Exploiting Bayesian Network Sensitivity Functions for Inference in Credal Networks

Janneke H. Bolt[1] and **Jasper De Bock**[2] and **Silja Renooij** [3]

Abstract. A Bayesian network is a concise representation of a joint probability distribution, which can be used to compute any probability of interest for the represented distribution. Credal networks were introduced to cope with the inevitable inaccuracies in the parametrisation of such a network. Where a Bayesian network is parametrised by defining unique local distributions, in a credal network sets of local distributions are given. From a credal network, lower and upper probabilities can be inferred. Such inference, however, is often problematic since it may require a number of Bayesian network computations exponential in the number of credal sets. In this paper we propose a preprocessing step that is able to reduce this complexity. We use sensitivity functions to show that for some classes of parameter in Bayesian networks the qualitative effect of a parameter change on an outcome probability of interest is independent of the exact numerical specification. We then argue that credal sets associated with such parameters can be replaced by a single distribution.

1 INTRODUCTION

Ever since the introduction of Bayesian networks [19], we have seen a growing interest for probabilistic graphical models in AI. A probabilistic graphical model concisely represents joint probability distributions over a set of stochastic variables, by combining the use of a graph to represent the independence relation among the variables with probability distributions over subsets of those variables. A *Bayesian network* defines a unique joint probability distribution by combining an acyclic directed graph with local discrete distributions, one for each node in the graph conditioned on its parents. A Bayesian network can be used to infer any probability of interest from the represented distribution.

Since in practice the specified probabilities, also called parameters, may be inaccurate, methods were developed to cope with such inaccuracies. A sensitivity analysis, for example, can be used to study the impact of inaccuracies on outcome probabilities of interest [16]. Alternatively, *credal networks* adopt the framework of imprecise probabilities [2, 23] by allowing the use of closed convex sets of distributions (also called credal sets), rather than defining unique local distributions per variable. Credal networks as such represent a set of Bayesian networks [1, 11, 12].

One of the main computational problems in a credal network is the computation of tight lower and upper bounds on the outcome probabilities that correspond to the represented Bayesian networks. These computations in essence require a number of Bayesian network inferences that is exponential in the number of imprecise local credal

sets, which are the local credal sets that consist of more than one distribution.

Although this combinatoric explosion can, in general, not be avoided, in many cases a reduction is possible. For example, similar to what can be done for Bayesian networks, it is possible to apply a preprocessing step in which the variables that are irrelevant for a specific problem, such as d-separated or barren nodes, are pruned in advance [11]. In this paper, we introduce a second type of preprocessing step to reduce the combinatoric complexity of credal network inference. Unlike the use of d-separation or the removal of barren nodes, our preprocessing step is tailored specifically to credal networks, and does not apply to Bayesian networks, as it involves replacing some of the local imprecise credal sets by a single distribution.

In order to prove the validity of our preprocessing step for credal network inference, we exploit results that we derive from Bayesian network sensitivity functions. More specifically, we use sensitivity functions to prove that for certain categories of distributions we can predict the qualitative effect of changing their parameters on an outcome probability, irrespective of the numerical specification of the network. In a credal network, this result allows us to identify, prior to the inference, local credal sets that can be replaced by a single distribution without changing the result of the inference. By applying this preprocessing step, the number of imprecise local credal sets and the corresponding combinatoric complexity of inference is reduced. Moreover, we demonstrate that for some widely used special classes of networks even all credal sets can be replaced by a single distribution, thereby essentially reducing the problem of computing a lower or upper probability in a credal network to an inference problem in a single Bayesian network.

This paper is organised as follows. In Section 2, we introduce the notation used throughout the paper. In Section 3, we start with a brief review of Bayesian networks and subsequently derive the properties of sensitivity functions for Bayesian networks that we want to exploit in the context of credal networks. Section 4 introduces the concept of a credal network, and then uses the results in Section 3 to establish the aforementioned preprocessing technique. We end the paper with conclusions and suggestions for future research.

2 NOTATION

This paper is concerned with graphical models for (sets of) joint probability distributions $\Pr(\mathbf{V})$ over a set of finite-valued random variables $\mathbf{V}$.

We use upper case V to denote a single random variable, writing $v \in V$ to indicate a value (lowercase v) of V. For binary-valued V, we use v and $\overline{v}$ to denote its two possible value assignments. Boldfaced letters are used to indicate sets (capitals), both of variables

[1] Utrecht University, The Netherlands, email: j.h.bolt@uu.nl
[2] Ghent University, Belgium, email: jasper.debock@ugent.be
[3] Utrecht University, The Netherlands, email: s.renooij@uu.nl

and of value assignments, or a joint value assignment (lowercase) to a set of variables; the distinction will be clear from the context. For example $V \in \mathbf{W}$ indicates that V is a variable in a set of variables $\mathbf{W}$, whereas $\mathbf{w} \in \mathbf{W}$ indicates that $\mathbf{w}$ is a joint value assignment of that same set of variables $\mathbf{W}$.

Two value assignments are said to be compatible, denoted by $\sim$, if they agree on the values of the shared variables; otherwise they are said to be incompatible, denoted by $\nsim$. For example, for three distinct variables U, V and W in $\mathbf{V}$, and value assignments $u, u' \in U$, $v \in V$ and $w \in W$ such that $u \neq u'$, we have that $uv \sim uvw$ and $uv \nsim u'vw$. Assignments to disjoint sets are always considered to be compatible.

The probabilistic graphical models that we consider are Bayesian networks and credal networks. The graphical part of these models consists of a directed graph, the nodes of which have a one-to-one correspondence with the random variables V in $\mathbf{V}$. For a given node V, π indicates its set of parents in the directed graph, and we will use π to indicate a joint instantiation of these parents. Without loss of generality, we assume that the set $\mathbf{V}$ of all random variables consists of the disjoint subsets $\mathbf{H}$ (hypothesis variables), $\mathbf{E}$ (evidence variables) and $\mathbf{R}$, with $\mathbf{V} = \mathbf{H} \cup \mathbf{E} \cup \mathbf{R}$. The sets of hypothesis variables and evidence variables that do not have parents in $\mathbf{R}$ are denoted by $\mathbf{H}_{|\mathbf{R}}$ and $\mathbf{E}_{|\mathbf{R}}$, respectively. Similarly, the set $\mathbf{H}_{|\mathbf{H}}$ consists of those hypothesis variables that have no parents in $\mathbf{H}$.

Finally, when referring to monotonic functions, we will use the term *increasing* to indicate both strictly increasing and non-decreasing functions; likewise *decreasing* refers to both strictly decreasing and non-increasing functions.

3 BAYESIAN NETWORKS' SENSITIVITY PROPERTIES

In this section, we first briefly review Bayesian networks and sensitivity analysis. Subsequently, we establish guarantees on the behaviour of sensitivity functions for some categories of parameters. Next, we use these results to define types of local distributions, which will then allow us in Section 4 to exploit our results in the context of credal networks.

3.1 Bayesian network preliminaries

A Bayesian network $\mathcal{B}$ is a concise representation of a joint probability distribution. It uses a directed acyclic graph to capture the (in)dependences among its variables $\mathbf{V}$ using the well-known d-separation criterion [19]. [4] Furthermore, for each variable $V \in \mathbf{V}$, it specifies exactly one local distribution $\Pr(V \mid \pi)$ over the values of V for each value assignment π to π; the separate probabilities in these distributions are termed the network's *parameters*. The joint probability distribution $\Pr$ factorises over the local distributions as follows:

$$\Pr(\mathbf{v}) = \prod_{V \in \mathbf{V}} \Pr(v \mid \pi)\big|_{v\pi \sim \mathbf{v}}$$

where the notation $|_{\text{prop}}$ is used to indicate the properties the arguments in the preceding formula adhere to. Outcome probabilities of the form $\Pr(\mathbf{h} \mid \mathbf{e})$ can be computed with various algorithms [15].

An example Bayesian network is shown in Figure 1. The figure shows a graph with hypothesis variables $\mathbf{H} = \{G, H, K\}$ (double circles), evidence variables $\mathbf{E} = \{E, F\}$ (shaded), and remaining variable $\mathbf{R} = \{R\}$. The figure also shows conditional probability tables (CPTs) that specify the local distributions. Actual numbers are not provided, but the top row in the table for $\Pr(H \mid G)$ could for example specify the parameters $\Pr(h \mid g) = 0.4$ and $\Pr(\overline{h} \mid g) = 0.6$.

The probabilities computed from a Bayesian network depend on the parameters specified in the network, which are inevitably inaccurate. To study the effects of such inaccuracies on an output probability of interest, a sensitivity analysis can be performed. One approach to performing a sensitivity analysis is to compute so-called sensitivity functions [16]. An n-way sensitivity function describes an output probability of interest $\Pr(\mathbf{h} \mid \mathbf{e})$ as a function of n network parameters $\vec{x} = \{x_1, \ldots, x_n\}$:

$$\Pr(\mathbf{h} \mid \mathbf{e})(\vec{x}) = \frac{\Pr(\mathbf{he})(\vec{x})}{\Pr(\mathbf{e})(\vec{x})}$$

where both the numerator and the denominator are known to be multi-linear expressions in $\vec{x}$. For $n = 1$, the one-way sensitivity function takes the general form

$$\Pr(\mathbf{h} \mid \mathbf{e})(x) = \frac{\tau_1 \cdot x + \tau_2}{\kappa_1 \cdot x + \kappa_2}$$

where the constants τ_1, τ_2, κ_1 and κ_2 are composed of non-varied network parameters. Throughout this paper we assume $\Pr(\mathbf{e})$ to be, and to remain, strictly positive.

Upon varying a parameter x of a distribution, the other parameters of the same distribution have to be co-varied to let the distribution sum to 1. If the distribution is associated with a binary variable, the co-varying parameter equals $1 - x$. If a variable is multi-valued, however, different co-variation schemes are possible [21].

Based upon graphical considerations alone, in computing a probability $\Pr(\mathbf{h} \mid \mathbf{e})$, and thus also for performing sensitivity analyses with respect to this output probability, a Bayesian network can be safely pruned by removing variables that are either d-separated from $\mathbf{H}$ given $\mathbf{E}$, or that are barren given these sets. [5] The remaining variables all have evidence or parameters that may be required for the computation of $\Pr(\mathbf{h} \mid \mathbf{e})$ [22]. This set of requisite variables coincides with the union of $\mathbf{E}$ and the so-called sensitivity set [18], which is the set of variables for which a change in one of its parameters may result in a change in $\Pr(\mathbf{h} \mid \mathbf{e})$ [10]. In our example network in Figure 1 all variables are in the sensitivity set of $\Pr(ghk \mid ef)$.

3.2 Guaranteed effects of parameter changes

In this section, we first consider one-way sensitivity functions $\Pr(\mathbf{h} \mid \mathbf{e})(x)$ and identify categories of parameters that are guaranteed to give monotonically increasing and parameters that are guaranteed to give monotonically decreasing one-way functions. We also identify parameters inside the sensitivity set of $\Pr(\mathbf{h} \mid \mathbf{e})$ that are not used in computing the outcome probability $\Pr(\mathbf{h} \mid \mathbf{e})$. Next, we use these results to provide guarantees on n-way functions $\Pr(\mathbf{h} \mid \mathbf{e})(\vec{x})$.

We begin by categorising the different parameters $x = \Pr(v \mid \pi)$ of a Bayesian network with respect to a specific outcome probability $\Pr(\mathbf{h} \mid \mathbf{e})$. We only categorise the subset of all parameters for which we can provide guarantees on their qualitative effect on the outcome probability. The categorisation is given in Table 1. For our example network in Figure 1, the categories are indicated with respect to the output probability $\Pr(ghk \mid ef)$.

---

[4] For any three disjoint sets of nodes X, Y, Z, the set Z is said to d-separate X and Y if there do not exist any active chains between X and Y given evidence for Z. A chain between two nodes is active if each of its head-to-head nodes is either instantiated or has an instantiated descendant, and none of its other nodes are instantiated. The variables captured by d-separated nodes are considered probabilistically independent.

[5] Barren variables are in the set $\mathbf{R}$ and have just barren descendants.
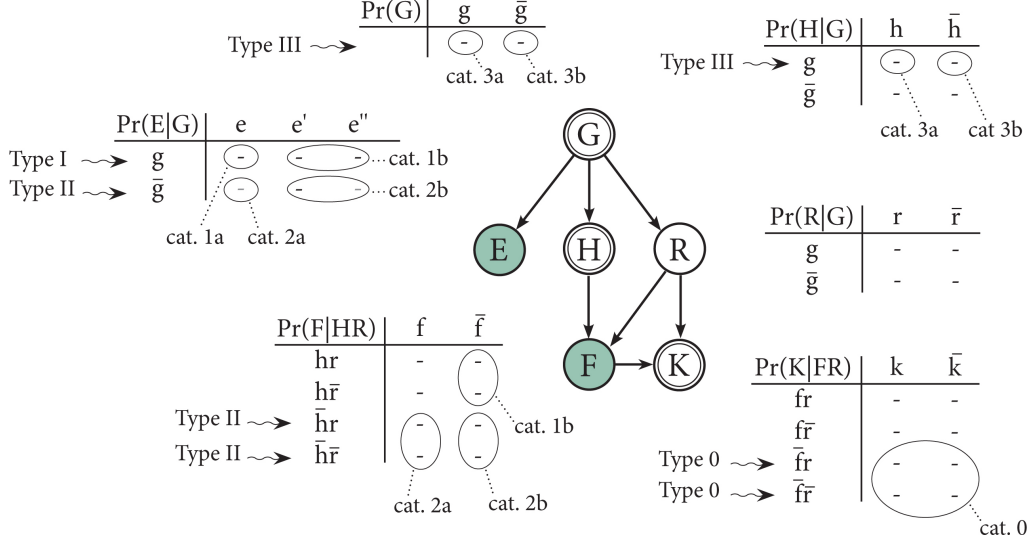
Figure 1: An example Bayesian—and credal—network. The CPTs are labelled with parameter categories (see Table 1) and distribution types (see Section 3.3) with respect to output probability $\Pr(ghk\,|\,ef)$.

Table 1: Categorisation of parameters $x = \Pr(v\,|\,\boldsymbol{\pi})$ with respect to the outcome probability $\Pr(\mathbf{h}\,|\,\mathbf{e})$.

	$\boldsymbol{\pi} \sim \mathbf{h}$	$\boldsymbol{\pi} \nsim \mathbf{h}$
$\boldsymbol{\pi} \nsim \mathbf{e}$	cat. 0	cat. 0
$V \in \mathbf{E}$, $v \nsim \mathbf{e}$ and $\boldsymbol{\pi} \sim \mathbf{e}$	cat. 1b	cat. 2b
$V \in \mathbf{E}_{\downarrow \mathbf{R}}$ and $v\,\boldsymbol{\pi} \sim \mathbf{e}$	cat. 1a	cat. 2a
$V \in \mathbf{E} \setminus \mathbf{E}_{\downarrow \mathbf{R}}$ and $v\,\boldsymbol{\pi} \sim \mathbf{e}$	-	cat. 2a
binary $V \in \mathbf{H}_{\downarrow \mathbf{R}}$ and $\boldsymbol{\pi} \sim \mathbf{e}$	cat. 3a ($v \sim \mathbf{h}$) cat. 3b ($v \nsim \mathbf{h}$)	-

The following proposition states that parameters from category 1a result in sensitivity functions that are monotonically increasing, whereas parameters from category 2a give monotonically decreasing sensitivity functions, *irrespective of the values of the non-varied network parameters and irrespective of the co-variation scheme that is used.*

Proposition 1 *Consider a Bayesian network $\mathcal{B}$ and a probability of interest $\Pr(\mathbf{h}\,|\,\mathbf{e})$. Let $x = \Pr(v\,|\,\boldsymbol{\pi})$ be a parameter of $V \in \mathbf{E}$ with $v\,\boldsymbol{\pi} \sim \mathbf{e}$. If $V \in \mathbf{E}_{\downarrow \mathbf{R}}$ and $\boldsymbol{\pi} \sim \mathbf{h}$, then the sensitivity function $\Pr(\mathbf{h}\,|\,\mathbf{e})(x)$ is monotonically increasing. If $\boldsymbol{\pi} \nsim \mathbf{h}$, then the sensitivity function $\Pr(\mathbf{h}\,|\,\mathbf{e})(x)$ is monotonically decreasing.*

Proof. We first consider the numerator of the sensitivity function:

$$\Pr(\mathbf{he}) = \sum_{\mathbf{r}\in\mathbf{R}} \Pr(\mathbf{rhe}) = \sum_{\mathbf{r}\in\mathbf{R}} \prod_{V^*\in\mathbf{V}} \Pr(v^*\,|\,\boldsymbol{\pi}^*)\big|_{v^*\boldsymbol{\pi}^*\sim\mathbf{rhe}} \quad (1)$$

Note that $x = \Pr(v\,|\,\boldsymbol{\pi})$ is a factor in this expression only if $\boldsymbol{\pi} \sim \mathbf{h}$. For $\boldsymbol{\pi} \nsim \mathbf{h}$, therefore, $\Pr(\mathbf{he})(x) = \tau_1$ for some constant $\tau_1 \geq 0$. Returning to the case where $\boldsymbol{\pi} \sim \mathbf{h}$, we observe that if $V \in \mathbf{E}_{\downarrow \mathbf{R}}$, we have that $\boldsymbol{\pi} \sim \mathbf{rhe}$ for all $\mathbf{r} \in \mathbf{R}$ and we thus find that

$$\Pr(\mathbf{he})(x) = x \cdot \left(\sum_{\mathbf{r}\in\mathbf{R}} \prod_{V^*\in\mathbf{V}\setminus\{V\}} \Pr(v^*\,|\,\boldsymbol{\pi}^*)\big|_{v^*\boldsymbol{\pi}^*\sim\mathbf{rhe}} \right)$$

Given $\boldsymbol{\pi} \sim \mathbf{h}$ and $V \in \mathbf{E}_{\downarrow \mathbf{R}}$, therefore, $\Pr(\mathbf{he})(x) = x \cdot \tau_1$ for some constant $\tau_1 \geq 0$.

Next, we consider the denominator of the sensitivity function:

$$\Pr(\mathbf{e}) = \sum_{\mathbf{r}\in\mathbf{R},\mathbf{h}^*\in\mathbf{H}} \Pr(\mathbf{rh}^*\mathbf{e})$$

$$= \sum_{\mathbf{r}\in\mathbf{R},\mathbf{h}^*\in\mathbf{H}} \prod_{V^*\in\mathbf{V}} \Pr(v^*\,|\,\boldsymbol{\pi}^*)\big|_{v^*\boldsymbol{\pi}^*\sim\mathbf{rh}^*\mathbf{e}} \quad (2)$$

We observe that $x = \Pr(v\,|\,\boldsymbol{\pi})$ is a factor in each term of the summation for which $\boldsymbol{\pi} \sim \mathbf{rh}^*$, and otherwise is absent. That is, no term includes x as a co-varying parameter. Therefore, $\Pr(\mathbf{e})(x) = \kappa_1 \cdot x + \kappa_2$, for constants $\kappa_1, \kappa_2 \geq 0$.

We are now able to determine the general form of the sensitivity function $\Pr(\mathbf{h}\,|\,\mathbf{e})(x)$ under the given conditions. Moreover, we can use the first derivative of this sensitivity function to determine whether the function is increasing or decreasing. If $\boldsymbol{\pi} \sim \mathbf{h}$ and $V \in \mathbf{E}_{\downarrow \mathbf{R}}$, we find that

$$\Pr(\mathbf{h}\,|\,\mathbf{e})'(x) = \frac{\tau_1 \cdot \kappa_2}{(\kappa_1 \cdot x + \kappa_2)^2}.$$

This first derivative is always non-negative which implies that $\Pr(\mathbf{h}\,|\,\mathbf{e})(x)$ is a monotone increasing function, irrespective of the values of the non-varied network parameters and irrespective of the co-variation scheme. If $\boldsymbol{\pi} \nsim \mathbf{h}$, we find that

$$\Pr(\mathbf{h}\,|\,\mathbf{e})'(x) = \frac{-\tau_1 \cdot \kappa_1}{(\kappa_1 \cdot x + \kappa_2)^2}.$$

This first derivative is always non-positive, which implies that $\Pr(\mathbf{h}\,|\,\mathbf{e})(x)$ is a monotone decreasing function, irrespective of the values of the non-varied network parameters and irrespective of the co-variation scheme. □

Our next proposition states that parameters from category 3a result in monotonically increasing sensitivity functions, whereas parameters from category 3b give monotonically decreasing sensitivity functions, again *irrespective of the values of the non-varied network*

parameters. The proposition concerns parameters for binary hypothesis variables, which implies that the co-variation scheme is fixed.

Proposition 2 *Consider a Bayesian network $\mathcal{B}$ and a probability of interest $\Pr(\mathbf{h}\,|\,\mathbf{e})$. Let $x = \Pr(v\,|\,\boldsymbol{\pi})$ be a parameter of a binary variable $V \in \mathbf{H}_{\not{\!\mathcal{R}}}$, with $\boldsymbol{\pi} \sim \mathbf{e}$ and $\boldsymbol{\pi} \sim \mathbf{h}$. If $v \sim \mathbf{h}$ then the sensitivity function $\Pr(\mathbf{h}\,|\,\mathbf{e})(x)$ is monotonically increasing. If $v \nsim \mathbf{h}$ then $\Pr(\mathbf{h}\,|\,\mathbf{e})(x)$ is monotonically decreasing.*

Proof. We first consider the parameters x with $v \sim \mathbf{h}$. As in the proof of Proposition 1, the numerator of the sensitivity function can be rewritten as

$$\Pr(\mathbf{he})(x) = x \cdot \Big(\sum_{\mathbf{r} \in \mathbf{R}} \prod_{V^* \in \mathbf{V} \setminus \{H\}} \Pr(v^*\,|\,\boldsymbol{\pi}^*)\big|_{v^*\boldsymbol{\pi}^* \sim \mathbf{rhe}} \Big)$$

because, under the given conditions, x is a factor in all terms of the summation. Therefore, we have that $\Pr(\mathbf{he})(x) = x \cdot \tau_1$, for some constant $\tau_1 \geq 0$.

The assumption that V is binary enforces the following co-variation scheme: $\Pr(\overline{v}\,|\,\boldsymbol{\pi}) = 1 - \Pr(v\,|\,\boldsymbol{\pi})$. Therefore, with respect to the denominator of the sensitivity function—see Equation (2)—we observe the following: 1) x itself is a factor in any term of the summation for which $\boldsymbol{\pi} \sim \mathbf{h}^*$ and $v \sim \mathbf{h}^*$, 2) x appears as $(1-x)$ for $\boldsymbol{\pi} \sim \mathbf{h}^*$ and $v \nsim \mathbf{h}^*$, and 3) x is absent whenever $\boldsymbol{\pi} \nsim \mathbf{h}^*$. Hence, $\Pr(\mathbf{e})(x) = x \cdot \kappa_1 + (1-x) \cdot \kappa_2 + \kappa_3$, with constants $\kappa_i \geq 0$, $i = 1, 2, 3$.

We now find that the first derivative of the sensitivity function equals the following non-negative function:

$$\Pr(\mathbf{h}\,|\,\mathbf{e})'(x) = \frac{(\kappa_2 + \kappa_3) \cdot \tau_1}{(\kappa_2 + \kappa_3 + \kappa_1 \cdot x - \kappa_2 \cdot x)^2}$$

This implies that $\Pr(\mathbf{h}\,|\,\mathbf{e})(x)$ is a monotonically increasing function irrespective of the values of the non-varied network parameters.

Next, we consider the parameters $x = \Pr(v\,|\,\boldsymbol{\pi})$ with $v \nsim \mathbf{h}$. Since we know that $\Pr(v\,|\,\boldsymbol{\pi}) = 1 - \Pr(\overline{v}\,|\,\boldsymbol{\pi})$ with $\overline{v} \sim \mathbf{h}$, an increase (decrease) of x is equivalent to a decrease (increase) of the parameter consistent with $\mathbf{h}$. Hence, using the first part of this proof, it follows that $\Pr(\mathbf{h}\,|\,\mathbf{e})(x)$ is a monotonically decreasing function irrespective of the values of the non-varied network parameters. $\square$

Although the above propositions are concerned with one-way sensitivity functions, we have shown that their monotone increasing or decreasing behaviour does not depend on the actual values of the non-varied parameters which determine the constants of the function. As a result, our observations extend to specific n-way sensitivity functions.

Theorem 1 *Consider a Bayesian network $\mathcal{B}$ and a probability of interest $\Pr(\mathbf{h}\,|\,\mathbf{e})$. Let $\vec{x}_+ = \{x_1^+, \ldots, x_k^+\}$ be a set of parameters from categories 1a and/or 3a, and let $\vec{x}_- = \{x_1^-, \ldots, x_l^-\}$ be a set of parameters from categories 2a and/or 3b. Then the $(k+l)$-way sensitivity function $\Pr(\mathbf{h}\,|\,\mathbf{e})(\vec{x}_+, \vec{x}_-)$ is monotonically increasing for increasing parameters x_i^+ and decreasing parameters x_i^-; the sensitivity function is monotonically decreasing for decreasing x_i^+ and increasing x_i^-.*

Proof. From Propositions 1 and 2 we have that a one-way sensitivity function $\Pr(\mathbf{h}\,|\,\mathbf{e})(x_i^+)$ increases with increasing x_i^+ and a one-way sensitivity function $\Pr(\mathbf{h}\,|\,\mathbf{e})(x_i^-)$ increases with decreasing x_i^-, irrespective of the values of the non-varied parameters, and irrespective of the co-variation scheme. Given a simultaneous increase of

parameters in $\vec{x}_+$ and decrease of parameters in $\vec{x}_-$, therefore, the probability $\Pr(\mathbf{h}\,|\,\mathbf{e})(\vec{x}_+, \vec{x}_-)$ will increase. A similar observation holds for decreasing x_i^+ and increasing x_i^-. $\square$

The above results describe an output probability of interest as a function of changing input parameters of categories 1a, 2a, 3a and 3b. The example network in Figure 1 also includes parameters of category 0, 1b and 2b. Parameters of these categories are not used in computing the output probability. For category 0, the parameters make up an entire local distribution (see e.g. the CPT for $\Pr(K\,|\,FR)$); for parameters of categories 1b and 2b, parameters are in a local distribution with a parameter of category 1a or 2a (see e.g. the CPTs for $\Pr(E\,|\,G)$ and $\Pr(F\,|\,HR)$).

Proposition 3 *Consider a Bayesian network $\mathcal{B}$ and a probability of interest $\Pr(\mathbf{h}\,|\,\mathbf{e})$. Let $\Pr(v\,|\,\boldsymbol{\pi})$ be a parameter. If $v \nsim \mathbf{e}$ and/or $\boldsymbol{\pi} \nsim \mathbf{e}$, then $\Pr(v\,|\,\boldsymbol{\pi})$ is not used in the computation of $\Pr(\mathbf{h}\,|\,\mathbf{e})$.*

Proof. As before, we have that

$$\Pr(\mathbf{h}\,|\,\mathbf{e}) = \frac{\Pr(\mathbf{he})}{\Pr(\mathbf{e})}$$

with numerator

$$\Pr(\mathbf{he}) = \sum_{\mathbf{r} \in \mathbf{R}} \Pr(\mathbf{rhe}) = \sum_{\mathbf{r} \in \mathbf{R}} \prod_{V^* \in \mathbf{V}} \Pr(v^*\,|\,\boldsymbol{\pi}^*)\big|_{v^*\boldsymbol{\pi}^* \sim \mathbf{rhe}}$$

The parameter $\Pr(v\,|\,\boldsymbol{\pi})$ can only be a factor in this numerator if $v \sim \mathbf{e}$ and $\boldsymbol{\pi} \sim \mathbf{e}$. A completely analogous observation holds for the denominator as well. $\square$

Note that we cannot conclude from the above proposition that the sensitivity function $\Pr(\mathbf{h}\,|\,\mathbf{e})(x)$ for parameters $x = \Pr(v\,|\,\boldsymbol{\pi})$ of categories 1b and 2b is constant. A change of such parameters may, by co-variation, induce a change in the parameter $\Pr(v'\,|\,\boldsymbol{\pi})$ of the same local distribution with $v' \sim \mathbf{e}$ which *is* used in the computation of $\Pr(\mathbf{h}\,|\,\mathbf{e})$. The actual value of a parameter with $v \nsim \mathbf{e}$, however, is irrelevant for $\Pr(\mathbf{h}\,|\,\mathbf{e})$. We *can* conclude that the sensitivity function $\Pr(\mathbf{h}\,|\,\mathbf{e})(x)$ for parameters of category 0 is a constant, however, since given a parameter with $\boldsymbol{\pi} \nsim \mathbf{e}$ this condition is fulfilled for all parameters in the same local distribution.

3.3 Defining local distribution types

In the previous subsection we have focused on properties of different categories of parameters in a Bayesian network. In order to apply our results in the context of credal networks, it will be useful to lift these properties to the level of distributions. To this end, we define four types of local distributions $\Pr(V\,|\,\boldsymbol{\pi})$, again with respect to a specific outcome probability of interest $\Pr(\mathbf{h}\,|\,\mathbf{e})$.

Definition 1 (Local distribution of Type 0) *Consider a Bayesian network $\mathcal{B}$ and a probability of interest $\Pr(\mathbf{h}\,|\,\mathbf{e})$. A local distribution $\Pr(V\,|\,\boldsymbol{\pi})$ is said to be of Type 0 if $\boldsymbol{\pi} \nsim \mathbf{e}$.*

All parameters of a local distribution of Type 0 are in category 0. From Proposition 3, we know that parameters from a local distribution of Type 0 are irrelevant for the computation of the outcome probability $\Pr(\mathbf{h}\,|\,\mathbf{e})$.

Definition 2 (Local distribution of Type I) *Consider a Bayesian network $\mathcal{B}$ and a probability of interest $\Pr(\mathbf{h}\,|\,\mathbf{e})$. A local distribution $\Pr(V\,|\,\boldsymbol{\pi})$ is said to be of Type I if $\boldsymbol{\pi} \sim \mathbf{e}$, $V \in \mathbf{E}_{\not{\!\mathcal{R}}}$ and $\boldsymbol{\pi} \sim \mathbf{h}$.*

The unique parameter of a distribution of Type I with $v \sim \mathbf{e}$ is in category 1a. Any other parameter $\Pr(v'|\mathbf{e})$ of the distribution is in category 1b and is therefore irrelevant for computing $\Pr(\mathbf{h}|\mathbf{e})$, since $v' \nsim \mathbf{e}$. Given changes in a local distribution of Type I, therefore, knowing the qualitative change of the parameter with $v \sim \mathbf{e}$ suffices to know the qualitative effect of the changes in the entire distribution on $\Pr(\mathbf{h}|\mathbf{e})$. In particular, an increase (decrease) of this parameter results in an increase (decrease) of $\Pr(\mathbf{h}|\mathbf{e})$.

Definition 3 (Local distribution of Type II) *Consider a Bayesian network $\mathcal{B}$ and a probability of interest $\Pr(\mathbf{h}|\mathbf{e})$. A local distribution $\Pr(V|\boldsymbol{\pi})$ is said to be of* Type II *if $\boldsymbol{\pi} \sim \mathbf{e}$, $V \in \mathbf{E}$ and $\boldsymbol{\pi} \nsim \mathbf{h}$.*

Analogous to distributions of Type I, knowing the qualitative change of the parameter with $v \sim \mathbf{e}$ is sufficient to know the qualitative effect of changes in the entire local distribution on $\Pr(\mathbf{h}|\mathbf{e})$. Now the unique parameter with $v \sim \mathbf{e}$ is in category 2a, which implies that an increase (decrease) of this parameter results in an decrease (increase) of $\Pr(\mathbf{h}|\mathbf{e})$.

Definition 4 (Local distribution of Type III) *Consider a Bayesian network $\mathcal{B}$ and a probability of interest $\Pr(\mathbf{h}|\mathbf{e})$. A local distribution $\Pr(V|\boldsymbol{\pi})$ is said to be of* Type III *if $\boldsymbol{\pi} \sim \mathbf{e}$, $\boldsymbol{\pi} \sim \mathbf{h}$, $V \in \mathbf{H}_{\not\perp\!\!\!\mathcal{R}}$ and $|V| = 2$.*

Distributions of Type III consist of exactly two parameters, one for v and one for $\overline{v}$, which are related by the fixed co-variation scheme: $\Pr(v|\boldsymbol{\pi}) = 1 - \Pr(\overline{v}|\boldsymbol{\pi})$. The parameter that is compatible with $\mathbf{h}$ is in category 3a, which implies that an increase (decrease) of this parameter results in an increase (decrease) of $\Pr(\mathbf{h}|\mathbf{e})$. The parameter that is not compatible with $\mathbf{h}$ is in category 3b, which implies that the concurrent decrease (increase) of this parameter results in an increase (decrease) of $\Pr(\mathbf{h}|\mathbf{e})$ as well.

By Theorem 1 and Proposition 3 we have that, given simultaneous changes in multiple local distributions, $\Pr(\mathbf{h}|\mathbf{e})$ will increase (decrease) given an increase (decrease) of the parameters with $v \sim \mathbf{e}$ in the distributions of Type I and III and a decrease (increase) of the parameters compatible with $v \sim \mathbf{e}$ in the distributions of Type II. The outcome probability $\Pr(\mathbf{h}|\mathbf{e})$ is not affected by changes in local distributions of Type 0.

3.4 Important special cases

In Section 3.3 we identified types of distributions for which we can guarantee the direction of change in $\Pr(\mathbf{h}|\mathbf{e})$ upon changes in certain parameters from those distributions. In general, as illustrated in Figure 1, a network will include distributions that do not belong to any of those types. We can, however, identify classes of networks for which the majority of distributions—or even all of them—can be classified as Type 0, I, II or III.

Consider for example a network obeying the following constraint: each observed variable only has hypothesis variables or other observed variables as parents, that is, $\mathbf{E} = \mathbf{E}_{\not\perp\!\!\!\mathcal{R}}$. It then follows from Definitions 1–3 that all local distributions of the observed variables are of type 0, I or II. Bayesian network classifiers with full evidence obey this constraint. Bayesian classifiers, a special type of Bayesian network, are widely used in practice (see [14] for an overview).

Now consider a network obeying the constraint that all hypothesis variables are binary valued and can only have observed variables as parents, that is, $\mathbf{H} = \mathbf{H}_{\not\perp\!\!\!\mathcal{R}} \cap \mathbf{H}_{\not\perp\!\!\!\mathcal{H}}$ and $|V| = 2$ for each V in $\mathbf{H}$. It then follows from Definitions 1 and 4 that all local distributions of

the hypothesis variables are of Type 0 or III.[6] This second constraint is for example obeyed by Bayesian network classifiers with a single binary class variable, or by Bayesian network classifiers with multiple binary class variables, provided that these class variables have no class parents.[7] Given full evidence, these two types of Bayesian network classifiers also satisfy the first constraint, it follows therefore that, given full evidence, for those networks all the distributions are of Type 0, I, II or III.

4 AN APPLICATION TO CREDAL NETWORKS

Since a sensitivity function is an object that is associated with a given Bayesian network, it might at first sight seem as if it has little to do with credal networks, because these are essentially sets of Bayesian networks. However, for the particular sensitivity properties that we developed in the previous section, this is not the case. As we will explain in this section, these sensitivity properties allow us to replace some of the local credal sets of a credal network by a single probability mass function, and in this way, the combinatoric nature of credal network inference can be reduced.

4.1 Credal network preliminaries

An important limitation of Bayesian networks is that they require the exact specification of a large number of probabilities. Since these probabilities have to be elicited from experts or estimated from data, this requirement is often unrealistic. By enforcing it anyway, as is conventionally done, we run the risk of ending up with a model whose precision is unwarranted, thereby making the resulting conclusions unreliable. In order to avoid these issues, credal networks explicitly drop the assumption that probabilities should be specified exactly, and instead adopt the framework of imprecise probabilities [2, 23, 25], thereby trading off precision for reliability.

Basically, a *credal network* [1, 11, 12] is just a Bayesian network of which the local models are partially specified, in the sense that they are only known to belong to some set of candidate distributions. Specifically, for every random variable $V \in \mathbf{V}$ and every instantiation $\boldsymbol{\pi}$ of its parents $\boldsymbol{\pi}$, the local distribution $\Pr(V|\boldsymbol{\pi})$ is replaced by a non-empty set $\mathcal{M}(V|\boldsymbol{\pi})$ of candidate distributions. This set is usually taken to be closed and convex, and is then called a *credal set*. The local credal sets of a credal network can be obtained in various ways [23], the most typical of which are to elicit them from experts [20], to learn them from data [4, 24] or to create a neighbourhood model around some distribution [3].

For example, in Figure 1, if an expert judges that the probability of g is at most $2/3$ but not less than $1/3$, then in particular, we are lead to consider the local credal set $\mathcal{M}(G)$ that is defined by

$$\Pr(G) \in \mathcal{M}(G) \iff \frac{1}{3} \leq \Pr(g) \leq \frac{2}{3}. \tag{3}$$

The same credal set can also be obtained by ϵ-contaminating the uniform probability distribution on $\{g, \overline{g}\}$ with $\epsilon = 1/3$ [3] or by learning it from a dataset in which each of the two possible outcomes—g and $\overline{g}$—is counted twice, using the Imprecise Dirichlet Model (IDM) with $s = 2$ [4, 24]. As another example, consider a situation in which, conditional on $G = g$, an expert judges that e is at least as

---

[6] $V \in \mathbf{H}_{\not\perp\!\!\!\mathcal{H}}$ implies that $\boldsymbol{\pi} \sim \mathbf{h}$

[7] Note that, by definition, in Bayesian classifiers class variables do not have feature parents, as a consequence, $\mathbf{H} = \mathbf{H}_{\not\perp\!\!\!\mathcal{R}}$ even in case of incomplete evidence.
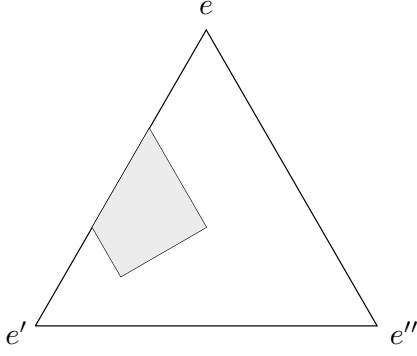
Figure 2: Every point in the above equilateral triangle of height one can be identified with a distribution $\Pr(E|g)$ on $\{e, e', e''\}$, by letting $\Pr(e|g)$ be equal to the perpendicular distance from that point to the edge that opposes the corner that corresponds to e, and similarly for $\Pr(e'|g)$ and $\Pr(e''|g)$. In this way, the grey area depicts the distributions in the credal set $\mathcal{M}(E|g)$ of Equation (4).

probable as e'' and that the probability of e' is at most $2/3$ and at least $1/3$. In that case, we are lead to consider the credal set $\mathcal{M}(E|g)$ that is defined by

$$\Pr(E|g) \in \mathcal{M}(E|g)$$
$$\Leftrightarrow \Pr(e|g) \ge \Pr(e''|g) \text{ and } \frac{1}{3} \le \Pr(e'|g) \le \frac{2}{3}, \quad (4)$$

as depicted in Figure 2. The other local credal sets can be obtained similarly. Depending on the amount of data and/or expert assessments that is available, some of these local credal sets—such as the two examples we have just provided—will be *imprecise*, meaning that they consist of multiple distributions, while others may be *precise*, meaning that they consist of a single distribution.

If all the local credal sets are precise, a credal network is exactly the same as a Bayesian network. However, if some of the credal sets $\mathcal{M}(V|\boldsymbol{\pi})$ are imprecise, then since the local distributions $\Pr(V|\boldsymbol{\pi})$ are only known to belong to these credal sets, they do not determine a unique Bayesian network. Therefore, in general, a credal network corresponds to a set of Bayesian networks $\mathbb{B}$, defined by[8]

$$\mathcal{B} \in \mathbb{B} \Leftrightarrow (\forall V \in \mathbf{V})(\forall \boldsymbol{\pi} \in \pi) \; \Pr(V|\boldsymbol{\pi}) \in \mathcal{M}(V|\boldsymbol{\pi}).$$

If all the local credal sets are precise, the set $\mathbb{B}$ consists of a single Bayesian network.

An important computational problem in credal networks consists in computing tight lower and upper bounds on the probabilities that correspond to the Bayesian networks $\mathcal{B}$ in $\mathbb{B}$, which are called lower and upper probabilities. In particular, we consider the problem of computing tight lower and upper bounds on probabilities of the form $\Pr(\mathbf{h}|\mathbf{e})$, defined by

$$\underline{\Pr}(\mathbf{h}|\mathbf{e}) := \min_{\mathcal{B} \in \mathbb{B}} \Pr(\mathbf{h}|\mathbf{e}) \text{ and } \overline{\Pr}(\mathbf{h}|\mathbf{e}) := \max_{\mathcal{B} \in \mathbb{B}} \Pr(\mathbf{h}|\mathbf{e}). \quad (5)$$

In order for these definitions to make sense, we require that $\Pr(\mathbf{e})$ is strictly positive for all $\mathcal{B} \in \mathbb{B}$.[9]

Computing the lower and upper probabilities in Equation (5) consists in optimising $\Pr(\mathbf{h}|\mathbf{e})$, which is a fraction of two multilinear functions of the network's parameters, under the constraint that each of the local distributions $\Pr(V|\boldsymbol{\pi})$ belongs to its credal set $\mathcal{M}(V|\boldsymbol{\pi})$. Obtaining the exact solution is only possible in special cases [13, 17] or for networks that are sufficiently small. Therefore, applied work on credal networks often resorts to approximate algorithms; see Reference [1] for a recent overview.

Most of the existing algorithms exploit the fact that for the purposes of computing $\underline{\Pr}(\mathbf{h}|\mathbf{e})$ or $\overline{\Pr}(\mathbf{h}|\mathbf{e})$, the—closed and convex—credal sets $\mathcal{M}(V|\boldsymbol{\pi})$ can be replaced by their set of extreme points[10] $\text{ext}(\mathcal{M}(V|\boldsymbol{\pi}))$ [13]:

$$\underline{\Pr}(\mathbf{h}|\mathbf{e}) = \min_{\mathcal{B} \in \mathbb{B}_{\text{ext}}} \Pr(\mathbf{h}|\mathbf{e}) \text{ and } \overline{\Pr}(\mathbf{h}|\mathbf{e}) = \max_{\mathcal{B} \in \mathbb{B}_{\text{ext}}} \Pr(\mathbf{h}|\mathbf{e}),$$

where $\mathbb{B}_{\text{ext}}$ is the set of Bayesian networks in $\mathbb{B}$ whose local distributions $\Pr(V|\boldsymbol{\pi})$ take values in $\text{ext}(\mathcal{M}(V|\boldsymbol{\pi}))$, defined by

$$\mathcal{B} \in \mathbb{B}_{\text{ext}} \Leftrightarrow (\forall V \in \mathbf{V})(\forall \boldsymbol{\pi} \in \pi) \; \Pr(V|\boldsymbol{\pi}) \in \text{ext}(\mathcal{M}(V|\boldsymbol{\pi})).$$

The reason why this is useful is because in practice, credal sets are usually *finitely generated*, meaning that they are defined by means of a finite number of linear constraints, or equivalently, that they have a finite number of extreme points. For example, the credal set $\mathcal{M}(G)$ of Equation (3) has two extreme points, which are characterised by $\Pr(g) = 1/3$ and $\Pr(g) = 2/3$, respectively. Similarly, the credal set $\mathcal{M}(E|g)$ in Figure 2 has four extreme points $\Pr(E|g) = (\Pr(e|g), \Pr(e'|g), \Pr(e''|g))$:

$$(\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}), \; (\tfrac{2}{3}, \tfrac{1}{3}, 0), \; (\tfrac{1}{3}, \tfrac{2}{3}, 0) \text{ and } (\tfrac{1}{6}, \tfrac{2}{3}, \tfrac{1}{6}). \quad (6)$$

If all the local credal sets are finitely generated, $\mathbb{B}_{\text{ext}}$ is finite, and computing lower and upper probabilities is then a combinatoric optimisation problem, the size of which is exponential in the number of local credal sets that are imprecise.

4.2 Reducing imprecise credal sets to precise ones

As we will now show, in many cases, we can use the sensitivity properties of Section 3.2 to reduce the computational complexity of inference in credal networks.

Basically, the idea is that, for the purposes of computing $\underline{\Pr}(\mathbf{h}|\mathbf{e})$ or $\overline{\Pr}(\mathbf{h}|\mathbf{e})$, some of the local credal sets of a credal network can be replaced by a singleton, that is, some imprecise local credal sets can be replaced by precise ones, without changing the result of the inference. In this way, the size of the combinatoric optimisation problem that needs to be solved can be reduced in advance, as a preprocessing step, before applying an inference algorithm of choice.

We introduce four types of local credal sets for which this is the case. Their definitions are completely analogous to the local distribution types that were defined in Section 3.3. For example, a local credal set $\mathcal{M}(V|\boldsymbol{\pi})$ is of Type 0 if and only if its distributions are of Type 0, and similarly for credal sets that are of type I, II or III.

---

[8] There are also other ways in which the local credal sets of a credal network can be combined to form a global model, depending on the type of credal network that is considered [11]; our definition corresponds to what is called a credal network under complete independence. Our results also apply to credal networks under strong independence, which replace $\mathbb{B}$ by its convex hull, but not to credal networks under epistemic irrelevance [5, 7, 9] or epistemic independence [8].

[9] For a detailed discussion on how to define $\underline{\Pr}(\mathbf{h}|\mathbf{e})$ and $\overline{\Pr}(\mathbf{h}|\mathbf{e})$ if $\Pr(\mathbf{e})$ is zero for some—but not all—$\mathcal{B}$ in $\mathbb{B}$, see Reference [6].

[10] An extreme point of a set is a point that cannot be written as a proper convex combination of two other points of that set. If a set is a closed and convex subset of a finite-dimensional space, then by Minkowski's finite-dimensional version of the Krein-Milman theorem, it is the convex hull of its extreme points.

Definition 5 (Local credal set types) *Consider a credal network $\mathbb{B}$ and a lower or upper probability of interest $\underline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$ or $\overline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$. A local credal set $\mathcal{M}(V\,|\,\boldsymbol{\pi})$ is then said to be of*

Type 0   *if $\boldsymbol{\pi} \nsim \mathbf{e}$;*

Type I   *if $\boldsymbol{\pi} \sim \mathbf{e}$, $V \in \mathbf{E}_{|\cancel{R}}$ and $\boldsymbol{\pi} \sim \mathbf{h}$;*

Type II   *if $\boldsymbol{\pi} \sim \mathbf{e}$, $V \in \mathbf{E}$ and $\boldsymbol{\pi} \nsim \mathbf{h}$;*

Type III   *if $\boldsymbol{\pi} \sim \mathbf{e}$, $\boldsymbol{\pi} \sim \mathbf{h}$, $V \in \mathbf{H}_{|\cancel{R}}$ and $|V| = 2$.*

Since the elements of a credal set $\mathcal{M}(V\,|\,\boldsymbol{\pi})$ of Type 0 are distributions of Type 0, it follows from Proposition 3 that this credal set is of no influence to the inference at hand. The following result is an immediate consequence of this fact.

Proposition 4 *Consider a credal network $\mathbb{B}$ and a lower or upper probability of interest $\underline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$ or $\overline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$. Then replacing a credal set $\mathcal{M}(V\,|\,\boldsymbol{\pi})$ of Type 0 with a singleton $\mathcal{M}'(V\,|\,\boldsymbol{\pi}) := \{\mathrm{Pr}(V\,|\,\boldsymbol{\pi})\}$ that consists of an arbitrary probability mass function $\mathrm{Pr}(V\,|\,\boldsymbol{\pi})$ on V will not change the result of the inference.*

Note that $\mathrm{Pr}(V\,|\,\boldsymbol{\pi})$ is not required to be an element of $\mathcal{M}(V\,|\,\boldsymbol{\pi})$, and that it can therefore be chosen as simple as possible.

As we are about to show, similar results can also be obtained for credal sets of Type I, II and III. However, in those cases, the distribution $\mathrm{Pr}(V\,|\,\boldsymbol{\pi})$ can no longer be chosen arbitrarily. We start with credal sets of Type I or II.

Since a distribution in a credal set of Type I or II is itself of Type I or II, we infer from the discussion in Section 3.3 that, for the purposes of computing $\mathrm{Pr}(\mathbf{h}\,|\,\mathbf{e})$, the only part of this distribution that is relevant is the parameter $\mathrm{Pr}(v\,|\,\boldsymbol{\pi})$, with v the unique instantiation of V that is compatible with $\mathbf{e}$. Furthermore, because of the monotone effect of changing this parameter—see Theorem 1—we know that the minimum and maximum value of $\mathrm{Pr}(\mathbf{h}\,|\,\mathbf{e})$ will be obtained by a Bayesian network $\mathcal{B} \in \mathbb{B}_{\mathrm{ext}}$ for which the parameter $\mathrm{Pr}(v\,|\,\boldsymbol{\pi})$ is minimal or maximal, in the sense that it is equal to the local lower probability

$$\underline{\mathrm{Pr}}(v\,|\,\boldsymbol{\pi}) := \min_{\mathrm{Pr}(V\,|\,\boldsymbol{\pi}) \in \mathcal{M}(V\,|\,\boldsymbol{\pi})} \mathrm{Pr}(v\,|\,\boldsymbol{\pi})$$

or the local upper probability

$$\overline{\mathrm{Pr}}(v\,|\,\boldsymbol{\pi}) := \max_{\mathrm{Pr}(V\,|\,\boldsymbol{\pi}) \in \mathcal{M}(V\,|\,\boldsymbol{\pi})} \mathrm{Pr}(v\,|\,\boldsymbol{\pi}),$$

and we can predict in advance which of these two extreme values it will be. Combined, these two observations allow us to replace the (possibly imprecise) credal set $\mathcal{M}(V\,|\,\boldsymbol{\pi})$ with a precise one. The following proposition makes this explicit.

Proposition 5 *Consider a credal network $\mathbb{B}$ and a lower or upper probability of interest $\underline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$ or $\overline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$. A credal set $\mathcal{M}(V\,|\,\boldsymbol{\pi})$ of Type I or II can then be replaced with a singleton $\mathcal{M}'(V\,|\,\boldsymbol{\pi}) := \{\mathrm{Pr}(V\,|\,\boldsymbol{\pi})\}$, without changing the result of the inference. If we let v be the unique value of V that is compatible with $\mathbf{e}$, then for $\underline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$, $\mathrm{Pr}(V\,|\,\boldsymbol{\pi})$ can be any probability mass function on V such that*

$$\mathrm{Pr}(v\,|\,\boldsymbol{\pi}) = \begin{cases} \underline{\mathrm{Pr}}(v\,|\,\boldsymbol{\pi}) & \text{if } \mathcal{M}(V\,|\,\boldsymbol{\pi}) \text{ is of Type I;} \\ \overline{\mathrm{Pr}}(v\,|\,\boldsymbol{\pi}) & \text{if } \mathcal{M}(V\,|\,\boldsymbol{\pi}) \text{ is of Type II,} \end{cases}$$

and for $\overline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$, $\mathrm{Pr}(V\,|\,\boldsymbol{\pi})$ can be any probability mass function on V such that

$$\mathrm{Pr}(v\,|\,\boldsymbol{\pi}) = \begin{cases} \overline{\mathrm{Pr}}(v\,|\,\boldsymbol{\pi}) & \text{if } \mathcal{M}(V\,|\,\boldsymbol{\pi}) \text{ is of Type I;} \\ \underline{\mathrm{Pr}}(v\,|\,\boldsymbol{\pi}) & \text{if } \mathcal{M}(V\,|\,\boldsymbol{\pi}) \text{ is of Type II.} \end{cases}$$

Proof. We only prove the result for $\underline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$; the proof for $\overline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$ is completely analogous.

If the local credal set $\mathcal{M}(V\,|\,\boldsymbol{\pi})$ is of Type I, then because of Theorem 1, the minimum value of $\mathrm{Pr}(\mathbf{h}\,|\,\mathbf{e})$ is obtained for a local distribution $\mathrm{Pr}(V\,|\,\boldsymbol{\pi})$ of which the parameter $\mathrm{Pr}(v\,|\,\boldsymbol{\pi})$ is minimal—equal to $\underline{\mathrm{Pr}}(v\,|\,\boldsymbol{\pi})$. Furthermore, due to Proposition 3, the other parameters of this local distribution are irrelevant. Hence, for the purposes of computing $\underline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$, $\mathcal{M}(V\,|\,\boldsymbol{\pi})$ can be replaced by a precise credal set $\mathcal{M}'(V\,|\,\boldsymbol{\pi}) := \{\mathrm{Pr}(V\,|\,\boldsymbol{\pi})\}$ that consists of a single distribution $\mathrm{Pr}(V\,|\,\boldsymbol{\pi})$ on V, with $\mathrm{Pr}(v\,|\,\boldsymbol{\pi}) = \underline{\mathrm{Pr}}(v\,|\,\boldsymbol{\pi})$. If the local credal set $\mathcal{M}(V\,|\,\boldsymbol{\pi})$ is of Type II, it follows from Theorem 1 that the minimum value of $\mathrm{Pr}(\mathbf{h}\,|\,\mathbf{e})$ is obtained for a local distribution $\mathrm{Pr}(V\,|\,\boldsymbol{\pi})$ of which the parameter $\mathrm{Pr}(v\,|\,\boldsymbol{\pi})$ is maximal—equal to $\overline{\mathrm{Pr}}(v\,|\,\boldsymbol{\pi})$—and therefore, $\mathcal{M}(V\,|\,\boldsymbol{\pi})$ can again be replaced by a singleton that consists of a single distribution $\mathrm{Pr}(V\,|\,\boldsymbol{\pi})$, now with $\mathrm{Pr}(v\,|\,\boldsymbol{\pi}) = \overline{\mathrm{Pr}}(v\,|\,\boldsymbol{\pi})$. $\qquad\square$

Note that $\mathrm{Pr}(V\,|\,\boldsymbol{\pi})$ is again not required to be an element of $\mathcal{M}(V\,|\,\boldsymbol{\pi})$, which allows us to choose it as simple as possible.

Consider for example the network in Figure 1, and assume that we want to compute $\underline{\mathrm{Pr}}(ghk\,|\,ef)$. The local credal set $\mathcal{M}(E\,|\,g)$—defined in Equation (4) and depicted in Figure 2—is then of Type I and can therefore be replaced by a singleton $\mathcal{M}'(E\,|\,g) = \{\mathrm{Pr}(E\,|\,g)\}$, where $\mathrm{Pr}(E\,|\,g)$ can be any probability mass function on $\{e, e', e''\}$ such that

$$\mathrm{Pr}(e\,|\,g) := \underline{\mathrm{Pr}}(e\,|\,g) = \frac{1}{6}.$$

The values of $\mathrm{Pr}(e'\,|\,g)$ and $\mathrm{Pr}(e''\,|\,g)$ are of no importance. If we want $\mathrm{Pr}(E\,|\,g)$ to belong to $\mathcal{M}(E\,|\,g)$, then the only possible choice is $\mathrm{Pr}(e'\,|\,g) = 2/3$ and $\mathrm{Pr}(e''\,|\,g) = 1/6$, which corresponds to one of the extreme points in Equation (6). However, this is not necessary. For example, choosing $\mathrm{Pr}(e'\,|\,g) = 0$ and $\mathrm{Pr}(e''\,|\,g) = 5/6$ works equally well.

Finally, for local credal sets that are of Type III, we obtain a result that is very similar to that of Proposition 5.

Proposition 6 *Consider a credal network $\mathbb{B}$ and a lower or upper probability of interest $\underline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$ or $\overline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$. A credal set $\mathcal{M}(V\,|\,\boldsymbol{\pi})$ of Type III can then be replaced with a singleton $\mathcal{M}'(V\,|\,\boldsymbol{\pi}) := \{\mathrm{Pr}(V\,|\,\boldsymbol{\pi})\}$, without changing the result of the inference. For $\underline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$, $\mathrm{Pr}(V\,|\,\boldsymbol{\pi})$ is defined by*

$$\mathrm{Pr}(v\,|\,\boldsymbol{\pi}) := \begin{cases} \underline{\mathrm{Pr}}(v\,|\,\boldsymbol{\pi}) & \text{if } v \sim \mathbf{h}; \\ \overline{\mathrm{Pr}}(v\,|\,\boldsymbol{\pi}) & \text{if } v \nsim \mathbf{h}, \end{cases} \tag{7}$$

and for $\overline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$, $\mathrm{Pr}(V\,|\,\boldsymbol{\pi})$ is defined by

$$\mathrm{Pr}(v\,|\,\boldsymbol{\pi}) := \begin{cases} \overline{\mathrm{Pr}}(v\,|\,\boldsymbol{\pi}) & \text{if } v \sim \mathbf{h}; \\ \underline{\mathrm{Pr}}(v\,|\,\boldsymbol{\pi}) & \text{if } v \nsim \mathbf{h}. \end{cases}$$

Proof. We only prove the result for $\underline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$; the proof for $\overline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$ is completely analogous.

Let v be the unique value of V that is compatible with $\mathbf{h}$. If the local credal set $\mathcal{M}(V\,|\,\boldsymbol{\pi})$ is of Type III, then because of Theorem 1, the minimum value of $\mathrm{Pr}(\mathbf{h}\,|\,\mathbf{e})$ is obtained for a local distribution $\mathrm{Pr}(V\,|\,\boldsymbol{\pi})$ of which the parameter $\mathrm{Pr}(v\,|\,\boldsymbol{\pi})$ is minimal—equal to $\underline{\mathrm{Pr}}(v\,|\,\boldsymbol{\pi})$—and the parameter $\mathrm{Pr}(\overline{v}\,|\,\boldsymbol{\pi})$ is maximal. The unique distribution for which this is the case is given by Equation (7). Hence, for the purposes of computing $\underline{\mathrm{Pr}}(\mathbf{h}\,|\,\mathbf{e})$, $\mathcal{M}(V\,|\,\boldsymbol{\pi})$ can be replaced by the singleton $\mathcal{M}'(V\,|\,\boldsymbol{\pi}) := \{\mathrm{Pr}(V\,|\,\boldsymbol{\pi})\}$. $\qquad\square$

Note that in this case, $\Pr(V\,|\,\boldsymbol{\pi})$ is unique and belongs to $\mathcal{M}(V\,|\,\boldsymbol{\pi})$.

For example, if we again consider the network in Figure 1, and revisit the problem of computing $\underline{\Pr}(ghk\,|\,ef)$, then the local credal set $\mathcal{M}(G)$—defined in Equation (3)—is of Type III and can therefore be replaced by a singleton $\mathcal{M}'(G) = \{\Pr(G)\}$, with

$$\Pr(g) := \underline{\Pr}(g) = {}^1\!/_3 \text{ and } \Pr(g) := \overline{\Pr}(g) = {}^2\!/_3.$$

In our example, in total, eight out of the fifteen credal sets of the credal network in Figure 1 can be replaced by a singleton, whereas other preprocessing techniques—such as the removal of barren or d-separated nodes—are not able to simplify the problem.

4.3 Special cases

Since a local credal set is of a given type if and only if its elements are, the special cases that were discussed in Section 3.4 carry over to credal networks. For example, for a credal network classifier with full evidence, the credal sets that correspond to the evidence variables are all of Type 0, I or II, and therefore, they can be replaced by singletons. It is essentially this feature that allows for tractable computations in basic credal classifiers such as the naive credal classifier [26] and the tree-augmented naive classifier [27]. In fact, many of the formulas in References [26] and [27] can be shown to follow from our results.

In the special case where we do not only have full evidence, but also know that each of the hypothesis variables $V \in \mathbf{H}$ is binary and has no parents in $\mathbf{H}$, then all the local credal sets of the credal network can be classified as one of the four types in Definition 5. This has far-reaching consequences: in this case, due to the results in the previous section, it follows that for the purpose of computing $\underline{\Pr}(\mathbf{h}\,|\,\mathbf{e})$ or $\overline{\Pr}(\mathbf{h}\,|\,\mathbf{e})$, all the local credal sets can be replaced by singletons, without changing the result of the inference. Hence, in this case, computing $\underline{\Pr}(\mathbf{h}\,|\,\mathbf{e})$ or $\overline{\Pr}(\mathbf{h}\,|\,\mathbf{e})$ reduces to the problem of computing $\Pr(\mathbf{h}\,|\,\mathbf{e})$ in a specific Bayesian network, the parameters of which can easily be determined by means of the methods in Section 4.2.

For example, in this special case, for the purpose of computing $\underline{\Pr}(\mathbf{h}\,|\,\mathbf{e})$, the local credal sets of a credal network can be replaced by precise distributions $\Pr_\bullet(V\,|\,\boldsymbol{\pi})$, the parameters of which are defined by

$$\Pr_\bullet(v\,|\,\boldsymbol{\pi}) := \begin{cases} \underline{\Pr}(v\,|\,\boldsymbol{\pi}) & \text{if } V \in \mathbf{H} \text{ and } v \sim \mathbf{h} \\ \overline{\Pr}(v\,|\,\boldsymbol{\pi}) & \text{if } V \in \mathbf{H} \text{ and } v \nsim \mathbf{h} \\ \underline{\Pr}(v\,|\,\boldsymbol{\pi}) & \text{if } V \in \mathbf{E}, v \sim \mathbf{e} \text{ and } \boldsymbol{\pi} \sim \mathbf{h} \\ \overline{\Pr}(v\,|\,\boldsymbol{\pi}) & \text{if } V \in \mathbf{E}, v \sim \mathbf{e} \text{ and } \boldsymbol{\pi} \nsim \mathbf{h} \\ ... & \text{if } V \in \mathbf{E} \text{ and } v \nsim \mathbf{e} \end{cases} \quad (8)$$

This definition does not distinguish between the cases $\boldsymbol{\pi} \sim \mathbf{e}$ and $\boldsymbol{\pi} \nsim \mathbf{e}$ because we know from Proposition 4 that a parameter for which $\boldsymbol{\pi} \nsim \mathbf{e}$ does not effect the result of the inference anyway. Similarly, the definition for the case with $V \in \mathbf{E}$ and $v \nsim \mathbf{e}$ is not provided because we know from Proposition 5 that these parameters do not influence the result of the inference either.

As an immediate consequence, for this special case, we obtain the following expression:

$$\underline{\Pr}(\mathbf{h}\,|\,\mathbf{e}) = \frac{\prod_{V \in \mathbf{V}} \Pr_\bullet(v\,|\,\boldsymbol{\pi})\big|_{v\boldsymbol{\pi}\sim\mathbf{he}}}{\sum_{\mathbf{h}^* \in \mathbf{H}} \prod_{V \in \mathbf{V}} \Pr_\bullet(v\,|\,\boldsymbol{\pi})\big|_{v\boldsymbol{\pi}\sim\mathbf{h}^*\mathbf{e}}}$$

Furthermore, by removing the parameters $\Pr_\bullet(v\,|\,\boldsymbol{\pi})$ for which V and its parents $\boldsymbol{\pi}$ belong to $\mathbf{E}$—since these are common factors of

the numerator and denominator—and then explicitly applying Equation (8) to the numerator, this expression can be simplified even more:

$$\underline{\Pr}(\mathbf{h}\,|\,\mathbf{e}) = \frac{\prod_{V \in \mathbf{K}} \underline{\Pr}(v\,|\,\boldsymbol{\pi})\big|_{v\boldsymbol{\pi}\sim\mathbf{he}}}{\sum_{\mathbf{h}^* \in \mathbf{H}} \prod_{V \in \mathbf{K}} \Pr_\bullet(v\,|\,\boldsymbol{\pi})\big|_{v\boldsymbol{\pi}\sim\mathbf{h}^*\mathbf{e}}}$$

where $\mathbf{K}$ is the set of variables that consist of the variables in $\mathbf{H}$ and their children variables.

Similar conclusions can be reached for the problem of computing $\overline{\Pr}(\mathbf{h}\,|\,\mathbf{e})$. In the special case of Section 3.4, this problem can again be reduced to computing $\Pr(\mathbf{h}\,|\,\mathbf{e})$ in a single Bayesian network. However, of course, it might not be the same Bayesian network as the one that is used to compute $\underline{\Pr}(\mathbf{h}\,|\,\mathbf{e})$. In this case, the local credal sets can be replaced by distributions $\Pr^\bullet(V\,|\,\boldsymbol{\pi})$, the parameters of which are defined by

$$\Pr^\bullet(v\,|\,\boldsymbol{\pi}) := \begin{cases} \overline{\Pr}(v\,|\,\boldsymbol{\pi}) & \text{if } V \in \mathbf{H} \text{ and } v \sim \mathbf{h} \\ \underline{\Pr}(v\,|\,\boldsymbol{\pi}) & \text{if } V \in \mathbf{H} \text{ and } v \nsim \mathbf{h} \\ \overline{\Pr}(v\,|\,\boldsymbol{\pi}) & \text{if } V \in \mathbf{E}, v \sim \mathbf{e} \text{ and } \boldsymbol{\pi} \sim \mathbf{h} \\ \underline{\Pr}(v\,|\,\boldsymbol{\pi}) & \text{if } V \in \mathbf{E}, v \sim \mathbf{e} \text{ and } \boldsymbol{\pi} \nsim \mathbf{h} \\ ... & \text{if } V \in \mathbf{E} \text{ and } v \nsim \mathbf{e} \end{cases}$$

5 CONCLUSIONS AND FUTURE WORK

A credal network represents a set of Bayesian networks thereby allowing for imprecisions in its parametrisation. One of the main computational problems in a credal network is the computation of lower and upper probabilities. Given a brute force approach, however, this requires a number of Bayesian network inferences that is exponential in the number of imprecisely specified local probability distributions.

The pruning of all variables irrelevant for a specific problem is a well-known approach to make inference more tractable. In this paper we proposed a different type of preprocessing step. Using Bayesian network sensitivity functions, we proved that for certain categories of parameters, we can predict the qualitative effect of their change on an outcome probability without needing any knowledge of the numerical specification of the network. This result allows for the identification of imprecisely specified local distributions that can be replaced by precisely specified ones, without affecting the outcome of the computation. Depending on the structure of the network and the specific problem at hand, our preprocessing step can be quite rewarding. We argued, for example, that for some classes of networks, even all imprecisely specified local distributions can be replaced by precise ones. In this case, credal network inference reduces to a single Bayesian network computation.

In future work, we would like to investigate the empirical impact of our preprocessing step on various existing algorithms. Moreover, we would like to use our results as a basis for the design of new algorithms. For example, we foresee that the algorithm in [13] could be extended to more general cases. Finally, we would like to extend our work to other types of inference, such as the computation of lower and upper expectations, choosing the most likely hypothesis, and maximising expected utility.

REFERENCES

[1] A. Antonucci, C. P. de Campos, M. Zaffalon. 2014. Probabilistic graphical models. In T. Augustin, F. P. A. Coolen, G. De Cooman, M. C. M. Troffaes (eds.), *Introduction to Imprecise Probabilities*, 207-229, John Wiley & Sons, Chichester.

[2] T. Augustin, F. P. A. Coolen, G. De Cooman, M. C. M. Troffaes (eds.). 2014. *Introduction to Imprecise Probabilities.* John Wiley & Sons.

[3] J. O. Berger. 1985. *Statistical decision theory and Bayesian analysis.* Springer series in statistics, Springer, New York.

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

[5] J. De Bock, G. De Cooman. 2015. Credal networks under epistemic irrelevance: the sets of desirable gambles approach. *International Journal of Approximate Reasoning* 56, 178207.

[6] J. De Bock, G. De Cooman. 2015. Conditioning, updating and lower probability zero. *International Journal of Approximate Reasoning* 67, 1-36.

[7] J. De Bock. 2015. *Credal networks under epistemic irrelevance: theory and algorithms*, PhD thesis, Ghent University, Faculty of Engineering and Architecture.

[8] C. P. de Campos, F. G. Cozman. 2007. Computing lower and upper expectations under epistemic independence. *International Journal of Approximate Reasoning* 44(3), 244-260.

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

[10] V. M. H. Coupé, L. C. van der Gaag. 2002. Properties of sensitivity analysis of Bayesian belief networks. *Annals of Mathematics and Artificial Intelligence*, 36(4), 323-356.

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

[12] F. G. Cozman. 2005. Graphical models for imprecise probabilities. *International Journal of Approximate Reasoning* 39(2), 167-184.

[13] E. Fagiuoli and M. Zaffalon. 1998. 2U: an exact interval propagation algorithm for polytrees with binary variables. *Artificial Intelligence*, 106(1), 77-107.

[14] M. J. Flores, J. A. Gámez, A. M Martínez. 2012. Supervised classification with Bayesian networks: A review on models and applications. *Intelligent Data Analysis for Real-Life Applications: Theory and Practice*, IGI Global, 72-102.

[15] F. V. Jensen, T. D. Nielsen. 2007. *Bayesian Networks and Decision Graphs*, second edition, Springer Verlag.

[16] U. Kjærulff, L. C. van der Gaag. 2000. Making sensitivity analysis computationally efficient. In C. Boutilier, M. Goldszmidt (eds.), *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, 317-325, Morgan Kaufmann Publishers, San Francisco.

[17] D. D. Mauá, C. P. de Campos, A. Benavoli, A. Antonucci. 2014. Probabilistic inference in credal networks: new complexity results. *Journal of Artificial Intelligence Research* 50, 603-637.

[18] M. Meekes, S. Renooij, L. C. van der Gaag. 2015. *Relevance of evidence in Bayesian networks*. In S. Destercke, T. Denoeux (eds.), *Proceedings of the 13th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, LNAI 9161, Springer, 366-375.

[19] J. Pearl. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Palo Alto.

[20] A. Piatti, A. Antonucci, M. Zaffalon. 2010. Building knowledge-based systems by credal networks: a tutorial. In A. R. Baswell (ed.), *Advances in Mathematics Research* 11, 227-279, Nova Science Publishers.

[21] S. Renooij. 2014. Co-variation for sensitivity analysis in Bayesian networks: Properties, consequences and alternatives. *International Journal of Approximate Reasoning*, 55, 1022-1042.

[22] R. Shachter. 1998. Bayes-Ball: The rational pastime (for determining irrelevance and requisite information in belief networks and influence diagrams). In G. Cooper, S. Moral (eds.), *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, 480-487, Morgan Kaufmann, San Francisco.

[23] P. Walley. 1991. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, New York.

[24] P. Walley. 1996. Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B* 58, 3-57.

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

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

[27] M. Zaffalon and E. Fagiuoli. 2003. Tree-Based Credal Networks for Classification. *Reliable Computing*, 9, 487-509.