

# Structure-based Categorisation of Bayesian Network Parameters

Janneke H. Bolt and S. Renooij

Department of Information and Computing Sciences, Utrecht University  
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands  
{j.h.bolt,s.renooij}@uu.nl

**Abstract.** Bayesian networks typically require thousands of probability parameters for their specification, many of which are bound to be inaccurate. Knowledge of the direction of change in an output probability of a network occasioned by changes in one or more of its parameters, i.e. the qualitative effect of parameter changes, has been shown to be useful both for parameter tuning and in pre-processing for inference in credal networks. In this paper we identify classes of parameter for which the qualitative effect on a given output of interest can be identified based upon graphical considerations.

## 1 Introduction

A *Bayesian network* defines a unique joint probability distribution over a set of discrete random variables [8]. It combines an acyclic directed graph, representing the independences among the variables, with a quantification of local discrete distributions. The individual probabilities of these local distributions are called the parameters of the network. A Bayesian network can be used to infer any probability from the represented distribution.

The effect of possible parameter inaccuracies on the output probabilities of a network can be studied with a sensitivity analysis. An output can be described as a fraction of two functions that are linear in any network parameter; the coefficients of the functions are determined by the non-varied parameters [6]. Depending on the coefficients, such so-called sensitivity functions are either monotone increasing or monotone decreasing functions in each parameter. Interestingly, as we showed in previous research, for some outputs and parameters, the sensitivity function is even always increasing (or decreasing) regardless of the specific values of the other network parameters [1]. That is, regardless of the specific quantification of a network, the coefficients of the sensitivity function for a certain parameter can be such that the gradient is always positive (or always negative). In such a case, the qualitative effect of a parameter change on an output probability can be predicted from properties of the network structure, without considering the values of the other network parameters.

Knowledge of the qualitative effect of parameter changes can be exploited for different purposes. Examples of applications can be found in pre-processing for inference in credal networks [1] and in multi-parameter tuning of Bayesian networks [2].

In this paper we present a complete categorisation of a network’s parameters with respect to their qualitative effect on some output, where we assume that the network is pruned before hand to a sub-network that is computationally relevant to the output. The paper extends the work in [1] in which only a partial categorisation of the network parameters was given. Compared to our previous results, the present results also enable a *meaningful* categorisation for a wider range of parameters.

## 2 Preliminaries

### 2.1 Bayesian Networks and Notation

A Bayesian network  $\mathcal{B} = (G, \text{Pr})$  represents a joint probability distribution  $\text{Pr}$  over a set of random variables  $\mathbf{W}$  as a factorisation of conditional distributions [5]. The independences underlying this factorisation are read from the directed acyclic graph  $G$  by means of the well-known d-separation criterion. In this paper we use upper case  $W$  to denote a single random variable, writing lowercase  $w \in W$  to indicate a value of  $W$ . For binary-valued  $W$ , we use  $w$  and  $\bar{w}$  to denote its two possible value assignments. Boldfaced capitals are used to indicate sets of variables or sets of value assignments, the distinction will be clear from the context; boldface lower cases are used to indicate a joint value assignment to a set of variables.

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  $\not\sim$ . We use  $\mathbf{W}_{pa(V)} = \mathbf{W} \cap pa(V)$  to indicate the subset of  $\mathbf{W}$  that is among the parents of  $V$ , and  $\mathbf{W}_{\bar{pa}(V)} = \mathbf{W} \setminus \mathbf{W}_{pa(V)}$  to indicate its complement in  $\mathbf{W}$ ; descendants of  $V$  are captured by  $de(V)$ . To conclude,  $\langle \mathbf{T}, \mathbf{U} | \mathbf{V} \rangle_d$ ,  $\mathbf{T}, \mathbf{U}, \mathbf{V} \subseteq \mathbf{W}$ , denotes that all variables in  $\mathbf{T}$  are d-separated from all variables in  $\mathbf{U}$  given the variables in  $\mathbf{V}$ , where we assume that  $\langle \mathbf{T}, \emptyset | \mathbf{V} \rangle_d = \text{True}$ .

A Bayesian network specifies for each variable  $W \in \mathbf{W}$  exactly one local distribution  $\text{Pr}(W | \pi)$  over the values of  $W$  per value assignment  $\pi$  to its parents  $pa(W)$  in  $G$ , such that

$$\text{Pr}(\mathbf{w}) = \prod_{W \in \mathbf{W}} \text{Pr}(w | \pi) \Big|_{w\pi \sim \mathbf{w}}$$

where the notation  $\Big|_{\text{prop}}$  is used to indicate the properties the arguments in the preceding formula adhere to. The individual probabilities in the local distributions are termed the network’s *parameters*. A Bayesian network allows for

computing any probability over its variables  $\mathbf{W}$ . A typical query is  $\Pr(\mathbf{h}|\mathbf{f})$ , involving two disjoint subsets of  $\mathbf{W}$ , often referred to as hypothesis variables ( $\mathbf{H}$ ) and evidence variables ( $\mathbf{F}$ );  $\mathbf{W}$  can also include variables not involved in the output query of interest.

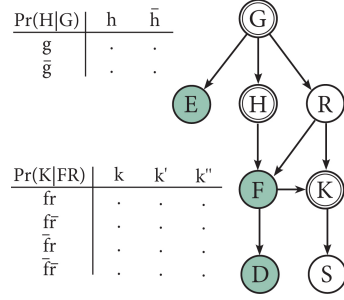


Fig. 1: An example Bayesian network (with just two of its CPTs).

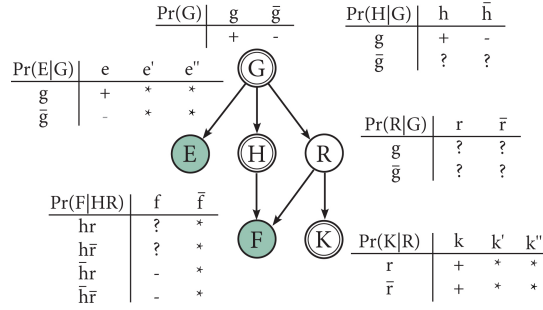


Fig. 2: The example network from Figure 1 after query dependent preprocessing given the output probability  $\Pr(ghk|def)$ .

An example Bayesian network is shown in Figure 1. For output  $\Pr(ghk|def)$  we identify hypothesis variables  $\mathbf{H} = \{G, H, K\}$  (double circles), evidence variables  $\mathbf{F} = \{D, E, F\}$  (shaded), and remaining variables  $\{R, S\}$ . In addition to the graph, conditional probability tables (CPTs) need to be specified for each node.

## 2.2 Query Dependent Pre-Processing

Prior to computing the result of a query, the Bayesian network can be pre-processed by removing parts of its specification that are easily identified as being irrelevant to the computations. Sets of nodes that can be removed based upon graphical considerations only are nodes d-separated from  $\mathbf{H}$  given  $\mathbf{F}$ , irrelevant evidence nodes (effects blocked by other evidence), and barren nodes, that is, nodes in  $\mathbf{W} \setminus (\mathbf{H} \cup \mathbf{F})$  which are leafs or have only barren descendants; the remaining nodes coincide with the so-called parameter sensitivity set [3,7]. In addition, evidence absorption can be applied, where the outgoing arcs of variables with evidence are removed and the CPTs of the former children are reduced by removing the parameters that are incompatible with the observed value(s) of their former parent(s) [4]. After evidence absorption, all variables with evidence correspond to leafs in the graph and the CPT parameters of their former children will all be compatible with the evidence.

From here on we consider Bayesian networks that are reduced to what we call its *query-dependent backbone*  $\mathcal{B}_q$ , using the above-mentioned pre-processing

options<sup>1</sup>.  $\mathcal{B}_q$ , tailored to the original query  $\Pr(\mathbf{h}|\mathbf{f})$ , is assumed to be a Bayesian network over variables  $\mathbf{V} \subseteq \mathbf{W}$  from which the now equivalent query  $\Pr(\mathbf{h}|\mathbf{e})$  is computed for evidence variables  $\mathbf{E} \subseteq \mathbf{F}$ ; the remaining variables  $\mathbf{V} \setminus (\mathbf{H} \cup \mathbf{E})$  will be denoted by  $\mathbf{R}$ .

The backbone given output  $\Pr(ghk|def)$  in our example network from Figure 1 is depicted in Figure 2. After evidence absorption, the arc from  $F$  to  $K$  and the last two rows of  $K$ 's CPT are removed. The node  $S$  is removed since it is barren, and  $D$  is removed, since it is d-separated from the variables in  $\mathbf{H}$  given  $\mathbf{F}$ . In the backbone network we have the hypothesis variables  $\mathbf{H} = \{G, H, K\}$ , the evidence variables  $\mathbf{E} = \{E, F\}$  and the remaining variable  $\mathbf{R} = \{R\}$ .

### 2.3 Relating Queries to Parameters

It is well-known that an output of a Bayesian network relates to a network parameter  $x$  as a fraction of two functions linear in that parameter:

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

where the constants  $\tau_1, \tau_2, \kappa_1$  and  $\kappa_2$  are composed of network parameters independent of  $x$  [3]. The above function can be generalised to multiple parameters [6] and is typically exploited in the context of sensitivity analysis, to determine how a change in one or more parameters affects  $\Pr(\mathbf{h}|\mathbf{e})$ . We note that 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 [9]. Sensitivity functions are monotonic functions in each parameter, and are either increasing or decreasing functions in such a parameter. Here we consider increasing (decreasing) in a non-strict sense, that is, increasing (decreasing) includes non-decreasing (non-increasing).

## 3 Categorisation of Parameters in a Backbone Network $\mathcal{B}_q$

We will discuss the parameters of the variables in  $\mathbf{R}$ ,  $\mathbf{H}$  and  $\mathbf{E}$  of a backbone network and categorise these parameters according to their qualitative effect on  $\Pr(\mathbf{h}|\mathbf{e})$  as summarised in Table 1. In the proofs of our propositions we repeatedly use the definition of conditional probability and the factorisation defined

<sup>1</sup> Note that the parameters of the local distributions that are in  $\mathcal{B}$  but not in  $\mathcal{B}_q$  do not affect the output of interest in any way.

by  $\mathcal{B}_q$ :

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

in which for the numerator  $\Pr(\mathbf{h}\mathbf{e})$  we find

$$\Pr(\mathbf{h}\mathbf{e}) = \sum_{\mathbf{r} \in \mathbf{R}} \Pr(\mathbf{r}\mathbf{h}\mathbf{e}) = \sum_{\mathbf{r} \in \mathbf{R}} \prod_{V^* \in \mathbf{V}} \Pr(v^* | \pi^*) \Big|_{v^* \pi^* \sim \mathbf{r}\mathbf{h}\mathbf{e}} \quad (1)$$

and for the denominator  $\Pr(\mathbf{e})$  we find that

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

Parameters which are not present in Equation (1) and (2) cannot affect the output directly and are categorised as ‘\*’. The effect of all other parameters is investigated by studying properties of their sensitivity functions  $\Pr(\mathbf{h}|\mathbf{e})(x)$ . Parameters that are guaranteed to give a monotone increasing sensitivity function are classified as ‘+’; parameters that are guaranteed to give monotone decreasing sensitivity functions as ‘-’. Parameters for which the sign of the derivative of the sensitivity function depends on the actual quantification of the network will be categorised as ‘?’’. Note that sensitivity functions for parameters of category ‘\*’ are not necessarily constant: *variation* of such a parameter may result in co-variation of a parameter from the same local distribution which *is* present in Equation (1) or (2). As such, parameters of category ‘\*’ may be indirectly affecting output  $\Pr(\mathbf{h}|\mathbf{e})$ , yet for its computation it suffices to know the values of parameters outside category ‘\*’ only. In the relation between parameter changes and output changes therefore these latter parameters are pivotal.

For the backbone network of our example in Figure 2, the categories of its parameters are indicated in the CPTs.

## 4 Categorisation of the Parameters of Variables in $\mathbf{R}$ and $\mathbf{H}$

### 4.1 Parameters of Variables in $\mathbf{R}$

For a variable  $R \in \mathbf{R}$  the qualitative effect of a change in one of its parameters  $x$  cannot be predicted without additional information: the sensitivity function  $\Pr(\mathbf{h}|\mathbf{e})(x)$  can be either monotone increasing or decreasing. Therefore, these parameters are categorised as ‘?’’.

Proof of the above claim, and of all further propositions concerning parameters in the category ‘?’’, is omitted due to space restrictions. All these proofs are based on demonstrating that additional knowledge—for example the specific network quantification—is required to determine whether the one-way sensitivity function is increasing or decreasing.

Table 1: Categorisation of parameters  $x = \Pr(v|\pi)$  of a Bayesian network  $\mathcal{B}_q$  with respect to the output probability  $\Pr(\mathbf{h}|\mathbf{e})$ .

$V \in \mathbf{R}$	cat '??'				
$V \in \mathbf{H}$	$\mathbf{E} \cap de(V) = \emptyset$	$v \sim \mathbf{h}, \pi \sim \mathbf{h}$	cat '+'		
		$v \approx \mathbf{h}$ or $\pi \approx \mathbf{h}$	cat '*'		
	$\mathbf{E} \cap de(V) \neq \emptyset$	non-binary $V$	cat '??'		
			$\pi \approx \mathbf{h}$	cat '??'	
		binary $V$	$\pi \sim \mathbf{h}$	$\neg \langle \mathbf{H}_{\bar{pa}(V)}, \mathbf{R}_{pa(V)}   \mathbf{E} \cup \mathbf{H}_{pa(V)} \cup V \rangle_d$	cat '??'
				$\langle \mathbf{H}_{\bar{pa}(V)}, \mathbf{R}_{pa(V)}   \mathbf{E} \cup \mathbf{H}_{pa(V)} \cup V \rangle_d$	$v \sim \mathbf{h}$ cat '+' $v \approx \mathbf{h}$ cat '-'
$V \in \mathbf{E}$	$v \approx \mathbf{e}$	cat '*'			
	$v \sim \mathbf{e}$	$\pi \approx \mathbf{h}$	cat '-'		
		$\pi \sim \mathbf{h}$	$\neg \langle \mathbf{H}_{\bar{pa}(V)}, \mathbf{R}_{pa(V)}   \mathbf{E} \cup \mathbf{H}_{pa(V)} \rangle_d$	cat '??'	
			$\langle \mathbf{H}_{\bar{pa}(V)}, \mathbf{R}_{pa(V)}   \mathbf{E} \cup \mathbf{H}_{pa(V)} \rangle_d$	cat '+'	

## 4.2 Parameters of Variables in $\mathbf{H}$ without Descendants in $\mathbf{E}$

The propositions in this section concern parameters  $x = \Pr(v|\pi)$  of nodes  $V \in \mathbf{H}$  without descendants in  $\mathbf{E}$ . The parameters of such a node which are fully compatible with  $\mathbf{h}$  have a monotone increasing sensitivity function  $\Pr(\mathbf{h}|\mathbf{e})(x)$  and therefore are classified as '+'. The parameters not fully compatible with  $\mathbf{h}$  are not used in the computation of  $\Pr(\mathbf{h}|\mathbf{e})$  and therefore are classified '\*'

**Proposition 1** Consider a query-dependent backbone Bayesian network  $\mathcal{B}_q$  with probability of interest  $\Pr(\mathbf{h}|\mathbf{e})$ . Let  $x = \Pr(v|\pi)$  be a parameter of a variable  $V \in \mathbf{H}$  such that  $de(V) \cap \mathbf{E} = \emptyset$ . If both  $v \sim \mathbf{h}$  and  $\pi \sim \mathbf{h}$ , then  $\Pr(\mathbf{h}|\mathbf{e})(x)$  is a monotone increasing function.

**Proof** Let  $\mathbf{r}_\pi$  denote the configuration of  $\mathbf{R}_{pa(V)}$  compatible with  $\pi$ . In addition, let  $\mathbf{h} = v\mathbf{h}_\pi\mathbf{h}_{\bar{\pi}}$ , where  $\mathbf{h}_\pi$  and  $\mathbf{h}_{\bar{\pi}}$  are assignments, compatible with  $\mathbf{h}$ , to  $\mathbf{H}_{pa(V)}$  and  $\mathbf{H}_{\bar{pa}(V)}$ , respectively. First consider the general form of  $\Pr(\mathbf{h}|\mathbf{e})$  given by Equation (1). We observe that under the given conditions we can write:

$$\Pr(\mathbf{h}|\mathbf{e})(x) = x \cdot \Pr(\mathbf{h}_{\bar{\pi}} | v\mathbf{h}_\pi\mathbf{e}\mathbf{r}_\pi) \cdot \frac{\Pr(v\mathbf{h}_\pi\mathbf{e}\mathbf{r}_\pi)}{\Pr(v|\pi)} + \sum_{\mathbf{r}^+ \in \mathbf{R}_{pa(V)}} \Pr(v\mathbf{h}_\pi\mathbf{h}_{\bar{\pi}}\mathbf{e}\mathbf{r}^+) \Big|_{\mathbf{r}^+ \neq \mathbf{r}_\pi}$$

where  $\Pr(v\mathbf{h}_\pi\mathbf{e}\mathbf{r}_\pi)/\Pr(v|\pi)$  represents a sum of products of parameters no longer including  $\Pr(v|\pi)$ . This expression thus is of the form  $x \cdot \tau_1 \cdot \tau_2 + \tau_3$ , for non-negative constants  $\tau_1, \tau_2, \tau_3$ .

For  $\Pr(\mathbf{e})$ , as given in Equation (2), we observe that since  $V$  has no descendants in  $\mathbf{E}$ , this node in fact is barren with respect to  $\Pr(\mathbf{e})$ . As a result, none of  $V$ 's parameters are relevant to the computation and  $\Pr(\mathbf{e})(x)$  therefore equals a constant  $\kappa_1 > 0$ .

The sensitivity function for parameter  $x$  thus is of the form  $\Pr(\mathbf{h}|\mathbf{e})(x) = (x \cdot \tau_1 \cdot \tau_2 + \tau_3) / \kappa_1$  with  $\tau_1, \tau_2, \tau_3 \geq 0$  and  $\kappa_1 > 0$ . The first derivative of this function equals  $(\tau_1 \cdot \tau_2) / \kappa_1$ , which is always non-negative.  $\square$

**Proposition 2** *Let  $\mathcal{B}_q$  and  $\Pr(\mathbf{h}|\mathbf{e})$  be as before. Let  $\Pr(v|\pi)$  be a parameter of a variable  $V \in \mathbf{H}$  such that  $de(V) \cap \mathbf{E} = \emptyset$ . If  $v \approx \mathbf{h}$  or  $\pi \approx \mathbf{h}$ , then  $\Pr(v|\pi)$  is not used in the computation of  $\Pr(\mathbf{h}|\mathbf{e})$ .*

**Proof** We again consider  $\Pr(\mathbf{h}\mathbf{e})$  as given by Equation (1) and observe that a parameter  $\Pr(v|\pi)$  with  $v \approx \mathbf{h}$  or  $\pi \approx \mathbf{h}$  is not included in this expression. Moreover, as argued in the proof of Proposition 1, no parameter of  $V$  is used in computing  $\Pr(\mathbf{e})$  from Equation (2).  $\Pr(v|\pi)$  is therefore not used in the computation of  $\Pr(\mathbf{h}|\mathbf{e})$ .  $\square$

### 4.3 Parameters of Variables in $\mathbf{H}$ with Descendants in $\mathbf{E}$

For a parameter of a non-binary variable  $V \in \mathbf{H}$  with at least one descendant in  $\mathbf{E}$ , we cannot predict without additional knowledge whether the sensitivity function  $\Pr(\mathbf{h}|\mathbf{e})(x)$  is monotone increasing or monotone decreasing. These parameters therefore are classified as '?'. The same observation applies to a parameter of a binary variable  $V \in \mathbf{H}$  with descendants in  $\mathbf{E}$  for which  $\pi \approx \mathbf{h}$  or for which  $\mathbf{H}_{\overline{pa}(V)}$  and  $\mathbf{R}_{pa(V)}$  are not d-separated given  $\mathbf{H}_{pa(V)}$ ,  $V$  itself and the evidence.

If  $V \in \mathbf{H}$  is binary,  $\pi \sim \mathbf{h}$  and  $\mathbf{H}_{\overline{pa}(V)}$  and  $\mathbf{R}_{pa(V)}$  are d-separated given  $\mathbf{H}_{pa(V)}$ ,  $V$  itself and the evidence, then we do have sufficient knowledge to determine the qualitative effect of varying parameter  $x = \Pr(v|\pi)$  of  $V$ . If  $v \sim \mathbf{h}$  then  $\Pr(\mathbf{h}|\mathbf{e})(x)$  is monotone increasing, and the parameter is classified as '+'. If  $v \approx \mathbf{h}$  then  $\Pr(\mathbf{h}|\mathbf{e})(x)$  is monotone decreasing, and the parameter is classified as '-'. These observations are captured by Proposition 3. This proposition extends Proposition 2 in [1] by replacing the condition that  $\mathbf{R}_{pa(V)} = \emptyset$  by the less strict d-separation condition mentioned above.

**Proposition 3** *Let  $\mathcal{B}_q$  and  $\Pr(\mathbf{h}|\mathbf{e})$  be as before. Let  $\Pr(v|\pi)$  with  $\pi \sim \mathbf{h}$  be a parameter of a binary variable  $V \in \mathbf{H}$  and let  $\langle \mathbf{H}_{\overline{pa}(V)}, \mathbf{R}_{pa(V)} | \mathbf{E} \cup \mathbf{H}_{pa(V)} \cup V \rangle_d$ . If  $v \sim \mathbf{h}$  then  $\Pr(\mathbf{h}|\mathbf{e})(x)$  is a monotone increasing function; if  $v \approx \mathbf{h}$  then  $\Pr(\mathbf{h}|\mathbf{e})(x)$  is a monotone decreasing function.*

**Proof** First consider the case where  $v \sim \mathbf{h}$ . Under the given conditions we have from the proof of Proposition 1 that  $\Pr(\mathbf{h}\mathbf{e})$  takes on the form  $\Pr(\mathbf{h}\mathbf{e})(x) = x \cdot \tau_1 \cdot \tau_2 + \tau_3$ , for constants  $\tau_1, \tau_2, \tau_3 \geq 0$ .

For  $\Pr(\mathbf{e})$  and binary  $V$  we note that Equation (2) can be written as

$$\begin{aligned} \Pr(\mathbf{e})(x) = & x \cdot \frac{\Pr(v\mathbf{h}_\pi\mathbf{r}_\pi\mathbf{e})}{\Pr(v|\pi)} + (1-x) \cdot \frac{\Pr(\bar{v}\mathbf{h}_\pi\mathbf{r}_\pi\mathbf{e})}{\Pr(\bar{v}|\pi)} + \sum_{\mathbf{r}^+ \in \mathbf{R}_{pa(V)}} \Pr(v\mathbf{h}_\pi\mathbf{r}^+\mathbf{e}) \Big|_{\mathbf{r}^+ \neq \mathbf{r}_\pi} \\ & + \sum_{\mathbf{r}^+ \in \mathbf{R}_{pa(V)}} \Pr(\bar{v}\mathbf{h}_\pi\mathbf{r}^+\mathbf{e}) \Big|_{\mathbf{r}^+ \neq \mathbf{r}_\pi} + \sum_{\mathbf{h}^+ \in \mathbf{H}_{pa(V)}} \Pr(\mathbf{h}^+\mathbf{e}) \Big|_{\mathbf{h}^+ \neq \mathbf{h}_\pi} \end{aligned}$$

which takes on the following form:  $\Pr(\mathbf{e})(x) = x \cdot \tau_2 + (1-x) \cdot \kappa_2 + \kappa_1 + \kappa_3 + \kappa_4$ , with constants  $\kappa_i \geq 0$ ,  $i = 1, \dots, 4$ .

The sign of the derivative of the sensitivity function is determined by the numerator  $\tau_1 \cdot \tau_2 \cdot (\kappa_1 + \kappa_2 + \kappa_3 + \kappa_4) - \tau_3 \cdot (\tau_2 - \kappa_2)$  of  $\Pr(\mathbf{h}|\mathbf{e})'(x)$ . We observe that given  $\langle \mathbf{H}_{\bar{pa}(V)}, \mathbf{R}_{pa(V)} | \mathbf{E} \cup \mathbf{H}_{pa(V)} \cup V \rangle_d$  we find that  $\tau_1 \cdot \kappa_1 = \tau_3$  which guarantees the derivative to be non-negative. This implies that, for  $v \sim \mathbf{h}$ ,  $\Pr(\mathbf{h}|\mathbf{e})(x)$  is a monotone increasing function.

Now consider the case where  $v \sim \mathbf{h}$ . Since  $V$  is binary, this implies that  $\bar{v} \sim \mathbf{h}$ . The proof for this case follows by replacing, in the above formulas, every occurrence of  $v$  by  $\bar{v}$  and, hence, every  $x$  with  $1-x$ . As a result we find that in this case  $\Pr(\mathbf{h}|\mathbf{e})(x)$  is a monotone decreasing function.  $\square$

## 5 Categorisation of the Parameters of the Variables in $\mathbf{E}$

### 5.1 Parameters $\Pr(v|\pi)$ of a Variable $V \in \mathbf{E}$ with $v \sim \mathbf{e}$

Recall that after evidence absorption all parameters with  $\pi \sim \mathbf{e}$  are removed from the network. For a parameter  $\Pr(v|\pi)$  of a variable in  $V \in \mathbf{E}$ , however, we may still find that  $v \sim \mathbf{e}$ ; these parameters are in category '\*’.

**Proposition 4** *Let  $\mathcal{B}_q$  and  $\Pr(\mathbf{h}|\mathbf{e})$  be as before. Let  $\Pr(v|\pi)$  be a parameter of a variable  $V \in \mathbf{E}$ . If  $v \sim \mathbf{e}$ , then  $\Pr(v|\pi)$  is not used in the computation of  $\Pr(\mathbf{h}|\mathbf{e})$ .*

**Proof** This proposition is equivalent to Proposition 3 in [1], but stated for  $\mathcal{B}_q$  rather than for  $\mathcal{B}$ .  $\square$

### 5.2 Parameters $\Pr(v|\pi)$ of a Variable $V \in \mathbf{E}$ with $v \sim \mathbf{e}$ and $\pi \sim \mathbf{h}$

We now consider the parameters  $\Pr(v|\pi)$  of  $V \in \mathbf{E}$ , with  $v \sim \mathbf{e}$  and  $\pi \sim \mathbf{h}$ . The one-way sensitivity functions of such parameters are monotone decreasing. These parameters therefore are categorised as '–’.

**Proposition 5** *Let  $\mathcal{B}_q$  and  $\Pr(\mathbf{h}|\mathbf{e})$  be as before. Let  $x = \Pr(v|\pi)$  be a parameter of  $V \in \mathbf{E}$  such that  $v \sim \mathbf{e}$ . If  $\pi \sim \mathbf{h}$ , then  $\Pr(\mathbf{h}|\mathbf{e})(x)$  is a monotone decreasing function.*



**Proof** This proposition is equivalent to Proposition 1 in [1], but stated for  $\mathcal{B}_q$  rather than for  $\mathcal{B}$ .  $\square$

### 5.3 Parameters $\Pr(v|\pi)$ of a Variable $V \in \mathbf{E}$ with $v \sim \mathbf{e}$ and $\pi \sim \mathbf{h}$

We now consider parameters  $\Pr(v|\pi)$  of  $V \in \mathbf{E}$  with  $v \sim \mathbf{e}$  and  $\pi \sim \mathbf{h}$ . The one-way sensitivity functions of such parameters are monotone increasing under the condition that  $\mathbf{H}_{\overline{pa}(V)}$  is d-separated from  $\mathbf{R}_{pa(V)}$  given  $\mathbf{H}_{pa(V)}$  and the evidence. Under this condition, these parameters therefore can be categorised as '+'. This proposition extends Proposition 1 in [1] by replacing the condition that  $\mathbf{R}_{pa(V)} = \emptyset$  by the less strict d-separation condition mentioned above.

**Proposition 6** *Let  $\mathcal{B}_q$  and  $\Pr(\mathbf{h}|\mathbf{e})$  be as before. Let  $x = \Pr(v|\pi)$  with  $v \sim \mathbf{e}$  and  $\pi \sim \mathbf{h}$  be a parameter of  $V \in \mathbf{E}$  and let  $\langle \mathbf{H}_{\overline{pa}(V)}, \mathbf{R}_{pa(V)} | \mathbf{H}_{pa(V)} \cup \mathbf{E} \rangle_d$ . Then  $\Pr(\mathbf{h}|\mathbf{e})(x)$  is a monotone increasing function.*

**Proof** For  $\Pr(\mathbf{h}\mathbf{e})$  we observe that Equation (1) can be written as the expression in the proof of Proposition 1, but with  $v$  included in  $\mathbf{e}$  rather than in  $\mathbf{h}$ . We therefore have that  $\Pr(\mathbf{h}\mathbf{e})(x) = x \cdot \tau_1 \cdot \tau_2 + \tau_3$  for constants  $\tau_1, \tau_2, \tau_3 \geq 0$ .

For  $\Pr(\mathbf{e})$ , given by Equation (2), we observe that we can write

$$\Pr(\mathbf{e})(x) = x \cdot \frac{\Pr(\mathbf{h}\pi\mathbf{r}\pi\mathbf{e})}{\Pr(v|\pi)} + \sum_{\mathbf{r}^+ \in \mathbf{R}_{pa(V)}} \Pr(\mathbf{h}\pi\mathbf{e}\mathbf{r}^+) \Big|_{\mathbf{r}^+ \neq \mathbf{r}_\pi} + \sum_{\mathbf{h}^+ \in \mathbf{H}_{pa(V)}} \Pr(\mathbf{h}^+\mathbf{e}) \Big|_{\mathbf{h}^+ \neq \mathbf{h}_\pi}$$

which is of the form  $x \cdot \tau_2 + \kappa_1 + \kappa_2$ , for constants  $\kappa_1, \kappa_2 \geq 0$ .

We now find that the *numerator* of the first derivative of the sensitivity function equals  $\tau_1 \cdot \tau_2 \cdot (\kappa_1 + \kappa_2) - \tau_2 \cdot \tau_3$ . We observe that given  $\langle \mathbf{H}_{\overline{pa}(V)}, \mathbf{R}_{pa(V)} | \mathbf{H}_{pa(V)} \cup \mathbf{E} \rangle_d$  we find that  $\tau_1 \cdot \kappa_1 = \tau_3$  which guarantees the derivative to be non-negative. This implies that  $\Pr(\mathbf{h}|\mathbf{e})(x)$  is a monotone increasing function.  $\square$

In case the above mentioned d-separation property does not hold, we need additional information to predict whether the sensitivity function  $\Pr(\mathbf{h}|\mathbf{e})(x)$  of a parameter  $x$  of a variable  $V \in \mathbf{E}$  with  $v \sim \mathbf{e}$  and  $\pi \sim \mathbf{h}$  is monotone increasing or monotone decreasing, without additional knowledge. If the property does not hold, therefore, these parameters are in category '?'.  $\square$

## 6 Discussion

In this paper we presented fundamental results concerning the qualitative effect of parameter changes on the output probabilities of a Bayesian network. Based on the graph structure and the query at hand, we categorised all network

parameters into one of four categories: parameters not included in the computation of the output, parameters with guaranteed monotone increasing sensitivity functions, parameters with guaranteed monotone decreasing sensitivity functions, and parameters of which the qualitative effect cannot be predicted without additional information. Previously we demonstrated that knowledge of the qualitative effects can be exploited in inference in credal networks [1] and in multiple-parameter tuning of Bayesian networks [2]. In our previous research only a partial categorisation of the parameters was given. Our present paper allocates a wider range of parameters into one of the meaningful categories '+', '-', and '\*'. For future research we would like to further study properties of the additional information required to predict the qualitative effect of the parameters in category '?'.

### Acknowledgements

This research was supported by the Netherlands Organisation for Scientific Research (NWO).

### References

1. J.H. Bolt, J. De Bock and S. Renooij. 2016. Exploiting Bayesian network sensitivity functions for inference in credal networks. *Proceedings of the 22nd European Conference on Artificial Intelligence*, 285, 646 - 654.
2. J.H. Bolt and L.C. van der Gaag. 2017. Balanced sensitivity functions for tuning multi-dimensional Bayesian network classifiers. *International Journal of Approximate Reasoning*, 80c, 361-376.
3. 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.
4. L.C. van der Gaag. 1996. On evidence absorption for belief networks *International Journal of Approximate Reasoning*, 15(3), 265-286.
5. F. V. Jensen, T. D. Nielsen. 2007. *Bayesian Networks and Decision Graphs*, second edition, Springer Verlag.
6. 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.
7. M. Meeke, 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.
8. J. Pearl. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Palo Alto.
9. S. Renooij. 2014. Co-variation for sensitivity analysis in Bayesian networks: Properties, consequences and alternatives. *International Journal of Approximate Reasoning*, 55, 1022-1042.