Reliable Uncertain Evidence Modeling in Bayesian Networks by Credal Networks

Sabina Marchetti
Sapienza University of Rome
Rome (Italy)
sabina.marchetti@uniroma1.it

Alessandro Antonucci
IDSIA
Lugano (Switzerland)
alessandro@idsia.ch

Abstract

A reliable modeling of uncertain evidence in Bayesian networks based on a set-valued quantification is proposed. Both soft and virtual evidences are considered. We show that evidence propagation in this setup can be reduced to standard updating in an augmented *credal* network, equivalent to a set of consistent Bayesian networks. A characterization of the computational complexity for this task is derived together with an efficient exact procedure for a subclass of instances. In the case of multiple uncertain evidences over the same variable, the proposed procedure can provide a set-valued version of the geometric approach to opinion pooling.

Introduction

Knowledge-based systems are used in AI to model relations among the variables of interest for a particular task, and provide automatic decision support by inference algorithms. This can be achieved by joint probability mass functions. When a subset of variables is observed, *belief updating* is a typical inference task that propagates such (fully reliable) evidence. Whenever the observational process is unable to clearly report a single state for the observed variable, we refer to *uncertain* evidence. This might take the form of a *virtual* instance, described by the relative likelihoods for the possible observation of every state of a considered variable (Pearl 1988). Also, *soft* evidence (Valtorta, Kim, and Vomlel 2002) denotes any observational process returning a probabilistic assessment, whose propagation induces a *revision* of the original model (Jeffrey 1965). Bayesian networks are often used to specify joint probability mass functions implementing knowledge-based systems (Koller and Friedman 2009). Full, or *hard* (Valtorta, Kim, and Vomlel 2002), observation of a node corresponds to its *instantiation* in the network, followed by belief updating. Given virtual evidence on some variable, the observational process can be modeled *à la Pearl* in Bayesian networks: an auxiliary binary child of the variable is introduced, whose conditional mass functions are proportional to the likelihoods (Pearl 1988). Instantiation of the auxiliary node yields propagation of virtual evidence, and standard inference algorithms for Bayesian networks can be used (Koller and Friedman 2009). Something similar can be done with soft evidence, but the quantification of the auxiliary node should be based on additional inferences in the original network (Chan and Darwiche 2005).

In the above classical setup, sharp probabilistic estimates are assumed for the parameters modeling an uncertain observation. We propose instead a generalized set-valued quantification, with interval-valued likelihoods for virtual evidence and sets of marginal mass functions for soft evidence. This offers a more robust modeling of observational processes leading to uncertain evidence. To this purpose, we extend the transformations defined for the standard case to the set-valued case. The original Bayesian network is converted into a *credal network* (Cozman 2000), equivalent to a set of Bayesian networks consistent with the set-valued specification. We characterize the computational complexity of the credal modeling of uncertain evidence in Bayesian networks, and propose an efficient inference scheme for a special class of instances. The discussion is indeed specialized to *opinion pooling* and our techniques used to generalize geometric functionals to support set-valued opinions.

Related Work

Model revision based on uncertain evidence is a classical topic in AI. Entropy-based techniques for the absorption of uncertain evidence were proposed in the Bayesian networks literature (Valtorta, Kim, and Vomlel 2002; Peng, Zhang, and Pan 2010), as well as for the pooling of convex sets of probability mass functions (Adamčík 2014). Yet, this approach was proved to fail standard postulates for revision operators in generalized settings (Grove and Halpern 1997). Uncertain evidence absorption has been also considered in the framework of generalized knowledge representation and reasoning (Dubois 2008). The discussion was specialized to evidence theory (Zhou, Wang, and Qin 2014; Ma et al. 2011), although revision based on uncertain instances with graphical models becomes more problematic and does not give a direct extension of the Bayesian networks formalism (Simon, Weber, and Levrat 2007). Finally, credal networks have been considered in the model revision framework (da Rocha, Guimaraes, and de Campos 2008). Yet, these authors consider the effect of a sharp quantification of the observation in a previously specified credal network, while we consider the opposite situation of a Bayesian network for which *credal* uncertain evidence is provided.

Background

Bayesian and Credal Networks

Let X be any discrete variable. Notation x and Ω_X is used, respectively, for a generic value and for the finite set of possible values of X. If X is binary, we set $\Omega_X := \{x, \neg x\}$. We denote as $P(X)$ a probability mass function (PMF) and as $K(X)$ a *credal set* (CS), defined as a set of PMFs over Ω_X. We remove *inner* points from CSs, i.e. those which can be obtained as convex combinations of other points, and assume the CS finite after this operation. CS $K_0(X)$, whose convex hull includes all PMFs over Ω_X is called *vacuous*.

Given another variable Y, define a collection of conditional PMFs as $P(X|Y) := \{P(X|y)\}_{y\in\Omega_Y}$. $P(X|Y)$ is called *conditional probability table* (CPT). Similarly, a credal CPT (CCPT) is defined as $K(X|Y) := \{K(X|y)\}_{y\in\Omega_Y}$. An *extensive* CPT (ECPT) is a finite collection of CPTs. A CCPT can be converted into an equivalent ECPT by considering all the possible combinations from the elements of the CSs.

Given a joint variable $\boldsymbol{X} := \{X_0, X_1, \ldots, X_n\}$, a Bayesian network (BN) (Pearl 1988) serves as a compact way to specify a PMF over $\boldsymbol{X}$. A BN is represented by a directed acyclic graph $\mathcal{G}$, whose nodes are in one-to-one correspondence with the variables in $\boldsymbol{X}$, and a collection of CPTs $\{P(X_i|\Pi_i)\}_{i=0}^n$, where Π_i is the joint variable of the parents of X_i according to $\mathcal{G}$. Under the *Markov condition*, i.e. each variable is conditionally independent of its non-descendants non-parents given its parents, the joint PMF $P(\boldsymbol{X})$ factorizes as $P(\boldsymbol{x}) := \prod_{i=0}^n P(x_i|\pi_i)$, where the values of x_i and π_i are those consistent with $\boldsymbol{x}$, for each $\boldsymbol{x} \in \Omega_{\boldsymbol{X}} = \times_{i=0}^n \Omega_{X_i}$.

A credal network (CN) (Cozman 2000) is a BN whose CPTs are replaced by CCPTs (or ECPTs). A CN specifies a joint CS $K(\boldsymbol{X})$, obtained by considering all the joint PMFs induced by the BNs with CPTs in the corresponding CCPTs (or ECPTs).

The typical inference task in BNs is *updating*, defined as the computation of the posterior probabilities for a variable of interest given hard evidence about some other variables. Without loss of generality, let the variable of interest and the observation be, respectively, X_0 and $X_n = x_n$. Standard belief updating corresponds to:

$$P(x_0|x_n) = \frac{\sum_{x_1,\ldots,x_{n-1}} \prod_{i=0}^n P(x_i|\pi_i)}{\sum_{x_0,x_1,\ldots,x_{n-1}} \prod_{i=0}^n P(x_i|\pi_i)} \,. \tag{1}$$

Updating is NP-hard in general BNs (Cooper 1990), although efficient computations can be performed in polytrees (Pearl 1988) by message propagation routines (Koller and Friedman 2009).

CN updating is similarly intended as the computation of lower and upper bounds of the updated probability in Eq. (1) with respect to $K(\boldsymbol{X})$. Notation $\underline{P}(x_0|x_n)$ ($\overline{P}(x_0|x_n)$) is used to denote lower (upper) bounds. CN updating extends BN updating and it is therefore NP-hard (de Campos and Cozman 2005). Contrary to the standard setting, inference in generic polytrees is still NP-hard (Mauá et al. 2014), with the notable exception of those networks whose variables are all binary (Fagiuoli and Zaffalon 1998).

Virtual and Soft Evidence

Eq. (1) gives the updated beliefs about queried variable X_0. The underlying assumption is that X_n has been the subject of a fully reliable observational process, and its actual value is known to be x_n. This is not always realistic. Evidence might result from a process which is unreliable and only the likelihoods for the possible values of the observed variable may be assessed (e.g., the precision and the false discovery rate for a positive medical test). *Virtual evidence* (VE) (Pearl 1988) applies to such type of observation. Notation $\lambda_{X_n} := \{\lambda_{x_n}\}_{x_n\in\Omega_{X_n}}$ identifies a VE, λ_{x_n} being the likelihood of the observation provided $(X_n = x_n)$. Given VE, the analogous of Eq. (1) is:

$$P_{\lambda_{X_n}}(x_0) := \frac{\sum_{x_n} \lambda_{x_n} P(x_0, x_n)}{\sum_{x_n} \lambda_{x_n} P(x_n)} \,, \tag{2}$$

where the probabilities in the right-hand side are obtained by marginalization of the joint PMF of the BN. Eq. (2) can be equivalently obtained by augmenting the BN with auxiliary binary node D_{X_n} as a child of X_n. By specifying $P(d_{X_n}|x_n) := \lambda_{x_n}$ for each $x_n \in \Omega_{X_n}$, it is easy to check that $P(x_0|d_{X_n}) = P_{\lambda_{X_n}}(x_0)$, i.e. Eq. (2) can be reduced to a standard updating in an augmented BN.

The notion of *soft evidence* (SE) refers to a different situation, in which the observational process returns an elicitation $P'(X_n)$ for the marginal PMF of X_n. See (Ben Mrad et al. 2015) for a detailed discussion on the possible situations producing SE. If this is the case, $P'(X_n)$ is assumed to replace the original beliefs about X_n by *Jeffrey's updating* (Jeffrey 1965), i.e.

$$P'_{X_n}(x_0) := \sum_{x_n} P(x_0|x_n) \cdot P'(x_n) \,. \tag{3}$$

Eq. (3) for SE reduces to Eq. (1) whenever $P'(X_n)$ assigns all the probability mass to a single value in Ω_{X_n}. The same happens for VE in Eq. (2), when all the likelihoods are zero apart from the one corresponding to the observed value. Although SE and VE refer to epistemologically different informational settings, the following result provides means for a unified approach to their modeling.

Proposition 1 ((Chan and Darwiche 2005))**.** *Absorption of a SE $P'(X_n)$ as in Eq.* (3) *is equivalent to Eq.* (2) *with a VE specified as:*

$$\lambda_{x_n} \propto \frac{P'(x_n)}{P(x_n)} \,, \tag{4}$$

for each $x_n \in \Omega_{X_n}$.[1]

Vice versa, absorption of a VE λ_{X_n} as in Eq. (2) *is equivalent to Eq.* (3) *with a SE specified as:*

$$P'(x_n) := \frac{\lambda_{x_n} P(x_n)}{\sum_{x_n} \lambda_{x_n} P(x_n)} \,, \tag{5}$$

for each $x_n \in \Omega_{X_n}$.

[1] VE is defined as a collection of likelihoods, which in turn are defined up to a multiplicative positive constant. This clearly follows from Eq. (2). The relation in Eq. (4) is proportionality and not equality just to make all the likelihoods smaller or equal than one.

In the above setup for SE, states that are impossible in the original BN cannot be revised, i.e. if $P(x_n) = 0$ for some $x_n \in \Omega_{X_n}$, then also $P'(x_n) = 0$ and any value can be set for λ_{x_n}. *Vice versa*, according to Eq. (5), a zero likelihood in a VE renders impossible the corresponding state of the SE. Thus, at least a non-zero likelihood should be specified in a VE. All these issues are shown in the following example.

Example 1. *Let X denote the actual color of a traffic light with $\Omega_X := \{g, y, r\}$. Assume g (green) more probable than r (red), and y (yellow) impossible. Thus, for instance, $P(X) = [4/5, 0, 1/5]$. We eventually revise $P(X)$ by a SE $P'(X)$, which keeps yellow impossible and assigns the same probability to the two other states, i.e. $P'(X) = [1/2, 0, 1/2]$. Because of Eq.* (4)*, this can be equivalently achieved by a VE $\lambda_X \propto \{1, 1, 4\}$. Vice versa, because of Eq.* (5)*, a VE $\tilde{\lambda}_X \propto \{1, 1, 5\}$ induces an updated $P_{\tilde{\lambda}}(X) = [4/9, 0, 5/9]$. Such PMF coincides with $P(X|d_X)$ in a two-node BN, with D_X child of X, CPT $P(D_X|X)$ with $P(d_X|X) = [1/10, 1/10, 1/2]$ and marginal PMF $P(X)$ as in the original specification.*

Credal Uncertain Evidence

Credal Virtual Evidence

We propose *credal* VE (CVE) as a robust extension of sharp virtual observations. Notation Λ_{X_n} is used here for the intervals $\{\underline{\lambda}_{x_n}, \overline{\lambda}_{x_n}\}_{x_n \in \Omega_{X_n}}$. CVE updating is defined as the computation of the bounds of Eq. (2) with respect to all VEs λ_{X_n} consistent with the interval constraints in Λ_{X_n}. Notation $\underline{P}_{\Lambda_{X_n}}(x_0)$ and $\overline{P}_{\Lambda_{X_n}}(x_0)$ is used to denote these bounds. CVE absorption in BNs is done as follows.

Transformation 1. *Given a BN over $\boldsymbol{X}$ and a CVE Λ_{X_n}, add a binary child D_{X_n} of X_n and quantify its CCPT $K(D_{X_n}|X_n)$ with constraints $\underline{\lambda}_{x_n} \leq P(d_{X_n}|x_n) \leq \overline{\lambda}_{x_n}$.*[2] *A CN with a single credal node results.*

By Tr. 1, CVE updating in a BN is reduced to CN updating.[3]

Theorem 1. *Given a CVE in a BN, consider the CN returned by Tr. 1. Then:*

$$\underline{P}(x_0|d_{X_n}) = \underline{P}_{\Lambda_{X_n}}(x_0)\,, \tag{6}$$

and analogously for the upper bounds.

Standard VE can be used to model partially reliable sensors or tests, whose quantification is based on sensitivity and specificity data. Since these data are not always promptly/easily available (e.g., a pregnancy test whose failure can be only decided later), a CVE with interval likelihoods can be quantified by the *imprecise Dirichlet model*[4] (Bernard 2005) as in the following example.

[2]For binary B, constraint $l \leq P(b) \leq u$ defines a CS $K(B)$ with elements $P_1(B) := [l, 1-l]$ and $P_2(B) := [u, 1-u]$.

[3]See `https://arxiv.org/abs/1802.05639` for the proofs of the theorems.

[4]Given N observations of X, if $n(x)$ of them reports x, the lower bound of $P(x)$ for to the imprecise Dirichlet model is $\frac{n(x)}{N+s}$, and the upper bound $\frac{n(x)+s}{N+s}$, with s effective prior sample size.

Example 2. *The reference standard for diagnosis of anterior cruciate legament sprains is arthroscopy. In a trial, 40 patients coming in with acute knee pain are examined using the* Declan test *(Cleland 2005). Every patient also has an arthroscopy procedure for a definitive diagnosis. Results are TP=17 (Declan positive, arthroscopy positive), FP=3 (Declan positive, arthroscopy negative), FN=6 (Declan negative, arthroscopy positive) and TN=14 (Declan negative, arthroscopy negative). Patients visiting a clinic have prior sprain probability $P(x) = 0.2$. Given a positive Declan, the imprecise Dirichlet model (see Footnote 4) with $s = 1$ corresponds to CVE $\underline{\lambda}_x = 17/23+1$, $\overline{\lambda}_x = 17+1/23+1$, $\underline{\lambda}_{\neg x} = 3/17+1$, $\overline{\lambda}_{\neg x} = 3+1/17+1$. The bounds of the updated sprain probability with respect to the above constraints are $\underline{P}_{\Lambda_X}(x) = 1/3$, $\overline{P}_{\Lambda_X}(x) \simeq 0.53$. A VE with frequentist estimates would have produced instead $P_{\lambda_X} \simeq 0.51$.*

Credal Soft Evidence

Analogous to CVE, *credal soft evidence* (CSE) on X_n can be specified by any CS $K'(X_n)$. Accordingly, CSE updating computes the bounds spanned by the updating of all SEs based on PMFs consistent with the CS, i.e.

$$\underline{P}'_{X_n}(x_0) := \min_{P'(X_n) \in K'(X_n)} \sum_{x_n} P(x_0|x_n) \cdot P'(x_n)\,, \tag{7}$$

and analogously for the upper bound $\overline{P}'_{X_n}(x_0)$.

The *shadow* of a CS $K(X)$ is a CS $\hat{K}(X)$ obtained from all the PMFs $\hat{P}(X)$ such that, for each $x \in \Omega_X$:

$$\min_{P(X) \in K(X)} P(x) \leq \hat{P}(x) \leq \max_{P(X) \in K(X)} P(x)\,. \tag{8}$$

A CS coinciding with its shadow is called *shady*. It is a trivial exercise to check that CSs over binary variables are shady.[5]

The following result extends Pr. 1 to the imprecise framework.

Theorem 2. *Absorption of a CSE with shady $K'(X_n)$ is equivalent to that of CVE Λ_{X_n} such that:*

$$\underline{\lambda}_{x_n} \propto \frac{\underline{P}'(x_n)}{P(x_n)}\,, \tag{9}$$

where $\underline{P}'(x_n) := \min_{P'(X_n) \in K'(X_n)} P'(x_n)$ and analogously for the upper bound. Vice versa absorption of a CVE Λ_{X_n} is equivalent to that of a CSE such that:

$$\underline{P}'(x_n) = \frac{P(x_n)\underline{\lambda}_{x_n}}{P(x_n)\underline{\lambda}_{x_n} + \sum_{x'_n \neq x_n} P(x'_n)\overline{\lambda}_{x'_n}}\,, \tag{10}$$

and analogously with a swap between lower and upper likelihoods for the upper bound.

By Th. 1 and 2, CSE updating in a BN is reduced to standard updating in a CN. This represents a generalization to the credal case of Pr. 1. For CSEs with non-shady CSs, the procedure is slightly more involved, as detailed by the following result.

[5]Following (Campos, Huete, and Moral 1994), a shadow is just the set of probability intervals induced by a generic CS.

Proposition 2. *Given a CSE $K'(X_n) := \{P'_i(X_n)\}_{i=1}^k$ in a BN, add a binary child D_{X_n} of X_n quantified by an ECPT $\{P_i(D_{X_n}|X_n)\}_{i=1}^k$ such that $P_i(d_{X_n}|x_n) \propto \frac{P'_i(x_n)}{P(x_n)}$ for each $i = 1, \ldots, k$ and $x_n \in \Omega_{X_n}$. Then:*

$$\underline{P}'_{X_n}(x_0) = \underline{P}(x_0|d_{X_n})\,. \tag{11}$$

To clarify these results, consider the following example.

Example 3. *Consider the same setup as in Ex. 1. Let us revise the original PMF $P(X)$ by a CSE based on the shady CS $K'(X) := \{P'_1(X), P'_2(X)\}$, with $P'_1(X) := [0.6, 0, 0.4]$ and $P'_2(X) := [0.4, 0, 0.6]$. Th. 2 can be used to convert such CSE in a CVE $\Lambda_X := \{$2-3 : 1 : 8-12$\}$. Vice versa, the beliefs induced by CVE $\tilde{\Lambda}_X := \{$3-5 : 1 : 8-10$\}$ are $\underline{P}_{\tilde{\Lambda}_X}(g) = 3/5$, $\overline{P}_{\tilde{\Lambda}_X}(g) = 2/3$, $\underline{P}_{\tilde{\Lambda}_X}(y) = \underline{P}_{\tilde{\Lambda}_X}(y) = 0$, and $\underline{P}_{\tilde{\Lambda}_X}(r) = 1/3$, $\overline{P}_{\tilde{\Lambda}_X}(r) = 2/5$. These bounds may be equivalently obtained in a two-node CN with D_X child of X and CCPT $K(D_X|X)$ such that $P(d_X|X = g) \in [0.6, 1]$, $P(d_X|X = y) = 1$, and $P(d_X|X = r) \in [0.8, 1]$. Alternatively, following Pr. 2, absorption of $K'(X)$ can be achieved by a ECCPT with two CPTs.*

We point out that *conservative updating* (CU), a credal updating rule for reliable treatment of missing non-MAR data (De Cooman and Zaffalon 2004), falls as a special case in our formalism. CU is defined as:

$$\underline{P}'_{X_n}(x_0) = \min_{x_n \in \Omega_{X_n}} P(x_0|x_n)\,, \tag{12}$$

and represents the most conservative approach to belief revision. A *vacuous* CCPT is specified, with $[0, 1]$ intervals for each value, either i) by Tr. 1, given CVE whose likelihoods take any value between zero and one [6], or ii) by straightforward application of Th. 2, if a *vacuous* CSE $K'_0(X_n)$ is provided. The resulting ECPT with $|\Omega_X|$ CPTs[7] corresponds to the CU implementation in (Antonucci and Zaffalon 2008). Also, Eq. (7) reduces to Eq. (12), given vacuous CSE. We can similarly proceed in the case of *incomplete* observations, i.e. some values of X_n are recognized as impossible, but no information can be provided about the other ones. If this is the case, we just replace Ω_{X_n} with $\Omega'_{X_n} \subset \Omega_{X_n}$.

Credal Probability Kinematics

Given two joint PMFs $P(\boldsymbol{X})$ and $P'(\boldsymbol{X})$, we say that the latter comes from the first by *probability kinematics* (PK) on the (coarse) partition of $\Omega_{\boldsymbol{X}}$ induced by X_n if and only if $P'(\boldsymbol{x}|x_n) = P(\boldsymbol{x}|x_n)$ for each $\boldsymbol{x} \in \Omega_{\boldsymbol{X}}$ and $x_n \in \Omega_{X_n}$ (Diaconis and Zabell 1982; Chan and Darwiche 2005).[8] This is the underlying assumption in Eq. (3). If $P'(\boldsymbol{X})$ is replaced by a CS, PK is generalized as follows.

Definition 1. *Let $P(\boldsymbol{X})$ and $K'(\boldsymbol{X})$ be, respectively, a joint PMF and a joint CS. We say that $K'(\boldsymbol{X})$ comes from $P(\boldsymbol{X})$ by* credal probability kinematics *(CPK) on the partition of $\Omega_{\boldsymbol{X}}$ induced by X_n if and only if it holds $\underline{P}'(\boldsymbol{x}|x_n) = \overline{P}'(\boldsymbol{x}|x_n) = P(\boldsymbol{x}|x_n)$, for each $\boldsymbol{x} \in \Omega_{\boldsymbol{X}}$ and $x_n \in \Omega_{X_n}$.*

That is, any revision process based on (generalized) PK guarantees invariance of the relevance of x_n, for each $x_n \in \Omega_{X_n}$, to any other possible event in the model, say x_0. The following consistency result holds for CSEs.

Theorem 3. *Given a BN over $\boldsymbol{X}$ and a shady CSE $K'(X_n)$, convert the CSE into a CVE as in Th. 2 and transform the BN into a CN by Tr. 1. Let $K'(\boldsymbol{X}, D_{X_n})$ be the joint CS associated to the CN. Then, $K'(\boldsymbol{X}|d_{X_n})$ comes from $P(\boldsymbol{X})$ by CPK on the partition induced by X_n. Moreover $K'(X_n|d_{X_n})$ coincides with the marginal CS in the CN.*

Multiple Evidences

So far, we only considered the updating of a *single* CVE or CSE. We call *uncertain credal updating* (UCU) of a BN the general task of computing updated/revised beliefs in a BN with an arbitrary number of CSEs, CVEs, and hard evidences as well. Here, UCU is intended as iterated application of the procedures outlined above. See for instance (Dubois 2008), for a categorization of iterated belief revision problems and their assumptions. When coping with multiple VEs in a BN, it is sufficient to add the necessary auxiliary children to the observed variables and quantify the CPTs as described. We similarly proceed with multiple CVEs.

The procedure becomes less straightforward when coping with multiple SEs or CSEs, since quantification of each auxiliary child by Eq. (4) requires a preliminary inference step. As a consequence, iterated revision might be not invariant with respect to the revision process scheme (Wagner 2002).

Additionally, with CSEs, absorption of the first CSE transforms the BN into a CN, and successive absorption of other CSEs requires further extension of the procedure in Th. 2. We leave such an extension as future work, and here we just consider simultaneous absorption of all evidences. If this is the case, multiple CSEs can be converted in CVEs and the inferences required for the quantification of the auxiliary children is performed in the original BN.

Algorithmic and Complexity Issues

ApproxLP (Antonucci et al. 2014) is an algorithm for general CN updating based on linear programming. It provides an inner approximation of the updated intervals with the same complexity of a BN inference on the same graph. Roughly, CN updating is reduced by ApproxLP to a sequence of linear programming tasks. Each is obtained by iteratively fixing all the local models to single elements of the corresponding CSs, while leaving a *free* single variable. It follows the algorithm efficiently produces exact inferences whenever a CN has all local CSs made of a single element apart from one. This is the case of belief updating with a single CVE/CSE.

Complexity Issues

Since standard BN updating of polytrees can be performed efficiently, the same happens with VEs and/or SEs, as Tr. 1

[6] As VE likelihoods are defined up to a positive multiplicative constant, we can set any positive $\overline{\lambda}_{x_n}$ provided that $\underline{\lambda}_{x_n} = 0$.

[7] The induced ECPT contains all $2^{|\Omega_{X_n}|}$ combinations of zero and ones in the CPTs. Yet, only those having a single one in the row associated to d_{X_n} remains after the convex hull.

[8] Full consistency of P' with the evidence inducing the revision process is not explicitly required. A more stringent characterization of PK was proposed, among others, by (Wagner 2002)

does not affect the topology (nor the treewidth) of the original network. Similarly, with multiply connected models, BN updating is exponential in the treewidth, and the same happens with models augmented by VEs and/or SEs.

As already noticed, with CNs, binary polytrees can be updated efficiently, while updating ternary polytrees is already NP-hard. An important question is therefore whether or not a similar situation holds for UCU in BNs. The (positive) answer is provided by the two following results.

Proposition 3. *UCU of polytree-shaped binary BNs can be solved in polynomial time.*

The proof of this proposition is trivial and simply follows from the fact that the auxiliary nodes required to model CVE and/or CSE are binary (remember that CSs over binary variables are always shady). The CN solving the UCU is therefore a binary polytree that can be updated by the exact algorithm proposed in (Fagiuoli and Zaffalon 1998).

Theorem 4. *UCU of non-binary polytree-shaped BNs is NP-hard.*

The proof of this theorem is based on a reduction to the analogous result for CNs (Mauá et al. 2014). This already concerns models whose variables have no more than three states and treewidth equal to two. In these cases, approximate inferences can be efficiently computed by ApproxLP.

Credal Opinion Pooling

Consider the generalized case of $m \geq 1$ overlapping probabilistic instances on X_n. For each $j = 1, \ldots, m$, let $P'_j(X_n)$ denote the SE reported by the j-th source. Straightforward introduction of m auxiliary nodes as outlined above would suffer *confirmational* dynamics, analogous to the well-known issue with posterior probability estimates in the naive Bayes classifier (Rish 2001). This might likely yield inconsistent revised beliefs, i.e. $\tilde{P}'(X_n)$ falls outside the convex hull of $\{P'_j(X_n)\}_{j=1}^m$.

A most conservative approach to prevent such inconsistency adopts the convex hull of all the opinions (Stewart and Quintana 2017). In our formalism, this is just the CS $K'(X_n) := \{P'_j(X_n)\}_{j=1}^m$. Yet, consider any small $\epsilon > 0$, and assume $P'_1(x_n) = \epsilon$, $P'_2(x_n) = 1 - \epsilon$, and $P'_j(x_n) = p \in (\epsilon, 1-\epsilon)$ for each $j = 3, \ldots, m$. Despite the consensus of all remaining sources on sharp value p, the conservative approach above would yield $K'(X_n) \simeq K_0(X_n)$. To what extent should this be preferred to the *confirmational* case is an open question.

A compromise solution might be offered by the *geometric pooling operator* (or LogOp) (Bacharach 1975). Given a collection of positive weights $\{\alpha_j\}_{j=1}^m$, with $\sum_{j=1}^m \alpha_j = 1$, the LogOp functional produces the PMF $\tilde{P}'(X_n)$ such that:

$$\tilde{P}'(x_n) \propto \prod_{j=1}^m P'_j(x)^{\alpha_j}\,, \tag{13}$$

for each $x_n \in \Omega_{X_n}$. $\tilde{P}'(X_n)$ belongs to the convex hull of $\{P'_j(X_n)\}_{j=1}^m$ for any specification of the weights (Adamčík 2014). The overlapping SEs associated to the PMF in Eq. (13) can be equivalently modeled by a collection of m VEs defined as follows.

Transformation 2. *Consider a BN over $\boldsymbol{X}$ and a collection of SEs on X_n, $\{P'_j(X_n)\}_{j=1}^m$. For each $j = 1, \ldots, m$, augment the BN with binary child $D_{X_n}^{(j)}$ of X_n whose CPT is such that $P(d_{X_n}^{(j)}|x_n) \propto \left[\frac{P'(x_n)}{P(x_n)}\right]^{\alpha_j}$, with $\sum_{j=1}^m \alpha_j = 1$.*

The transformation is used for the following result.

Proposition 4. *Consider the same inputs as in Tr. 2. Then:*

$$\tilde{P}'_{X_n}(x_0) = P(x_0|d_{X_n}^{(1)}, \ldots, d_{X_n}^{(m)})\,, \tag{14}$$

where the probability on the left-hand side is obtained by the direct revision induced by $\tilde{P}'(X_n)$, while the probability on the right-hand side of Eq. (14) *has been computed in the BN returned by Tr. 2.*

The proof follows from the conditional independence of the auxiliary nodes given X_n. Also, note how our proposal simultaneously performs pooling and absorption of overlapping SEs.

Suppose m sources provide generalized CSEs about X_n, say $\{K'_j(X_n)\}_{j=1}^m$. Let $\tilde{K}'(X_n)$ denote the CS induced by LogOp as in Eq. (13), for each $P'_j(X_j) \in K'_j(X_n)$, $j = 1, \ldots, m$ (Adamčík 2014). We generalize Tr. 2 as follows:

Transformation 3. *Consider a BN over $\boldsymbol{X}$ and the collection of CSEs $\{K'_j(X_n)\}_{j=1}^m$. For each $j = 1, \ldots, m$, augment the BN with binary child $D_{X_n}^{(j)}$ of X_n, whose CCPT is such that $\underline{P}(d_{X_n}^{(j)}|x_n) \propto \left[\frac{\underline{P}'(x_n)}{P(x_n)}\right]^{\alpha_j}$ and $\overline{P}(d_{X_n}^{(j)}|x_n) \propto \left[\frac{\overline{P}'(x_n)}{P(x_n)}\right]^{\alpha_j}$.*

This transformation returns a CN. A result analogous to Pr. 4 can be derived.

Theorem 5. *Consider the same inputs as in Tr. 3. Then:*

$$\underline{\tilde{P}}'_{X_n}(x_0) = \underline{P}(x_0|d_{X_n}^{(1)}, \ldots, d_{X_n}^{(m)})\,, \tag{15}$$

where the lower probability on the left-hand side has been computed by absorption of the single CSE $\tilde{K}'(X_n)$ and the probability on the right-hand side has been computed in the CN returned by Tr. 3. The same relation also holds for the corresponding upper probabilities.

Conclusions

Credal, or set-valued, modeling of uncertain evidence has been proposed within the framework of Bayesian networks. Such procedure generalizes standard updating. More importantly, our proposal allows to reduce the task of absorption of uncertain evidence to standard updating in credal networks. Complexity results, specific inference schemes, and generalized pooling procedures have been also derived.

As a future work we intend to evaluate the proposed technique with knowledge-based decision-support systems based on Bayesian network to model unreliable observational processes. Moreover the proposed procedure should be extended to the framework of credal networks, thus reconciling the orthogonal viewpoints considered in this paper and in (da Rocha, Guimaraes, and de Campos 2008), and tackling the case of non-simultaneous updating.

References

[Adamčík 2014] Adamčík, M. 2014. The information geometry of Bregman divergences and some applications in multi-expert reasoning. *Entropy* 16(12):6338–6381.

[Antonucci and Zaffalon 2008] Antonucci, A., and 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.

[Antonucci et al. 2014] Antonucci, A.; de Campos, C.; Zaffalon, M.; and Huber, D. 2014. Approximate credal network updating by linear programming with applications to decision making. *International Journal of Approximate Reasoning* 58:25–38.

[Bacharach 1975] Bacharach, M. 1975. Group decisions in the face of differences of opinion. *Management Science* 22(2):182–191.

[Ben Mrad et al. 2015] Ben Mrad, A.; Delcroix, V.; Piechowiak, S.; Leicester, P.; and Abid, M. 2015. An explication of uncertain evidence in Bayesian networks: likelihood evidence and probabilistic evidence. *Applied Intelligence* 43(4):802–824.

[Bernard 2005] Bernard, J.-M. 2005. An introduction to the imprecise Dirichlet model for multinomial data. *International Journal of Approximate Reasoning* 39(2-3):123–150.

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

[Chan and Darwiche 2005] Chan, H., and Darwiche, A. 2005. On the revision of probabilistic beliefs using uncertain evidence. *Artificial Intelligence* 163(1):67–90.

[Cleland 2005] Cleland, J. 2005. *Orthopaedic Clinical Examination: An Evidence-Based Approach for Physical Therapists*. Saunders.

[Cooper 1990] Cooper, G. F. 1990. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence* 42:393–405.

[Cozman 2000] Cozman, F. G. 2000. Credal networks. *Artificial Intelligence* 120:199–233.

[da Rocha, Guimaraes, and de Campos 2008] da Rocha, J.; Guimaraes, A.; and de Campos, C. 2008. Dealing with soft evidence in credal networks. In *Proceedings of* Conferencia Latino-Americana de Informatica.

[de Campos and Cozman 2005] de Campos, C. P., and Cozman, F. G. 2005. The inferential complexity of Bayesian and credal networks. In *Proceedings of IJCAI '05*, 1313–1318.

[De Cooman and Zaffalon 2004] De Cooman, G., and Zaffalon, M. 2004. Updating beliefs with incomplete observations. *Artificial Intelligence* 159(1-2):75–125.

[Diaconis and Zabell 1982] Diaconis, P., and Zabell, S. L. 1982. Updating subjective probability. *Journal of the American Statistical Association* 77(380):822–830.

[Dubois 2008] Dubois, D. 2008. Three scenarios for the revision of epistemic states. *Journal of Logic and Computation* 18(5):721–738.

[Fagiuoli and Zaffalon 1998] Fagiuoli, E., and Zaffalon, M. 1998. 2U: an exact interval propagation algorithm for polytrees with binary variables. *Artificial Intelligence* 106:77–107.

[Grove and Halpern 1997] Grove, A. J., and Halpern, J. Y. 1997. Probability update: conditioning vs. cross-entropy. In *Proceedings of the Thirteenth conference on Uncertainty in Artificial Intelligence*, 208–214. Morgan Kaufmann Publishers Inc.

[Jeffrey 1965] Jeffrey, R. C. 1965. Ethics and the logic of decision. *The Journal of Philosophy* 62(19):528–539.

[Koller and Friedman 2009] Koller, D., and Friedman, N. 2009. *Probabilistic graphical models: principles and techniques*. MIT Press.

[Ma et al. 2011] Ma, J.; Liu, W.; Dubois, D.; and Prade, H. 2011. Bridging Jeffrey's rule, AGM revision and Dempster conditioning in the theory of evidence. *International Journal on Artificial Intelligence Tools* 20(04):691–720.

[Mauá et al. 2014] Mauá, D.; de Campos, C.; Benavoli, A.; and Antonucci, A. 2014. Probabilistic inference in credal networks: new complexity results. *Journal of Artificial Intelligence Research* 50:603–637.

[Pearl 1988] Pearl, J. 1988. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. San Mateo, California: Morgan Kaufmann.

[Peng, Zhang, and Pan 2010] Peng, Y.; Zhang, S.; and Pan, R. 2010. Bayesian network reasoning with uncertain evidences. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 18(05):539–564.

[Rish 2001] Rish, I. 2001. An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in Artificial Intelligence*, volume 3, 41–46. IBM New York.

[Simon, Weber, and Levrat 2007] Simon, C.; Weber, P.; and Levrat, E. 2007. Bayesian networks and evidence theory to model complex systems reliability. *Journal of Computers* 2(1):33–43.

[Stewart and Quintana 2017] Stewart, R. T., and Quintana, I. O. 2017. Probabilistic opinion pooling with imprecise probabilities. *Journal of Philosophical Logic* 1–29.

[Valtorta, Kim, and Vomlel 2002] Valtorta, M.; Kim, Y.-G.; and Vomlel, J. 2002. Soft evidential update for probabilistic multiagent systems. *International Journal of Approximate Reasoning* 29(1):71–106.

[Wagner 2002] Wagner, C. G. 2002. Probability kinematics and commutativity. *Philosophy of Science* 69(2):266–278.

[Zhou, Wang, and Qin 2014] Zhou, C.; Wang, M.; and Qin, B. 2014. Belief-kinematics Jeffrey's rules in the theory of evidence. In *Proceedings of UAI 2014*, 917–926.