Journal of Artificial Intelligence Research 50 (2014) 603-637 Submitted 02/14; published 07/14

# Probabilistic Inference in Credal Networks: New Complexity Results

**Denis Deratani Mauá** DENIS.MAUA@USP.BR
*Escola Politécnica, Universidade de São Paulo*
*Av. Prof. Luciano Gualberto, travessa 3, 380*
*São Paulo, 05508-010 Brazil*

**Cassio Polpo de Campos** CASSIO@IDSIA.CH
**Alessio Benavoli** ALESSIO@IDSIA.CH
**Alessandro Antonucci** ALESSANDRO@IDSIA.CH
*Istituto Dalle Molle di Studi sull'Intelligenza Artificiale*
*Galleria 2*
*Manno, 6928 Switzerland*

## Abstract

Credal networks are graph-based statistical models whose parameters take values in a set, instead of being sharply specified as in traditional statistical models (e.g., Bayesian networks). The computational complexity of inferences on such models depends on the irrelevance/independence concept adopted. In this paper, we study inferential complexity under the concepts of epistemic irrelevance and strong independence. We show that inferences under strong independence are NP-hard even in trees with binary variables except for a single ternary one. We prove that under epistemic irrelevance the polynomial-time complexity of inferences in credal trees is not likely to extend to more general models (e.g., singly connected topologies). These results clearly distinguish networks that admit efficient inferences and those where inferences are most likely hard, and settle several open questions regarding their computational complexity. We show that these results remain valid even if we disallow the use of zero probabilities. We also show that the computation of bounds on the probability of the future state in a hidden Markov model is the same whether we assume epistemic irrelevance or strong independence, and we prove a similar result for inference in naive Bayes structures. These inferential equivalences are important for practitioners, as hidden Markov models and naive Bayes structures are used in real applications of imprecise probability.

## 1. Introduction

Bayesian networks are multivariate probabilistic models where stochastic independence assessments are compactly represented by an acyclic directed graph whose nodes are identified with variables (Pearl, 1988). In addition to its acyclic directed graph, the specification of a Bayesian network requires the specification of a conditional probability distribution for every variable and every assignment of its parents. When information is costly to acquire, specifying these conditional probabilities can be a daunting task, whether they are estimated from data or elicited from experts. This causes the inferences drawn with the model to contain imprecisions and arbitrarinesses (Kwisthout & van der Gaag, 2008).

An arguably more principled approach to coping with the imprecision in the numerical parameters is by incorporating it into the formalism. One way of doing so is by means of closed and convex sets of probability distributions, which are called *credal sets* (Levi, 1980).[1] Bayesian networks whose numerical parameters are specified by conditional credal sets are known as *credal networks* (Cano, Cano, & Moral, 1994; Cozman, 2000, 2005). Credal networks have been successfully applied to robust pattern recognition,[2] and to knowledge-based systems, where it has been argued that allowing parameters to be imprecisely specified facilitates elicitation from experts.[3]

A Bayesian network provides a concise representation of the (single) joint probability distribution that is consistent with the network parameters and satisfies (at least) the set of stochastic independences encoded in its underlying graph. Analogously, a credal network provides a concise representation of the credal set of joint distributions that are consistent with the local credal sets and satisfy (at least) the irrelevances encoded in its underlying graph. The precise characterization of that joint credal set depends however on the concept of irrelevance adopted.

The two most commonly used irrelevance concepts in the literature are *strong independence* and *epistemic irrelevance*. Two variables $X$ and $Y$ are *strongly independent* if their joint credal set can be regarded as originating from a number of precise probability distributions under each of which the two variables are stochastically independent. Strong independence follows a sensitivity analysis interpretation, which regards imprecision in the model as arising out of partial ignorance about an ideal precise model. Epistemic irrelevance, on the other hand, is defined irrespective of the existence of any ideal precise model. A variable $X$ is *epistemically irrelevant* to a variable $Y$ if the marginal credal set of $Y$ coincides with the conditional credal set of $Y$ given $X$. Unlike strong independence, epistemic irrelevance is an asymmetric concept and cannot in general be characterized by the properties of the elements of the credal set alone (de Cooman et al., 2010).

If on the one hand the flexibility provided by credal sets arguably facilitates model building, on the other, it imposes a great burden on the computation of inferences. For example, whereas the posterior probability of a variable is polynomial-time computable in polytree-shaped Bayesian networks, the analogous task of computing upper and lower bounds on the posterior probability of a given variable in a polytree-shaped credal networks is an NP-hard task (de Campos & Cozman, 2005). There are however exceptional cases, such as the case of inference in polytree-shaped credal networks with binary variables, which can be solved in polynomial time under strong independence (Fagiuoli & Zaffalon, 1998). Like in Bayesian networks, the theoretical and practical tractability of inferences in credal networks depends strongly on the network topology and the cardinality of the variable domains. Credal networks however include another dimension in the parametrized complexity of inference, given by the type of irrelevance concept adopted, which in the

1. Other approaches include random sets (Kendall, 1974), evidence theory (Shafer, 1976; Shenoy & Shafer, 1988), possibility theory (Zadeh, 1978), conditional plausibility measures (Halpern, 2001), and coherent lower previsions (Walley, 1991; de Cooman & Miranda, 2012), the last one being largely equivalent to credal sets (there is an one-to-one correspondence between credal sets and coherent lower previsions).
2. For example, see the works of Zaffalon, Wesnes, and Petrini (2003), Zaffalon (2005), de Campos, Zhang, Tong, and Ji (2009), Antonucci, Brühlmann, Piatti, and Zaffalon (2009), Corani, Giusti, Migliore, and Schmidhuber (2010), Antonucci, de Rosa, and Giusti (2011), de Campos and Ji (2011).
3. For example, see the works of Walley (2000), Antonucci, Piatti, and Zaffalon (2007), Salvetti, Antonucci, and Zaffalon (2008), de Campos and Ji (2008), Antonucci et al. (2009), Piatti, Antonucci, and Zaffalon (2010), Antonucci, Huber, Zaffalon, Luginbühl, Chapman, and Ladouceur (2013).

Bayesian case is usually fixed. For instance, computing probability bounds in tree-shaped credal networks under the concept of epistemic irrelevance can be performed in polynomial time (de Cooman et al., 2010), whereas we show here that the same task is NP-hard under strong independence.

In the rest of this paper, we properly define credal networks and the inference problem we address (sect. 2), and investigate the parametrized theoretical computational complexity of inferences in credal networks (sect. 3), both under strong independence and epistemic irrelevance. We show that a particular type of inference in imprecise hidden Markov models (i.e., hidden Markov models with uncertainty quantified by local credal sets) is invariant to the choice of either irrelevance concept, being thus polynomial-time computable (as this is known to be the case under epistemic irrelevance). We obtain as corollaries of that result that inferences under strong independence and epistemic irrelevance coincide also in tree-shaped networks if no evidence is given, and in naive Bayes structures. We also show that even in tree-shaped credal networks inferences are NP-hard if we assume strong independence, and that this is the same complexity of inference in polytree-shaped credal networks for both irrelevance concepts, even if we assume that all variables are (at most) ternary. We prove that the so-called precise-vacuous models, that is, credal networks that have vacuous root nodes and precise non-root nodes, lead to the same inferences whether we assume epistemic irrelevance or strong independence, and that the same is true (apart from an arbitrarily small error) when vacuous nodes are replaced by near-vacuous ones, avoiding the problematic case of zero probabilities. This last fact proves that our hardness results hold true even in cases where the lower probability of any possible event is strictly positive.

## 2. Updating Credal Networks

In this section, we review the necessary concepts and definitions, and formalize the problem of inference in credal networks.

### 2.1 Bayesian Networks

Consider a finite set $\mathbf{X} = \{X_1, \dots, X_n\}$ of categorical variables, and let $\mathbf{Z} \subseteq \mathbf{X}$ be some set of variables. A *probability distribution* $p$ of $\mathbf{Z}$ is a non-negative real-valued function on the space of assignments of $\mathbf{Z}$ such that $\sum_{\mathbf{z}\sim\mathbf{Z}} p(\mathbf{z}) = 1$, where the notation $\mathbf{z} \sim \mathbf{Z}$ entails that $\mathbf{z}$ is an arbitrary (joint) assignment or configuration of the variables in $\mathbf{Z}$. Any joint probability distribution $p$ induces a probability measure $\mathrm{P}_p$ on the sigma-field of all subsets of assignments of $\mathbf{Z}$.

Let $G$ be an acyclic directed graph (DAG) with nodes $N = \{1, \dots, n\}$. We denote the *parents* of a node $i$ in $G$ by $\mathrm{Pa}(i)$. The set of *non-descendants* of $i$, written $\mathrm{Nd}(i)$, contains the nodes not reachable from $i$ by a directed path. Note that $\mathrm{Pa}(i) \subseteq \mathrm{Nd}(i)$. Fix a probability measure P on the sigma-field of subsets of $\mathbf{X}$ and associate every node $i$ with a variable $X_i$. The DAG $G$ represents the following set of stochastic independence assessments known as *local Markov conditions*:

$$\mathrm{P}(X_i = x_i | \mathbf{X}_{\mathrm{Nd}(i)} = \mathbf{x}_{\mathrm{Nd}(i)}) = \mathrm{P}(X_i = x_i | \mathbf{X}_{\mathrm{Pa}(i)} = \mathbf{x}_{\mathrm{Pa}(i)}) \,, \tag{1}$$

for all $i \in N$, and $\mathbf{x} \sim \mathbf{X}$. In words, every variable is stochastically independent from its non-descendant non-parents given its parents under a suitable measure P.

A Bayesian network is a triple $(\mathbf{X}, G, Q)$, where $Q$ is the set of conditional probability assessments

$$\mathrm{P}(X_i = x_i | \mathbf{X}_{\mathrm{Pa}(i)} = \mathbf{x}_{\mathrm{Pa}(i)}) = q(x_i | \mathbf{x}_{\mathrm{Pa}(i)}) \,, \tag{2}$$

for all $i \in N$, $x_i \sim X_i$ and $\mathbf{x}_{\mathrm{Pa}(i)} \sim \mathbf{X}_{\mathrm{Pa}(i)}$, where $q(X_i|\mathbf{x}_{\mathrm{Pa}(i)})$ is a probability distribution of $X_i$. By assumption, a Bayesian network defines a joint probability distribution $p$ of $\mathbf{X}$ by

$$p(\mathbf{x}) = \prod_{i \in N} q(x_i | \mathbf{x}_{\mathrm{Pa}(i)}) \,, \tag{3}$$

for all $\mathbf{x} \sim \mathbf{X}$. It is not difficult to see that (1) and (2) imply (3) by the Chain Rule using a topological ordering of variables. That (3) and (2) imply (1) is a bit more intricate to see but also true (Cowell, Dawid, Lauritzen, & Spiegelhalter, 2007). Thus, these two seemingly different approaches to specifying a probability measure are virtually equivalent. To be more explicit: given a Bayesian network, the probability measure that satisfies (1) and (2) is the same probability measure that satisfies (3) and (2), and we can choose any pair of assumptions to define a (single) measure for a network. As we shall see, an analogous equivalence is not observed when probabilities are imprecisely specified, which leads to different definitions of credal networks with different computational complexity.

#### 2.1.1 Probabilistic Inference In Bayesian Networks

An essential task in many applications of probabilistic modeling is to compute a certain probability value implied by a Bayesian network. We call such a computational task the BN-INF problem, and define it as follows.

> BN-INF
> *Input*: A Bayesian network $(\mathbf{X}, G, Q)$, a *target* node $t$, a target value $x_t$ of $X_t$, a (possibly empty) set of evidence nodes $O$, and an assignment $\mathbf{x}_O$ to $\mathbf{X}_O$.
> *Output*: The conditional probability $\mathrm{P}(X_t = x_t | \mathbf{X}_O = \mathbf{x}_O)$, where P is the probability measure specified by the network.

In the problem above we assume that when $\mathrm{P}(\mathbf{X}_O = \mathbf{x}_O) = 0$ the output is a special symbol (e.g. $\perp$) indicating the problem solution is undefined.

Roth (1996) showed that BN-INF is #P-hard, defining a lower bound to the complexity of the problem. All known (exact) algorithms take time at least exponential in the treewidth of the network in the worst case. The treewidth is a measure of the resemblance of a network to a tree; small treewidth suggests a tree-like structure and the treewidth of a tree is one and is minimal (Koller & Friedman, 2009). Recently, Kwisthout, Bodlaender, and van der Gaag (2010) proved that contingent on the hypothesis that satisfiability of Boolean formulas takes exponential time in the worst-case (known as ETH) this is the best performance an algorithm for BN-INF can achieve. As we shall see in the next section, Bayesian networks are particular instances of credal networks. As such, these complexity results set lower bounds on the complexity of inference in credal networks.

A DAG is said to be *singly connected* if there is at most one *undirected* path connecting any two nodes in the graph; it is a *tree* if additionally each node has at most one parent. If a graph is not singly connected, we say it is *multiply connected*. Singly connected directed graphs are also called *polytrees*. Pearl's belief propagation algorithm (Pearl, 1988) computes BN-INF in polynomial time in singly connected Bayesian networks. More generally, the junction tree propagation algorithm (Cowell et al., 2007) solves BN-INF in polynomial time in any network of bounded treewidth, which includes singly connected networks of bounded in-degree (i.e., maximum number of parents).

## 2.2 Credal Networks

In this section we describe credal sets, irrelevance concepts, credal networks and probabilistic inference in credal networks.

### 2.2.1 Credal Sets

Let $\mathbf{Z} \subseteq \mathbf{X}$. A *credal set* $M$ is a closed and convex set of joint probability distributions on the same domain, say $\mathbf{z} \sim \mathbf{Z}$ (Levi, 1980). The *vacuous credal set* of $\mathbf{Z}$ is the largest credal set on that domain, and is denoted by $V(\mathbf{Z})$. An *extreme distribution* of a credal set is an element of the set that cannot be written as a convex combination of other elements in the same set. We denote the set of extreme distributions of a credal set $M$ by $\operatorname{ext} M$. A credal set is *finitely generated* if it contains a finite number of extreme distributions. A finite representation of a finitely generated credal set by means of its extreme distributions is called *vertex-based*. Any finitely generated credal set of $\mathbf{Z}$ defines a (bounded) polytope in the probability simplex of distributions of $\mathbf{Z}$, and can be specified through a finite set of linear inequalities of the form

$$\mathrm{E}_p(f_l) \stackrel{\text{def}}{=} \sum_{\mathbf{z}\sim\mathbf{Z}} f_l(\mathbf{z})p(\mathbf{z}) \leq 0\,, \tag{4}$$

where $\{f_l\}$ is a finite collection of real-valued functions of $\mathbf{Z}$ (Cozman, 2000). The converse is also true: any finite set of linear inequalities of the form above determines a (bounded) polytope in the probability simplex (Boyd & Vandenberghe, 2004, ch. 2), and hence a finitely generated credal set. Thus, an alternative finite representation of a credal set is by means of a finite set of functions defining linear inequalities of the type above. Such a representation is called *constraint-based*.

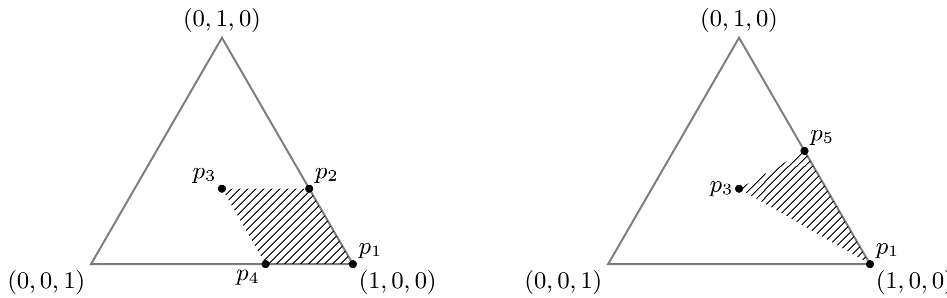
**Example 1.** *Consider $\mathbf{X} = \{X_1, X_2\}$, where $X_1$ takes values in $\{0, 1, 2\}$ and $X_2$ takes values in $\{0, 1\}$. The vacuous set of $X_1$ is the probability simplex on the plane, drawn as a triangle with vertices $(p(0), p(1), p(2)) = (0, 0, 1)$, $(0, 1, 0)$ and $(1, 0, 0)$ in Figure 1. Let*

$$M(X_1|X_2\!=\!0) = \{p \in V(X_1) : p(k) \leq 1/3, k = 1, 2\}$$

*and*

$$M(X_1|X_2\!=\!1) = \{p \in V(X_1) : p(0) \geq p(1) \geq p(2)\}$$

*be conditional credal sets for $X_1$ given $X_2$, and $M(X_2)$ be the singleton containing the distribution $p$ of $X_2$ such that $p(0) = p(1) = 1/2$. The first two sets are depicted in Figure 1.*




(a) $M(X_1|X_2=0)$ (b) $M(X_1|X_2=1)$

Figure 1: Barycentric coordinate-system visualization of the conditional credal sets in Example 1 (hatched regions) and their corresponding extreme distributions (black circles).

*Let us represent a generic function $f$ on $\{0,\ldots,m\}$ by the $m$-tuple $(f(0),\ldots,f(m))$, and define*

$$
\begin{aligned}
p_1 &= (1,0,0)\,, & p_2 &= (2/3,1/3,0)\,,\\
p_3 &= (1/3,1/3,1/3)\,, & p_4 &= (2/3,0,1/3)\,,\\
p_5 &= (1/2,1/2,0)\,, & f_1 &= (-1,2,-1)\,,\\
f_2 &= (-1,-1,2)\,, & f_3 &= (1,-1,0)\,,\\
f_4 &= (0,1,-1)\,, & f_5 &= (1,-1)\,.
\end{aligned}
$$

*Then the set $M(X_1|X_2=0)$ can be represented in vertex- and constraint-based form, respectively, as $M(X_1|X_2=0) = \mathrm{co}\{p_1,p_2,p_3,p_4\}$ (co denotes the convex hull operator) and $M(X_1|X_2=0) = \{p \in V(X_1) : \mathrm{E}_p(f_1) \le 0, \mathrm{E}_p(f_2) \le 0\}$, while the set $M(X_1|X_2=1)$ is represented in vertex- and constraint-based forms as $M(X_1|X_2=1) = \mathrm{co}\{p_1,p_3,p_5\}$ and $M(X_1|X_2=1) = \{p \in V(X_1) : \mathrm{E}_p(f_3) \le 0, \mathrm{E}_p(f_4) \le 0\}$, respectively. Similarly, $M(X_2)$ can be represented as $M(X_2) = \{(1/2,1/2)\}$ in vertex-based form, and as $M(X_2) = \{p \in V(X_2) : \mathrm{E}_p(f_5) \le 0, \mathrm{E}_p(-f_5) \le 0\}$ in constraint-based form.* ☐

Vertex- and constraint-based representations of the same credal set can have very different sizes. To see this, consider a single variable $X$ taking values in $\{0,\ldots,m\}$, and let $M = \{p \in V(X) : p(k) \le 1/(m+1), k = 1,\ldots,m\}$. The set $M$ is isomorphic to an $m$-dimensional hypercube, and therefore has $2^m$ extreme distributions,[4] whereas the same set can be represented in constraint-based form by $m$ degenerate functions of $X$ translated by $1/(m+1)$. Moving from a vertex-based to a constraint-based representation can also result in an exponential increase in the size of the input. Consider a variable $X$ taking values in $\{0,\ldots,m\}$ and let $M = \{f(X) \ge 0 : \sum_{k=1}^{m} |f(k) - 1/(2m)| \le 1/(2m), \sum_{k=0}^{m} f(k) = 1\}$. The set $M$ is affinely equivalent to the $m$-dimensional cross-polytope, and hence requires $2^m$ inequalities to be described in constraint-based form, while it can be represented in

4. For any non-negative integer $k$ not greater than $m$ and any (potentially empty) subset $S$ of $\{1,\ldots,m\}$ of cardinality $k$, any distribution that assigns mass $(m+1-k)/(m+1)$ to $p(0)$, mass $1/(m+1)$ to $p(j)$ such that $j$ is in $S$, and zero mass elsewhere, is in $M$, since it satisfies all the constraints in $M$ and is a valid distribution. There are $2^m$ such distributions, and each one cannot be written as a convex combination of any other distribution in the set.

vertex-based form by its $2m$ extreme distribution (Kalai & Ziegler, 2000, p. 11).[5] Tessem (1992) and de Campos, Huete, and Moral (1994) studied the representation of credal sets defined by linear constraints of the form $l_x \leq p(x) \leq u_x$, where $l_x$ and $u_x$ are real numbers, and showed that these credal sets can have exponentially many extreme distributions in the number of constraints. Wallner (2007) proved an attainable upper bound of $m!$ extreme distributions on credal sets more generally defined by a coherent lower probability function of an $m$-ary variable. More recently, Miranda and Destercke (2013) investigated the number of extreme distributions in credal sets defined by linear constraints of the form $p(x) \leq p(x')$ for $x \neq x'$, and proved an attainable upper bound of $2^{m-1}$ extreme distributions for the case of an $m$-ary variable. Importantly, both vacuous credal sets (of variables of any cardinality) and credal sets of binary variables can be succinctly represented in either vertex- or constraint-based form. Some of the complexity results we obtain later on use vacuous credal sets and/or binary variables and are thus representation independent.

#### 2.2.2 Graph-Based Representation

So far we have only considered the explicit representation of a finitely generated credal set by a finite number of functions representing either the vertices of the set or a set of linear inequalities. Our final goal is however to be able to specify credal sets on large domains $\mathbf{x} \sim \mathbf{X}$. For a large set $\mathbf{X}$, such an explicit representation is both too difficult to obtain and too large to manipulate in a computer. Thus, analogously to the more efficient graph-based representation of a large probability distribution given by a Bayesian network, a large joint credal set is usually more efficiently represented implicitly as the credal set that satisfies all irrelevances encoded in a given graph while agreeing on its projection with all local credal sets, where the latter are credal sets that can be efficiently represented (either in vertex- or constraint-based form) explicitly by functions of only small subsets of $\mathbf{X}$.

A (separately specified) *credal network* $\mathcal{N}$ is a triple $(\mathbf{X}, G, Q)$, where $G$ is a DAG with nodes $N$, and $Q$ is a set of imprecise probability assessments

$$\forall f : \sum_{x_i \sim X_i} f(x_i) \mathrm{P}(X_i = x_i | \mathbf{X}_{\mathrm{Pa}(i)} = \mathbf{x}_{\mathrm{Pa}(i)}) \geq \min_{q \in Q(X_i | \mathbf{x}_{\mathrm{Pa}(i)})} \sum_{x_i \sim X_i} f(x_i) q(x_i) \,, \tag{5}$$

one for every $i \in N$ and $\mathbf{x}_{\mathrm{Pa}(i)} \sim \mathbf{X}_{\mathrm{Pa}(i)}$, where each $Q(X_i | \mathbf{x}_{\mathrm{Pa}(i)})$ is a credal set of $X_i$ and $f$ is an arbitrary real-valued function of $X_i$. Note that we left unspecified how these credal sets are represented.

**Example 2.** *Consider the credal network $\mathcal{N}$ over variables $X_1, X_2$ and $X_3$ that take values in $\{0,1\}$, and with graph structure as shown in Figure 2. The local credal sets are*

$$Q(X_1) = \{p \in V(X_1) : 0.5 \leq p(1) \leq 0.6\} = \mathrm{co}\{(0.4, 0.6), (0.5, 0.5)\} \,,$$
$$Q(X_2) = \{p \in V(X_2) : 0.5 \leq p(1) \leq 0.6\} = \mathrm{co}\{(0.4, 0.6), (0.5, 0.5)\} \,,$$

*and $Q(X_3 | X_1 = i, X_2 = j) = \{p^{ij}\}$ for any $i$ and $j$, where $p^{ij}$ is the probability distribution on $\{0,1\}$ such that $p^{ij}(1) = 0$ if $i = j$ and $p^{ij}(1) = 1$ otherwise.* □

5. The $m$-dimensional cross-polytope is the set $\{f(X) : \sum_x |f(x)| \leq 1\}$, whose extreme distributions are $\{\pm e_1, \ldots, \pm e_m\}$, where $e_k$ $(k = 1, \ldots, m)$ is the degenerate distribution placing all mass at $X = k$. The fact that that set cannot be written with less than $2^m$ inequalities follows from it being the dual of the $m$-dimensional hypercube.

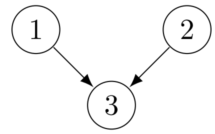



Figure 2: DAG of the credal network in Example 2.

A node $i$ in the network and its associated variable $X_i$ are said to be *precise* (i.e., precisely specified) if the corresponding conditional credal sets $Q(X_i|\mathbf{x}_{\mathrm{Pa}(i)})$ are all singletons, otherwise they are said to be *imprecise* (i.e., imprecisely specified). In Example 2 variable $X_3$ and its corresponding associated node are precise, while $X_1$ and $X_2$ are imprecise. If all of its local credal sets are vacuous, the node and its corresponding variable are said to be vacuous. A Bayesian network is simply a credal network with all nodes precise.

The DAG $G$ of a credal network specifies a set of conditional irrelevances between sets of variables which generalize the Markov condition in Bayesian networks. More specifically, for any node $i$ in $G$, the set $\mathbf{X}_{\mathrm{Nd}(i)\setminus\mathrm{Pa}(i)}$ of non-descendant non-parent variables of $X_i$ is assumed irrelevant to $X_i$ conditional on its parent variables $\mathbf{X}_{\mathrm{Pa}(i)}$. The precise definition of this statement requires the definition of an irrelevance concept. For instance, if stochastic independence is adopted as irrelevance concept, then the DAG $G$ describes a set of Markov conditions as a Bayesian network (stochastic irrelevance implies stochastic independence). In the credal network formalism, the two most common irrelevance concepts used are *strong independence* and *epistemic irrelevance*.

Fix a joint credal set $M$ of probability distributions of $\mathbf{X}$, and consider subsets $\mathbf{Y}$, $\mathbf{Z}$ and $\mathbf{W}$ of $\mathbf{X}$. We say that a set of variables $\mathbf{Y}$ is *strongly independent* of a set of variables $\mathbf{Z}$ given variables $\mathbf{W}$ if $\mathbf{Y}$ and $\mathbf{Z}$ are stochastically independent conditional on $\mathbf{W}$ under every extreme distribution $p \in \operatorname{ext} M(\mathbf{X})$, which implies for every $\mathbf{y}$, $\mathbf{z}$ and $\mathbf{w}$ that $\mathrm{P}_p(\mathbf{Y} = \mathbf{y}|\mathbf{Z} = \mathbf{z}, \mathbf{W} = \mathbf{w}) = \mathrm{P}_p(\mathbf{Y} = \mathbf{y}|\mathbf{W} = \mathbf{w})$. We say that a set of variables $\mathbf{Z}$ is *epistemically irrelevant* to a set of variables $\mathbf{Y}$ conditional on variables $\mathbf{W}$ if for every function $f$ on $\mathbf{y} \sim \mathbf{Y}$ and assignments $\mathbf{z}$ and $\mathbf{w}$ it follows that

$$\min_{p\in M} \sum_{\mathbf{y}\sim\mathbf{Y}} f(\mathbf{y})\mathrm{P}_p(\mathbf{Y}=\mathbf{y}|\mathbf{Z}=\mathbf{z}, \mathbf{W}=\mathbf{w}) = \min_{p\in M} \sum_{\mathbf{y}\sim\mathbf{Y}} f(\mathbf{y})\mathrm{P}_p(\mathbf{Y}=\mathbf{y}|\mathbf{W}=\mathbf{w})\,, \tag{6}$$

which is equivalent to say that the projection of $M$ on $\mathbf{Y}$ conditioned on $\mathbf{W}=\mathbf{w}$ and $\mathbf{Z}=\mathbf{z}$ equals the projection of $M$ on $\mathbf{Y}$ conditioned only on $\mathbf{W}=\mathbf{w}$. It is an immediate conclusion that strong independence implies epistemic irrelevance (and the converse is not necessarily true) (Cozman, 2000; de Cooman & Troffaes, 2004). Variables $\mathbf{Y}$ and $\mathbf{Z}$ are *epistemically independent* conditional on $\mathbf{W}$ if, given any assignment $\mathbf{w}$, $\mathbf{Y}$ and $\mathbf{Z}$ are epistemically irrelevant to each other (Walley, 1991, ch. 9).

The *strong extension* of a credal network $\mathcal{N} = (\mathbf{X}, G, Q)$ is the largest credal set $K_S$ of distributions of $\mathbf{X}$ whose extreme distributions satisfy the strong independence assessments in $G$ (viz. that every variable is strongly independent of its non-descendant non-parents given its parents), and whose projections on local domains lie inside the local credal sets specified in $Q$, that is, $K_S$ is the convex hull of the set of distributions of $\mathbf{X}$ whose induced measure satisfies (1) and (5). One can show that the strong extension can be equivalently defined as (Antonucci & Zaffalon, 2008; Antonucci, de Campos, & Zaffalon, 2014)

$$K_S \stackrel{\text{def}}{=} \text{co}\left\{ p \in V(\mathbf{X}) : p(\mathbf{x}) = \prod_{i \in N} q_i^{\mathbf{x}_{\text{Pa}(i)}}(x_i),\ q_i^{\mathbf{x}_{\text{Pa}(i)}} \in \text{ext}\, Q(X_i|\mathbf{x}_{\text{Pa}(i)}) \right\}. \tag{7}$$

The *epistemic extension* of a credal network is the largest joint credal set $K_E$ of $\mathbf{X}$ that satisfies the epistemic irrelevance assessments in $G$ (viz. the non-descendant non-parents are irrelevant to a variable given its parents), and the assessments in $Q$. Equivalently, the epistemic extension is the credal set $K_E$ defined by the set of probability distributions $p$ of $\mathbf{X}$ such that

$$\sum_{x_i} f(x_i) \mathrm{P}_p(x_i|\mathbf{x}_{\text{Nd}(i)}) \geq \min_{q \in Q(X_i|\mathbf{x}_{\text{Pa}(i)})} \sum_{x_i} f(x_i) q(x_i), \tag{8}$$

for all functions $f$ of $X_i$, and assignment $\mathbf{x}_{\text{Nd}(i)}$. Note that these inequalities can be turned into linear inequalities of the form (4) by multiplying both sides by $\mathrm{P}_p(\mathbf{x}_{\text{Nd}(i)})$ and rearranging terms.

The following example is largely based on Example 9.3.4 of the work of Walley (1991).

**Example 3.** *Consider the network in Example 2, and represent a function $f$ of a binary variable as the pair $(f(0), f(1))$. The strong extension $K_S$ is the credal set whose extreme distributions are the four joint probability distributions $p \in V(X_1, X_2, X_3)$ such that*

$$p(x_1, x_2, x_3) = p_1(x_1) p_2(x_2) p_3^{x_1 x_2}(x_3) \quad \textit{for } x_1, x_2, x_3 \in \{0, 1\},$$

*where*

$$p_1 \in \{(0.4, 0.6), (0.5, 0.5)\}, \qquad p_2 \in \{(0.4, 0.6), (0.5, 0.5)\},$$
$$p_3^{00} = p_3^{11} = (1, 0), \qquad p_3^{01} = p_3^{10} = (0, 1).$$

*Note that the strong extension contains four extreme distributions. The epistemic extension $K_E$ is the set of joint probability distributions $p \in V(X_1, X_2, X_3)$ that satisfies the system of linear inequalities*

$$\begin{aligned}
&0.5 = \min_{q \in Q(X_1)} q(1) \leq \mathrm{P}_p(X_1 = 1|x_2) \leq \max_{q \in Q(X_1)} q(1) = 0.6 && [x_2 = 0, 1], \\
&0.5 = \min_{q \in Q(X_2)} q(1) \leq \mathrm{P}_p(X_2 = 1|x_1) \leq \max_{q \in Q(X_2)} q(1) = 0.6 && [x_1 = 0, 1], \\
&\mathrm{P}_p(X_3 = 1|X_1 = x, X_2 = x) = 0 && [x = 0, 1], \\
&\mathrm{P}_p(X_3 = 1|X_1 = 0, X_2 = 1) = \mathrm{P}_p(X_3 = 1|X_1 = 1, X_2 = 0) = 1.
\end{aligned}$$

*One can verify that the set $K_E$ has the following six extreme distributions:*

$$\begin{aligned}
p1 &= (0.25, 0, 0, 0.25, 0, 0.25, 0.25, 0), & p2 &= (0.16, 0, 0, 0.36, 0, 0.24, 0.24, 0), \\
p3 &= (0.2, 0, 0, 0.3, 0, 0.2, 0.3, 0), & p4 &= (0.2, 0, 0, 0.3, 0, 0.3, 0.2, 0), \\
p5 &= (2/9, 0, 0, 3/9, 0, 2/9, 2/9, 0), & p6 &= (2/11, 0, 0, 3/11, 0, 3/11, 3/11, 0),
\end{aligned}$$

*where the tuples on the right-hand side represent distributions $p(x_1, x_2, x_3)$ by*

$$(p(0,0,0), p(1,0,0), p(0,1,0), p(1,1,0), p(0,0,1), p(1,0,1), p(0,1,1), p(1,1,1)).$$

*We observe that distributions $p_1$ to $p_4$ are extreme distributions of the strong extension, whereas $p_5$ and $p_6$ are not in the strong extension.* □

The example above shows an interesting and well-known relation between epistemic and strong extensions, namely, that the latter is always contained in the former, and thus produces more precise results (Walley, 1991, ch. 9.2).

We have already discussed how the choice of a representation of credal sets can affect the complexity. The following result connects vertex- and constraint-based credal networks under strong independence.

**Proposition 1.** *Any vertex-based (separately specified) credal network can be efficiently reduced to a constraint-based credal network over a larger set of variables that induces the same strong extension when projected on the original set of variables.*

*Proof.* Let $X_i$ be a variable whose local credal set $Q(X_i|\mathbf{x}_{\mathrm{Pa}(i)})$ is specified by the extreme distributions $p_1, \ldots, p_m$, for a given assignment of the parents. Insert a new vacuous variable $X_\alpha$ taking values in $\{1, \ldots, m\}$, and with $X_i$ as its child and $\mathbf{X}_{\mathrm{Pa}(i)}$ as its parents, and redefine $Q(X_i|\mathbf{x}_{\mathrm{Pa}(i)})$ as the singleton that contains the conditional distribution $q(x_i|\mathbf{x}_{\mathrm{Pa}(i)}, x_\alpha = k) = p_k(x_i)$. One can verify that the strong extension of the new network after marginalizing $X_\alpha$ coincides with the original strong extension. □

The result above cannot be applied to derive the complexity of singly connected networks since the reduction used in the proof inserts (undirected) cycles in the network. Thus, it is not true that hardness results obtained for vertex-based singly connected networks immediately extend to constraint-based singly connected networks, even though this is always the case in the results we present here (for instance, we only use credal sets that are easily translated from one representation to the other in our hardness results). Conversely, tractability of constraint-based singly connected networks does not immediately extend to vertex-based singly connected networks. It is unclear whether constraint-based networks can be efficiently reduced to vertex-based form by inserting new variables, but we conjecture that this is true.

### 2.2.3 Probabilistic Inference

Similarly to Bayesian networks, a primary use of credal networks is in deriving bounds for probabilities implied by the model. The precise characterization depends on the choice of an irrelevance concept. We define the inference problem under strong independence as follows.

STRONG-INF
*Input:* A credal network $(\mathbf{X}, G, Q)$, a target node $t$, an assignment $x_t$ of $X_t$, a (possibly empty) set of evidence nodes $O$, and an assignment $\mathbf{x}_O$ of $\mathbf{X}_O$.
*Output*: The numbers

$$\min_{p \in K_S} \mathrm{P}_p(X_t = x_t | \mathbf{X}_O = \mathbf{x}_O) \text{ and } \max_{p \in K_S} \mathrm{P}_p(X_t = x_t | \mathbf{X}_O = \mathbf{x}_O)\,,$$

where $K_S$ is the strong extension of the network.

An analogous inference problem can be defined for epistemic irrelevance, simply by replacing the strong extension in the output of the problem above by the epistemic extension:

EPISTEMIC-INF
*Input:* A credal network $(\mathbf{X}, G, Q)$, a target node $t$, an assignment $x_t$ of $X_t$, a (possibly empty) set of evidence nodes $O$, and an assignment $\mathbf{x}_O$ of $\mathbf{X}_O$.
*Output*: The numbers

$$\min_{p \in K_E} \mathrm{P}_p(X_t = x_t | \mathbf{X}_O = \mathbf{x}_O) \text{ and } \max_{p \in K_E} \mathrm{P}_p(X_t = x_t | \mathbf{X}_O = \mathbf{x}_O)\,,$$

where $K_E$ is the epistemic extension of the network.

We assume in both problems that when the lower probability of the evidence is zero (i.e., $\min_p \mathrm{P}_p(\mathbf{X}_O = \mathbf{x}_O) = 0$), then any value is a solution (that is, the minimization may achieve zero and the maximization one).[6] For a recent treatment of the zero probability case, see the work of de Bock and de Cooman (2013). We emphasize that our complexity results hold true regardless of how zero probabilities are treated, because we take the appropriate care to avoid any conditioning event with zero probability in our reductions (as it will become clear later on).

**Example 4.** *Consider again the network in Example 2, and assume that the target node is $t = 3$, $x_t = 0$ and $\mathbf{X}_O$ is the empty set. The strong extension $K_S$ has been defined in Example 3. The outcome of STRONG-INF is*

$$\begin{aligned}\min_{p \in K_S} \mathrm{P}(X_3 = 0) &= \min \sum_{x_1, x_2} p_1(x_1) p_2(x_2) p_3^{x_1 x_2}(0) \\ &= 1 + \min\{2p_1(0)p_2(0) - p_1(0) - p_2(0)\} \\ &= 1 - (2 \cdot 1/2 \cdot 1/2 - 1/2 - 1/2) = 1/2\,,\end{aligned}$$

*where the minimizations on the right are performed over $p_1$ and $p_2$, and*

$$\begin{aligned}\max_{p \in K_S} \mathrm{P}(X_3 = 0) &= \max \sum_{x_1, x_2} p_1(x_1) p_2(x_2) p_3^{x_1 x_2}(0) \\ &= 1 + \max\{2p_1(0)p_2(0) - p_1(0) - p_2(0)\} \\ &= 1 - (2 \cdot 0.4 \cdot 0.4 - 0.4 - 0.4) = 0.52\,.\end{aligned}$$

*The outcome of EPISTEMIC-INF are the values of the solutions of the linear programs*

$$\min\{p(0,0,0) + p(1,1,0) : p \in \mathrm{K}_E\} = 5/11 < 1/2$$

*and*

$$\max\{p(0,0,0) + p(1,1,0) : p \in \mathrm{K}_E\} = 5/9 > 0.52\,,$$

*where $K_E$ is the epistemic extension defined in Example 3.* □

The fact that the lower bound (resp., upper bound) of EPISTEMIC-INF in the example above is smaller (resp., greater) than the lower bound (upper bound) of STRONG-INF is a direct consequence of the fact that the strong extension is contained in the epistemic extension.

6. If the upper probability of the evidence is positive, the regular extension can be used to compute non-vacuous inferences (see Walley, 1991, Appendix J).

Let also $\delta_{\mathbf{z}}$ denote the *Kronecker's delta function* at $\mathbf{z}$, which returns one at $\mathbf{z}$ and zero elsewhere. Since there is an one-to-one mapping between expectation and probability, we can state the inference problems in a slightly different but equivalent way.

STRONG-INF2
*Input:* A credal network $(\mathbf{X}, G, Q)$, a target node $t$, an assignment $x_t$ of $X_t$, a (possibly empty) set of evidence nodes $O$, and an assignment $\mathbf{x}_O$ of $\mathbf{X}_O$.
*Output*: The solution $\mu$ of $\text{opt}_{p \in K_S} \mathrm{E}_p([\mu - \delta_{x_t}]\delta_{\mathbf{x}_O}) = 0$, where $K_S$ is the strong extension of the network, with $\text{opt} \in \{\min, \max\}$.

EPISTEMIC-INF2
*Input:* A credal network $(\mathbf{X}, G, Q)$, a target node $t$, an assignment $x_t$ of $X_t$, a (possibly empty) set of evidence nodes $O$, and an assignment $\mathbf{x}_O$ of $\mathbf{X}_O$.
*Output*: The solution $\mu$ of $\text{opt}_{p \in K_E} \mathrm{E}_p([\mu - \delta_{x_t}]\delta_{\mathbf{x}_O}) = 0$, where $K_E$ is the epistemic extension of the network, and $\text{opt} \in \{\min, \max\}$.

The main advantage of these reformulations is that a linear-fractional programming problem is transformed into a linear programming problem, which facilitates obtaining some results. We will refer to both reformulations interchangeably, having in mind their equivalence.

## 3. Complexity Results

In this section we study the complexity of inference in credal networks with respect to the irrelevance concept adopted, the network topology and the variable domain cardinality.

### 3.1 Previously Known Results

Computing STRONG-INF is notoriously a hard task, whose complexity strongly depends on the topology of the DAG and the cardinality of the variable domains. Cozman et al. (2004) proved this problem to be $\text{NP}^{\text{PP}}$-hard. De Campos and Cozman (2005) studied the parametrized complexity and concluded that the problem is NP-hard even on singly connected networks of bounded treewidth. A long-known positive result is the 2U algorithm of Fagiuoli and Zaffalon (1998), which solves the problem in polynomial time if the underlying graph is a polytree and all variables are binary. For networks assuming a naive Bayes topology (i.e., containing a single root variable which is the single parent of all remaining variables), Zaffalon (2002) showed that STRONG-INF can be computed efficiently when the query is the root variable. Zaffalon and Fagiuoli (2003) showed that the problem is polynomial-time solvable in tree-shaped networks if there is no evidence. De Campos and Cozman (2005) showed that obtaining approximate solutions with a provably maximum error bound is impossible unless P equals NP, even in polytrees. On the other hand, Mauá, de Campos, and Zaffalon (2013) showed that when both variable cardinalities and treewidth are assumed bounded, there is a fully polynomial-time approximation scheme that finds solutions withing a given error $\epsilon$ in time polynomial in the input and in $1/\epsilon$. This is the only known positive result regarding the complexity of approximate inference with credal networks when the exact solution is NP-hard.

Much fewer is known about the complexity of EPISTEMIC-INF. A positive result was given by de Cooman et al. (2010), who developed a polynomial-time algorithm for computations in credal trees.

### 3.2 Outline of the Contributions

Our first contribution (sect. 3.3) is the development of credal networks with polynomial-time computable numbers. We show that such networks allow us to approximate arbitrarily well any network while still inducing positive lower probabilities on any event. This is useful to extend our subsequent complexity results (some of them involving events of lower probability zero) to the case where the lower probability of any event is strictly positive.

We then proceed to derive complexity results about STRONG-INF and EPISTEMIC-INF. The first new complexity result concerns precise-vacuous networks, which are credal networks comprised of vacuous root nodes and precise non-root nodes (sect. 3.4). Any credal network can be transformed into a precise-vacuous network on which STRONG-INF provides the same results as the original network; moreover, STRONG-INF is known to be $\mathrm{NP}^{\mathrm{PP}}$-hard in such models. We show that the solutions of STRONG-INF and EPISTEMIC-INF coincide in precise-vacuous networks, which implies the $\mathrm{NP}^{\mathrm{PP}}$-hardness of EPISTEMIC-INF. The hardness result holds even in the case of binary variables (and multiply connected networks).

We next show that both problems remain NP-hard in singly connected credal networks even if we constraint variables to take on at most three values and bound the treewidth on two (sect. 3.5). We show in the sequence that STRONG-INF is NP-hard already in credal trees (sect. 3.6); as discussed, EPISTEMIC-INF is polynomial-time computable in this case.

Imprecise hidden Markov models are tree-shaped credal networks that extend standard (precise) Hidden Markov Models (HMMs) to allow for imprecisely specified parameters in the form of credal sets. HMMs are commonly used to represent time-dependent process and have wide applicability. Since imprecise HMMs are particular instances of credal trees, EPISTEMIC-INF can be performed in polynomial time in such models. We show in Section 3.7 that when the target node is the last node in the longest directed path of the network inferences under strong independence and epistemic irrelevance coincide (in the context of time-series prediction, such an inference is known as filtering). As a consequence, STRONG-INF is also polynomial-time computable for such queries. This is despite the fact that the strong and the epistemic extensions might disagree on such models, as we show by a counter-example with a different type of query. We leave open the complexity of more general inferences in imprecise HMMs under strong independence.

As corollaries of the equivalence of a certain type of inference in HMMs, we obtain that STRONG-INF and EPISTEMIC-INF also coincide in marginal inference (i.e., with no evidence) in tree-shaped networks, in last-node inference in imprecise Markov chains (sect. 3.8) and in naive Bayes structures (sect. 3.9). These results have been previously obtained independently for each irrelevance concept and their tractability was thought to be coincidental.

We organize the presentation of the above mentioned results according to the complexity of the underlying DAG, listing results from the most complex structures to the simplest ones. The reason to proceed in this fashion is to allow obtaining some results for simpler models such as imprecise Markov chains and naive Bayes structures as corollaries of results for more complex models such as imprecise hidden Markov models.

### 3.3 Networks Specified with Computable Numbers

Before presenting the complexity results, we need to introduce the concept of *polynomial-time computable numbers*, and to discuss some properties of networks specified with such numbers. This will be used within later proofs in an essential way, including (but not only) showing that the hardness results hold true even if we assume that any possible event has positive probability.

A number $r$ is polynomial-time computable if there exists a transducer Turing machine $M_r$ that, for an integer input $b$ (represented through a binary string), runs in time at most $\mathrm{poly}(b)$ (the notation $\mathrm{poly}(b)$ denotes an arbitrary polynomial function of $b$, and might indicate a different polynomial at each time it is used) and outputs a rational number $r'$ (represented through its numerator and denominator in binary strings) such that $|r - r'| < 2^{-b}$. Of special relevance to us are numbers of the form $(2^{-v_1} - 2^{-v_2})/(1 + 2^{-v_3})$, with $v_1$, $v_2$ and $v_3$ being non-negative rationals no greater than two. For any rational $v$ between zero and two we can build a machine that outputs a rational $r'$ that approximates $2^{-v}$ with precision $b$ in time $\mathrm{poly}(b)$ by computing the Taylor expansions of $2^{-v}$ around zero with sufficiently many terms (depending on the value of $b$) similar to the proof of Lemma 4 of the work of Mauá et al. (2013). The desired numbers can then be obtained by the corresponding fractional expression. The following lemmas ensure that the outcome of STRONG-INF on networks specified with polynomial-time computable numbers can be approximated arbitrarily well using a network specified only with positive rational numbers. It allows us to specify the desired precision and for which nodes of the network the numerical parameters will be "approximated" by positive rational numbers.

**Lemma 1.** *Consider a vertex-based credal network $\mathcal{N}$ whose numerical parameters are specified with polynomial-time computable numbers encoded by their respective machines (or directly given as rational numbers), and let $b$ be the size of the encoding of $\mathcal{N}$. Given a subset of the nodes $N' \subseteq N$ of $\mathcal{N}$ and a rational number $1 \geq \varepsilon \geq 2^{-\mathrm{poly}(b)}$, we can construct in time $\mathrm{poly}(b)$ a vertex-based credal network $\mathcal{N}'$ over the same variables whose numerical parameters that specify the credal sets of nodes $N'$ are all rational numbers greater than $2^{-\mathrm{poly}(b)}$ (numerical parameters related to nodes not in $N'$ are kept unchanged), and such that there is a polynomial-time computable surjection $(p, p')$ that associates any extreme $p$ of the strong extension of $\mathcal{N}$ with an extreme $p'$ of the strong extension of $\mathcal{N}'$ satisfying*

$$\max_{\mathbf{x}_A} |\mathrm{P}_{p'}(\mathbf{X}_A = \mathbf{x}_A) - \mathrm{P}_p(\mathbf{X}_A = \mathbf{x}_A)| \leq \varepsilon\,,$$

*for any subset $\mathbf{X}_A \subseteq \mathbf{X}$ of the variables.*

*Proof.* Take $\mathcal{N}'$ to be equal to $\mathcal{N}$ except that each computable number $r$ used in the specification of $\mathcal{N}$ for nodes $N'$ is replaced by a rational $r'$ such that $|r' - r| < 2^{-(n+1)(v+1)-1}\varepsilon$, where $n$ is the number of variables, and $v$ is the maximum cardinality of the domain of any variable in $\mathcal{N}$. Because $\varepsilon \geq 2^{-\mathrm{poly}(b)}$, we can run the Turing machine $M_r$ used to represent $r$ on input $\mathrm{poly}(b)+(n+1)(v+1)+1$ to obtain $r'$ in time $O(\mathrm{poly}(\mathrm{poly}(b) + (n+1)(v+1)+1))$, which is $O(\mathrm{poly}(b))$. After obtaining $r'$, add to it $2^{-(n+1)(v+1)-1}\varepsilon$ to ensure that $r < r' <$

$r + 2^{-(n+1)(v+1)}\varepsilon$, that is, the approximation is from above. However, exactly one of the probability values in each distribution used to represent an extreme of a local credal set in $\mathcal{N}'$ is not approximated in that way but is computed as one minus the sum of the other numbers to ensure that its distribution adds up exactly to one; we can choose the greatest value for that (by trying each of the (at most) $v$ states, which probability value is at least $1/v$), and its error with respect to the corresponding original computable number will be at most $(v-1) \cdot 2^{-(n+1)(v+1)}\varepsilon < 2^{-n(v+1)}\varepsilon$. This construction ensures that every created rational number is greater than $2^{-(n+1)(v+1)-1}\varepsilon > 2^{-\mathrm{poly}(b)}$ and have an "error" of at most $2^{-n(v+1)}\varepsilon$ to the original corresponding number.

Let $q_i(x_i|\mathbf{x}_{\mathrm{Pa}(i)})$ and $q'_i(x_i|\mathbf{x}_{\mathrm{Pa}(i)})$ denote, respectively, the parameters of $\mathcal{N}$ and $\mathcal{N}'$ (i.e. they are corresponding extreme distributions of the local credal sets $Q(X_i|\mathbf{x}_{\mathrm{Pa}(i)})$ in the two networks) such that $q'_i(x_i|\mathbf{x}_{\mathrm{Pa}(i)})$ is the approximated version computed from $q_i(x_i|\mathbf{x}_{\mathrm{Pa}(i)})$ as explained. Consider an assignment $\mathbf{x}$ to all variables in $\mathcal{N}$ (or in $\mathcal{N}'$). Let also $p$ be an extreme of the strong extension of $\mathcal{N}$. Then $p$ factorizes as $p(\mathbf{x}) = \prod_{i\in N} q_i(x_i|\mathbf{x}_{\mathrm{Pa}(i)})$, for some combination of extreme distributions $q_i(\cdot|\mathbf{x}_{\mathrm{Pa}(i)})$ from $Q(X_i|\mathbf{x}_{\mathrm{Pa}(i)})$, $i \in N$. Finally, let $p'$ be an extreme distribution in the strong extension of $\mathcal{N}'$ that satisfies $p'(\mathbf{x}) = \prod_{i\in N} q'_i(x_i|\mathbf{x}_{\mathrm{Pa}(i)})$. By construction, $|q'_i(x_i|\mathbf{x}_{\mathrm{Pa}(i)}) - q_i(x_i|\mathbf{x}_{\mathrm{Pa}(i)})| \le 2^{-n(v+1)}\varepsilon$. It follows from the binomial expansion of the factorization of $p'(\mathbf{x})$ on any $\mathbf{x}$ that

$$\begin{aligned} p'(\mathbf{x}) = \prod_{i\in N} q'_i(x_i|\mathbf{x}_{\mathrm{Pa}(i)}) &\le \prod_{i\in N}\left(2^{-n(v+1)}\varepsilon + q_i(x_i|\mathbf{x}_{\mathrm{Pa}(i)})\right) \\ &= \sum_{A\subseteq N} (2^{-n-vn}\varepsilon)^{n-|A|} \prod_{i\in A} q_i(x_i|\mathbf{x}_{\mathrm{Pa}(i)}) \\ &\le 2^n 2^{-n-vn}\varepsilon + \prod_{i\in N} q_i(x_i|\mathbf{x}_{\mathrm{Pa}(i)}) \\ &= p(\mathbf{x}) + 2^{-nv}\varepsilon\,. \end{aligned}$$

The second inequality follows from the fact that there is one term for $p(\mathbf{x})$ in the expansion and $2^n - 1$ terms that can be written as a product of $2^{-n(v+1)}\varepsilon$ by non-negative numbers less than or equal to one. With a similar reasoning, we can show that

$$p'(\mathbf{x}) \ge \prod_{i\in N}\left(q_i(x_i|\mathbf{x}_{\mathrm{Pa}(i)}) - 2^{-n(v+1)}\varepsilon\right) \ge p(\mathbf{x}) - 2^{-nv}\varepsilon\,.$$

Thus, $\max_{\mathbf{x}} |p'(\mathbf{x}) - p(\mathbf{x})| \le 2^{-nv}\varepsilon$. Now consider a subset of the variables $\mathbf{X}_A$ and an assignment $\mathbf{x}_A \sim \mathbf{X}_A$. Since

$$\mathrm{P}_{p'}(\mathbf{X}_A = \mathbf{x}_A) = \sum_{\mathbf{x}':\mathbf{x}'_A = \mathbf{x}_A} p'(\mathbf{x}')\,,$$

each term $p'(\mathbf{x}')$ in that sum satisfies $p'(\mathbf{x}') \le p(\mathbf{x}') + 2^{-nv}\varepsilon$, and because there are less than $v^n \le 2^{vn}$ terms being summed, it follows that

$$\mathrm{P}_{p'}(\mathbf{X}_A = \mathbf{x}_A) \le \sum_{\mathbf{x}':\mathbf{x}'_A = \mathbf{x}_A} \left(p(\mathbf{x}) + 2^{-vn}\varepsilon\right) \le \mathrm{P}_p(\mathbf{X}_A = \mathbf{x}_A) + \varepsilon\,.$$

An analogous argument can be used to show that $\mathrm{P}_{p'}(\mathbf{X}_A = \mathbf{x}_A) \ge \mathrm{P}_p(\mathbf{X}_A = \mathbf{x}_A) - \varepsilon$. Note that the obtained mapping $(p, p')$ is a surjection by construction. $\square$

The above lemma has the following direct consequence on the computation of the STRONG-INF with polynomial-time computable numbers. The only restriction for its application is that the computable numbers must be either zero or greater than some $\rho$ that is not "exponentially close to zero".

**Corollary 1.** *Consider a vertex-based credal network $\mathcal{N}$ whose numerical parameters are specified with polynomial-time computable numbers encoded by their respective machines (or directly given as rational numbers), such that no such number lies properly in $]0, \rho[$, for some $0 \leq \rho \leq 1$. Let $b$ be the size of the encoding of the network. Given a subset of the nodes $N' \subseteq N$ of $\mathcal{N}$ and a rational number $\varepsilon$ with $2^{-\text{poly}(b)} \leq \varepsilon < \rho$, we can construct in time $\text{poly}(b)$ a vertex-based credal network $\mathcal{N}'$ over the same variables whose numerical parameters defining credal sets related to nodes $N'$ are all strictly positive rational numbers greater than $2^{-\text{poly}(b)}$ (numbers defining credal sets of nodes not in $N'$ are kept unchanged), and such that* [7]

$$|\textit{STRONG-INF}(\mathcal{N}', t, x_t, O, \mathbf{x}_O) - \textit{STRONG-INF}(\mathcal{N}, t, x_t, O, \mathbf{x}_O)| \leq \varepsilon\,,$$

*for any query $t, x_t, O, \mathbf{x}_O$ such that either $O = \emptyset$ or $\min_p \mathrm{P}_p(\mathbf{x}_O) > 0$ in $\mathcal{N}$.*

*Proof.* According to Lemma 1, there is a polynomial-time computable network $\mathcal{N}'$ whose numerical parameters that specify the credal sets related to nodes $N'$ are positive rational numbers and a polynomial-time computable surjection $(p, p')$ such that $p$ and $p'$ are, respectively, extreme distributions of the strong extension of $\mathcal{N}$ and $\mathcal{N}'$, and satisfy $|\mathrm{P}_{p'}(\mathbf{x}_A) - \mathrm{P}_p(\mathbf{x}_A)| \leq \varepsilon^{n+1}/3$ for all $\mathbf{X}_A \subseteq \mathbf{X}$ and $\mathbf{x}_A \sim \mathbf{X}_A$. It follows that

$$\mathrm{P}_{p'}(x_t|\mathbf{x}_O) = \frac{\mathrm{P}_{p'}(x_t, \mathbf{x}_O)}{\mathrm{P}_{p'}(\mathbf{x}_O)} \geq \frac{\mathrm{P}_p(x_t, \mathbf{x}_O) - \varepsilon^{n+1}/3}{\mathrm{P}_p(\mathbf{x}_O) + \varepsilon^{n+1}/3}\,,$$

where $p'$ is the image of $p$ according to the surjection. If $\mathrm{P}_p(x_t, \mathbf{x}_O) = 0$, this equation is useless and vanishes. Otherwise, by Lemma 7 of the work of de Campos and Cozman (2013), we have that

$$\frac{\mathrm{P}_p(x_t, \mathbf{x}_O) - \varepsilon^{n+1}/3}{\mathrm{P}_p(\mathbf{x}_O) + \varepsilon^{n+1}/3} \geq \mathrm{P}_p(x_t|\mathbf{x}_O) - \frac{2\varepsilon^{n+1}/3}{\rho^n} \geq \mathrm{P}_p(x_t|\mathbf{x}_O) - \varepsilon\,.$$

The other side of the inequality is obtained analogously (using once more Lemma 7 of the work of de Campos & Cozman, 2013, except for the case of $\mathrm{P}_p(x_t, \mathbf{x}_O) = 0$, when the following reasoning is valid without the need of that lemma):

$$\begin{aligned}
\mathrm{P}_{p'}(x_t|\mathbf{x}_O) &\leq \frac{\mathrm{P}_p(x_t, \mathbf{x}_O) + \varepsilon^{n+1}/3}{\mathrm{P}_p(\mathbf{x}_O) - \varepsilon^{n+1}/3} \leq \mathrm{P}_p(x_t|\mathbf{x}_O) + \frac{2\varepsilon^{n+1}/3}{\rho^n - \varepsilon^{n+1}/3} \\
&\leq \mathrm{P}_p(x_t|\mathbf{x}_O) + \frac{2\varepsilon}{3 - \varepsilon} \leq \mathrm{P}_p(x_t|\mathbf{x}_O) + \varepsilon.
\end{aligned}$$

Hence, $|\mathrm{P}_{p'}(x_t|\mathbf{x}_O) - \mathrm{P}_p(x_t|\mathbf{x}_O)| \leq \varepsilon$. Let $p$ be an extreme distribution in the strong extension of $\mathcal{N}$ such that $\mathrm{P}_p(x_t|\mathbf{x}_O) = \min_{q \in K_S} \mathrm{P}_q(x_t|\mathbf{x}_O)$, where $K_S$ denotes the strong extension of

7. We abuse notation of STRONG-INF, as it is defined for opt $\in \{\min, \max\}$. We intend to mean that the equation is valid for both options of opt.

$\mathcal{N}$. We have that

$$\min_{q \in K_S} \mathrm{P}_q(x_t|\mathbf{x}_O) = \mathrm{P}_p(x_t|\mathbf{x}_O) \geq \mathrm{P}_{p'}(x_t|\mathbf{x}_O) - \varepsilon \geq \min_{q' \in K'_S} \mathrm{P}_{q'}(x_t|\mathbf{x}_O) - \varepsilon\,,$$

where $p'$ in the first inequality is the image of $p$ according to the surjection, and $K'_S$ in the last inequality is the strong extension of $\mathcal{N}'$. It follows from the above that

$$\min_{q' \in K'_S} \mathrm{P}_{q'}(x_t|\mathbf{x}_O) - \min_{q \in K_S} \mathrm{P}_q(x_t|\mathbf{x}_O) \leq \varepsilon\,.$$

The other side comes by contradiction. Suppose that

$$\min_{q \in K_S} \mathrm{P}_q(x_t|\mathbf{x}_O) - \min_{q' \in K'_S} \mathrm{P}_{q'}(x_t|\mathbf{x}_O) > \varepsilon \iff \min_{q \in K_S} \mathrm{P}_q(x_t|\mathbf{x}_O) - \varepsilon > \min_{q' \in K'_S} \mathrm{P}_{q'}(x_t|\mathbf{x}_O)\,.$$

Hence there would exist an extreme $q' \in K'_S$ such that $|\mathrm{P}_{q'}(x_t|\mathbf{x}_O) - \mathrm{P}_q(x_t|\mathbf{x}_O)| > \varepsilon$ for any $q \in K_S$, which is impossible because the mapping $(q, q')$ is a surjection. An analogous proof works for showing that the upper bounds according to the two networks do not differ by more than $\varepsilon$. $\square$

### 3.4 Precise-Vacuous Networks

To show that EPISTEMIC-INF is $\mathrm{NP}^{\mathrm{PP}}$-hard in arbitrary networks, we need the following result, which shows that inferences under strong independence and epistemic irrelevance coincide on precise-vacuous networks.

**Proposition 2.** *Consider a credal network whose root nodes are vacuous and non-root nodes are precise. Let $t$ be a non-root node, $\tilde{x}_t$ an arbitrary value of $X_t$ and $O = \emptyset$. Then STRONG-INF equals EPISTEMIC-INF.*

*Proof.* Let $\mathbf{X}_R$ be the vacuous variables associated with the root nodes (hence to vacuous local credal sets), and $\mathbf{X}_I$ denote the remaining variables (which are precise). For every precise node $i$ in $I$, let $q_i^{\mathbf{x}_{\mathrm{Pa}(i)}}(x_i)$ be the single distribution in the associated credal set $Q(X_i|\mathbf{x}_{\mathrm{Pa}(i)})$. Consider an arbitrary distribution $p$ in the epistemic extension $K_E$, and let $<$ be a topological ordering of the nodes. For an assignment $\mathbf{x}$ to $\mathbf{X}$, we write $\mathbf{x}_{<i}$ to denote the coordinates $j < i$ of $\mathbf{x}$ according to the topological ordering. For every node $i$ the set $\{j \in N : j < i\}$ is a subset of $\mathrm{Nd}(i)$, and it follows from the definition of epistemic extension that $\mathrm{P}_p(x_i|\mathbf{x}_{<i}) = q_i^{\mathbf{x}_{\mathrm{Pa}(i)}}(x_i)$ for every precise node $i$ and assignments $x_i$ and $\mathbf{x}_{<i}$. By the Chain Rule we have that

$$\forall \mathbf{x} \sim \mathbf{X}: \quad \mathrm{P}_p(\mathbf{x}) = \mathrm{P}_p(\mathbf{x}_R) \prod_{i \in I} \mathrm{P}_p(x_i|\mathbf{x}_{<i}) = q(\mathbf{x}_R) \prod_{i \in I} q_i^{\mathbf{x}_{\mathrm{Pa}(i)}}(x_i)\,,$$

where $q$ is any distribution of $\mathbf{X}_R$ (since these nodes are vacuous, any distribution satisfies the constraints in $K_E$ for them). As stated, let $\tilde{x}_t$ be the value of interest of $X_t$. The result of EPISTEMIC-INF is thus given by

$$\begin{aligned}\min_{p\in \mathrm{K}_E} \mathrm{P}_p(X_t\!=\!\tilde{x}_t) &= \min_{q\in V(\mathbf{X}_R)} \sum_{\mathbf{x}} \delta_{\tilde{x}_t}(x_t)\cdot q(\mathbf{x}_R)\cdot \prod_{i\in I} q_i^{\mathbf{x}_{\mathrm{Pa}(i)}}(x_i)\\ &= \min_{q\in V(\mathbf{X}_R)} \sum_{\mathbf{x}_R} q(\mathbf{x}_R) \sum_{\mathbf{x}_I}\prod_{i\in I} q_i^{\mathbf{x}_{\mathrm{Pa}(i)}}(x_i)\delta_{\tilde{x}_t}(x_t)\\ &= \min_{q\in V(\mathbf{X}_R)} \sum_{\mathbf{x}_R} q(\mathbf{x}_R)\cdot g(\mathbf{x}_R)\,,\end{aligned}$$

where $g(\mathbf{x}_R) \overset{\text{def}}{=} \sum_{\mathbf{x}_I}\prod_{i\in I}\delta_{\tilde{x}_t}(x_t)q_i^{\mathbf{x}_{\mathrm{Pa}(i)}}(x_i)$. According to the last equality, the lower marginal probability of $X_t\!=\!\tilde{x}_t$ is a convex combination of $g(\mathbf{x}_R)$. Hence,

$$\min_{p\in \mathrm{K}_E} \mathrm{P}_p(X_t\!=\!\tilde{x}_t) \geq \min_{\mathbf{x}_R} \sum_{\mathbf{x}_{I\setminus\{q\}}}\prod_{i\in I} q_i^{\mathbf{x}_{\mathrm{Pa}(i)}}(x_i)\,.$$

The rightmost minimization is exactly the value of the lower marginal probability returned by STRONG-INF, and since the strong extension is contained in the epistemic extension, the inequality above is tight. An analogous result can be obtained for the upper probability by substituting minimizations with maximization and inverting the inequality above. ☐

The class of networks considered in the result above might seem restrictive at first sight. However, Antonucci and Zaffalon (2008) showed that STRONG-INF in any credal network whose local credal sets are represented in vertex-based form can be reduced in linear time to the same problem in a credal network containing only vacuous and precise nodes. Such a network can then be transformed in linear time into a precise-vacuous network (i.e., one in which root nodes are vacuous and non-root nodes are precise) by applying Transformation 6 in the work of Mauá, de Campos, and Zaffalon (2012a),[8] which increases the treewidth of the network by at most three. Hence, any vertex-based credal network can be reduced in polynomial time into a precise-vacuous network for which STRONG-INF provides the same result as for the original network (and whose treewidth remains bounded, if it originally were). The hardness of EPISTEMIC-INF in precise-vacuous credal networks follows immediately from the hardness of inference under strong independence and Proposition 2, as the following corollary shows.

**Corollary 2.** *STRONG-INF and EPISTEMIC-INF are NP$^{PP}$-hard even if all variables are binary and all numerical parameters are strictly positive.*

*Proof.* Cozman et al. (2004) used a reduction from E-MAJSAT to STRONG-INF without evidence in a binary credal network whose root nodes are vacuous and non-root nodes are precise to show that such inference is NP$^{\mathrm{PP}}$-hard. Since according to Proposition 2 the result of EPISTEMIC-INF is the same, EPISTEMIC-INF is also NP$^{\mathrm{PP}}$-hard. In order to show that the result is valid also if all numerical parameters are strictly positive, we will only sketch the proof so as to avoid repeating all the formulation for the E-MAJSAT problem. Using Lemma 1 with epsilon $\varepsilon = 2^{-\mathrm{poly}(b)}$ smaller than the precision of any number involved in

8. Strictly speaking, the work of Mauá et al. (2012a) deals with influence diagrams; the link between those and credal networks was established by Antonucci and Zaffalon (2008) and de Campos and Ji (2008).

any calculation, we build a new network where all numerical parameters are strictly positive and the variation in the result of STRONG-INF is negligible such that it can still decide E-MAJSAT (further details on how small $\varepsilon$ has to be are omitted for simplicity, but the gap between instances is large enough that $2^{-O((n+m)^2)}$ will suffice, where $n, m$ are the number of variables and clauses in the specification of E-MAJSAT, see Park & Darwiche, 2004, Thm. 2). Because EPISTEMIC-INF contains STRONG-INF, its result after applying Lemma 1 will be between the results of STRONG-INF in the new network and in the old network (the latter equals that of EPISTEMIC-INF). Hence, EPISTEMIC-INF in the new network with strictly positive numerical parameters also decides E-MAJSAT. □

Note that the result holds irrespective of how the local credal sets are represented, since vacuous and precise nodes can be mapped from constraint-based to vertex-based form and vice-versa in polynomial time.

### 3.5 Singly Connected Networks

We now turn our attention to singly connected networks. A first result is a direct consequence of Proposition 2 is the NP-hardness of EPISTEMIC-INF in singly connected credal networks, since STRONG-INF is NP-hard in singly connected networks, even if we admit imprecision only on root nodes:

**Corollary 3.** *EPISTEMIC-INF is NP-hard in singly connected credal networks.*

*Proof.* In the work of de Campos and Cozman (2005) it has been shown that STRONG-INF is NP-hard in precise-vacuous singly connected networks. Since Proposition 2 shows that EPISTEMIC-INF can be reduced to STRONG-INF on the same input, the result follows. □

The proof of NP-hardness of STRONG-INF provided by de Campos and Cozman (2005) requires the variable domain cardinalities to be unbounded. We present here the stronger result of NP-hardness of credal inference in networks where imprecise variables are binary and precise ones are at most ternary. We can now show NP-hardness of credal inference in singly connected networks with bounded variable cardinality.

**Theorem 1.** *STRONG-INF and EPISTEMIC-INF are NP-hard even if the network is singly connected and has treewidth at most two, all imprecise variables are binary, and all precise variables are (at most) ternary. Moreover, all numerical parameters in the network are strictly positive.*

*Proof.* We defer the treatment of zero numerical parameters to the final part. We build a singly connected credal network with underlying graph as in Figure 3. The variables (associated with nodes) on the upper row are binary and vacuous, namely $X_1, \dots, X_k$, while the remaining variables $X_{k+1}, \cdots, X_{2k+1}$ are ternary and precise. The local credal sets associated with precise nodes are singletons such that $Q(X_{k+1})$ contains a uniform distribution $q(x_{k+1}) = 1/3$, and, for $i = k+2, \dots, 2k+1$, $Q(X_i|x_{i-1}, x_{i-k-1})$ contains the conditional distribution $q(X_i|x_{i-1}, x_{i-k-1})$ specified in Table 1. The rational numbers $v_i$ in the table shall be defined later on. Consider an extreme distribution $p(\mathbf{x})$ of the strong

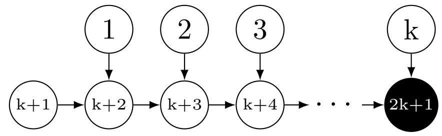



Figure 3: Credal network structure used to prove Theorem 1. The shaded node is the target.

| $q(x_i \vert x_{i-1}, x_{i-k-1})$ | $x_i=1$ | $x_i=2$ | $x_i=3$ |
|---|---|---|---|
| $x_{i-1}=1, x_{i-k-1}=1$ | $2^{-v_i}$ | 0 | $1-2^{-v_i}$ |
| $x_{i-1}=2, x_{i-k-1}=1$ | 0 | 1 | 0 |
| $x_{i-1}=3, x_{i-k-1}=1$ | 0 | 0 | 1 |
| $x_{i-1}=1, x_{i-k-1}=0$ | 1 | 0 | 0 |
| $x_{i-1}=2, x_{i-k-1}=0$ | 0 | $2^{-v_i}$ | $1-2^{-v_i}$ |
| $x_{i-1}=3, x_{i-k-1}=0$ | 0 | 0 | 1 |

Table 1: Local probability distributions used to prove Theorem 1

extension of the network. It follows for all $\mathbf{x}$ that

$$p(\mathbf{x}) = q(x_{k+1}) \prod_{i=k+2}^{2k+1} q(x_i|x_{i-1}, x_{i-k-1}) \prod_{i \in A} \delta_1(x_i) \prod_{i \notin A} \delta_0(x_i)\,,$$

for some $A \subseteq \{1, \ldots, k\}$. This is because the extreme distributions of local vacuous sets over binary variables are $\delta_0$ and $\delta_1$, and each choice of a local extreme for a root node can be associated with a choice of either including or excluding its corresponding node in/from $A$. Let $\neg A \stackrel{\text{def}}{=} \{1, \ldots, k\} \setminus A$ denote the complement of a set $A$ with respect to $\{1, \ldots, k\}$. We have that

$$p(\mathbf{x}) = \begin{cases} \frac{1}{3} \prod_{i=k+2}^{2k+1} q(x_i|x_{i-1}, x_{i-k-1}), & \text{if } \mathbf{x}_A = 1 \text{ and } \mathbf{x}_{\neg A} = 0\,; \\ 0, & \text{otherwise.} \end{cases}$$

It follows that

$$\mathrm{P}_p(X_{2k+1}=1) = \sum_{\mathbf{x}} p(\mathbf{x}) \delta_1(x_{2k+1}) = \frac{2^{-\sum_{i \in A} v_i}}{3} \text{ and } \mathrm{P}_p(X_{2k+1}=2) = \frac{2^{\sum_{i \in A} v_i - 2}}{3}\,.$$

We show the NP-hardness of credal inference by reducing the NP-complete PARTITION problem (Garey & Johnson, 1979) to the computation of $\max_{p \in K_S} \mathrm{P}_p(X_{2n+1} = 3)$. We define PARTITION as follows.

PARTITION
*Input*: List of positive integers $z_1, \ldots, z_k$.
*Output*: Is there a subset $A \subseteq \{1, \ldots, k\}$ such that

$$\sum_{i \in A} z_i = \sum_{i \in \neg A} z_i \quad ?$$

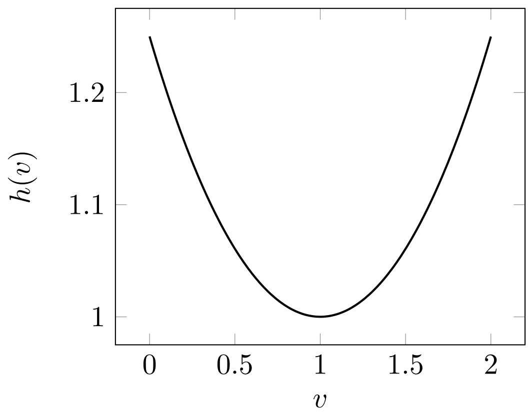



Figure 4: Function used in the reduction in the proof of Theorem 1.

Notice that the above equality is equivalent to

$$\sum_{i \in A} z_i/z = 1\,, \quad \text{where } z = \frac{1}{2}\sum_{i=1}^{k} z_i\,.$$

Define the exponents in Table 1 as $v_i \stackrel{\text{def}}{=} z_i/z$, and let $v_A \stackrel{\text{def}}{=} \sum_{i \in A} v_i$. It follows for any $A$ that $v_A = 2 - \sum_{i \in \neg A} v_i$. If an instance of the PARTITION is a *yes-instance* (i.e., if the output of PARTITION is yes), then there is $A$ for which $v_A = 1$, whereas if it is a *no-instance* (i.e., if the output is no), then for any $A$, it follows that $|v_A - 1| \geq 1/(2z)$ because the numbers in the input are integers and hence the sums of two different sets are either equal or differ by at least one. Consider the function

$$h(v_A) = \frac{2^{-(v_A-1)} + 2^{v_A-1}}{2}\,.$$

The graph of the function is depicted in Figure 4. Seen as a function of a continuous variable $v_A \in [0, 2]$, the function above is strictly convex, symmetric around one, and achieves the minimum value of one at $v_A = 1$. Thus, if PARTITION returns yes then $\min_A h(v_A) = 1$, while if it returns no we have that

$$\min_A h(v_A) \geq 2^{-1/(2z)-1} + 2^{1/(2z)-1} \geq 2^{(2z)^{-4}} > 1 + (2z)^{-4}/2 = 1 + 1/(32z^4)\,,$$

where the second inequality is due to Lemma 24 in the work of Mauá et al. (2012a), and the strict inequality follows from the first-order Taylor expansion of $2^{(2z)^{-4}}$. Let $\alpha \stackrel{\text{def}}{=} (1 + z^{-4}/64)/3$. By computing STRONG-INF with query $X_{2n+1} = 3$ and no evidence, we can decide PARTITION, as

$$1 - \max \mathrm{P}_p(X_{2k+1} = 3) = \min_p \left(\mathrm{P}_p(X_{2k+1} = 1) + \mathrm{P}_p(X_{2k+1} = 2)\right) = \min_A \frac{h(v_A)}{3} \leq \alpha$$

if and only if the result of PARTITION is yes. It remains to show that we can polynomially encode the numbers $2^{-z_i/z}$. This is done by applying Lemma 1 with a small enough $\varepsilon$

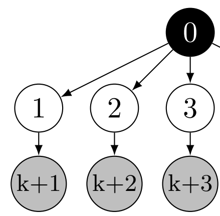



Figure 5: DAG of the credal tree used to prove Theorem 2.

computable in time polynomial in the size of the partition problem: $\varepsilon = 1/(3 \cdot 64z^4)$ suffices. Note that if we only apply Lemma 1 to the non-root nodes and leave the root nodes untouched as vacuous, then according to Proposition 2, the outcome of EPISTEMIC-INF is the same, proving also its NP-hardness. Now, by further applying Lemma 1 to root vacuous nodes, we ensure all numerical parameters are strictly positive and still yield a result that can be used to decide PARTITION. Because EPISTEMIC-INF contains STRONG-INF, its result after applying Lemma 1 will be between the results of STRONG-INF in the new network and in that network with vacuous root nodes. Hence, EPISTEMIC-INF in the new network with strictly positive numerical parameters also decides PARTITION. ☐

### 3.6 Credal Trees

The previous complexity results showed that, from a theoretical standpoint, computing the EPISTEMIC-INF is just as difficult as solving STRONG-INF. When the underlying graph is a tree, de Cooman et al. (2010) showed that EPISTEMIC-INF can be computed efficiently, and it was previously unknown whether a similar result could be obtained for STRONG-INF. The next result shows that in this case the equivalence on the tractability under the two different irrelevance concepts does not hold unless P equals NP.

**Theorem 2.** *STRONG-INF in tree-shaped credal networks is NP-hard, even if only one variable is ternary and precise and all the rest are binary, and even if all numerical parameters are strictly positive.*

*Proof.* We show hardness by a reduction from PARTITION as defined previously. As before, we define $v_i \stackrel{\text{def}}{=} z_i/z$, and $v_A \stackrel{\text{def}}{=} \sum_{i \in A} v_i$, and note that $v_A = 2 - \sum_{i \in \neg A} v_i$. We also let $h(v_A)$ to be as before (thus $h$ is strictly convex on $[0, 2]$, symmetric around one, and achieves the minimum value of one at $v_A = 1$). Given an instance of PARTITION (i.e., a list of integers), we build a credal tree $\mathcal{N}$ over variables $X_0, \ldots, X_{2k}$ with DAG as in Figure 5. The root variable $X_0$ takes values in $\{1, 2, 3\}$, and is precise and uniformly distributed (i.e., its local credal set contains only the distribution $q_0(x_0) = 1/3$). The remaining variables are all binary and take values in $\{0, 1\}$. For $i = 1, \ldots, k$, we specify the local conditional credal sets $Q(X_i|x_0)$ as singletons $\{q_i^{x_0}\}$ such that

$$q_i^{x_0}(1) = \begin{cases} 2^{-v_i}/(1 + 2^{-v_i}), & \text{if } x_0 = 1, \\ 1/(1 + 2^{-v_i}), & \text{if } x_0 = 2, \\ 1/2, & \text{if } x_0 = 3. \end{cases}$$

For $i = 1+k, \dots, 2k$ we specify the local credal sets $Q(X_i|x_{i-k}) = \{p \in V(X_i) : \epsilon \leq p(1) \leq 1\}$, where $\epsilon = 2^{-k-3}/(64z^4)$. Each of these local credal sets can be represented either in vertex-based form by two extreme distributions or by a couple of constraints.

Let

$$\mu \stackrel{\text{def}}{=} \max_{p \in K_S} \mathrm{P}_p(X_0 = 3 | \mathbf{X}_O = \mathbf{x}_O) = \max_{p \in K_S} \frac{\mathrm{P}_p(X_0 = 3, \mathbf{X}_O = \mathbf{x}_O)}{\mathrm{P}_p(\mathbf{X}_O = \mathbf{x}_O)} .$$

Hence, $\mu$ is the solution of

$$\max_{p \in K_S} \left[ \sum_{x_0} (\delta_3(x_0) - \mu) \mathrm{P}_p(X_0 = x_0, \mathbf{X}_O = \mathbf{x}_O) \right] = 0 .$$

By definition, any extreme distribution $p$ in the strong extension $\mathrm{K}_S$ satisfies for $\mathbf{x} \sim \mathbf{X}$ such that $x_{k+1} = x_{k+2} = \cdots = x_{2k} = 1$ the equality

$$p(\mathbf{x}) = q_0(x_0) \prod_{i=1}^{k} q_i^{x_0}(x_i) \alpha_i^{x_i} ,$$

where each $\alpha_i^{x_i}$ is a number in $[\epsilon, 1]$. Let $O = \{k+1, \dots, 2k\}$ and $\mathbf{x}_O = (1, \dots, 1)$. It follows that $\mu$ is the solution of

$$\max \sum_{x_0, \dots, x_k} (\delta_3(x_0) - \mu) q_0(x_0) \prod_{i=1}^{k} q_i^{x_0}(x_i) \alpha_i^{x_i} = 0 ,$$

where the maximization is performed on $\alpha_i^0, \alpha_i^1$, for $i = 1, \dots, k$. Consider $j \in \{1, \dots, k\}$ and let

$$\beta_j^{x_j} \stackrel{\text{def}}{=} \sum_{x_0, \dots, x_{j-1}} \sum_{x_{j+1}, \dots, x_k} (\delta_3(x_0) - \mu)\, q_0(x_0) q_j^{x_0}(x_j) \prod_{i=1, i \neq j}^{k} [q_i^{x_0}(x_i) \alpha_i^{x_i}] .$$

Then,

$$\max \sum_{x_0, \dots, x_k} (\delta_3(x_0) - \mu)\, q_0(x_0) \prod_{i=1}^{k} [q_i^{x_0}(x_i) \alpha_i^{x_i}] = \max \left( \alpha_j^0 \beta_j^0 + \alpha_j^1 \beta_j^1 \right) .$$

Since $\alpha_j^0$ and $\alpha_j^1$ are both positive, the maximization in the right-hand side above equals zero only if both $\beta_j^0$ and $\beta_j^1$ are zero or they have different signs. In the former case, any value of $\alpha_j^0$ and $\alpha_j^1$ maximizes the expression, and we can assume that $(\alpha_j^0, \alpha_j^1)$ equals $(\epsilon, 1)$ or $(1, \epsilon)$. In the latter case, $\beta_j^0 < \beta_j^1$ implies that $(\alpha_j^0, \alpha_j^1)$ equals $(\epsilon, 1)$ in order to maximize the expression, and $(\alpha_j^0, \alpha_j^1) = (1, \epsilon)$ would do it otherwise. Since we selected $j$ arbitrarily, the result holds for all $j$. Thus, the maximization is equivalent to selecting, for $i = 1, \dots, k$, a value $y_i$ in $\{0, 1\}$ such that $\alpha_i^0 = \epsilon^{1-y_i}$ and $\alpha_i^1 = \epsilon^{y_i}$. It follows that

$$\max \sum_{x_0, \dots, x_k} (\delta_3(x_0) - \mu)\, q_0(x_0) \prod_{i=1}^{k} [q_i^{x_0}(x_i) \alpha_i^{x_i}] =$$

$$\max_{\mathbf{y} \in \{0,1\}^k} \sum_{x_0, \dots, x_k} (\delta_3(x_0) - \mu)\, q_0(x_0) \prod_{i=1}^{k} [q_i^{x_0}(x_i) \epsilon^{(1-x_i)(1-y_i)} \epsilon^{x_i y_i}] .$$

By rearranging terms, we obtain

$$\max_{\mathbf{y}\in\{0,1\}^k} \sum_{x_0} (\delta_3(x_0) - \mu)\, q_0(x_0) \prod_{i=1}^{k} \left[ q_i^{x_0}(0)\epsilon^{1-y_i} + q_i^{x_0}(1)\epsilon^{y_i} \right] ,$$

which by construction equals

$$\max_{\mathbf{y}\in\{0,1\}^k} -\left( \frac{\mu}{3}\gamma \prod_{i=1}^{k} [\epsilon^{1-y_i} \; + \; 2^{-v_i}\epsilon^{y_i}] \; + \; \frac{\mu}{3}\gamma \prod_{i=1}^{k} [2^{-v_i}\epsilon^{1-y_i} \; + \; \epsilon^{y_i}] \; + \; \frac{\mu-1}{3} \prod_{i=1}^{k} \frac{1+\epsilon}{2} \right),$$

where $\gamma = \prod_{i=1}^{k} q_i^2(1)$ (recall that $q_i^2(1)$ is the conditional probability value of $X_i = 1$ given $X_0 = 2$). The binary vector $\mathbf{y}$ can be seen as the characteristic vector of a subset $A \subset \{1, \ldots, k\}$. Define

$$b_A \stackrel{\text{def}}{=} \prod_{i \in A} (2^{-v_i} + \epsilon) \prod_{i \in \neg A} (1 + 2^{-v_i}\epsilon)$$

for every subset $A$. The optimization on $\mathbf{y}$ can be rewritten as the following optimization over subsets $A$: find $\mu$ such that

$$\left( -\frac{\mu-1}{3} \left( \frac{1+\epsilon}{2} \right)^k + \max_A -\frac{\mu}{3}\gamma\, (b_A + b_{\neg A}) \right) = 0 \, .$$

Solving the expression above for $\mu$, we get to

$$\mu = \left( 1 + \left( \frac{2}{1+\epsilon} \right)^k \gamma \min_A (b_A + b_{\neg A}) \right)^{-1} .$$

Define the function $g(a)$ as

$$g(a) \stackrel{\text{def}}{=} 1 + \left( \frac{2}{1+\epsilon} \right)^k \gamma(1+a)$$

for any real number $a$, and let $a_A \stackrel{\text{def}}{=} b_A + b_{\neg A} - 1$ for any $A \subset \{1, \ldots, k\}$. Now $\mu = -(\min_A g(a_A))^{-1}$. Note that $g(a_A) > 1 + (1 + a_A)2^{-k}$, because $\gamma > 2^{-k}$ (this will be used later). It follows from the Binomial Theorem that the value of $b_A$ is very close to the value of $2^{-v_A}$ from above.

$$\begin{aligned} 2^{-v_A} \le b_A &\le (2^{-v_A} + 2^k\epsilon)(1+\epsilon)^k \\ &\le (2^{-v_A} + 2^k\epsilon)(1 + 2k\epsilon) \\ &\le 2^{-v_A} + 2^{k+2}\epsilon \, , \end{aligned}$$

where we have used the inequality $(1+r/c)^c \le 1+2r$ valid for $r \in [0,1]$ and positive integer $c$ (Mauá, de Campos, & Zaffalon, 2011, Lemma 37). Thus we conclude that the value $a_A$ is very close (again from above) to $h(v_A) - 1$.

$$h(v_A) - 1 \le a_A \le h(v_A) - 1 + 2^{k+3}\epsilon = h(v_A) - 1 + 1/(64z^4) \, .$$

Now, if the partition problem is a yes-instance, then $h(v_A) = 1$ (recall the behavior of $h$ from the proof of Theorem 3) and thus $a_A \leq 1/(64z^4)$, while if it is a no-instance, then $h(v_A) > 1 + 1/(32z^4)$ and thus $a_A > 1/(32z^4)$. Hence, there is a gap of at least $1/(64z^4)$ in the value of $a_A$ between yes- and no-instances, and we can decide the partition problem by verifying whether $\mu \leq -g(3/(128z^4))^{-1}$. This proof shall be completed with the guarantee that we can approximate in polynomial time the irrational numbers used to specify the credal tree and $g(a)$ well enough so that $-g(3/(128z^4))^{-1}$ falls exactly in the middle of the gap between the values of $\mu$ for yes- and no-instances (because $g$ is linear in $a$). First, note that

$$g\left(\frac{1}{32z^4}\right) - g\left(\frac{1}{64z^4}\right) = \frac{1}{64z^4}\left(\frac{2}{1+\epsilon}\right)^k \gamma\,,$$

which is greater than $2^{-k}/(64z^4)$ (since $\gamma > 2^{-k}$). The gap in the value of $\mu$ is at least

$$\begin{aligned}\frac{1}{g(1/(64z^4))} - \frac{1}{g(1/(32z^4))} &= \frac{g(1/(32z^4)) - g(1/(64z^4))}{g(1/(64z^4))g(1/(32z^4))} \\ &> \frac{g(1/(32z^4)) - g(1/(64z^4))}{g(1/(32z^4))^2} \\ &> \frac{2^{-k}/(64z^4)}{(1+(1+\frac{1}{32z^4})2^{-k})^2} > \frac{2^{-k}}{4\cdot 64z^4}\,.\end{aligned}$$

So we apply Corollary 1 with $\varepsilon = \frac{1}{2}\frac{2^{-k}}{4\cdot 64z^4}$ to obtain from $\mathcal{N}$ a network $\mathcal{N}'$ made only of positive rational numbers. Such $\varepsilon$ guarantees that the separation between yes-instances and no-instances of PARTITION will continue to exist. □

The credal network used in the reduction that proves the previous result is in a sense the simplest tree-shaped network on which solving STRONG-INF is hard, since the problem would be polynomial-time solvable if the root node were replaced with a binary variable. It is also interesting as it describes a naive Bayes structure with a single layer of latent variables, a useful topology for robust classification problems on non-linearly separable feature spaces.

### 3.7 Imprecise Hidden Markov Models

An imprecise hidden Markov model (HMM) is a credal tree whose nodes can be partitioned into *hidden* and *manifest* nodes such that the hidden nodes form a chain (i.e., a sequence of nodes with one node linking to the next and to no other in the sequence), and manifest nodes are leaves of the graph. HMMs are widely used to represent discrete dynamic systems whose output at any given time step can be stochastically determined by the current state of the system, which is assumed to be only partially observable.

Since an HMM is simply a credal tree, the algorithm of de Cooman et al. (2010) can be used to efficiently compute EPISTEMIC-INF in HMMs, while 2U can be used to solve STRONG-INF if all variables are binary. For networks with variables taking on more than two values, no polynomial-time is known for STRONG-INF. In this section, we show that when there is no evidence on variables farther (in the sense of number of nodes in the path) from the root node than the target variable, the outcomes of the STRONG-INF and EPISTEMIC-INF coincide. On these cases, we can run de Cooman et al.'s (2010) algorithm

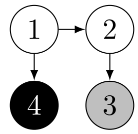


Figure 6: Credal HMM in Example 5.

to compute STRONG-INF in polynomial time. This is however not always true, that is, there are types of queries in which the results of STRONG-INF and EPISTEMIC-INF differ, as the following example shows.

**Example 5.** *Consider an HMM of length two whose topology is depicted in Figure 6. All variables are binary and take values in $\{0,1\}$. Variables $X_1$ and $X_2$ are hidden, while variables $X_3$ and $X_4$ are manifest. The local credal sets are given by $Q(X_1) = Q(X_2|0) = Q(X_4|0) = \{p \in V(X_4) : p(1) = 1/4\}$, $Q(X_2|1) = Q(X_4|1) = \{p \in V(X_4) : p(1) = 3/4\}$, and $Q(X_3|0) = \{p \in V(X_3) : 1/2 \leq p(1) \leq 3/4\}$ and $Q(X_3|1) = \{p \in V(X_3) : 1/4 \leq p(1) \leq 1/2\}$. Thus, variable $X_3$ is imprecise, and the remaining variables are precise. Consider a query with target $X_4 = 0$ and evidence $X_3 = 0$. The lower bound of STRONG-INF is the value of $\mu$ that solves the equation*

$$\min \sum_{x_2} q_3^{x_2}(0) g_\mu(x_2) = \sum_{x_2} \min q_3^{x_2}(0) g_\mu(x_2) = 0\,,$$

*where the minimizations are performed over $q_3^{x_2} \in Q(X_3|x_2)$, $x_2 = 0, 1$, and*

$$g_\mu(x_2) \overset{def}{=} \sum_{x_1,x_4} (\delta_0(x_4) - \mu)\, q_1(x_1) q_2^{x_1}(x_2) q_4^{x_1}(x_4)\,,$$

*with $q_1 = q_2^0 = q_4^0 = (3/4, 1/4)$ and $q_2^1 = q_4^1 = (1/4, 3/4)$. The values of $q_3^{x_2}(0)$ depend only on the signs of $g_\mu(0)$ and $g_\mu(1)$, which ought to be different for the expression to vanish. Solving for $\mu$ for each of the four possibilities, and taking the minimum value of $\mu$, we find that $\mu = 4/7 > 1/2$.*

*The lower bound of EPISTEMIC-INF is the value of $\mu$ that solves*

$$\min \sum_{x_1,x_2,x_4} q_1(x_1) q_2^{x_1}(x_2) q_4^{x_1}(x_4) q_{x_1,x_2,x_4}(0) h_\mu(x_4) =$$
$$(1-\mu) \sum_{x_1,x_2} q_1(x_1) q_2^{x_1}(x_2) q_4^{x_1}(0) \min q_{x_1,x_2,0}(0)$$
$$- \mu \sum_{x_1,x_2} q_1(x_1) q_2^{x_1}(x_2) q_4^{x_1}(1) \max q_{x_1,x_2,1}(0) = 0\,,$$

*where $h_\mu(x_4) = \delta_0(x_4) - \mu$, $q_1$, $q_2^{x_1}$ and $q_4^{x_1}$ are defined as before, and $q_{x_1,x_2,x_4} \in Q(X_3|x_2)$ for every $x_1, x_2, x_4$. Solving the equation above for $\mu$ we obtain $\mu = 13/28 < 1/2$.* □

The above example shows that STRONG-INF and EPISTEMIC-INF might differ, even in the simple case of HMMs with binary variables. It is currently unknown whether this type of inference is hard for STRONG-INF. The following result shows that at least for a particular case, the computations of the STRONG-INF and EPISTEMIC-INF in HMMs coincide.

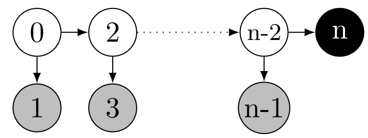



Figure 7: DAG of the HMM considered in Theorem 3.

**Theorem 3.** *Consider a separately specified HMM over variables $X_0, \ldots, X_n$. The variables associated with odd numbers are manifest, and the remaining variables are hidden (see Figure 7). Consider also the target hidden node $X_n = \tilde{x}_n$, and evidence $\mathbf{X}_O = \tilde{\mathbf{x}}_O$ on a subset $O$ of the manifest nodes. Then the outcomes of* STRONG-INF *and* EPISTEMIC-INF *are the same.*

*Proof.* Define $f_\mu(x_n) \overset{\text{def}}{=} \delta_{\tilde{x}_n}(x_n) - \mu$ for any given $\mu$, and consider the distribution $p$ in the epistemic extension $K_E$ that minimizes

$$\sum_{\mathbf{x}\sim\mathbf{X}:\mathbf{x}_O=\tilde{\mathbf{x}}_O} f_\mu(x_n)p(\mathbf{x})\,.$$

Let $<$ be any topological ordering of the nodes. By the Chain Rule, we have for all $\mathbf{x}$ that $p$ factorizes as $p(\mathbf{x}) = \mathrm{P}_p(x_0)\prod_{i=1}^n \mathrm{P}_p(x_i|\mathbf{x}_{<i})$, where $\mathbf{x}_{<i}$ denotes the coordinates $x_j$ of $\mathbf{x}$ with $j < i$ according to the topological ordering (we also write $\mathbf{x}_{\geq i}$ and $\mathbf{x}_{>i}$ to denote analogous projections). Assume that for some non-negative integer $i$ less than or equal to $n$ it holds that

$$\sum_{\mathbf{x}:\mathbf{x}_O=\tilde{\mathbf{x}}_O} f_\mu(x_n)p(\mathbf{x}) \geq \sum_{\mathbf{x}:\mathbf{x}_O=\tilde{\mathbf{x}}_O} f_\mu(x_n)\prod_{j<i}\mathrm{P}_p(x_j|\mathbf{x}_{<j})\prod_{j\geq i} p_j^{\mathbf{x}_{\mathrm{Pa}(j)}}(x_j)\,,$$

where each $p_j^{\mathbf{x}_{\mathrm{Pa}(j)}}$ is recursively defined as the extreme distribution of the local credal set $Q(X_j|\mathbf{x}_{\mathrm{Pa}(j)})$ that minimizes either

$$\sum_{x_j} p_j^{\mathbf{x}_{\mathrm{Pa}(j)}}(x_j)\sum_{\mathbf{x}_{>j}} f_\mu(x_n)\prod_{k>j} p_k^{\mathbf{x}_{\mathrm{Pa}(k)}}(x_k)\,,$$

if $j$ is not in $O$, or

$$p_j^{x_{\mathrm{Pa}(j)}}(\tilde{x}_j)\sum_{\mathbf{x}_{>j}} f_\mu(x_n)\prod_{k>j} p_k^{\mathbf{x}_{\mathrm{Pa}(k)}}(x_k)\,,$$

if $j$ is in $O$, where $\tilde{x}_j$ is the value of $X_j$ compatible with $\tilde{\mathbf{x}}_O$. We will show by induction in $i = n, \ldots, 0$ that the assumption is true. If $i-1$ is not in $O$ then

$$\begin{aligned}
&\sum_{\mathbf{x}:\mathbf{x}_O=\tilde{\mathbf{x}}_O} f_\mu(x_n)\prod_{j<i}\mathrm{P}_p(x_j|\mathbf{x}_{<j})\prod_{j\geq i} p_j^{\mathbf{x}_{\mathrm{Pa}(j)}}(x_j) = \\
&\quad\sum_{\mathbf{x}_{<i-1}:\mathbf{x}_O=\tilde{\mathbf{x}}_O}\prod_{j<i-1}\mathrm{P}_p(x_j|\mathbf{x}_{<j})\sum_{x_{i-1}}\mathrm{P}_p(x_{i-1}|\mathbf{x}_{<i-1})\sum_{\mathbf{x}_{\geq i}} f_\mu(x_n)\prod_{k\geq i} p_k^{\mathbf{x}_{\mathrm{Pa}(k)}}(x_k) \geq \\
&\quad\sum_{\mathbf{x}_{<i-1}:\mathbf{x}_O=\tilde{\mathbf{x}}_O}\prod_{j<i-1}\mathrm{P}_p(x_j|\mathbf{x}_{<j})\min_{q\in Q(X_{i-1}|\mathbf{x}_{\mathrm{Pa}(i-1)})}\sum_{x_{i-1}} q(x_{i-1})\sum_{\mathbf{x}_{\geq i}} f_\mu(x_n)\prod_{k>i} p_k^{\mathbf{x}_{\mathrm{Pa}(k)}}(x_k) = \\
&\qquad\qquad\sum_{\mathbf{x}:\mathbf{x}_O=\tilde{\mathbf{x}}_O} f_\mu(x_n)\prod_{j<i-1}\mathrm{P}_p(x_j|\mathbf{x}_{<j})\prod_{j\geq i-1} p_j^{\mathbf{x}_{\mathrm{Pa}(j)}}(x_j)\,,
\end{aligned}$$

where the inequality follows from the definition of epistemic extension, which implies that $\sum_{x_i} h(x_i)\mathrm{P}_p(x_i|\mathbf{x}_{\mathrm{Nd}(i)}) \geq \min_{q\in Q(X_i|x_{\mathrm{Pa}(i)})} \sum_{x_i} h(x_i)q(x_i)$ for any function $h$ on $x_i$ (note that $\mathrm{Nd}(i) \supseteq \{j < i\}$, and that the minimization on the right is constant w.r.t. values $\mathbf{x}_{j<i:j\notin\mathrm{Pa}(i)}$). The case of a node $i$ in $O$ is analogous with the sum substituted by a single term. For $i = n$, it follows that

$$\begin{aligned}\sum_{\mathbf{x}:\mathbf{x}_O=\tilde{\mathbf{x}}_O} f_\mu(x_n)p(\mathbf{x}) &= \sum_{\mathbf{x}_{j<n}:\mathbf{x}_O=\tilde{\mathbf{x}}_O} \prod_{j\leq n} \mathrm{P}_p(x_j|\mathbf{x}_{<j}) \sum_{x_t} f_\mu(x_n)\mathrm{P}_p(x_t|\mathbf{x}_{<n}) \\ &\geq \sum_{\mathbf{x}:\mathbf{x}_O=\tilde{\mathbf{x}}_O} f_\mu(x_n)p_n^{x_{n-2}}(x_n) \prod_{j<n} \mathrm{P}_p(x_j|\mathbf{x}_{<j})\,,\end{aligned}$$

so that the basis of the induction holds. For $i = 0$, we have that

$$\sum_{\mathbf{x}:\mathbf{x}_O=\tilde{\mathbf{x}}_O} f_\mu(x_n)p(\mathbf{x}) \geq \sum_{\mathbf{x}:\mathbf{x}_O=\tilde{\mathbf{x}}_O} f_\mu(x_n)p_0(x_0) \prod_{i\in N} p_j^{x_{\mathrm{Pa}(j)}}(x_j)\,,$$

which is the lower bound of STRONG-INF. Thus, since the epistemic extension contains the strong extension, the inequality above is tight. In particular, the equality holds if $\mu$ is the lower bound of EPISTEMIC-INF, and it follows that

$$\min_{p\in K_S} \sum_{\mathbf{x}:\mathbf{x}_O=\tilde{\mathbf{x}}_O} f_\mu(x)p(\mathbf{x}) = \min_{p\in K_E} \sum_{\mathbf{x}:\mathbf{x}_O=\tilde{\mathbf{x}}_O} f_\mu(x)p(\mathbf{x}) = 0\,,$$

where $K_S$ denotes the strong extension. An analogous proof shows that also the upper bounds coincide. □

The previous result shows that at least for the particular case where one seeks the probability of the "last" variable, STRONG-INF can be computed in polynomial time. Although restrictive, this type of inference is highly relevant, as it corresponds to predicting the future state of a partially observable dynamic system whose future state depends in some level only on its current (unknown) state. There is also another type of inference in trees which is insensitive to the irrelevance concept adopted, which is the case of marginal inferences:

**Corollary 4.** *Consider a tree-shaped network $\mathcal{N}$ and a target $X_t = x_t$. Then*

$$\textit{STRONG-INF}(\mathcal{N}, t, x_t, \emptyset, \cdot) = \textit{EPISTEMIC-INF}(\mathcal{N}, t, x_t, \emptyset, \cdot)\,.$$

*Proof.* We say that a node is *barren* if it is not an ancestor of any target or evidence node. It is well-known that removing barren nodes from a Bayesian network does not affect the outcome of BN-INF (Koller & Friedman, 2009). Since inference under strong independence can be seen as (exponentially many) inferences in Bayesian networks, the result of STRONG-INF is also unaltered if remove barren nodes. Moreover, since $\mathcal{N}$ is a tree, removing barren nodes leaves with a chain of ancestors of $t$. According to Theorem 15 of the work of Cozman (2000), the epistemic extension of $\mathcal{N}$ projected on the ancestors of $t$, that is, the set of marginal distributions on $\mathbf{x}_A \sim \mathbf{X}_A$ induced from joint distributions in the epistemic extension, where $A$ denotes the ancestors of $t$, is the epistemic extension of the network we get by removing nodes not in $A$. This implies that barren nodes can be discarded also in inferences under epistemic irrelevance, and the result follows. □

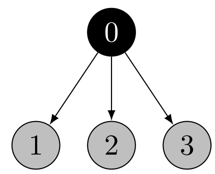



Figure 8: DAG of a naive Bayes with 3 feature variables.

Zaffalon and Fagiuoli (2003) developed a linear-time algorithm to compute marginal inferences under strong independence in trees as a by-product of their work on imprecise tree-augmented naive Bayes classifiers. The result above shows that the same algorithm can be used to compute marginal inferences in trees under epistemic irrelevance; conversely, de Cooman et al.'s (2010) algorithm for epistemic trees can be used to compute marginal inferences in strong trees.

### 3.8 Imprecise Markov Chains

The simplest DAG structure forming a connected graph is that of a chain, that is, of a network in which each variable has at most one parent and one child. Credal chains are more usually known as (imprecise) Markov chains. As a chain is also a tree, computing EPISTEMIC-INF can be done in polynomial time; this is also the case for STRONG-INF on chains of binary variables, as this is a subcase of binary polytrees. A chain can be seen as an HMM where the values of the manifest variables are deterministically determined by the values of the hidden variables. As such, the equivalence of both types of inference in certain types of HMMs extends to chains:

**Corollary 5.** *Consider a credal chain $X_0 \to \cdots \to X_n$, a target $X_n = x_n$ on the single leaf variable of a separately specified (imprecise) Markov chain, and some evidence $\mathbf{X}_O = \tilde{\mathbf{x}}_O$ on arbitrary non-leaf variables. Then the outcomes of STRONG-INF and EPISTEMIC-INF coincide.*

*Proof.* The same proof of Theorem 3 applies here, if we omit manifest nodes. □

### 3.9 Imprecise Naive Bayes

A widely used DAG structure is the naive Bayes, where a node (usually called *class*) has all other nodes (called *features*) as its children, and no other arc is present. Figure 8 depicts a naive Bayes structure with class variable $X_0$ and features $X_1$, $X_2$ and $X_3$. Such a DAG constitutes the structure behind the Naive Bayes and the Naive Credal Classifiers (Zaffalon, 2002). As it is a tree, computing EPISTEMIC-INF can be done in polynomial time; this is also the case for STRONG-INF when the target node is the class variable (Zaffalon, 2002). We show next that this similar tractability is not coincidental: both inferences yield the same result, even if the target is not the class node. We achieve such a result by building an HMM where the first hidden variable is the class and all other hidden variables have the same state space as the class and are deterministically determined by its value, while the manifest variables are the features in the naive Bayes structure. As such, the equivalence of inferences under both types of irrelevance extends to queries in any node of a naive Bayes structure:

| MODEL | STRONG-INF | EPISTEMIC-INF |
|---|---|---|
| *Naive Bayes | P | P |
| *Imprecise HMM (query on last node) | P | P |
| Imprecise HMMs | Unknown | P |
| *Credal trees (no evidence) | P | P |
| Credal trees | NP-hard | P |
| Credal polytrees with binary variables | P | Unknown |
| Credal polytrees with ternary variables | NP-hard | NP-hard |
| Bounded treewidth networks | NP-hard | NP-hard |
| Credal networks | $\mathrm{NP}^{\mathrm{PP}}$-hard | $\mathrm{NP}^{\mathrm{PP}}$-hard |
| *Precise-vacuous | $\mathrm{NP}^{\mathrm{PP}}$-hard | $\mathrm{NP}^{\mathrm{PP}}$-hard |

Table 2: Parametrized complexity of the inference in credal networks.

**Corollary 6.** *Consider a credal network $\mathcal{N}$ with the naive Bayes structure $X_0 \to X_1, X_0 \to X_2, \cdots, X_0 \to X_n$, a target $X_t = x_t$ on a node $t$, and some evidence $\mathbf{X}_O = \tilde{\mathbf{x}}_O$ on arbitrary features (leaf variables). Then the outcomes of STRONG-INF and EPISTEMIC-INF coincide.*

*Proof.* Let $X'_0 = X_0$ and $X'_1, \cdots, X'_n$ be precise variables with the same state space as the class $X'_0$ and probability distributions $q(x'_i|x'_{i-1}) = 1$ if $x'_i = x'_{i-1}$ and zero otherwise, for all $i = 1, \ldots, n$. Define $Q(X_i|X'_i = x'_i)$ by using the credal set $Q(X_i|X_0 = x_0)$ of the original network $\mathcal{N}$, whenever $x'_i = x_0$, for $i = 1, \ldots, n$. Without loss of generality, assume that $t = n$ (if the query was in $X_0$, then use $X'_n$ as query instead of $X_n$). This procedure creates an imprecise HMM with hidden nodes $X'_0, \ldots, X'_n$, manifest nodes $X_1, \ldots, X_{n-1}$, and a final query $t = n$ with $X_t = x_t$. This HMM clearly yields the same inferential result as does the naive Bayes network $\mathcal{N}$ for STRONG-INF. By Theorem 3, the results of STRONG-INF and EPISTEMIC-INF coincide in this HMM, hence the result of EPISTEMIC-INF in this HMM is equal to the result of STRONG-INF in $\mathcal{N}$. By construction, EPISTEMIC-INF in this HMM contains the result of EPISTEMIC-INF in $\mathcal{N}$ (that is, the latter is equal or lies inside the former). Because EPISTEMIC-INF always contains STRONG-INF, and in particular in the naive Bayes structure, they must all coincide. ☐

### 3.10 Summary of the Complexity Results

The complexity results obtained in this section suggest that inference in credal networks is computationally difficult for a wide variety of model structures and dimensionalities. This is the case, for instance, in precise-vacuous networks and singly connected networks with ternary variables, according to the negative results we have shown. A few importation exceptions have been obtained including last-node inference (filtering) in imprecise HMMs and Markov chains and inference in naive Bayes structures. These few positive results are important as such structures have applications in pattern recognition tasks such as activity recognition (Antonucci et al., 2011) and robust classification (Zaffalon et al., 2003). The previously known and new inferential complexity results are summarized in Table 2. The star indicates models in which inferences under both irrelevance concepts coincide.

Yet another irrelevance concept adopted in imprecisely specified models is Kuznetsov independence. We can define the Kuznetsov extension analogously to the definitions of strong and epistemic extensions, and define the problem of inference under Kuznetsov independence accordingly. It is known that the Kuznetsov extension lies between the epistemic and the strong extensions (Cozman & de Campos, 2014). This implies that the outcomes of inferences under Kuznetsov independence coincide with those under strong independence and epistemic irrelevance whenever the last two coincide. We hence get as corollaries of the results shown here that inference under Kuznetsov independence is $\text{NP}^{\text{PP}}$-hard in precise-vacuous networks, NP-hard in singly connected networks with ternary variables, polynomial-time computable in HMMs and Markov chains if the target variable is the last-node, and polynomial-time computable in naive Bayes structures if the target variable is the root node.

## 4. Conclusion

Credal networks generalize Bayesian networks to allow for the representation of uncertain knowledge in the form of credal sets, closed and convex sets of probability distributions. The use of credal sets arguably facilitates the constructions of complex models, but presents a challenge to the computation of inferences with the model.

In this paper we studied the theoretical complexity of inferences in credal networks, in what concerns the topology of the network, the semantics of the arcs (i.e., whether epistemic irrelevance or strong independence is assumed), and the cardinality of variable domains. In a nutshell, computing with credal networks is NP-hard except in the cases of tree-shaped models under epistemic irrelevance, and binary polytree-shaped models under strong independence. A notable exception is the computation of probability bounds on the value of the last variable in a imprecise hidden Markov models, in which case we have shown that inferences under epistemic irrelevance and strong independence coincide, which implies that the latter is polynomial-time computable. We leave as an open question the complexity of generic inferences in imprecise HMMs under strong independence.

Another possible avenue for future research is investigating the complexity of approximate inference. De Campos and Cozman (2005) showed that approximating inference under strong independence is NP-hard, even if we consider only singly connected networks of bounded treewidth. This is however not the case if variables are binary, as in this case we can run the 2U algorithm to obtain the exact value. Mauá, de Campos, and Zaffalon (2012b) showed that for any network of bounded treewidth whose variables have bounded cardinality there exists a fully polynomial time approximation scheme for performing inference under strong independence, that is, an algorithm that given a rational $\epsilon > 0$ finds solutions which are within a factor $1 + \epsilon$ of the true value in time polynomial in the input size and in $1/\epsilon$. Apart from its tractability on credal trees, nothing is known about the complexity of approximate inference under epistemic irrelevance, unless for the case of precise-vacuous networks, which we showed here to provide the same inferences under strong independence or epistemic irrelevance, so the results for approximate inference under the former extend to the latter.

## Acknowledgments

The first author received financial support from PNPD/CAPES and the São Paulo Research Foundation (FAPESP) grant no. 2013/23197-4. The second and third authors received financial support from the Swiss National Science Foundation grants no. 200021-146606/1 and no. 200020-137680/1, respectively. A shorter version of this paper appeared in (Mauá, de Campos, Benavoli, & Antonucci, 2013).

## References

Antonucci, A., Brühlmann, R., Piatti, A., & Zaffalon, M. (2009). Credal networks for military identification problems. *International Journal of Approximate Reasoning*, *50*, 666–679.

Antonucci, A., de Campos, C., & Zaffalon, M. (2014). Probabilistic graphical models. In Augustin, T., Coolen, F., de Cooman, G., & Troffaes, M. (Eds.), *Introduction to Imprecise Probabilities*, pp. 207–229. John Wiley & Sons.

Antonucci, A., de Rosa, R., & Giusti, A. (2011). Action recognition by imprecise hidden Markov models. In *Proceedings of the 2011 International Conference on Image Processing, Computer Vision and Pattern Recognition (IPCV)*, pp. 474–478.

Antonucci, A., Huber, D., Zaffalon, M., Luginbühl, P., Chapman, I., & Ladouceur, R. (2013). CREDO: a military decision-support system based on credal networks. In *Proceedings of the 16th Conference on Information Fusion (FUSION 2013)*.

Antonucci, A., Piatti, A., & Zaffalon, M. (2007). Credal networks for operational risk measurement and management. In *International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES)*, Vol. LNCS 4693, pp. 604–611.

Antonucci, A., & Zaffalon, M. (2008). Decision-theoretic specification of credal networks: a unified language for uncertain modeling with sets of Bayesian networks. *International Journal of Approximate Reasoning*, *49*(2), 345–361.

Boyd, S. P., & Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Cano, A., Cano, J. E., & Moral, S. (1994). Convex sets of probabilities propagation by simulated annealing. In *Proceedings of the Fith International Conference on Information Processing and Management of Uncertainty in Knowledge Based Systems (IPMU)*, pp. 4–8.

Corani, G., Giusti, A., Migliore, D., & Schmidhuber, J. (2010). Robust texture recognition using credal classifiers. In *Proceedings of the British Machine Vision Conference (BMVA)*, pp. 78.1–78.10.

Cowell, R., Dawid, P., Lauritzen, S., & Spiegelhalter, D. (2007). *Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks*. Statistics for Engineering and Information Science Series. Springer.

Cozman, F. G. (2000). Credal networks. *Artificial Intelligence*, *120*(2), 199–233.

Cozman, F. G. (2005). Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning*, *39*(2–3), 167–184.

Cozman, F. G., de Campos, C. P., Ide, J. S., & da Rocha, J. C. F. (2004). Propositional and relational Bayesian networks associated with imprecise and qualitative probabilistic assessments. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 104–111.

Cozman, F., & de Campos, C. (2014). Kuznetsov independence for interval-valued expectations and sets of probability distributions: Properties and algorithms. *International Journal of Approximate Reasoning*, *55*(2), 666–682.

de Bock, J., & de Cooman, G. (2013). Allowing for probability zero in credal networks under epistemic irrelevance. In *Proceedings of the 8th International Symposium on Imprecise Probabilty: Theories and Applications (ISIPTA)*.

de Campos, C., & Ji, Q. (2011). Bayesian networks and the imprecise Dirichlet model applied to recognition problems. In Liu, W. (Ed.), *Symbolic and Quantitative Approaches to Reasoning With Uncertainty*, Vol. 6717 of *Lecture Notes in Computer Science*, pp. 158–169. Springer, Berlin / Heidelberg.

de Campos, C. P., & Cozman, F. G. (2005). The inferential complexity of Bayesian and credal networks. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1313–1318.

de Campos, C. P., & Cozman, F. G. (2013). Complexity of inferences in polytree-shaped semi-qualitative probabilistic networks. In *Proceedings of the 27th AAAI Conference on Advances in Artificial Intelligence*, pp. 217–223.

de Campos, C. P., & Ji, Q. (2008). Strategy selection in influence diagrams using imprecise probabilities. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 121–128.

de Campos, C. P., Zhang, L., Tong, Y., & Ji, Q. (2009). Semi-qualitative probabilistic networks in computer vision problems. *Journal of Statistical Theory and Practice*, *3*(1), 197–210.

de Campos, L., Huete, J., & Moral, S. (1994). Probability intervals: a tool for uncertain reasoning. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *2*, 167–196.

de Cooman, G., Hermans, F., Antonucci, A., & Zaffalon, M. (2010). Epistemic irrelevance in credal nets: the case of imprecise Markov trees. *International Journal of Approximate Reasoning*, *51*(9), 1029–1052.

de Cooman, G., & Miranda, E. (2012). Irrelevant and independent natural extension for sets of desirable gambles. *Journal of Artificial Intelligence Research*, *45*, 601–640.

de Cooman, G., & Troffaes, M. C. M. (2004). Coherent lower previsions in systems modelling: Products and aggregation rules. *Reliability Engineering & System Safety*, *85*(1–3), 113–134.

Fagiuoli, E., & Zaffalon, M. (1998). 2U: An exact interval propagation algorithm for polytrees with binary variables. *Artificial Intelligence*, *106*(1), 77–107.

Garey, M. R., & Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman.

Halpern, J. (2001). Conditional plausibility measures and Bayesian networks. *Journal of Artificial Intelligence Research*, *14*, 359–389.

Kalai, G., & Ziegler, G. N. M. (2000). *Polytopes: Combinatorics and Computation*. DMV Seminar. Birkhäuser Verlag.

Kendall, D. G. (1974). Foundations of a theory of random sets. In Harding, E., & Kendall, D. G. (Eds.), *Stochastic Geometry*, pp. 322–376. John Wiley & Sons.

Koller, D., & Friedman, N. (2009). *Probabilistic Graphical Models*. MIT press.

Kwisthout, J., & van der Gaag, L. C. (2008). The computational complexity of sensitivity analysis and parameter tuning. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 349–356.

Kwisthout, J. H. P., Bodlaender, H. L., & van der Gaag, L. C. (2010). The necessity of bounded treewidth for efficient inference in Bayesian networks. In *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI)*, pp. 237–242.

Levi, I. (1980). *The Enterprise of Knowledge*. MIT Press.

Mauá, D. D., de Campos, C. P., Benavoli, A., & Antonucci, A. (2013). On the complexity of strong and epistemic credal networks. In *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 391–400.

Mauá, D. D., de Campos, C. P., & Zaffalon, M. (2011). Solving limited memory influence diagrams. *CoRR*, *abs/1109.1754*.

Mauá, D. D., de Campos, C. P., & Zaffalon, M. (2012a). Solving limited memory influence diagrams. *Journal of Artificial Intelligence Research*, *44*, 97–140.

Mauá, D. D., de Campos, C. P., & Zaffalon, M. (2012b). Updating credal networks is approximable in polynomial time. *International Journal of Approximate Reasoning*, *53*(8), 1183–1199.

Mauá, D. D., de Campos, C. P., & Zaffalon, M. (2013). On the complexity of solving polytree-shaped limited memory influence diagrams with binary variables. *Artificial Intelligence*, *205*, 30–38.

Miranda, E., & Destercke, S. (2013). Extreme points of the credal sets generated by elementary comparative probabilities. In *Proceedings of the 12th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU)*, Vol. 7958, pp. 424–435.

Park, J. D., & Darwiche, A. (2004). Complexity results and approximation strategies for MAP explanations. *Journal of Artificial Intelligence Research*, *21*, 101–133.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

Piatti, A., Antonucci, A., & Zaffalon, M. (2010). *Building Knowledge-Based Systems by Credal Networks: A Tutorial*. Nova Science.

Roth, D. (1996). On the hardness of approximate reasoning. *Artificial Intelligence*, *82*(1–2), 273–302.

Salvetti, A., Antonucci, A., & Zaffalon, M. (2008). Spatially distributed identification of debris flow source areas by credal networks. In *Transactions of the 4th International Congress on Environmental Modelling and Software Integrating Sciences and Information Technology for Environmental Assessment and Decision Making (iEMSs)*, pp. 380–387.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press.

Shenoy, P. P., & Shafer, G. (1988). Axioms for probability and belief-function propagation. In *Proceedings of the Fourth Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 169–198.

Tessem, B. (1992). Interval probability propagation. *International Journal of Approximate Reasoning*, *7*(3–4), 95–120.

Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall.

Walley, P. (2000). Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, *24*(2–3), 125–148.

Wallner, A. (2007). Extreme points of coherent probabilities in finite spaces. *International Journal of Approximate Reasoning*, *44*(3), 339–357.

Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, *1*(1), 3–28.

Zaffalon, M. (2002). The naive credal classifier. *Journal of Statistical Planning and Inference*, *105*(1), 5–21.

Zaffalon, M. (2005). Credible classification for environmental problems. *Environmental modelling and software*, *20*(8), 1003–1012.

Zaffalon, M., & Fagiuoli, E. (2003). Tree-based credal networks for classification. *Reliable Computing*, *9*(6), 487–509.

Zaffalon, M., Wesnes, K., & Petrini, O. (2003). Reliable diagnoses of dementia by the naive credal classifier inferred from incomplete cognitive data. *Artificial Intelligence in Medicine*, *29*(1–2), 61–79.