in [\underline{p}_{\tilde{i}}^A, \bar{p}_{\tilde{i}}^A]$ is unambiguously determined. In order to achieve the minimum in (12) the values of x_i need to be as high as possible when \underline{p}_j^i is small and vice versa. Therefore, the vector $(x_i)_{i=1}^n$ that realizes the minimum is $(\bar{p}_1^A, \dots, \bar{p}_{\tilde{i}-1}^A, x_{\tilde{i}}, \underline{p}_{\tilde{i}+1}^A, \dots, \underline{p}_n^A)$, where \tilde{i} is taken as small as possible.

Let us rewrite inequality (12) using (8) and the definition of standard good mass:

$$\sum_{i=1}^n \underline{p}_i^A \underline{p}_j^i + \sum_{i=1}^n m_A(E_A \setminus \{a_i\}) \prod_{h \neq i} \underline{p}_j^h + \Delta_A \prod_{i=1}^n \underline{p}_j^i \leq \sum_{i < \tilde{i}} \bar{p}_i^A \underline{p}_j^i + x_{\tilde{i}} \underline{p}_j^{\tilde{i}} + \sum_{i > \tilde{i}} \underline{p}_i^A \underline{p}_j^i.$$

Therefore, we need to prove that:

$$\sum_{i < \tilde{i}} (\bar{p}_i^A - \underline{p}_i^A) \underline{p}_j^i + (x_{\tilde{i}} - \underline{p}_{\tilde{i}}^A) \underline{p}_j^{\tilde{i}} \geq \sum_{i=1}^n m_A(E_A \setminus \{a_i\}) \prod_{h \neq i} \underline{p}_j^h + \Delta_A \prod_{i=1}^n \underline{p}_j^i =: \star.$$

Let us note that due to the condition of coherency [Coh2], it follows that $\tilde{i} \geq 2$ for every $n \geq 3$. Using that $\underline{p}_j^1 \geq \prod_{h \neq i} \underline{p}_j^h$ for all $i \neq 1$ and $\underline{p}_j^2 \geq \prod_{h \neq 1} \underline{p}_j^h$, we get

$$\star \leq \underline{p}_j^1 \left(\sum_{i \neq 1} m_A(E_A \setminus \{a_i\}) + m_A(E_A) \right) + \underline{p}_j^2 m_A(E_A \setminus \{a_1\})$$

Let us observe that it follows from the definition of standard good mass that

$$\sum_{i \neq 1} m_A(E_A \setminus \{a_i\}) + m_A(E_A) = \bar{p}_1^A - \underline{p}_1^A.$$

Therefore, it is enough to prove that

$$\sum_{2 \leq i < \tilde{i}} (\bar{p}_i^A - \underline{p}_i^A) \underline{p}_j^i + (x_{\tilde{i}} - \underline{p}_{\tilde{i}}^A) \underline{p}_j^{\tilde{i}} \geq \underline{p}_j^2 m_A(E_A \setminus \{a_1\})$$

and this inequality immediately follows from the fact that

$$\begin{aligned} \sum_{2 \leq i < \tilde{i}} (\bar{p}_i^A - \underline{p}_i^A) + (x_{\tilde{i}} - \underline{p}_{\tilde{i}}^A) &= 1 - \underline{p}_1^A - (\bar{p}_1^A - \underline{p}_1^A) \\ &= m_A(E_A \setminus \{a_1\}). \end{aligned}$$

The proof for (13) follows a procedure similar to the one just presented. It is also easy to check that (12) and (13) also hold for the particular case $n = 2$. \square

This result is straightforwardly extended to chains of arbitrary length. Indeed, if we consider the chain $A \rightarrow B \rightarrow C$ where $E_A = \{a_1, \dots, a_n\}$, $E_B = \{b_1, \dots, b_m\}$, $E_C = \{c_1, \dots, c_l\}$, and

$$\begin{aligned} \underline{p}_i^A &\leq \mathbb{P}(A = a_i) \leq \bar{p}_i^A \\ \underline{p}_j^i &\leq \mathbb{P}(B = b_j | A = a_i) \leq \bar{p}_j^i \\ \underline{p}_k^j &\leq \mathbb{P}(C = c_k | B = b_j) \leq \bar{p}_k^j, \end{aligned}$$

it also holds for all $k = 1, \dots, l$ that

$$\underline{p}_k^C \leq \min_{i,j} \sum \mathbb{P}(a_i) \mathbb{P}(b_j | a_i) \mathbb{P}(c_k | b_j), \quad (14)$$

and similarly for the upper bound \bar{p}_k^C . The quantity \underline{p}_k^C is obtained by applying (8), (9) twice—first to obtain $(\underline{p}_k^B, \bar{p}_k^B)$, then to combine *that* with the $(\underline{p}_k^j, \bar{p}_k^j)$ —whence it follows from Theorem 3.3 that

$$\underline{p}_k^C \leq \min_{\substack{\mathbb{P}(b_j) \leq \bar{\mathbb{P}}(b_j) \\ \sum_j \mathbb{P}(b_j) = 1}} \sum_j \mathbb{P}(b_j) \underline{p}_k^j =: \diamond$$

with $\mathbb{P}(b_j)$ and $\bar{\mathbb{P}}(b_j)$ as before. Moreover, we clearly also have that $\diamond \leq \min_{i,j} \sum \mathbb{P}(a_i) \mathbb{P}(b_j | a_i) \underline{p}_k^j$, which is all that is required to obtain (14). A similar argument can be applied to prove that the bound holds for \bar{p}_k^C and for chains of arbitrary length.

3.3. Fixing bad probability intervals. Suppose we are considering a CN $X_1 \rightarrow \dots \rightarrow X_k$ and suppose we are given good probability intervals for the random variable X_1 and for each conditional distribution $X_l | (X_{l-1} = x)$ of

every step of the chain. Then, we may apply formulas (8) and (9) step-by-step starting from X_1 to eventually infer a probability interval for X_k , $(\underline{p}^{X_k}, \overline{p}^{X_k})$. At each step, after applying (8) and (9), we revert the probability interval into a mass function and proceed along the chain. We call this process *belief inference* on the credal network $X_1 \rightarrow \dots \rightarrow X_k$. This probability interval on X_k is the same as the one associated with the mass function

$$(m_{X_1} \otimes m_{X_2|X_1} \otimes \dots \otimes m_{X_k|X_{k-1}})^{\downarrow X_k}, \quad (15)$$

thanks to the local computation property [13] of mass functions, which states that

$$(m_{X_1} \otimes m_{X_2|X_1} \otimes m_{X_3|X_2})^{\downarrow (X_2, X_3)} = (m_{X_1} \otimes m_{X_2|X_1})^{\downarrow X_2} \otimes m_{X_3|X_2}.$$

Directly computing (15) is computationally expensive as m_{X_1, \dots, X_k} gives mass to an exponential number of focal sets namely $\mathcal{O}(2^{kn})$, where $|X_i| = n$ for all $i = 1, \dots, k$.

The step-by-step process we described requires only $\mathcal{O}(kn^3)$ operations when employing standard good masses. However, this method is not always applicable since there is no guarantee that, after one step of inference, we can find a mass function associated to the inferred probability interval. One of the possible solutions to this issue is to consider the closest good outer approximation to a bad probability interval.

If we are considering the credal network $A \rightarrow B$ and we are given a bad probability interval for the random variable A , a simple way to fix $(\underline{p}^A, \overline{p}^A)$ consists in slightly enlarging the intervals, by distributing the quantity $-\Delta_A = (n-1) - \overline{S}_A - (n-2)\underline{S}_A > 0$ over the $\{\underline{p}_i^A\}_{i=1}^n$; it follows from Equation (10) that the resulting interval will be good. The bottom line of this approach is that we are bartering precision for computational speed and tractability.

Fixing a bad probability interval for A does not break inequalities (12) and (13), as the following corollary states.

Corollary 3.1. *If $(\underline{p}^A, \overline{p}^A)$ is a bad probability interval and m_A is the standard good mass associated to an arbitrary fixing of $(\underline{p}^A, \overline{p}^A)$, then inequalities (12) and (13) still hold.*

The proof of this result is straightforward and omitted due to page limit constraints.

If some probability interval for the conditional random distributions $B \mid (A = a_i)$ are bad, we may compute \underline{p}_j^B and \overline{p}_j^B using expressions (8) and (9), treating the bad probability intervals as if they were good. This occurs because increasing the upper bounds $\{\overline{p}_j^i\}_{j=1}^m$ to fix $(\underline{p}^i, \overline{p}^i)$ does not affect \underline{p}_j^B in expression (8). When

calculating \overline{p}_j^B for a fix j , there always exists a fixing for every $(\underline{p}^i, \overline{p}^i)$ such that \overline{p}_j^i remains unchanged for all $i = 1, \dots, n$. In order to prove this statement, it is enough to show that we can always distribute $-\Delta_i = (m-1) - \overline{S}_i - (m-2)\underline{S}_i$ over $\{\overline{p}_k^i\}_{k \neq j}$. Each \overline{p}_k^i can increase by at most $r_k^i = 1 - \underline{S}_i - (\overline{p}_k^i - \underline{p}_k^i)$ to preserve the condition of coherency [Coh3]. Therefore

$$\begin{aligned} \sum_{k \neq j} r_k^i &= \sum_{k \neq j} (1 - \underline{S}_i - \overline{p}_k^i + \underline{p}_k^i) \\ &= (m-1) - (m-1)\underline{S}_i - \overline{S}_i + \overline{p}_j^i + \underline{S}_i - \underline{p}_j^i \\ &= (m-1) - \overline{S}_i - (m-2)\underline{S}_i + \overline{p}_j^i - \underline{p}_j^i \\ &= -\Delta_i + \overline{p}_j^i - \underline{p}_j^i \geq -\Delta_i. \end{aligned}$$

This confirms that if we are provided with any coherent probability interval for $B \mid (A = a_i)$ for all i , inference can be performed without fixing these intervals into good ones.

Ad-hoc fixing. Consider the credal network $A \rightarrow B$ where we are given a bad probability interval for the random variable A , $(\underline{p}^A, \overline{p}^A)$, and coherent probability intervals for the conditional distributions, namely $\{(\underline{p}^i, \overline{p}^i)\}_{i=1}^n$. Since inequality (12) holds for every fixing of $(\underline{p}^A, \overline{p}^A)$, we shall consider the distribution of $-\Delta_A$ that maximizes the value of \underline{p}_j^B thus minimizing the width of the inferred interval. Let $t = \{t_i\}_{i=1}^n$ be the values added to $\{\underline{p}_i^A\}_{i=1}^n$ to transform the bad probability interval into a good one. It necessarily holds that $0 \leq t_i \leq r_i^A$ and $\sum_i t_i = -\Delta_A$. Let \underline{t}^{*j} be the distribution of such values that maximizes \underline{p}_j^B , namely

$$\begin{aligned} \underline{t}^{*j} &= \arg \max_{\substack{0 \leq t_i \leq r_i^A \\ \sum_i t_i = -\Delta_A}} \sum_{i=1}^n \left((1 - (\overline{p}_i^A + t_i) - \sum_{h \neq i} \underline{p}_h^A) \prod_{h \neq i} \underline{p}_h^B \right) \\ &= \arg \min_{\substack{0 \leq t_i \leq r_i^A \\ \sum_i t_i = -\Delta_A}} \sum_{i=1}^n \left(t_i \prod_{h \neq i} \underline{p}_h^B \right). \end{aligned} \quad (16)$$

Similarly, we shall define \overline{t}^{*j} , the distribution over $\{\overline{p}_i^A\}_{i=1}^n$ that minimizes \overline{p}_j^B .

$$\overline{t}^{*j} = \arg \min_{\substack{0 \leq t_i \leq r_i^A \\ \sum_i t_i = -\Delta_A}} \sum_{i=1}^n \left(t_i \prod_{h \neq i} (1 - \overline{p}_h^B) \right). \quad (17)$$

By first fixing $(\underline{p}^A, \overline{p}^A)$ with \underline{t}^{*j} to compute \underline{p}_j^B , and then fixing $(\underline{p}^A, \overline{p}^A)$ with \overline{t}^{*j} to compute \overline{p}_j^B , we obtain the

smallest interval — among those generated using the standard good mass — for $\mathbb{P}(B = b_j)$. We shall note that inequalities (12) and (13) remain valid, as they hold for any possible fixing of the probability interval $(\underline{p}^A, \overline{p}^A)$. This operation can be repeated for each j , resulting in a probability interval $(\underline{p}^B, \overline{p}^B)$ for B . We refer to this method of fixing $(\underline{p}^A, \overline{p}^A)$ to perform inference as *ad-hoc fixing*.

Example 3.2. Consider an urn with $N = 100$ balls coloured in red (R), green (G), blue (B) or yellow (Y). We do not know the proportion of balls for each colour and we have the following estimates:

$$\begin{aligned} 6 \leq N(R) \leq 33 & & 10 \leq N(G) \leq 42 \\ 15 \leq N(B) \leq 32 & & 25 \leq N(Y) \leq 58. \end{aligned}$$

We possess four coins, one for each colour, which we know are unbalanced yet we are still uncertain about the exact mean value:

$$\begin{aligned} 0.8 \leq \mathbb{P}(H|col = R) \leq 0.9 & & 0.6 \leq \mathbb{P}(H|col = G) \leq 0.7 \\ 0.4 \leq \mathbb{P}(H|col = B) \leq 0.5 & & 0.2 \leq \mathbb{P}(H|col = Y) \leq 0.3. \end{aligned}$$

We sample uniformly at random a ball from the urn and flip the associated coin. The inference task of calculating the lower probability of getting heads (H) can be computed using belief inference with ad-hoc fixing. The probability interval associated with the random variable col representing the colour of the sampled ball is

$$\begin{aligned} 0.06 \leq \mathbb{P}(R) \leq 0.33 & & 0.1 \leq \mathbb{P}(G) \leq 0.42 \\ 0.15 \leq \mathbb{P}(B) \leq 0.32 & & 0.25 \leq \mathbb{P}(Y) \leq 0.58. \end{aligned}$$

This probability interval is bad ($\Delta = -0.23$) thus requiring fixing. We can widen this probability interval by at most $r = (0.17, 0.12, 0.27, 0.11)$ while preserving the conditions of coherency. The values of $\prod_{h \neq i} \underline{p}_j^h$ in (16) are (0.048, 0.064, 0.096, 0.192). Therefore, the best way to fix the probability interval for col in order to maximize \underline{p}_{-H} , the lower probability of getting heads, is $\underline{t}_H^* = (0.17, 0.06, 0, 0)$ and we find

$$\underline{p}_{-H} \approx 0.253 \leq 0.328 = \underline{\mathbb{P}}(H)$$

where this last value represents the lower probability obtained via classical sensitivity analysis. \diamond

Ad-hoc mass functions. Recalling that our goal is to estimate credal intervals as accurately as possible, one potential approach is to explore alternatives beyond relying solely on the standard good mass. Specifically, we could determine the mass function on A that generates the smallest possible probability interval for B . For a

fixed $j = 1, \dots, m$, we may consider the following mass function on A :

$$\underline{m}_j^{max} := \arg \max_{\substack{0 \leq m_A(V) \leq 1 \\ Bel_{m_A}(\{a_i\}) = \underline{p}_i^A \\ Pl_{m_A}(\{a_i\}) = \overline{p}_i^A}} \sum_{V \in 2^{E_A}} \left(m_A(V) \prod_{i: x_i \in V} \underline{p}_j^i \right). \quad (18)$$

This mass function yields the largest possible value for \underline{p}_{-j}^B . However, as mentioned at the beginning of Section 3, there is no guarantee that this value of \underline{p}_{-j}^B is always less than the lower bound of the corresponding credal interval, given by $\min_{\underline{p}_i^A \leq x_i \leq \overline{p}_i^A, \sum_i x_i = 1} \sum_{i=1}^n \underline{p}_j^i x_i$. Example 3.1

shows one such example.

Referring to the proof of Theorem 3.3, we observe that the use of the standard good mass ensures that every set V with $|V| \geq 2$, always contains either $\{a_1\}$ or $\{a_2\}$. This property does not hold for an arbitrary mass function on A , preventing the use of the inequality $\prod_{i: a_i \in V} \underline{p}_j^i \leq \underline{p}_j^2$, which is fundamental to complete the proof of Theorem 3.3. It is important to recall that by \underline{p}_{-j}^2 we mean the second smallest value among $\{\underline{p}_{-j}^i\}_{i=1}^n$.

For instance, in Example 3.1, the set $\{a_3, a_4\}$ is a focal set for m_A and it holds that $\underline{p}_{-j}^3 \underline{p}_{-j}^4$ is greater than \underline{p}_{-j}^2 .

Therefore, if we assign zero mass to sets V with $|V| \geq 2$ such that $\prod_{i: a_i \in V} \underline{p}_j^i > \underline{p}_j^2$, inference using this mass function on A always guarantees that \underline{p}_{-j}^B is smaller than the corresponding credal lower bound.

Let us introduce \underline{m}_j^* , a mass function on A designed to ensure that inequality (12) holds:

$$\underline{m}_j^* := \arg \max_{\substack{0 \leq m_A(V) \leq 1 \\ Bel_{m_A}(\{a_i\}) = \underline{p}_i^A \\ Pl_{m_A}(\{a_i\}) = \overline{p}_i^A \\ m_A(V) = 0 \text{ if } |V| \geq 2 \wedge \prod_{i: a_i \in V} \underline{p}_j^i > \underline{p}_j^2}} \sum_{V \in 2^{E_A}} \left(m_A(V) \prod_{i: x_i \in V} \underline{p}_j^i \right). \quad (19)$$

Similarly, we may define a mass function that minimizes the upper bound \overline{p}_{-j}^B while ensuring that inequality (13) holds:

$$\overline{m}_j^* := \arg \max_{\substack{0 \leq m_A(V) \leq 1 \\ Bel_{m_A}(\{a_i\}) = \underline{p}_i^A \\ Pl_{m_A}(\{a_i\}) = \overline{p}_i^A \\ m_A(V) = 0 \text{ if } |V| \geq 2 \wedge \prod_{i: a_i \in V} (1 - \overline{p}_j^i) > 1 - \overline{p}_j^2}} \sum_{V \in 2^{E_A}} \left(m_A(V) \prod_{i: x_i \in V} (1 - \overline{p}_j^i) \right). \quad (20)$$

We refer to these mass functions on A as the *ad-hoc mass functions* for computing the interval $(\underline{p}_{-j}^B, \overline{p}_{-j}^B)$. The operation of inferencing with ad-hoc mass functions can

be repeated for each j , resulting in a probability interval $(\underline{p}^B, \overline{p}^B)$ for B . Practically, computing each of the $2m$ ad-hoc mass functions consists of solving a linear programming problem.

Example 3.3. Let us try to use ad-hoc mass functions to compute the lower probability of getting heads in Example 3.2. Solving the linear programming problem of (19) we obtain

$$\begin{aligned} \underline{m}_H^*({R}) &= 0.06 & \underline{m}_H^*({G}) &= 0.1 & \underline{m}_H^*({B}) &= 0.15 \\ \underline{m}_H^*({Y}) &= 0.25 & \underline{m}_H^*({R, B}) &= 0.054 \\ \underline{m}_H^*({G, B}) &= 0.048 & \underline{m}_H^*({R, G, B}) &= 0.138 \\ \underline{m}_H^*({R, G, Y}) &= 0.174 & \underline{m}_H^*({R, G, B, Y}) &= 0.006. \end{aligned}$$

The lower probability calculated using the ad-hoc mass function \underline{m}_H^* in (8) is 0.2861, which is a slight improvement over the ad-hoc fixing method. \diamond

Let us note that belief inference with ad-hoc mass functions may generate probability intervals that cannot be represented by an ad-hoc mass function, as the argmax in definitions (19) and (20) might be taken over unfeasible constraints. A practical solution to this issue is presented in the following section.

4. NUMERICAL ANALYSIS

In Section 3.2 it was shown that, under specific assumptions, intervals generated using belief inference are more conservative than their credal counterparts. Here, we aim to address the question of how much larger these intervals become.

Let us present some numerical results about the chain $X_1 \rightarrow \dots \rightarrow X_k$. In order to perform numerical simulations on the chain, we fixed the cardinality of each random variable $|X_i| = n$ and sampled uniformly at random (using a hit&run algorithm [16]) $n(k-1) + 1$ probability intervals; one for each conditional distribution $X_i | (X_{1-i} = x_i)$, $i = 2, \dots, k$ and $i = 1, \dots, n$, and one for X_1 . Then, we implemented different methods to perform inference, namely credal inference, belief inference with ad-hoc mass functions and belief inference using the ad-hoc fixing method. We compared the results of these simulations using the average width of inferred intervals as a metric.

Figure 1 illustrates the average width of intervals as the inference process, using coherent probability intervals sampled uniformly at random, advances along the chain. Let us first note that when $n = 2$ or $n = 3$ the two belief inference methods we presented coincide, as the mass function associated to a probability interval is then unique (if it exists). We observe that the process of belief inference becomes increasingly uninformative as the cardinality of the random variables grows. Already with $n = 5$ belief inference generates intervals that are

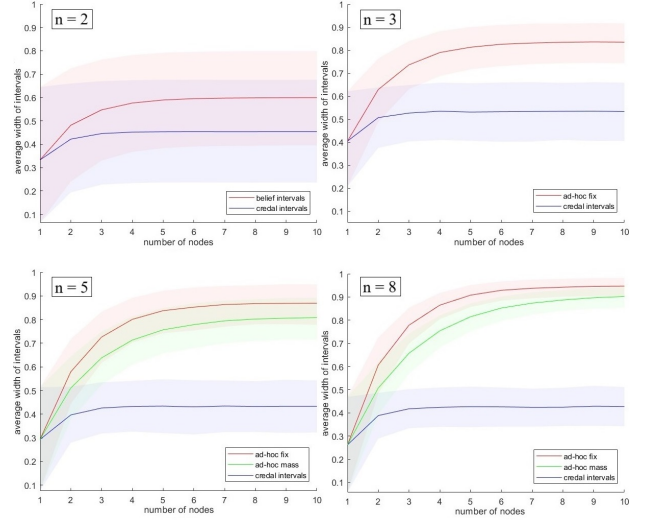


Figure 1. Evolution of the average width of intervals as inference progresses along the chain $X_1 \rightarrow \dots \rightarrow X_{10}$, with each graph corresponding to a different cardinality n of random variables. The colours represent different inference methods: blue for credal intervals, red for belief inference with standard good masses and ad-hoc fixing, green for belief inference with ad-hoc mass functions. Each coloured region contains the width values of 75% of inferred intervals.

twice as large as their credal counterparts, leading to a significant loss of precision.

On the other hand, in the binary case ($|X_i| = 2$ for all random variables), we observe that intervals generated using belief inference are asymptotically about 40% larger than their credal counterparts. Moreover, after a single step of inference the average enlargement is approximately 0.0592; equivalent to 19% of the credal interval, as shown in Table 1. These observations suggest that belief inference may be effectively applied to more complex binary credal networks with minimal information loss.

n	2	3	4	5	6	8
RIW ad-hoc fix	19%	26%	39%	49%	55%	60%
RIW ad-hoc mass	19%	26%	29%	30%	31%	32%

Table 1. Approximate Relative Increase in Width (RIW) after one step of inference in the CN $A \rightarrow B$, with $|A| = |B| = n$, comparing intervals generated using belief inference and credal inference.

It is worth noting that the average width of a coherent probability interval can be quite large, meaning that sampling them uniformly at random may not accurately represent real-world scenarios where intervals tend to be

narrower. To address this limitation, the following section examines the performance of belief inference when restricted to smaller intervals.

4.1. Using small intervals. This section focuses on the analysis of belief inference using conditional and prior probability intervals with limited width.

We empirically observe that this setup, when using standard good masses, generates a higher number of bad probability intervals as we advance along the chain, thus requiring fixing in order to proceed with the inference process. This leads to larger intervals, as we outer-approximate a bad probability interval with a good one, requiring additional steps to reach asymptotic behaviour. This phenomenon is illustrated in Figure 2. Furthermore,

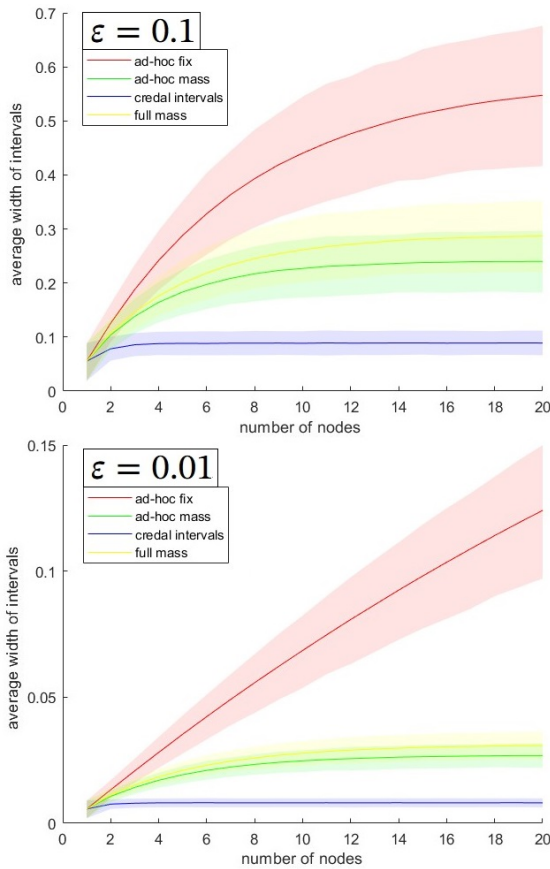


Figure 2. Evolution of the average width of intervals as we make inference along the chain $X_1 \rightarrow \dots \rightarrow X_{20}$, where $|X_i| = 5$ for all i and we vary the value of ϵ , the largest possible width of a conditional interval. Each coloured region contains the width values of 75% of inferred intervals.

in this setup it is highly likely that belief inference with ad-hoc mass function generates probability intervals that cannot be represented by an ad-hoc mass function, as mentioned in Section 3.3. In our experiments we addressed this issue by uniformly enlarging the probability

interval until becoming representable with a mass function. It has been empirically observed that this enlargement is often minimal and not impactful in the overall performance.

Looking at Figure 2, we observe that the ad-hoc fixing method performs significantly worse than other analysed inference approaches. The need to enlarge intervals to proceed with inference lead to excessively wide intervals and it is a great limitation of this method.

On the other hand, we observe that belief inference with ad-hoc mass functions ensures that the asymptotic width of intervals almost never exceeds 3ϵ , keeping intervals narrow and representative of the inherent uncertainty. This result suggest that this approach might be extended to more complex graphical models without a significant loss of information.

Inferencing keeping the full mass function without converting it into a probability interval has shown average worse performance than the ad-hoc mass approach. Most importantly, this method does not guarantee the bounds stated in Theorem 3.3. In other words, probability intervals produced by this approach lack any inherent connection to credal intervals, which prevents us from assigning them a meaningful interpretation.

5. CONCLUSION

This paper explored belief inference in credal networks through Dempster-Shafer theory, proposing a framework for uncertainty propagation. Key contributions include the formalization of belief-based methods, novel computational techniques for marginal lower and upper probabilities, and theoretical bounds with standard credal inference. While the theoretical framework guarantees conservative intervals, numerical results reveal significant limitations. Belief inference methods often yield overly conservative intervals, especially in scenarios with higher cardinality. Belief inference in binary chains and the ad-hoc mass function approach appear to be the only frameworks that reliably ensure a close outer-approximation of credal intervals in chains. It remains open whether this will carry on to more complex graphical models, where achieving precise approximations is often more challenging, in view of the growth in complexity for algorithms on non-binary credal networks.

A. APPENDIX

Proof of Theorem 3.1. Let us first observe that focal sets of $m_{B|A}$ have the form of $\bigcup_{i=1}^n (a_i \times T^i)$ where $T^i \in 2^{E_B}$ for all $i = 1, \dots, n$. It holds that

$$m_{B|A} \left(\bigcup_{i=1}^n (a_i \times T^i) \right) = \prod_{i=1}^n m_{B_{a_i}}(T^i).$$

It follows that focal sets of $m_{A,B}$ have the form $\bigcup_{i: a_i \in V} (a_i \times T^i)$ for all $V \in 2^{E_A}$ and $T^i \in 2^{E_B}$. Recalling

that $m_{A,B} = m_A \otimes (\bigotimes_i m_{B|a_i})$ and applying Dempster's rule of combination, we obtain

$$\begin{aligned} m_{A,B} \left(\bigcup_{i: a_i \in V} (a_i \times T^i) \right) &= \\ &= m_A(V) \sum_{h: a_h \notin V} \left(\sum_{T^h \in 2^{E_B}} m_{B|A} \left(\bigcup_{i=1}^n (a_i \times T^i) \right) \right) \\ &= m_A(V) \sum_{h: a_h \notin V} \left(\sum_{T^h \in 2^{E_B}} \prod_{i=1}^n m_{B|a_i}(T^i) \right) \\ &= m_A(V) \prod_{i: a_i \in V} m_{B|a_i}(T^i), \end{aligned}$$

where the last equality follows from the fact that $\sum_{T^h \in 2^{E_B}} m_{B|a_h}(T^h) = 1$ which implies that indices $h : a_h \notin V$ disappear from the product. Expression (8) immediately follows from the definitions, and we get

$$\begin{aligned} \underline{p}_j^B &= Bel_{m_B}(\{b_j\}) = m_B(\{b_j\}) = \\ &= \sum_{V \in 2^{E_A}} m_{A,B}(V \times \{b_j\}) \\ &= \sum_{V \in 2^{E_A}} m_A(V) \prod_{i: a_i \in V} m_{B|a_i}(\{b_j\}) \\ &= \sum_{V \in 2^{E_A}} m_A(V) \prod_{i: a_i \in V} \underline{p}_j^i. \end{aligned}$$

The proof for (9) is similar: the key is to prove that

$$Bel_{m_B}(E_B \setminus \{b_j\}) = \sum_{V \in 2^{E_A}} m_A(V) \prod_{i: a_i \in V} (1 - \underline{p}_j^i)$$

which follows from $Bel_{m_{B|a_i}}(E_B \setminus \{b_j\}) = 1 - \underline{p}_j^i$. \square

Proof of Theorem 3.2. The fact that $m_A^M(\{a_i\}) = \underline{p}_i^A$ for all $i = 1, \dots, n$ follows immediately from the condition of coherency [Coh2]. To prove that every subset with cardinality $2 \leq k \leq n-2$ has zero mass we use strong induction on k . If $k = 2$, then

$$\begin{aligned} m_A^M(\{a_i, a_j\}) &= Bel_{m_A^M}(\{a_i, a_j\}) - m_A^M(\{a_i\}) - m_A^M(\{a_j\}) \\ &= \max\{\underline{p}_i^A + \underline{p}_j^A, 1 - \sum_{h \neq i, j} \bar{p}_h^A\} - \underline{p}_i^A - \underline{p}_j^A. \end{aligned}$$

Hence, it is enough to prove that $\underline{p}_i^A + \underline{p}_j^A \geq 1 - \sum_{h \neq i, j} \bar{p}_h^A$, which is the same as showing that

$$\begin{aligned} \bar{S}_A - 1 - (\bar{p}_i^A - \underline{p}_i^A) - (\bar{p}_j^A - \underline{p}_j^A) &\geq 0 \\ \Leftrightarrow (n-2)(1 - \underline{S}_A) - (\bar{p}_i^A - \underline{p}_i^A) - (\bar{p}_j^A - \underline{p}_j^A) &\geq 0 \end{aligned} \quad (21)$$

where the implication follows from the fact that the probability interval is good thus $\Delta_A \geq 0$. To prove (21) we just note that $n \geq 4$ and

$$\bar{p}_h^A - \underline{p}_h^A \leq 1 - \underline{S}_A \quad \text{for all } h = 1, \dots, n \quad (22)$$

which is just a rewriting of the condition of coherency [Coh3]. Therefore, due to the arbitrariness of i and j , every set of cardinality 2 has zero mass.

Consider $V \in 2^{E_A}$ such that $2 < |V| \leq n-2$. From the definition of belief we have that

$$\begin{aligned} m_A^M(V) &= Bel_{m_A^M}(V) - \sum_{W \subset V} m_A^M(W) \\ &= Bel_{m_A^M}(V) - \sum_{i: a_i \in V} m_A^M(\{a_i\}) \\ &= \max\left\{ \sum_{i: a_i \in V} \underline{p}_i^A, 1 - \sum_{i: a_i \notin V} \bar{p}_i^A \right\} - \sum_{i: a_i \in V} \underline{p}_i^A \end{aligned} \quad (23)$$

where (23) follows from the induction hypothesis. Therefore, as in the previous case, it suffices to prove that

$$\sum_{i: a_i \in V} \underline{p}_i^A \geq 1 - \sum_{i: a_i \notin V} \bar{p}_i^A.$$

This inequality is implied by

$$(n-2)(1 - \underline{S}_A) - \sum_{i: a_i \in V} (\bar{p}_i^A - \underline{p}_i^A) \geq 0$$

due to the fact that $(\bar{p}^A, \underline{p}^A)$ is a good probability interval. This last inequality follows from (22) and the fact that $|V| \leq n-2$. Therefore, due to the arbitrariness of V , every set of cardinality $2 < k \leq n-2$ has mass m_A^M equal to zero.

To complete the proof, it remains to show that the mass of sets with cardinality $n-1$ is equivalent to the standard good mass of these same sets. Using the definition of belief, we have

$$\begin{aligned} m_A^M(E_A \setminus \{a_i\}) &= Bel_{m_A^M}(E_A \setminus \{a_i\}) - \sum_{W \subset E_A \setminus \{a_i\}} m_A^M(W) \\ &= \max\left\{ \sum_{h \neq i} \underline{p}_h^A, 1 - \bar{p}_i^A \right\} - \sum_{h \neq i} m_A^M(\{a_h\}) \end{aligned} \quad (24)$$

$$= 1 - \bar{p}_i^A - \sum_{h \neq i} \underline{p}_h^A \quad (25)$$

where (24) follows from the fact that every other subset has zero mass, and (25) follows from the condition of coherency [Coh3].

The fact that $m_A^M(E_A) = \Delta_A$ follows from the fact that $\sum_{V \in 2^{E_A}} m_A^M(V) = 1$.

$$\begin{aligned} m_A^M(E_A) &= 1 - \sum_{i=1}^n m_A^M(\{a_i\}) - \sum_{i=1}^n m_A^M(E_A \setminus \{a_i\}) \\ &= \bar{S}_A + (n-2)\underline{S}_A - (n-1). \end{aligned}$$

\square

ADDITIONAL AUTHOR INFORMATION

Acknowledgements. This work has been partly supported by the PersOn project (P21-03), which has received funding from Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).

REFERENCES

- [1] Alessandro Antonucci, Cassio Polpo de Campos, and Marco Zaffalon. “Probabilistic graphical models”. In: *Introduction to Imprecise Probabilities*. Ed. by Thomas Augustin, Frank P. A. Coolen, Gert De Cooman, and Matthias C. M. Troffaes. Wiley, 2014. Chap. 9. DOI: [10.1002/9781118763117](https://doi.org/10.1002/9781118763117).
- [2] Adnan Darwiche. *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, 2014. DOI: [10.1017/CBO9780511811357](https://doi.org/10.1017/CBO9780511811357).
- [3] Jasper De Bock. “Reintroducing Credal Networks under Epistemic Irrelevance”. In: *Proceedings of the Eighth International Conference on Probabilistic Graphical Models*. Ed. by Alessandro Antonucci, Giorgio Corani, and Cassio de Campos. Vol. 52. Proceedings of Machine Learning Research. Lugano, Switzerland: PMLR, 2016, pp. 123–135.
- [4] Cassio Polpo De Campos and Fabio Gagliardi Cozman. “The inferential complexity of Bayesian and credal networks”. In: *Proceedings of the 19th International Joint Conference on Artificial Intelligence*. IJCAI’05. Edinburgh, Scotland: Morgan Kaufmann Publishers Inc., 2005, pp. 1313–1318. URL: <http://www.ijcai.org/papers/0841.pdf>.
- [5] Arthur Pentland Dempster. “Upper and lower probabilities induced by a multivalued mapping”. In: *Annals of Mathematical Statistics* 38 (1967), pp. 325–339. DOI: [10.1214/aoms/1177698950](https://doi.org/10.1214/aoms/1177698950).
- [6] Arthur Pentland Dempster. “A Generalization of Bayesian Inference”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 30.2 (1968), pp. 205–247. DOI: [10.1111/j.2517-6161.1968.tb00722.x](https://doi.org/10.1111/j.2517-6161.1968.tb00722.x).
- [7] Sébastien Destercke and Didier Dubois. “Special Cases”. In: *Introduction to Imprecise Probabilities*. Ed. by Thomas Augustin, Frank P. A. Coolen, Gert De Cooman, and Matthias C. M. Troffaes. Wiley, 2014. Chap. 4. DOI: [10.1002/9781118763117](https://doi.org/10.1002/9781118763117).
- [8] Denis Deratani Maua, Cassio Polpo de Campos, Alessio Benavoli, and Alessandro Antonucci. “Probabilistic Inference in Credal Networks: New Complexity Results”. In: *Journal of Artificial Intelligence Research* 50 (2014), pp. 603–637. DOI: [10.1613/jair.4355](https://doi.org/10.1613/jair.4355).
- [9] Denis Deratani Mauá and Fabio Gagliardi Cozman. “Thirty years of credal networks: Specification, algorithms and complexity”. In: *International Journal of Approximate Reasoning* 126 (2020), pp. 133–157. DOI: [10.1016/j.ijar.2020.08.009](https://doi.org/10.1016/j.ijar.2020.08.009).
- [10] Enrique Miranda and Arthur Van Camp. “The law of iterated expectation and imprecise probabilities”. In: *International Fuzzy Systems Association* 504 (Mar. 2025), p. 109258.
- [11] Serafin Moral-García and Joaquín Abellán. “Credal sets representable by reachable probability intervals and belief functions”. In: *International Journal of Approximate Reasoning* 129 (2021), pp. 84–102. DOI: [10.1016/j.ijar.2020.11.007](https://doi.org/10.1016/j.ijar.2020.11.007).
- [12] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.
- [13] Prakash P. Shenoy. “Making Inferences in Incomplete Bayesian Networks: A Dempster-Shafer Belief Function Approach”. In: *International Journal of Approximate Reasoning* 160 (Sept. 2023), p. 108967. DOI: [10.1016/j.ijar.2023.108967](https://doi.org/10.1016/j.ijar.2023.108967).
- [14] Philippe Smets. *Un modèle mathématico-statistique simulant le processus du diagnostic médical*. Université libre de Bruxelles, 1978.
- [15] Philippe Smets. “Belief functions: The disjunctive rule of combination and the generalized Bayesian theorem”. In: *International Journal of Approximate Reasoning* 9.1 (Aug. 1993), pp. 1–35.
- [16] Robert L. Smith. “Efficient Monte Carlo Procedures for Generating Points Uniformly Distributed over Bounded Regions”. In: *Operations Research* (1984), pp. 1195–1384. DOI: [10.1287/opre.32.6.1296](https://doi.org/10.1287/opre.32.6.1296).