Equivalence between Bayesian and Credal Nets on an Updating Problem *

Alessandro Antonucci and Marco Zaffalon

Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA)
Galleria 2, CH-6928 Manno (Lugano), Switzerland
{alessandro,zaffalon}@idsia.ch

We establish an intimate connection between Bayesian and credal nets. Bayesian nets are precise graphical models, credal nets extend Bayesian nets to imprecise probability. We focus on traditional belief updating with credal nets, and on the kind of belief updating that arises with Bayesian nets when the reason for the missingness of some of the unobserved variables in the net is unknown. We show that the two updating problems are formally the same.

1 Introduction

Imagine the following situation. You want to use a graphical model to formalize your uncertainty about a domain. You prefer precise probabilistic models and so you choose the *Bayesian network* (BN) formalism [5] (see Sect. 2.1). You take care to precisely specify the graph and all the conditional mass functions required. At this point you are done with the modelling phase, and start updating beliefs about a target variable conditional on the observation of some variables in the net. The remaining variables are not observed, i.e., they are *missing*. You know that some of the missing variables are simply *missing at random* (MAR), and so they can easily be dealt with by traditional approaches. Yet, there is a subset of missing variables for which you do not know the process originating the missingness.

This innocuous-looking detail is going to change the very nature of your model: while you think you are working with BNs, what you are actually using are credal networks. *Credal networks* (CNs, see Sect. 2.2) are graphical models that generalize Bayesian nets to sets of distributions [3], i.e., to *imprecise probability* [6].

The implicit passage from Bayesian to credal nets is based on two steps. First, the above conditions, together with relatively weak assumptions, give rise to a specific way to update beliefs called *conservative inference rule* (CIR,

* This research was partially supported by the Swiss NSF grant 200020-109295/1.

see Sect. 3) [7]. CIR is an imprecise-probability rule: it leads, in general, to imprecise posterior probabilities for the target variable, even if the original model is precise. The second step is done in this paper: we show the formal equivalence between CIR-based updating in BNs, and traditional credal-network updating (see Sect. 4 and App. A) based on the popular notion of *strong independence* [3].

CIR and CNs have been proposed with quite different motivations in the literature: CIR as an updating rule for the case of partial ignorance about the missingness (or incompleteness) process; CNs as a way to relax the strict modelling requirements imposed by precise graphical models. The main interest in our result is just the established connection between two such seemingly different worlds. But the result appears also to be a basis to use algorithms for CNs to solve CIR-based updating problems.

2 Bayesian and Credal Networks

In this section we review the basics of Bayesian networks and their extension to convex sets of probabilities, i.e., credal networks. Both the models are based on a collection of random variables $\mathbf{X}$, which take values in finite sets, and a directed acyclic graph $\mathcal{G}$, whose nodes are associated to the variables of $\mathbf{X}$.

For both models, we assume the *Markov condition* to make $\mathcal{G}$ represent probabilistic independence relations between the variables in $\mathbf{X}$: every variable is independent of its non-descendant non-parents conditional on its parents. What makes BNs and CNs different is a different notion of independence and a different characterization of the conditional mass functions for each variable given the possible values of the parents, which will be detailed next.

Regarding notation, for each $X_i \in \mathbf{X}$, Ω_{X_i} is the possibility space of X_i, x_i a generic element of Ω_{X_i}, $P(X_i)$ a mass function for X_i and $P(x_i)$ the probability that $X_i = x_i$. A similar notation with uppercase subscripts (e.g., X_E) denotes arrays (and sets) of variables in $\mathbf{X}$. The parents of X_i, according to $\mathcal{G}$, are denoted by Π_i and for each $\pi_i \in \Omega_{\Pi_i}$, $P(X_i|\pi_i)$ is the conditional mass function for X_i given the joint value π_i of the parents of X_i.

2.1 Bayesian Networks

In the case of Bayesian networks, the modelling phase involves specifying a conditional mass function $P(X_i|\pi_i)$ for each $X_i \in \mathbf{X}$ and $\pi_i \in \Omega_{\Pi_i}$; and the standard notion of probabilistic independence is assumed in the Markov condition. A BN can therefore be regarded as a joint probability mass function over $\mathbf{X} \equiv (X_1, \ldots, X_n)$, that factorizes as follows: $P(\mathbf{x}) = \prod_{i=1}^{n} P(x_i|\pi_i)$, for each $\mathbf{x} \in \Omega_{\mathbf{X}}$, because of the Markov condition. The updated belief about a queried variable X_q, given some evidence $X_E = x_E$, is:

$$P(x_q|x_E) = \frac{\sum_{x_M} \prod_{i=1}^{n} P(x_i|\pi_i)}{\sum_{x_M, x_q} \prod_{i=1}^{n} P(x_i|\pi_i)}, \tag{1}$$

where $X_M \equiv \mathbf{X} \setminus (\{X_q\} \cup X_E)$, the domains of the arguments of the sums are left implicit and the values of x_i and π_i are consistent with (x_q, x_M, x_E). Despite its hardness in general, Eq. (1) can be efficiently solved for polytree-shaped BNs with standard propagation schemes based on local computations and message propagation [5]. Similar techniques apply also for general topologies with increased computational time.

2.2 Credal Networks

CNs relax BNs by allowing for imprecise probability statements. There are many kinds of CNs. We stick to those consistent with the following:[1]

Definition 1. *Consider a finite set of BNs with the same graph $\mathcal{G}$, over the same variables $\mathbf{X}$, i.e., a pair $\langle \mathcal{G}, \mathbf{P}(\mathbf{X}) \rangle$, where $\mathbf{P}(\mathbf{X})$ is the array of the joint mass functions associated to the set of BNs. Define a* credal network *as the convex hull of such mass functions: i.e., $K(\mathbf{X}) \equiv \mathrm{CH}\{\tilde{P}(\mathbf{X})\}_{\tilde{P} \in \mathbf{P}}$.*

We define a *credal set* as the convex hull of a collection of mass functions over a vector of variables. In this paper we assume this collection to be finite; therefore a credal set can be geometrically regarded as a *polytope*. Such a credal set contains an infinite number of mass functions, but only a finite number of *extreme mass functions*: those corresponding to the *vertices* of the polytope. It is possible to show that inference based on a credal set is equivalent to that based only on its vertices [6]. Clearly $K(\mathbf{X})$ is a credal set over $\mathbf{X}$ [we similarly denote by $K(X)$ a credal set over X]. The vertices of $K(\mathbf{X})$ are generally a subset of the original set of BNs and the CN is said equivalent to this *finite set* of BNs.

Note that $K(\mathbf{X})$ in Def. 1 is not specified via local pieces of probabilistic information, and so we say that the corresponding CN is *globally specified*. When the construction of $K(\mathbf{X})$ emphasizes locality, we talk of *locally specified* CNs. We can specify CNs locally in two ways. In the first, each probability mass function $P(X_i|\pi_i)$ is defined to belong to a finite set of mass functions [whose convex hull $K(X_i|\pi_i)$ is a credal set by definition, which is said to be local]. We talk of *separately specified* credal nets in this case. In the second, the generic probability table $P(X_i|\Pi_i)$, i.e., a function of both X_i and Π_i, is defined to belong to a finite set of tables, denoted by $K(X_i|\Pi_i)$. In this case we talk of *extensive* specification. In both cases, the multiplicity of local mass functions or tables gives rise to a multiplicity of joint mass functions over $\mathbf{X}$ by simply taking all the combinations of the local pieces of knowledge. Such joint mass functions are just those making up the finite set of BNs in Def. 1.

Belief updating with CNs is defined as the computation of the posterior credal set for a queried variable X_q, conditionally on evidence about some other variables X_E:

[1] By Def. 1 we are implicitly assuming the notion of *strong independence* in the Markov condition for CNs, see [3]. $K(\mathbf{X})$ is usually called the *strong extension*.

$$K(X_q|x_E) \equiv \mathrm{CH}\left\{\tilde{P}(X_q|x_E)\right\}_{\tilde{P}\in\mathbf{P}} . \tag{2}$$

3 Conservative Inference Rule

The most popular approach to missing data in the literature and in the statistical practice is based on the so-called *missing-at-random* assumption. MAR allows missing data to be neglected, thus turning the incomplete data problem into one of complete data. Unfortunately, MAR embodies the idea that the process responsible for the missingness (i.e., the *missingness process*) is not selective, which is not realistic in many cases. De Cooman and Zaffalon have developed an inference rule based on much weaker assumptions than MAR, which deals with near-ignorance about the missingness process [4]. This result has been expanded by Zaffalon [7] to the case of mixed knowledge about the missingness process: for some variables the process is assumed to be nearly unknown, while it is assumed to be MAR for the others. The resulting updating rule is called *conservative inference rule* (CIR).

To show how CIR-based updating works, we partition the variables in $\mathbf{X}$ in four classes: (i) the queried variable X_q, (ii) the observed variables X_E, (iii) the unobserved MAR variables X_M, and (iv) the variables X_I made missing by a process that we basically ignore. CIR leads to the following credal set as our updated beliefs about the queried variable:

$$K(X_q||^{X_I} x_E) \equiv \mathrm{CH}\left\{P(X_q|x_E, x_I)\right\}_{x_I\in\Omega_{X_I}} , \tag{3}$$

where the superscript on the double conditioning bar is used to denote beliefs updated with CIR and to specify the set of missing variables X_I assumed to be non-MAR, and clearly $P(X_q|x_E, x_I) = \sum_{x_M} P(X_q, x_M|x_E, x_I)$.

4 Equivalence between CIR-based Updating in Bayesian Nets and Credal Nets Updating

In this section we prove the formal equivalence between updating with CIR on BNs and standard updating on CNs, defining two distinct mappings from a generic instance of the first problem in a corresponding instance of the second (see Sect. 4.1) and *vice versa* (see Sect. 4.2). Fig. 1 reports the correspondence scheme with the names of the mappings that will be introduced next. According to CIR assumptions [7], we focus on the case of BNs assigning positive probability to each event.

4.1 From Bayesian to Credal Networks

First let us define the B2C transformation, mapping a BN $\langle\mathcal{G}, P(\mathbf{X})\rangle$, where a subset X_I of $\mathbf{X}$ is specified, in a CN. For each variable $X \in X_I$, B2C prescribes

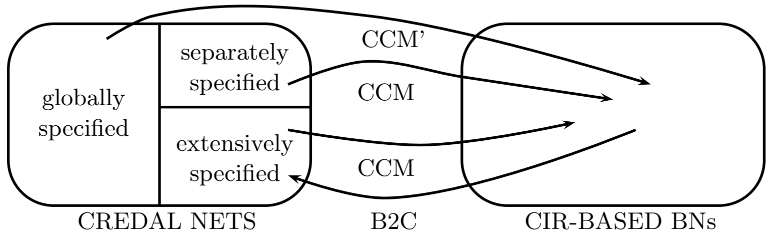

Fig. 1. Relations between updating on CNs and CIR-updating in BNs.

to: (i) add to X an *auxiliary child node*[2] X', associated to a binary variable with possible values x' and $\neg x'$; and (ii) extensively specify the probability table $P(X'|X)$, to belong to the following set of $|\Omega_X|$ tables:

$$\left\{ \begin{bmatrix} 1\,0\,0\ldots0 \\ 0\,1\,1\ldots1 \end{bmatrix}, \ldots, \begin{bmatrix} 0\ldots0\,1\,0\ldots0 \\ 1\ldots1\,0\,1\ldots1 \end{bmatrix}, \ldots, \begin{bmatrix} 0\,0\,0\ldots0\,1 \\ 1\,1\,1\ldots1\,0 \end{bmatrix} \right\}. \tag{4}$$

Each table in Eq. (4) specifies a conditional probability for the state x' of X' (corresponding to the first row of the table), which is zero conditionally on any state of X except a single one, different for any table. The B2C transformation, clearly linear in the input size, is the basis for the following:

Theorem 1. *Consider a CIR instance on a Bayesian network $\langle \mathcal{G}, P(\mathbf{X})\rangle$. Let X_I be the array of the unobserved non-MAR variables. Let $K(X_q||^{X_I}x_E)$ be the credal set returned by CIR for a queried variable X_q given the evidence $X_E = x_E$. If $K(X_q|x_E, x'_I)$ is the posterior credal set for X_q in the credal net $\langle \mathcal{G}', \mathbf{P}(\mathbf{X}')\rangle$ obtained from $\langle \mathcal{G}, P(\mathbf{X})\rangle$ by a B2C transformation with the nodes X_I specified, conditional on the evidences $X_E = x_E$ and $X'_I = x'_I$, then:*[3]

$$K(X_q||^{X_I}x_E) = K(X_q|x_E, x'_I). \tag{5}$$

4.2 From Credal to Bayesian Networks

For globally specified CNs we define a transformation that returns a BN given a CN $\langle \mathcal{G}, \mathbf{P}(\mathbf{X})\rangle$ as follows. The BN is obtained: (i) adding a *transparent* node X'' that is parent of all the nodes in $\mathbf{X}$ (see Fig. 2 left) and such that there is a one-to-one correspondence between the elements of $\Omega_{X''}$ and those of $\mathbf{P}$; and (ii) setting for each $X_i \in \mathbf{X}$ and $x'' \in \Omega_{X''}$: $P(X_i|\Pi_i, x'') \equiv \tilde{P}(X_i|\Pi_i)$, where Π_i are the parents of X_i in the CN and $\tilde{P}$ is the element of $\mathbf{P}$ corresponding to x''.

[2] This transformation takes inspirations from Pearl's prescriptions about boundary conditions for propagation [5, Sect. 4.3].

[3] Th. 1 can be extended also to CIR instances modeling incomplete observations where the value of the observed variable is know to belong to a generic subset of the possibility space, rather than missing observations for which the universal space is considered. We skip this case for lack of space.

In the case of locally specified CNs, we consider a slightly different transformation, where: (i) we add a transparent node X_i'' for each $X_i \in \mathbf{X}$, that is parent only of X_i (see Fig. 2 right) and such that there is a one-to-one correspondence between the elements of $\Omega_{X_i''}$ and the probability tables $P(X_i|\Pi_\mathrm{i})$ in the extensive[4] specification of $K(X_i|\Pi_\mathrm{i})$; and (ii) we set for each $X_i \in \mathbf{X}$: $P(X_i|\Pi_\mathrm{i}, x_i'') \equiv \tilde{P}(X_i|\Pi_\mathrm{i})$, where Π_i are the parents of X_i in the CN and $\tilde{P}(X_i|\Pi_\mathrm{i})$ is the probability table of $K(X_i|\Pi_\mathrm{i})$ relative to x_i''. Note that no prescriptions are given about the unconditional mass functions for the transparent nodes in both the transformations, because irrelevant for the results we will obtain. The second is the so-called CCM transformation [1] for CNs, while the first is simply an extension of CCM to the case of globally specified CNs and will be denoted as CCM'. These transformations are the basis for the following:

Theorem 2. *Let $K(X_q|x_E)$ be the posterior credal set of a queried variable X_q, given some evidence $X_E = x_E$, for a CN $\langle \mathcal{G}, \mathbf{P}(\mathbf{X})\rangle$. Let also $\langle \mathcal{G}', P'(\mathbf{X}')\rangle$ be the BN obtained from $\langle \mathcal{G}, \mathbf{P}(\mathbf{X})\rangle$ through CCM' (or CCM if the CN is not globally specified). Denote as $K(X_q||^{X''} x_E)$ the CIR-based posterior credal set for X_q in the BN obtained assuming what follows: the nodes in X_E instantiated to the values x_E, the transparent nodes, denoted as X'' also if CCM is used, to be not-MAR and the remaining nodes MAR. Then:*

$$K(X_q|x_E) = K(X_q||^{X''} x_E). \qquad (6)$$

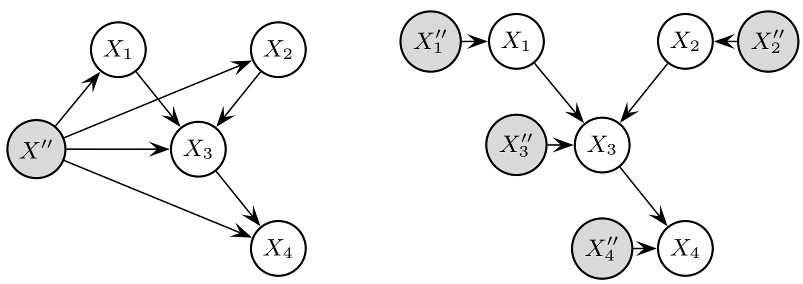

Fig. 2. The Bayesian networks returned by CCM' (left) and CCM (right). Transparent nodes are gray, while the nodes of the original CN are white.

[4] Separately specified credal sets can be extensively specified, considering all the probability tables obtained from the combinations of the vertices of the original credal sets. Although correct, this transformation gives rise to an exponential explosion of the number of tables. An alternative transformation, described in [2], might avoid this problem, as suggested by a reviewer.

5 Conclusions and Outlook

We have proved the formal equivalence between two updating problems on different graphical models: CIR-based updating on BNs and traditional updating with CNs. The result follows easily via simple transformations of the graphical models. An important consequence of the established link between BNs and CNs is that under realistic conditions of partial ignorance about the missingness process, working with BNs is actually equivalent to working with CNs. This appears to make CNs even more worthy of investigation than before.

The result makes it also possible in principle to solve CIR-based updating on BNs, for which there are no algorithms at presents, by means of algorithms for CNs. Unfortunately, the main corpus of algorithms for CNs considers the case of separately specified CNs, while CIR problems on BNs correspond to extensively specified CNs (see Fig. 1). Future work should therefore involve developing generalizations of the existing algorithms for CNs to the extensive case.

A Proofs of the Theorems

Proof of Theorem 1 *According to Eq. (3) and Eq. (2) respectively, we have:*

$$K(X_q||^{X_I}x_E) = \mathrm{CH}\{P(X_q|x_E,\tilde{x}_I)\}_{\tilde{x}_I\in\Omega_{X_I}} \tag{7}$$

$$K(X_q|x_E,x'_I) = \mathrm{CH}\{\tilde{P}(X_q|x_E,x'_I)\}_{\tilde{P}\in\mathbf{P}}. \tag{8}$$

An obvious isomorphism holds between $\mathbf{P}$ *and* Ω_{X_I}*: that follows from the correspondence, for each* $X_i \in X_I$*, between the conditional probability tables for* $P(X'_i|X_i)$ *as in Eq. (4) and the elements of* Ω_{X_i}*. Accordingly, we denote by* $\tilde{x}_I$ *the element of* Ω_{X_I} *corresponding to* $\tilde{P}\in\mathbf{P}$*. The thesis will be proved by showing, for each* $\tilde{P}\in\mathbf{P}$*,* $\tilde{P}(X_q|x_E) = P(X_q|x_E,\tilde{x}_I)$*. For each* $x_q\in\Omega_{X_q}$*:*

$$P(x_q|x_E,\tilde{x}_I) = \sum_{x_M} P(x_q,x_M|x_E,\tilde{x}_I) \propto \sum_{x_M} P(x_q,x_M,x_E,\tilde{x}_I) \tag{9}$$

$$\tilde{P}(x_q|x_E,x'_I) = \sum_{x_M,x_I} \tilde{P}(x_q,x_M,x_I|x_E,x'_I) \propto \sum_{x_M,x_I} \tilde{P}(x_q,x_M,x_I,x_E,x'_I). \tag{10}$$

According to the Markov condition:

$$\tilde{P}(x_q,x_M,x_I,x_E,x'_I) = \prod_{i:X_i\in X_I}\left[\tilde{P}(x'_i|x_i)\cdot\tilde{P}(x_i|\pi_i)\right]\cdot\prod_{j:X_j\in\mathbf{X}'\setminus(X_I\cup X'_I)}\tilde{P}(x_j|\pi_j), \tag{11}$$

with the values of x'_i*,* x_i*,* π_i*,* x_j *and* π_j *consistent with* (x_q,x_M,x_E,x_I,x'_I)*.*

According to Eq. (4), $P(x'_i|x_i)$ *is zero for each* $x_i\in\Omega_{X_i}$ *except for the value* $\tilde{x}_i$*, for which is one. The sum over* $x_i\in\Omega_{X_i}$ *of the probabilities in*

Eq. (11) is therefore reduced to a single non-zero term. Thus, taking all the sums over X_i with $X_i \in X_I$:

$$\sum_{x_I} \tilde{P}(x_q, x_M, x_I, x_E, x'_I) = \prod_{i: X_i \in X_I} P(\tilde{x}_i|\pi_i) \cdot \prod_{j: X_j \in \mathbf{X} \setminus X_I} P(x_j|\pi_j) = P(x_q, x_M, x_E, \tilde{x}_I), \tag{12}$$

with the values of π_i, x_j and π_j consistent with $(x_q, x_M, x_E, \tilde{x}_I)$. But Eq. (12) allows us to rewrite Eq. (9) as Eq. (10) and conclude the thesis. □

Proof of Theorem 2 *Consider a globally specified CN, for which CCM' should be used and X'' denotes a single transparent node. According to Eq. (3):*

$$K(X_q||^{X''} x_E) = \mathrm{CH}\{P(X_q|x_E, x'')\}_{x'' \in \Omega_{X''}}. \tag{13}$$

Setting $X_M \equiv \mathbf{X} \setminus (X_E \cup \{X_q\})$, for each $x_q \in \Omega_{X_q}$:

$$P(x_q|x_E, x'') = \sum_{x_M} P(x_q, x_M|x_E, x'') \propto \sum_{x_M} P(x_q, x_M, x_E, x''). \tag{14}$$

According to the Markov condition and CCM' definition, we have:

$$P(x_q, x_M, x_E, x'') = P(x'') \cdot \prod_{i=1}^{n} P(x_i|\pi_i, x'') \propto \prod_{i=1}^{n} \tilde{P}(x_i|\pi_i) = \tilde{P}(x_q, x_M, x_E), \tag{15}$$

where $\tilde{P}$ is the element of $\mathbf{P}$ associated to $x'' \in \Omega_{X''}$. The sum over x_M of the probabilities in Eq. (15) is proportional to $\tilde{P}(x_q|x_E)$. Thus, $\tilde{P}(X_q|x_E) = P(X_q|x_E, x'')$ for each $(\tilde{P}, x'') \in \mathbf{P} \times \Omega_{X''}$, that proves the thesis. Analogous considerations can be done for locally defined CNs transformed by CCM. □

References

1. A. Cano, J. Cano, and S. Moral. Convex sets of probabilities propagation by simulated annealing on a tree of cliques. In *Proceedings of the Fifth International Conference (IPMU '94)*, pages 978–983, Paris, 1994.
2. A. Cano and S. Moral. Using probability trees to compute marginals with imprecise probabilities. *Int. J. Approx. Reasoning*, 29(1):1–46, 2002.
3. F.G. Cozman. Graphical models for imprecise probabilities. *Int. J. Approx. Reasoning*, 39(2-3):167–184, 2005.
4. G. de Cooman and M. Zaffalon. Updating beliefs with incomplete observations. *Artificial Intelligence*, 159:75–125, 2004.
5. J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, 1988.
6. P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, New York, 1991.
7. M. Zaffalon. Conservative rules for predictive inference with incomplete data. In F.G. Cozman, R. Nau, and T. Seidenfeld, editors, *Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications (ISIPTA '05)*, pages 406–415, Pittsburgh, 2005.