Properties of Sensitivity Analysis of Bayesian Belief Networks*

Veerle M.H. Coupé
Center for Clinical Decision Sciences, Erasmus University Rotterdam,
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
coupe@mgz.fgg.eur.nl

Linda C. van der Gaag
Utrecht University, Department of Computer Science,
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands
linda@cs.ruu.nl

Abstract

The assessments obtained for the various conditional probabilities of a Bayesian belief network inevitably are inaccurate. The inaccuracies involved influence the reliability of the network's output. By subjecting the belief network to a *sensitivity analysis* with respect to its conditional probabilities, the reliability of the output can be investigated. Unfortunately, straightforward sensitivity analysis of a Bayesian belief network is highly time-consuming. In this paper, we show that, by qualitative considerations, several analyses can be identified as being uninformative as the conditional probabilities under study cannot affect the network's output. In addition, we show that the analyses that are informative comply with simple mathematical functions; more specifically, we show that the network's output can be expressed as a quotient of two functions that are linear in a conditional probability under study. These properties allow for considerably reducing the computational burden of sensitivity analysis of Bayesian belief networks, as will be illustrated by means of various examples and experiments.

1 Introduction

During the last decades much effort in artificial-intelligence research has focused on modelling and reasoning with uncertainty in knowledge-based systems. As the oldest, well-founded mathematical theory of uncertainty, probability theory plays a prominent role in this research effort. Unfortunately, straightforward application of probability theory in a knowledge-based system leads to prohibitively high computational costs. Over the years, various attempts have been made to settle this problem, leading, in the late 1980s, to the framework of *Bayesian belief networks*. Bayesian belief networks by now have become widely accepted as intuitively appealing probabilistic models that are highly valuable in addressing real-life problems in complex domains. Practical applications of the framework of belief networks are being developed for various problem domains, most notably in the field of medical diagnosis and prognostic assessment [Andreassen *et al.*, 1987, Heckerman *et al.*, 1992].

A Bayesian belief network basically is a concise representation of a joint probability distribution on a set of statistical variables [Pearl, 1988]. It consists of a qualitative part and an associated quantitative part. The qualitative part of a belief network encodes, in a directed graph, the variables under study, along with their probabilistic interrelationships. The nodes in the digraph represent the statistical variables. The digraph's arcs with each other serve to capture

*The investigations were (partly) supported by the Netherlands Computer Science Research Foundation with financial support from the Netherlands Organization for Scientific Research (NWO).

the independences among these variables: absence of an arc between two nodes indicates that the corresponding variables do not influence each other directly and, hence, are (conditionally) independent. The quantitative part of the belief network is a set of conditional probabilities that describe the strengths of the dependences between the variables represented in the qualitative part: with each node are associated conditional probabilities describing the joint influence of values of the node's predecessors on the probabilities of the values of the node itself. A belief network's qualitative and quantitative part with each other provide enough information to uniquely define a joint probability distribution on the statistical variables under study. A Bayesian belief network thus allows for computing any (prior or posterior) probability of interest [Pearl, 1988].

Bayesian belief networks are generally constructed with the help of experts from the domain of application. Experience shows that, although it may require considerable effort, building the qualitative part of a belief network is quite practicable. In fact, as it has parallels to designing a domain model for a more traditional knowledge-based system, well-known knowledge-engineering techniques can be employed. Assessing the conditional probabilities for the quantitative part of a Bayesian belief network, however, is generally found to be a much harder task, not in the least because of the large number of assessments required [Druzdzel & Van der Gaag, 1995]. In general, various different sources of information can be exploited for probability assessment, ranging from databases and literature to human experts. The assessments obtained from these sources, however, are inevitably inaccurate, due to incompleteness of data and partial knowledge of the problem under study. Particularly assessments obtained from experts are known to be highly inaccurate [Tversky *et al.*, 1982].

The inaccuracies in the probability assessments for a Bayesian belief network influence the reliability of the network's output. In a medical application, for example, erroneous diagnoses or non-optimal treatment recommendations may result from building upon inaccurate assessments. The reliability of the output of a belief network can be investigated by studying its robustness. Robustness pertains to the extent to which the network's conditional probabilities influence the output when deviations from the specified assessments are assumed. For gaining detailed insight in output robustness, a Bayesian belief network can be subjected to a sensitivity analysis. In general, *sensitivity analysis* of a mathematical model amounts to investigating the effects of the inaccuracies in the model's parameters on its output; to this end, the values of the model's parameters are varied systematically [Morgan & Henrion, 1990, Habbema *et al.*, 1990]. For a belief network, sensitivity analysis amounts to varying the assessments for one or more conditional probabilities of the network's quantitative part simultaneously and investigating the effects on a probability of interest or, for example, on a diagnosis or decision based upon this probability of interest [Laskey, 1995, Coupé *et al.*, 1999a]. Upon such an analysis, some conditional probabilities will show a considerable effect, while others will hardly reveal any influence.

Straightforward sensitivity analysis of a Bayesian belief network, unfortunately, is highly time-consuming. In the simplest type of sensitivity analysis, for example, for every single conditional probability of the network's quantitative part, a number of deviations from the specified assessment are investigated. For every value under study, the probability of interest is computed from the network. Even for a rather small belief network, the analysis thus easily requires tens of thousands of network computations. By restricting the sensitivity analysis to the conditional probabilities that are expected to be influential, as indicated for example by a domain expert, the computational effort required can be reduced. The computational burden still remains considerable, however, and, in fact, is prohibitive when sensitivity analysis is to be used for verifying the robustness of a network's output in, for example, daily medical practice. To be of practical use, therefore, more efficient methods for sensitivity analysis of belief networks are indispensable.

In this paper, we present an efficient method for sensitivity analysis of Bayesian belief networks that requires considerably less computational effort than straightforward variation of conditional probabilities. Our method builds to a large extent on the qualitative part of a belief network. As the digraph of a network represents the independences among the statistical variables involved, it allows for identifying conditional probabilities that upon variation cannot influence the probability of interest. Analyses with respect to these conditional probabilities are uninformative and can therefore be excluded from the overall analysis. Experiments on randomly generated belief networks indicate that the number of analyses that can be thus excluded may be considerable. In addition, we show that the analyses that are informative comply with simple mathematical functions. More in specific, we show that the probability of interest of a belief network can be expressed as a quotient of two functions that are linear in a conditional probability under study. The constants in this fractional function determine the sensitivity of the probability of interest to the conditional probability concerned. We show that computing the constants from the network requires just a small number of network computations. These properties with each other allow for considerably reducing the computational burden and thus for improving upon the practicability of sensitivity analysis of Bayesian belief networks.

The paper is organised as follows. In Section 2 we briefly review the framework of Bayesian belief networks and detail some of the concepts that will be used throughout the paper. We then present the various properties of sensitivity analysis of belief networks outlined above. In doing so, we focus on a one-way sensitivity analysis, that is, an analysis in which a network's conditional probabilities are investigated one at a time. In Section 3, we discuss the identification of a belief network's conditional probabilities that upon variation cannot influence the probability of interest. In Section 4, we detail the functional relation that holds between a network's probability of interest and a single conditional probability under study. In Section 5, we comment on results obtained from experiments with one-way sensitivity analysis of randomly generated belief networks. In Section 6, we compare our results with previous work on sensitivity analysis of Bayesian belief networks. The paper ends with our conclusions and directions for further research in Section 7.

2 The belief-network framework

A *Bayesian belief network* basically is a concise representation of a joint probability distribution on a set of statistical variables. In a belief network, information about the independences holding among the variables is explicitly separated from the numerical quantities involved in the distribution. To this end, the network comprises a qualitative part and an associated quantitative part. In this section, we briefly review the formalism of belief networks; for further details, we refer the reader to [Pearl, 1988].

The qualitative part of a Bayesian belief network is a graphical representation of the independences holding among the variables in the probability distribution that is being represented. It takes the form of an *acyclic directed graph*, or *digraph*, for short. In this digraph G, each node represents a statistical variable that can take one of a finite set of values. The digraph's set of arcs models the independences among the represented variables. Informally speaking, we take an arc $V_i \rightarrow V_j$ to represent a direct influential or causal relationship between the variables V_i and V_j; the arc's direction designates V_j as the effect or consequence of the cause V_i. Absence of an arc between two nodes means that the corresponding variables do not influence each other directly and, hence, are (conditionally) independent. In the sequel, we will use $\pi_G(V_i)$ to denote the set of (immediate) predecessors, or causes, of node V_i in G and use $\pi^*_G(V_i)$ to denote the set of nodes composed of V_i and all its ancestors; we will use $\sigma_G(V_i)$ to denote the set of (immediate) successors, or effects, of node V_i in G and use $\sigma^*_G(V_i)$ to denote the set of nodes composed of V_i and all its descendants. The following definitions review the probabilistic meaning that is assigned to the digraph of a Bayesian belief network more formally.

Definition 2.1 *Let $G = (V(G), A(G))$ be an acyclic digraph and let s be a chain in G between the nodes V_i and V_j. We say that s is* blocked *by the set of nodes $Y \subseteq V(G)$, if either V_i or V_j is included in Y, or s contains three consecutive nodes X_1, X_2, X_3, for which one of the following conditions holds:*

1. *arcs $X_1 \leftarrow X_2$ and $X_2 \rightarrow X_3$ are on the chain s, and $X_2 \in Y$;*
2. *arcs $X_1 \rightarrow X_2$ and $X_2 \rightarrow X_3$ are on the chain s, and $X_2 \in Y$;*
3. *arcs $X_1 \rightarrow X_2$ and $X_2 \leftarrow X_3$ are on the chain s and $\sigma^*_G(X_2) \cap Y = \varnothing$.*

In reviewing the concept of a blocked chain, we have distinguished between three conditions. Figure 1 serves as a reference for these conditions; in the two chains representing the conditions 1 and 2, node X_2 is drawn with shading to indicate that it is comprised in the blocking set Y for the chain at hand.

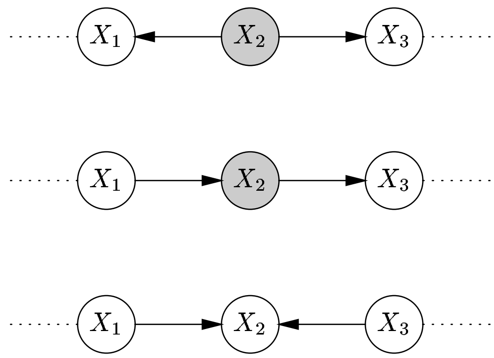

Figure 1: The three conditions for chain blocking.

Building upon the concept of blocking, we review the well-known *d-separation criterion* for sets of chains.

Definition 2.2 *Let $G = (V(G), A(G))$ be an acyclic digraph and let $X, Y, Z \subseteq V(G)$. The set of nodes Y is said to* d-separate *the sets of nodes X and Z in G, denoted $\langle X \mid Y \mid Z \rangle^d_G$, if for each node $V_i \in X$ and each node $V_j \in Z$, every chain from V_i to V_j in G is blocked by Y.*

The following definition relates the d-separation criterion to the concept of independence.

Definition 2.3 *Let $G = (V(G), A(G))$ be an acyclic digraph and let* Pr *be a joint probability distribution on $V(G)$. Then, G is called an* I-map *for* Pr *if for all sets of variables $X, Y, Z \subseteq V(G)$, we have: if $\langle X \mid Y \mid Z \rangle^d_G$, then X and Z are conditionally independent given Y in* Pr.

The d-separation criterion thus provides for reading independences from a belief network's digraph without having to resort to probabilistic computations. We would like to note that the criterion of d-separation generally is defined for mutually exclusive sets of nodes only. We have extended the definition to apply to overlapping set of nodes as well, to provide for reading from a digraph independences for instantiated nodes. We take an instantiated node to be d-separated from any other node. Our extension has been inspired by previous work on informational independence [Van der Gaag & Meyer, 1998].

Associated with the qualitative part of a Bayesian belief network are numerical quantities that describe the strengths of the dependences among the represented variables. With each node V_i of the network's digraph G is associated a set of conditional probabilities $p(V_i \mid \pi_G(V_i))$ describing the joint influence of the various values for the node's (immediate) predecessors $\pi_G(V_i)$ on the probabilities of the values of V_i itself. These probabilities with each other constitute the quantitative part of the belief network.

We review the concept of Bayesian belief network more formally.

Definition 2.4 *A* Bayesian belief network *is a tuple $B = (G, P)$ where*

- *$G = (V(G), A(G))$ is an acyclic digraph with nodes $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$, and arcs $A(G)$;*
- *P is a set of conditional probabilities $p(V_i \mid \pi_G(V_i))$, for all $V_i \in V(G)$.*

We illustrate the concept of Bayesian belief network by means of an example that will be used for our running example throughout the paper.

Example 2.5 We consider the well-known ALARM-network [Beinlich *et al.*, 1989]. The digraph of the network is reproduced in Figure 2; for the examples in the remainder of the paper, we have indicated the node of interest, *LV failure*, by a double circle and the network's observable

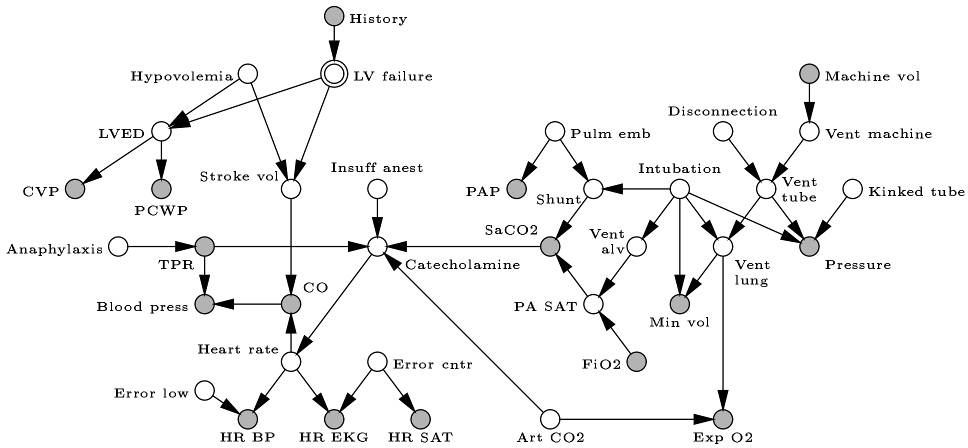

Figure 2: The digraph of the ALARM belief network.

nodes by shading. From the network's digraph, various independences are read. For example, the variable *LV failure* is independent of the variable *Insuff anest*, if no information is available yet; the two variables become dependent, however, when, for example, the value of the variable *Blood press* becomes available. The variables *Pulm emb* and *Heart rate*, on the other hand, are dependent, but become independent once a value for *SaCO2* is observed. Associated with the nodes of the network are conditional probabilities. For example, for the node *Stroke vol*, the following conditional probabilities are specified:

$$\begin{aligned}
p(Stroke\ vol = low \mid Hypovolemia = false \wedge LV\ failure = false) &= 0.05 \\
p(Stroke\ vol = normal \mid Hypovolemia = false \wedge LV\ failure = false) &= 0.90 \\
p(Stroke\ vol = high \mid Hypovolemia = false \wedge LV\ failure = false) &= 0.05 \\[1em]
p(Stroke\ vol = low \mid Hypovolemia = true \wedge LV\ failure = false) &= 0.5 \\
p(Stroke\ vol = normal \mid Hypovolemia = true \wedge LV\ failure = false) &= 0.49 \\
p(Stroke\ vol = high \mid Hypovolemia = true \wedge LV\ failure = false) &= 0.01 \\[1em]
p(Stroke\ vol = low \mid Hypovolemia = false \wedge LV\ failure = true) &= 0.95 \\
p(Stroke\ vol = normal \mid Hypovolemia = false \wedge LV\ failure = true) &= 0.04 \\
p(Stroke\ vol = high \mid Hypovolemia = false \wedge LV\ failure = true) &= 0.01 \\[1em]
p(Stroke\ vol = low \mid Hypovolemia = true \wedge LV\ failure = true) &= 0.98 \\
p(Stroke\ vol = normal \mid Hypovolemia = true \wedge LV\ failure = true) &= 0.01 \\
p(Stroke\ vol = high \mid Hypovolemia = true \wedge LV\ failure = true) &= 0.01
\end{aligned}$$

As for this paper, the specific assessments for the various conditional probabilities are not of interest, we refrain from further detailing them. □

The following proposition states that the conditional probabilities of a Bayesian belief network provide all information necessary for uniquely defining a joint probability distribution on the variables discerned that respects the independences portrayed by the network's qualitative part; henceforth, we will call this distribution the joint probability distribution *defined* by the network.

Proposition 2.6 *Let $B = (G, P)$ be a Bayesian belief network. Then,*

$$\Pr(V(G)) = \prod_{V_i \in V(G)} p(V_i \mid \pi_G(V_i))$$

defines a joint probability distribution Pr *on $V(G)$ such that G is an I-map for* Pr.

Since the digraph of a Bayesian belief network and its associated conditional probabilities with each other define a unique joint probability distribution on the variables discerned, any (prior or posterior) probability of interest can be computed from the network. For this purpose various algorithms are available [Pearl, 1988, Lauritzen & Spiegelhalter, 1988].

3 Uninfluential probabilities in a sensitivity analysis

Sensitivity analysis is a general technique for studying the effects of the inaccuracies in the parameters of a mathematical model on this model's output [Habbema *et al.*, 1990, Morgan & Henrion, 1990]. Sensitivity analysis basically amounts to systematically varying the values of the parameters of the model under study. In a one-way sensitivity analysis, the values of the parameters are varied one at a time while keeping the values of all other parameters fixed. For a Bayesian belief network, a one-way sensitivity analysis amounts to varying the assessment for a single conditional probability of the network's quantitative part. The analysis provides for studying the effects of the inaccuracy in the specified assessment on a probability of interest [Coupé *et al.*, 1999a].

In essence, in a one-way sensitivity analysis of a Bayesian belief network, the sensitivity of the network's probability of interest is investigated with respect to every single conditional probability. Various conditional probabilities of a belief network, however, are known beforehand not to affect the probability of interest upon variation, for example because this probability of interest is shielded from their influence by available observations. These uninfluential probabilities can be readily identified by inspection of the network's digraph, that is, without extensive probabilistic computations. We say that the probability of interest is *algebraically independent* of these uninfluential conditional probabilities. For abbreviation, we will write $p \not\sim q$ to denote that the probability p is algebraically independent of the probability q. We would like to note that the phrase algebraic independence is used to refer to the absence of any effect of varying the assessment for a conditional probability under study on a probability of interest, *as induced by the network's digraph.* Also note that the phrase applies to probabilities whereas the phrase probabilistic independence pertains to variables. Now, in a one-way sensitivity analysis of a Bayesian belief network, for a conditional probability of which the network's probability of interest is algebraically independent, no further investigation is required. The sensitivity analysis of the network can therefore be restricted to the conditional probabilities of which the probability of interest is algebraically *dependent.* The nodes to which these conditional probabilities refer constitute the *sensitivity set* for the node of interest.

We define the concept of sensitivity set more formally.

Definition 3.1 *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$. Let $V_r \in V(G)$ be the network's node of interest and let $O \subseteq V(G)$ be the set of observed nodes in G. Now, let G^* be the digraph that is constructed from G by adding an auxiliary predecessor X_i to every node $V_i \in V(G)$. Then, the* sensitivity set *for V_r given O, denoted $Sen(V_r, O)$, is the set of all nodes $V_i \in V(G)$ for which $\neg\langle\{X_i\} \mid O \mid \{V_r\}\rangle^d_{G^*}$.*

From the previous definition we have that the sensitivity set for a node of interest V_r is computed from the digraph of a belief network under consideration by adding an auxiliary predecessor X_i to every node V_i and thereupon exploiting the d-separation criterion. The auxiliary predecessor X_i of node V_i can be looked upon as capturing the presence of inaccuracy in the probability assessments for V_i. If the presence of inaccuracy in V_i's assessments is not d-separated from the node of interest or, in other words, if V_r is not shielded from the inaccuracy by the available evidence, then varying the assessments for V_i's conditional probabilities may influence the probabilities of the values of

V_r. V_i is therefore included in V_r's sensitivity set. We would like to note that the basic idea of capturing the presence of inaccuracy by means of auxiliary nodes has been exploited before [Spiegelhalter, 1989]. We further note that we capture the *presence* of inaccuracy rather than the inaccuracy itself by auxiliary nodes.

The following example illustrates our concept of sensitivity set.

Example 3.2 We consider once again the ALARM-network, the digraph of which is shown in Figure 2. We are interested in the diagnostic variable *LV failure*; our probability of interest is the probability that *LV failure* = *true*. We consider the sensitivity set for the node *LV failure* given various different sets of observed nodes.

If the set of observed nodes is empty, that is, when no observations are available, the sensitivity set for the node *LV failure* equals

$$Sen(LV\ failure, \varnothing) = \{LV\ failure, History\}$$

Upon performing a one-way sensitivity analysis of the a priori belief network, only the conditional probabilities of these two nodes need be investigated; the conditional probabilities of all other nodes in the network upon variation cannot influence the probability of interest.

Now, suppose that we would like to evaluate the sensitivity of the network's probability of interest in view of observations for the nodes in the set $O_1 = \{History, CVP, TPR, Blood\ press, CO\}$. The sensitivity set for *LV failure* given O_1 equals

$$\begin{aligned} Sen(LV\ failure, O_1) = \{ & LV\ failure, Hypovolemia, LVED, CVP, Stroke\ vol, CO, Insuff \\ & anest, Catecholamine, Heart\ rate, Art\ CO2, SaCO2, PA\ SAT, \\ & FiO2, Vent\ alv, Shunt, Intubation, Pulm\ emb\} \end{aligned}$$

From the 37 nodes included in the belief network, the conditional probabilities of only 17 nodes need be investigated in the analysis. We would like to note that, in general, a sensitivity set does *not* coincide with the set of non-d-separated nodes for the node of interest. From the sensitivity set for the node *LV failure* given O_1, for example, it is readily seen that a sensitivity set can include both non-d-separated nodes (such as the node *Stroke vol*) and d-separated nodes (such as the node *CO*); also, the set of nodes that are not comprised in the sensitivity set can include non-d-separated nodes (such as the node *PCWP*) as well as d-separated nodes (such as *Blood press*).

Now, if in addition to observations for the nodes in the set O_1 an observation is assumed for the node *SaCO2*, yielding O_2 for the new set of observed nodes, the sensitivity set for *LV failure* reduces in size from 17 nodes to 10 nodes:

$$\begin{aligned} Sen(LV\ failure, O_2) = \{ & LV\ failure, Hypovolemia, LVED, CVP, Stroke\ vol, CO, Insuff \\ & anest, Catecholamine, Heart\ rate, Art\ CO2\} \end{aligned}$$

Note that, when a value for the node of interest *LV failure* is available, every node in the auxiliary network for determining the sensitivity set is d-separated from *LV failure*. The sensitivity set then is empty. □

In order to prove the claims we have made so far with respect to a sensitivity set, we will partition a belief network's set of nodes that are not included in a sensitivity set under study into three mutually exclusive sets of nodes. We will then show that, for various different reasons, the conditional probabilities for the nodes included in these sets upon variation have no effect on the network's probability of interest.

Definition 3.3 *Let B be a Bayesian belief network with the digraph G, let V_r be the network's node of interest, and let O the set of observed nodes, as before. We define the sets of nodes $Insen_1(V_r, O)$, $Insen_2(V_r, O)$, and $Insen_3(V_r, O)$, respectively, as*

- *for every node $V_i \in \pi^*_G(V_r)$, if $\langle(\{V_i\} \cup \pi_G(V_i)) \mid O \mid \{V_r\}\rangle^d_G$, then $V_i \in Insen_1(V_r, O)$;*

- *for every node* $V_i \in V(G) \setminus \pi^*_G(V_r)$, *if* $\langle(\{V_i\} \cup \pi_G(V_i)) \mid O \mid \{V_r\}\rangle^d_G$ *and* $\sigma^*_G(V_i) \cap O \neq \varnothing$, *then* $V_i \in \mathit{Insen}_2(V_r, O)$;
- *for every node* $V_i \in V(G) \setminus \pi^*_G(V_r)$, *if* $\sigma^*_G(V_i) \cap O = \varnothing$, *then* $V_i \in \mathit{Insen}_3(V_r, O)$.

The sets $\mathit{Insen}_1(V_r, O)$, $\mathit{Insen}_2(V_r, O)$, and $\mathit{Insen}_3(V_r, O)$ include nodes to whose conditional probabilities a belief network's probability of interest is insensitive. Before illustrating the three sets of nodes for our running example, we informally address their meaning. In doing so, we begin by considering the ancestors V_i of the node of interest V_r. We observe that any unblocked chain from V_i to V_r, be it a direct chain or a chain via a predecessor of V_i, provides for conveying an influence from V_i's probability assessments to V_r. If no such chain is present, therefore, varying the assessments for node V_i can have no influence on the probability of interest. The set $\mathit{Insen}_1(V_r, O)$ now includes all ancestors V_i of V_r such that V_r is d-separated by the available observations from both V_i and V_i's predecessors. We now consider the non-ancestors of the node of interest. We observe that the probability assessments for a non-ancestor V_i of V_r cannot influence the probability of interest if there are no observations available. Only an observed descendant of V_i that induces an influence on V_r through V_i, can cause varying V_i's probability assessments to affect the probability of interest. The set $\mathit{Insen}_3(V_r, O)$ now includes all non-ancestors of V_r that do not have any observed descendants. The set $\mathit{Insen}_2(V_r, O)$, to conclude, includes the non-ancestors of V_r that happen to have observed descendants yet whose influence is shielded from V_r by the available observations: the set includes all non-ancestors V_i of V_r with at least one observed descendant such that V_r is d-separated from both V_i and V_i's predecessors.

We illustrate the various sets of nodes defined above by means of our running example.

Example 3.4 We consider again the ALARM-network, the digraph of which is shown in Figure 2. We are once more interested in the variable *LV failure*; for our probability of interest, we take the probability that *LV failure* is *true*. We recall from Example 3.2 that, if the set of observed nodes is empty, the sensitivity set for the node of interest *LV failure* equals $\mathit{Sen}(\mathit{LV\ failure}, \varnothing) = \{\mathit{LV\ failure}, \mathit{History}\}$. The set of all remaining nodes, that is, the set of all nodes, *LV failure* and *History* excluded, is partitioned into three sets as defined above. Of these, the sets $\mathit{Insen}_1(\mathit{LV\ failure}, \varnothing)$ and $\mathit{Insen}_2(\mathit{LV\ failure}, \varnothing)$ are empty; the set $\mathit{Insen}_3(\mathit{LV\ failure}, \varnothing)$ includes any node that is not comprised in the sensitivity set. We consider, as an example, the node *Stroke vol*. From Figure 2, we see that *Stroke vol* is not an ancestor of the node of interest *LV failure*; furthermore, it does not have any observed descendants. From Definition 3.3, we conclude that the node *Stroke vol* belongs to the set $\mathit{Insen}_3(\mathit{LV\ failure}, \varnothing)$. Informally speaking, as the node *Stroke vol* is not observed and does not have any observed descendants, it cannot exert nor pass on any diagnostic influence on the probabilities for *LV failure*. The probability of interest $\Pr(\mathit{LV\ failure} = \mathit{true})$ therefore is algebraically independent of the conditional probabilities for *Stroke vol*. A similar observation applies to any other node from the set $\mathit{Insen}_3(\mathit{LV\ failure}, \varnothing)$.

We now assume that observations are obtained for the nodes in the set $O_1 = \{\mathit{History}, \mathit{CVP}, \mathit{TPR}, \mathit{Blood\ press}, \mathit{CO}\}$. The sets $\mathit{Insen}_1(\mathit{LV\ failure}, O_1)$, $\mathit{Insen}_2(\mathit{LV\ failure}, O_1)$, and $\mathit{Insen}_3(\mathit{LV\ failure}, O_1)$ equal

$$\begin{aligned}
\mathit{Insen}_1(\mathit{LV\ failure}, O_1) &= \{\mathit{History}\} \\
\mathit{Insen}_2(\mathit{LV\ failure}, O_1) &= \{\mathit{Blood\ press}, \mathit{TPR}, \mathit{Anaphylaxis}\} \\
\mathit{Insen}_3(\mathit{LV\ failure}, O_1) &= \{\mathit{PCWP}, \mathit{Error\ low}, \mathit{HR\ BP}, \mathit{HR\ EKG}, \mathit{HR\ SAT}, \mathit{Error\ cntr}, \\
&\quad\ \mathit{Exp\ O2}, \mathit{Min\ vol}, \mathit{Vent\ lung}, \mathit{Pressure}, \mathit{Vent\ tube}, \mathit{Kinked\ tube}, \\
&\quad\ \mathit{Disconnection}, \mathit{Vent\ machine}, \mathit{Machine\ vol}, \mathit{PAP}\}
\end{aligned}$$

We consider, as an example, the node *History*. This node is a predecessor of the node of interest *LV failure*. It is d-separated from *LV failure* and does not have any immediate predecessors that are not d-separated from *LV failure*. From Definition 3.3, therefore, we have that *History* is included in the set $\mathit{Insen}_1(\mathit{LV\ failure}, O_1)$. Informally speaking, as a value for the node *History* is available, its prior probabilities are irrelevant to the probabilities for its successor *LV failure*. The probability of *LV failure* = *true* given the available observations for O_1 therefore is algebraically independent of the prior probabilities for the node *History*. To conclude our example, we consider the node *TPR*. From Figure 2, we observe that *TPR* is not an ancestor of the node of interest *LV failure*. The node *TPR* itself as well as its immediate predecessor *Anaphylaxis* are d-separated from *LV failure* given the available observations. Furthermore, the descendant *Blood press* of *TPR* is observed. By definition, we have that the node *TPR* is included in the set $\mathit{Insen}_2(\mathit{LV\ failure}, O_1)$. Informally speaking, from *TPR* and its predecessor *Anaphylaxis* being d-separated from the node of interest *LV failure*, we find that any diagnostic influence originating from *TPR* is shielded from *LV failure* by the available observations. Therefore, the probability of interest is algebraically independent of the conditional probabilities for the node *TPR*. A similar observation applies to any other node from the set $\mathit{Insen}_2(\mathit{LV\ failure}, O_1)$. □

We would like to note that for a node of interest V_r and any set of observed nodes O, the three sets $\mathit{Insen}_1(V_r, O)$, $\mathit{Insen}_2(V_r, O)$, and $\mathit{Insen}_3(V_r, O)$, and the sensitivity set $\mathit{Sen}(V_r, O)$ are mutually exclusive and collectively exhaustive; for a formal proof of this property, the reader is referred to the appendix.

In the remainder of this section, we will show that the probability of interest of a Bayesian belief network is indeed algebraically independent of the conditional probabilities of any node that is not included in the sensitivity set $\mathit{Sen}(V_r, O)$ under study. To this end, we investigate the three sets $\mathit{Insen}_1(V_r, O)$, $\mathit{Insen}_2(V_r, O)$, and $\mathit{Insen}_3(V_r, O)$ separately and provide for each of these sets a lemma stating algebraic independence of the probability of interest for the conditional probabilities of the nodes in the set at hand. Our main result is then stated in Proposition 3.11, building upon these lemmas. The proofs of the three lemmas, although not complicated, are rather elaborate; the full proofs therefore are deferred to the appendix.

In the first lemma, we state that a belief network's probability of interest for a node V_r given observations for nodes O is algebraically independent of the conditional probabilities of any node from the set $\mathit{Insen}_3(V_r, O)$.

Lemma 3.5 *Let B be a Bayesian belief network and let* Pr *be the joint probability distribution defined by B. Let O be the set of observed nodes and let o denote the corresponding observations. Let V_r be the network's node of interest. Then, for any value v_r of V_r, we have that* $\Pr(v_r \mid o) \not\sim p(V_i \mid \pi(V_i))$ *for every node* $V_i \in \mathit{Insen}_3(V_r, O)$.

Proof (Sketch). The probability of interest $\Pr(v_r \mid o)$ for the belief network B equals

$$\Pr(v_r \mid o) = \frac{\Pr(v_r \wedge o)}{\Pr(o)}$$

We recall from Section 2 that the joint probability distribution Pr that is defined by B, can be written as a product of the network's conditional probabilities. From the basic property of marginalisation, we now have that both the numerator and the denominator can be written as a sum of products of conditional probabilities. In these sums, for every unobserved leaf node, there appear as many products as there are values for this node that differ in this node's probability only. Summing over these products amounts to summing out the leaf node by marginalisation. The same argument applies recursively to all unobserved non-ancestors of V_r that do not have any observed descendants, that is, the argument applies to every node from the set $\mathit{Insen}_3(V_r, O)$. We conclude that the probability of interest is algebraically independent of the conditional probabilities of any node from this set. □

We illustrate the property stated in the previous lemma by means of an example.

Example 3.6 We consider the belief network from Figure 3, which is a small fragment of the ALARM-network. The possible values of the node *LV failure* are *fail* and *no fail*; the possible values for each of the nodes *LVED*, *CVP*, and *PCWP* are *low*, *normal*, and *high*. Our node of

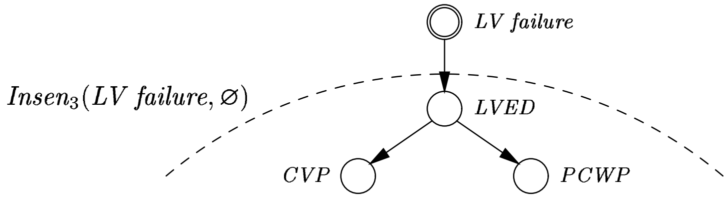

Figure 3: An example belief network, illustrating the property stated in Lemma 3.5 for the node of interest *LV failure* and the empty set of observed nodes; the set $\mathit{Insen}_3(\mathit{LV\ failure}, \varnothing)$ consists of the nodes *LVED*, *CVP*, and *PCWP*.

interest once again is the node *LV failure*, indicated in the figure by a double circle. We now address the situation where no observations are available yet and investigate the probability of interest $\Pr(\mathit{fail})$. From Definition 3.3, we find that the set $\mathit{Insen}_3(\mathit{LV\ failure}, \varnothing)$ consists of the three nodes *LVED*, *CVP*, and *PCWP*. For the probability of interest, we find that

$$\Pr(\mathit{fail}) =$$

$$= \sum_{\substack{\{LVED,\, CVP,\\ PCWP\}}} p(PCWP \mid LVED) \cdot p(CVP \mid LVED) \cdot p(LVED \mid \mathit{fail}) \cdot p(\mathit{fail}) =$$

$$= \sum_{\{LVED, CVP\}} \Big(p(\mathit{low}\ PCWP \mid LVED) + p(\mathit{normal}\ PCWP \mid LVED) + p(\mathit{high}\ PCWP \mid LVED) \Big) \cdot$$
$$\cdot\, p(CVP \mid LVED) \cdot p(LVED \mid \mathit{fail}) \cdot p(\mathit{fail}) =$$

$$= \sum_{\{LVED, CVP\}} p(CVP \mid LVED) \cdot p(LVED \mid \mathit{fail}) \cdot p(\mathit{fail}) =$$

$$= \sum_{\{LVED\}} \Big(p(\mathit{low}\ CVP \mid LVED) + p(\mathit{normal}\ CVP \mid LVED) + p(\mathit{high}\ CVP \mid LVED) \Big) \cdot$$
$$\cdot\, p(LVED \mid \mathit{fail}) \cdot p(\mathit{fail}) =$$

$$= \sum_{\{LVED\}} p(LVED \mid \mathit{fail}) \cdot p(\mathit{fail}) =$$

$$= \Big(p(\mathit{low}\ LVED \mid \mathit{fail}) + p(\mathit{normal}\ LVED \mid \mathit{fail}) + p(\mathit{high}\ LVED \mid \mathit{fail}) \Big) \cdot p(\mathit{fail}) =$$

$$= p(\mathit{fail})$$

From this derivation, it is readily seen that the probability of interest $\Pr(\mathit{fail})$ is algebraically independent of the conditional probabilities of the three nodes included in the set $\mathit{Insen}_3(\mathit{LV\ failure}, \varnothing)$.

We now address the situation where the value *high* is observed for the node *PCWP*. This situation is depicted in Figure 4, where the node *PCWP* is drawn with shading to indicate that its value has been observed. The set $\mathit{Insen}_3(\mathit{LV failure}, \{PCWP\})$ is composed of the node *CVP* only. For our probability of interest $\Pr(\mathit{fail} \mid \mathit{high}\ PCWP)$, we now find that

$$\Pr(\mathit{fail} \mid \mathit{high}\ PCWP) = \frac{\Pr(\mathit{fail} \wedge \mathit{high}\ PCWP)}{\Pr(\mathit{high}\ PCWP)}$$

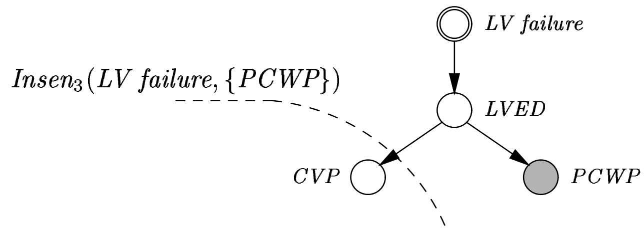

Figure 4: An example belief network, illustrating the property stated in Lemma 3.5 for the node of interest *LV failure* and the set of observed nodes $\{PCWP\}$; the set $Insen_3(LV\ failure, \{PCWP\})$ consists of the single node CVP.

The numerator in this equation equals

$$\Pr(fail \wedge high\ PCWP) =$$

$$= \sum_{\{LVED, CVP\}} p(CVP \mid LVED) \cdot p(high\ PCWP \mid LVED) \cdot p(LVED \mid fail) \cdot p(fail) =$$

$$= \sum_{\{LVED\}} \Big(p(low\ CVP \mid LVED) + p(normal\ CVP \mid LVED) + p(high\ CVP \mid LVED)\Big) \cdot$$
$$\cdot\, p(high\ PCWP \mid LVED) \cdot p(LVED \mid fail) \cdot p(fail) =$$

$$= \sum_{\{LVED\}} p(high\ PCWP \mid LVED) \cdot p(LVED \mid fail) \cdot p(fail)$$

The denominator equals

$$\Pr(high\ PCWP) =$$

$$= \sum_{\substack{\{LV\ failure,\\ LVED,\ CVP\}}} p(CVP \mid LVED) \cdot p(high\ PCWP \mid LVED) \cdot p(LVED \mid LV\ failure) \cdot p(LV\ failure) =$$

$$= \sum_{\substack{\{LV\ failure,\\ LVED\}}} \Big(p(low\ CVP \mid LVED) + p(normal\ CVP \mid LVED) + p(high\ CVP \mid LVED)\Big) \cdot$$
$$\cdot\, p(high\ PCWP \mid LVED) \cdot p(LVED \mid LV failure) \cdot p(LV failure) =$$

$$= \sum_{\substack{\{LV\ failure,\\ LVED\}}} p(high\ PCWP \mid LVED) \cdot p(LVED \mid LV\ failure) \cdot p(LV\ failure)$$

We conclude that the probability of interest equals

$$\Pr(fail \mid high\ PCWP) =$$

$$= \frac{\displaystyle\sum_{\{LVED\}} p(high\ PCWP \mid LVED) \cdot p(LVED \mid fail) \cdot p(fail)}{\displaystyle\sum_{\substack{\{LV\ failure,\\ LVED\}}} p(high\ PCWP \mid LVED) \cdot p(LVED \mid LV\ failure) \cdot p(LV\ failure)}$$

From this derivation, it is readily seen that the probability of interest $\Pr(fail \mid high\ PCWP)$ is algebraically independent of the conditional probabilities of *CVP*, the only node included in the set $Insen_3(LV\ failure, \{PCWP\})$. □

So far, we have shown that a belief network's probability of interest for a node V_r given observations for nodes O is algebraically independent of the conditional probabilities of any node from the set $Insen_3(V_r, O)$. We now proceed by observing that this probability of interest is also algebraically independent of the conditional probabilities of the nodes from the set $Insen_2(V_r, O)$.

Lemma 3.7 *Let B be a Bayesian belief network and let* $\Pr$ *be its joint probability distribution. Let O be the set of observed nodes with observations o, as before. Let V_r be the network's node of interest. Then, for any value v_r of V_r, we have that $\Pr(v_r \mid o) \not\sim p(V_i \mid \pi(V_i))$ for every node $V_i \in Insen_2(V_r, O)$.*

Proof (Sketch). The probability of interest $\Pr(v_r \mid o)$ for the belief network B equals

$$\Pr(v_r \mid o) = \frac{\Pr(v_r \wedge o)}{\Pr(o)}$$

Both the numerator and the denominator of this equation can be written as a sum of products of conditional probabilities from the network. From Definition 3.3, we know that the nodes from the set $Insen_2(V_r, O)$ as well as their immediate predecessors are d-separated from the node of interest V_r by the available observations; more in specific, we know that a predecessor of any node from $Insen_2(V_r, O)$ either belongs to $Insen_2(V_r, O)$ itself or is an observed node. In both the numerator and the denominator of the above equation, therefore, a term can be isolated that includes all the nodes from $Insen_2(V_r, O)$ and no other nodes that are not observed. As this term appears in the numerator as well as in the denominator, it cancels out. The conditional probabilities of the nodes from $Insen_2(V_r, O)$ upon variation therefore do not affect the probability of interest. □

We illustrate the property stated in Lemma 3.7 by means of an example.

Example 3.8 We consider the belief network from Figure 5, which again is a small fragment of the

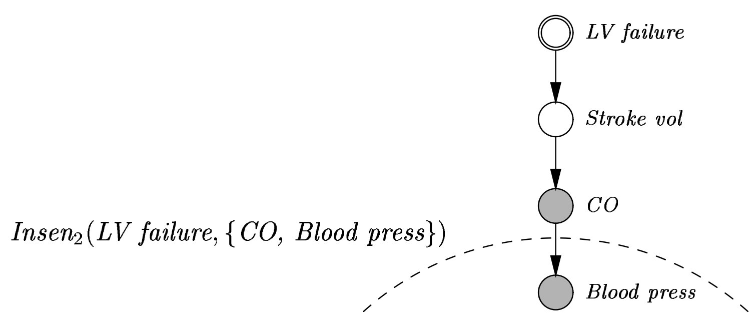

Figure 5: An example belief network, illustrating the property stated in Lemma 3.7 for the node of interest *LV failure* and the set of observed nodes $\{CO, Blood\ press\}$; the set $Insen_2(LV\ failure, \{CO, Blood\ press\})$ consists of the node *Blood press* only.

ALARM-network. The possible values of the node *LV failure* are *fail* and *no fail*; the possible values for each of the other nodes are *low*, *normal*, and *high*. Our node of interest once again is the node *LV failure*. We now address the situation where the value *low* has been observed for both the nodes *CO* and *Blood press*, and investigate the probability of interest $\Pr(fail \mid low\ CO \wedge low\ Blood\ press)$. From Definition 3.3, we have that the set $Insen_2(LV\ failure, \{CO, Blood\ press\})$ comprises the node *Blood press* only. Note that $Insen_3(LV\ failure, \{CO, Blood\ press\}) = \varnothing$. For the probability of interest, we find that

$$\Pr(fail \mid low\ CO \wedge low\ Blood\ press) = \frac{\Pr(fail \wedge low\ CO \wedge low\ Blood\ press)}{\Pr(low\ CO \wedge low\ Blood\ press)}$$

The numerator in this equation equals

$$\Pr(\mathit{fail} \wedge \mathit{low}\ \mathit{CO} \wedge \mathit{low}\ \mathit{Blood}\ \mathit{press}) =$$

$$= \sum_{\{Stroke\ vol\}} p(low\ Blood\ press \mid low\ CO) \cdot p(low\ CO \mid Stroke\ vol) \cdot p(Stroke\ vol \mid fail) \cdot p(fail) =$$

$$= p(low\ Blood\ press \mid low\ CO) \cdot \cdot \left(\sum_{\{Stroke\ vol\}} p(low\ CO \mid Stroke\ vol) \cdot p(Stroke\ vol \mid fail) \cdot p(fail) \right)$$

The denominator in the equation equals

$$\Pr(low\ CO \wedge low\ Blood\ press) =$$

$$= \sum_{\substack{\{LV\ failure, \\ Stroke\ vol\}}} p(low\ Blood\ press \mid low\ CO) \cdot p(low\ CO \mid Stroke\ vol) \cdot p(Stroke\ vol \mid LV\ failure) \cdot \cdot p(LV\ failure) =$$

$$= p(low\ Blood\ press \mid low\ CO) \cdot \cdot \left(\sum_{\substack{\{LV\ failure, \\ Stroke\ vol\}}} p(low\ CO \mid Stroke\ vol) \cdot p(Stroke\ vol \mid LV\ failure) \cdot p(LV\ failure) \right)$$

We now conclude that the probability of interest equals

$$\Pr(fail \mid low\ CO \wedge low\ Blood\ press) =$$

$$= \frac{\sum_{\{Stroke\ vol\}} p(low\ CO \mid Stroke\ vol) \cdot p(Stroke\ vol \mid fail) \cdot p(fail)}{\sum_{\substack{\{LV\ failure, \\ Stroke\ vol\}}} p(low\ CO \mid Stroke\ vol) \cdot p(Stroke\ vol \mid LV\ failure) \cdot p(LV\ failure)}$$

From this derivation, it is readily seen that the probability of interest is algebraically independent of the conditional probabilities of *Blood press*, the only node that is included in the set $Insen_2(LV\ failure, \{CO, Blood\ press\})$. □

So far, we have shown that a belief network's probability of interest for a node V_r given observations for nodes O is algebraically independent of the conditional probabilities of any node from the sets $Insen_3(V_r, O)$ and $Insen_2(V_r, O)$. To conclude, we now state that this probability of interest is also algebraically independent of the conditional probabilities of the nodes from the set $Insen_1(V_r, O)$.

Lemma 3.9 *Let B be a Bayesian belief network and let* Pr *be its joint probability distribution. Let O be the set of observed nodes with observations o, as before. Let V_r be the network's node of interest. Then, for any value v_r of V_r, we have that $\Pr(v_r \mid o) \not\sim p(V_i \mid \pi(V_i))$ for every node $V_i \in Insen_1(V_r, O)$.*

Proof (Sketch). The probability of interest $\Pr(v_r \mid o)$ for the belief network B once again equals

$$\Pr(v_r \mid o) = \frac{\Pr(v_r \wedge o)}{\Pr(o)}$$

As before, both the numerator and the denominator of this equation can be written as a sum of products of conditional probabilities from the network. The proof is now based on canceling out terms from the numerator and the denominator as in the proof of the previous lemma. From Definition 3.3, we know that the nodes from the set $\mathit{Insen}_1(V_r, O)$ as well as their immediate predecessors are d-separated from the node of interest V_r by the available observations; more in specific, we know that a predecessor of any node from $\mathit{Insen}_1(V_r, O)$ either belongs to $\mathit{Insen}_1(V_r, O)$ itself or is an observed node. In both the numerator and the denominator of the above equation, therefore, a term can be isolated that includes all the nodes from $\mathit{Insen}_1(V_r, O)$ and no other nodes that are not observed. As this term appears in the numerator as well as in the denominator, it once again cancels out. The conditional probabilities of the nodes from $\mathit{Insen}_1(V_r, O)$ upon variation therefore do not affect the probability of interest. □

We illustrate the property stated in the previous lemma by means of an example.

Example 3.10 We consider the belief network from Figure 6, which again is a small fragment of the ALARM-network. The possible values of the node *LV failure* once more are *fail* and *no fail*; the values of the node *History* are *history* and *no history*. Our node of interest again is the node *LV failure*. We now address the situation where the value *history* is observed for the node *History* and investigate the probability of interest $\Pr(\mathit{fail} \mid \mathit{history})$. From Definition 3.3, we have that the set $\mathit{Insen}_1(\mathit{LV\ failure}, \{\mathit{History}\})$ consists of the node *History* only. Note that both $\mathit{Insen}_2(\mathit{LV\ failure}, \{\mathit{History}\})$ and $\mathit{Insen}_3(\mathit{LV\ failure}, \{\mathit{History}\})$ are empty. For the probability of interest, we find that

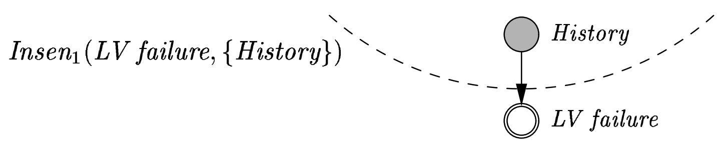

Figure 6: An example belief network, illustrating the property stated in Lemma 3.9 for the node of interest *LV failure* and the set of observed nodes $\{\mathit{History}\}$; the set $\mathit{Insen}_1(\mathit{LV\ failure}, \{\mathit{History}\})$ consists of just the node *History*.

$$\begin{aligned}
\Pr(\mathit{fail} \mid \mathit{history}) &= \frac{\Pr(\mathit{fail} \wedge \mathit{history})}{Pr(\mathit{history})} = \\
&= \frac{p(\mathit{fail} \mid \mathit{history}) \cdot p(\mathit{history})}{\sum\limits_{\{LV\ failure\}} \Big(p(LV\ failure \mid \mathit{history}) \cdot p(\mathit{history}) \Big)} = \\
&= \frac{p(\mathit{fail} \mid \mathit{history}) \cdot p(\mathit{history})}{p(\mathit{fail} \mid \mathit{history}) \cdot p(\mathit{history}) + p(\mathit{no\ fail} \mid \mathit{history}) \cdot p(\mathit{history})} = \\
&= \frac{p(\mathit{fail} \mid \mathit{history}) \cdot p(\mathit{history})}{\Big(p(\mathit{fail} \mid \mathit{history}) + p(\mathit{no\ fail} \mid \mathit{history}) \Big) \cdot p(\mathit{history})} = \\
&= p(\mathit{fail} \mid \mathit{history})
\end{aligned}$$

From this derivation, it is readily seen that the probability of interest $\Pr(\mathit{fail} \mid \mathit{history})$ is algebraically independent of the prior probabilities of *History*, the only node that is included in the set $\mathit{Insen}_1(\mathit{LV\ failure}, \{\mathit{History}\})$. □

Building upon the three preceding lemmas, we now state our main result.

Proposition 3.11 *Let B be a Bayesian belief network and let* Pr *be its joint probability distribution. Let O be the set of observed nodes with the observations o, as before. Let V_r be the network's node of interest and let $Sen(V_r, O)$ be the sensitivity set for V_r given O. Then, for any value v_r of V_r, we have that $\Pr(v_r \mid o) \not\sim p(V_i \mid \pi(V_i))$ for every node $V_i \notin Sen(V_r, O)$.*

Proposition 3.11 states that a belief network's probability of interest is algebraically independent of the conditional probabilities of any node that is not included in the sensitivity set under study. From this property, we have that sensitivity analyses with respect to these conditional probabilities are uninformative as they will reveal no effect whatsoever on the probability of interest. These sensitivity analyses, therefore, are to no avail and can be excluded from the overall analysis. The number of analyses that can be thus excluded may be considerable, as will be demonstrated in Section 5.

4 Functional relations in a sensitivity analysis

In the previous section, we have argued that a sensitivity analysis of a Bayesian belief network can be restricted to the conditional probabilities of the nodes in a sensitivity set under study: we know that the conditional probabilities of any other node do not contribute to the probability of interest and upon variation will not show any effect on this probability. To gain insight into the sensitivity of the probability of interest to the various conditional probabilities of the nodes that *are* included in the sensitivity set, further analysis is required. In essence, for every such conditional probability, the effect on the probability of interest can be studied by investigating a number of deviations from the specified assessment. Now, the curve yielded by such an analysis is not arbitrarily shaped, but instead is strongly constrained by the independences that are portrayed by the digraph of the network. In fact, the network's probability of interest relates as a quotient of two linear functions to a conditional probability under study. As we will argue presently, knowledge of this mathematical function renders systematic variation of conditional probabilities in a sensitivity analysis unnecessary.

Proposition 4.1 *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$ and let* Pr *be the joint probability distribution defined by B. Let $O \subseteq V(G)$ be the set of observed nodes in G and let o denote the corresponding observations. Let V_r be the network's node of interest and let $Sen(V_r, O)$ be the sensitivity set for V_r given O. Then, for any value v_r of V_r, we have that*

$$\Pr(v_r \mid o) = \frac{a \cdot x + b}{c \cdot x + d}$$

for every conditional probability $x = p(v_s \mid \pi')$ of every node $V_s \in Sen(V_r, O)$, where a, b, c, and d are constants that are dependent upon the values v_s of V_s and π' of $\pi_G(V_s)$.

Proof (Sketch). The probability of interest $\Pr(v_r \mid o)$ for the belief network B equals

$$\Pr(v_r \mid o) = \frac{\Pr(v_r \wedge o)}{\Pr(o)}$$

We recall that the joint probability distribution Pr, that is defined by the network, can be written as a product of the network's conditional probabilities. From the basic property of marginalisation, we further have that both the numerator and the denominator can be written as a sum of products of conditional probabilities. By separating, in these sums, the terms that specify the conditional probability x under study and those that do not, it is readily seen that $\Pr(v_r \wedge o)$ as well as $\Pr(o)$ relate linearly to x. □

We illustrate the property stated in Proposition 4.1 by means of an example.

Example 4.2 We consider the Bayesian belief network from Figure 7, which again is a small fragment of the ALARM-network. The possible values of the node *Shunt* are *normal* and *high*; the possible values of the node *Pulm emb* are *pulm emb* and *no pulm emb*, and the possible values of the node *PAP* are *low*, *normal*, and *high*. Our node of interest is the node *Shunt*, indicated in the figure by a double circle. We address the situation where the value *high* has been observed for

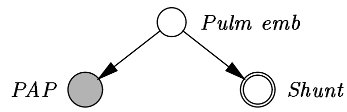

Figure 7: An example belief network, illustrating the property stated in Proposition 4.1 for the probability of interest Pr(*normal Shunt* | *high PAP*) and the conditional probability under study $p(high\ PAP \mid no\ pulm\ emb)$.

the node *PAP*, indicated by shading, and consider the probability of interest Pr(*normal Shunt* | *high PAP*). From Definition 3.1, we have that the sensitivity set $Sen(Shunt, \{PAP\})$ comprises all three nodes from the network. We now investigate the functional relation between the probability of interest and the conditional probability $x = p(high\ PAP \mid no\ pulm\ emb)$ for the node $PAP \in Sen(Shunt, \{PAP\})$. For our probability of interest, we find that

$$\Pr(normal\ Shunt \mid high\ PAP) = \frac{\Pr(normal\ Shunt \wedge high\ PAP)}{\Pr(high\ PAP)}$$

The numerator in this equation equals

$$\Pr(normal\ Shunt \wedge high\ PAP) =$$

$$= \sum_{\{Pulm\ emb\}} p(high\ PAP \mid Pulm\ emb) \cdot p(normal\ Shunt \mid Pulm\ emb) \cdot p(Pulm\ emb) =$$

$$= p(high\ PAP \mid no\ pulm\ emb) \cdot p(normal\ Shunt \mid no\ pulm\ emb) \cdot p(no\ pulm\ emb) +$$

$$+ p(high\ PAP \mid pulm\ emb) \cdot p(normal\ Shunt \mid pulm\ emb) \cdot p(pulm\ emb) =$$

$$= a \cdot x + b$$

where a equals

$$a = p(normal\ Shunt \mid no\ pulm\ emb) \cdot p(no\ pulm\ emb)$$

and b equals

$$b = p(high\ PAP \mid pulm\ emb) \cdot p(normal\ Shunt \mid pulm\ emb) \cdot p(pulm\ emb)$$

The denominator of the probability of interest equals

$$\Pr(high\ PAP) =$$

$$= \sum_{\substack{\{Shunt,\\ Pulm\ emb\}}} p(high\ PAP \mid Pulm\ emb) \cdot p(Shunt \mid Pulm\ emb) \cdot p(Pulm\ emb) =$$

$$= \sum_{\{Pulm\ emb\}} p(high\ PAP \mid Pulm\ emb) \cdot p(Pulm\ emb) =$$

$$= p(high\ PAP \mid no\ pulm\ emb) \cdot p(no\ pulm\ emb) + \ p(high\ PAP \mid pulm\ emb) \cdot p(pulm\ emb) =$$

$$= c \cdot x + d$$

where c equals

$$c = p(no\ pulm\ emb)$$

and d equals

$$d = p(high\ PAP \mid pulm\ emb) \cdot p(pulm\ emb)$$

From the previous derivations, it is readily seen that both the numerator $\Pr(normal\ Shunt \wedge high\ PAP)$ and the denominator $\Pr(high\ PAP)$ of the probability of interest $\Pr(normal\ Shunt \mid high\ PAP)$ relate linearly to the conditional probability $p(high\ PAP \mid no\ pulm\ emb)$. The probability of interest therefore relates as a quotient of two linear functions to this conditional probability. The sensitivity of the probability of interest with regard to the conditional probability under study is now uniquely determined by the values of the constants a, b, c, and d. These values are computed from the assessments for the appropriate conditional probabilities in the network:

$$\begin{aligned} p(high\ PAP \mid pulm\ emb) &= 0.8 \\ p(high\ PAP \mid no\ pulm\ emb) &= 0.05 \\ \\ p(normalShunt \mid pulm\ emb) &= 0.096 \\ p(normalShunt \mid no\ pulm\ emb) &= 0.905 \\ \\ p(pulm\ emb) &= 0.01 \end{aligned}$$

We find that

$$\begin{aligned} a &= 0.896 \\ b &= 0.00076 \\ c &= 0.99 \\ d &= 0.008 \end{aligned}$$

The mathematical function relating the probability of interest $Pr(normal\ Shunt \mid high\ PAP)$ to the conditional probability $x = p(high\ PAP \mid no\ pulm\ emb)$ therefore equals

$$\Pr(normal\ Shunt \mid high\ PAP) = \frac{0.896 \cdot x + 0.00076}{0.99 \cdot x + 0.008}$$

The function is depicted in Figure 8. Note that the probability of interest shows a high sensitivity for the conditional probability under study at the specified assessment 0.05. □

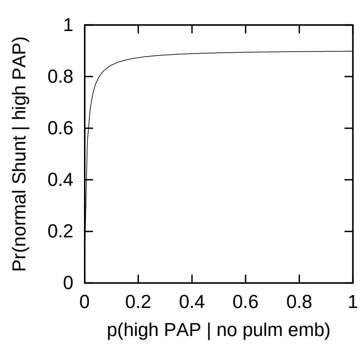

Figure 8: The function relating the probability of interest $\Pr(normal\ Shunt \mid high\ PAP)$ to the conditional probability under study $p(high\ PAP \mid no\ pulm\ emb)$.

So far, we have shown that a belief network's probability of interest relates as a quotient of two linear functions to a conditional probability under study. For a conditional probability that pertains to a node from the sensitivity set that does not have any observed descendants, this functional relation reduces to a *linear function.* The following proposition states this property more formally.

Proposition 4.3 *Let B be a Bayesian belief network and let* Pr *be its joint probability distribution. Let O be the set of observed nodes with the corresponding observations o, as before. Let V_r be the network's node of interest and let $Sen(V_r, O)$ be the sensitivity set for V_r given O. Let $V_s \in Sen(V_r, O)$ with $\sigma^*(V_s) \cap O = \varnothing$. Then, for any value v_r of V_r, we have that*

$$\Pr(v_r \mid o) = a \cdot x + b$$

for every conditional probability $x = p(v_s \mid \pi')$ of V_s, where a and b are constants that are dependent upon the values v_s of V_s and π' of $\pi_G(V_s)$.

Proof (Sketch). The probability of interest $\Pr(v_r \mid o)$ for the belief network B once more equals

$$\Pr(v_r \mid o) = \frac{\Pr(v_r \wedge o)}{\Pr(o)}$$

From the proof of Proposition 4.1, we have that the numerator $\Pr(v_r \wedge o)$ in this equation relates linearly to the conditional probability x under study. Now, with regard to the probability $\Pr(o)$, we recall from the previous section that, if no observations are available for descendants of a non-ancestor, the probability of interest is algebraically independent of the conditional probabilities of this node. Likewise, the probability of a combination of observations is algebraically independent of the conditional probabilities of any non-ancestor without observed descendants. From this property, we have that the probability $\Pr(o)$ is algebraically independent of the conditional probability x under study. We conclude that $\Pr(o)$ is a constant with respect to x. □

We illustrate the property stated in the previous proposition by means of an example.

Example 4.4 We consider again the belief network from Figure 7. Once more, we address the situation where the value *high* has been observed for the node *PAP*, and consider the probability of interest $\Pr(\textit{normal Shunt} \mid \textit{high PAP})$. As mentioned in Example 4.2, the sensitivity set $Sen(\textit{Shunt}, \{\textit{PAP}\})$ comprises all three nodes from the network. We now investigate the functional relation between the probability of interest and the conditional probability $p(\textit{normal Shunt} \mid \textit{pulm emb})$ for the node $\textit{Shunt} \in Sen(\textit{Shunt}, \{\textit{PAP}\})$. Note that the node *Shunt* does not have any observed descendants. For our probability of interest, we once again find that

$$\Pr(\textit{normal Shunt} \mid \textit{high PAP}) = \frac{\Pr(\textit{normal Shunt} \wedge \textit{high PAP})}{\Pr(\textit{high PAP})}$$

The numerator in this equation equals

$$\Pr(\textit{normal Shunt} \wedge \textit{high PAP}) =$$

$$= \sum_{\{\textit{Pulm emb}\}} p(\textit{high PAP} \mid \textit{Pulm emb}) \cdot p(\textit{normal Shunt} \mid \textit{Pulm emb}) \cdot p(\textit{Pulm emb}) =$$

$$= p(\textit{high PAP} \mid \textit{pulm emb}) \cdot p(\textit{normal Shunt} \mid \textit{pulm emb}) \cdot p(\textit{pulm emb}) +$$

$$+ p(\textit{high PAP} \mid \textit{no pulm emb}) \cdot p(\textit{normal Shunt} \mid \textit{no pulm emb}) \cdot p(\textit{no pulm emb}) =$$

$$= a' \cdot x + b'$$

where a' equals

$$a' = p(\textit{high PAP} \mid \textit{pulm emb}) \cdot p(\textit{pulm emb})$$

and b' equals

$$b' = p(\mathit{high\ PAP} \mid \mathit{no\ pulm\ emb}) \cdot p(\mathit{normal\ Shunt} \mid \mathit{no\ pulm\ emb}) \cdot p(\mathit{no\ pulm\ emb})$$

The denominator of the probability of interest equals

$$\Pr(\mathit{high\ PAP}) =$$

$$= \sum_{\substack{\{\mathit{Shunt},\\ \mathit{Pulm\ emb}\}}} p(\mathit{high\ PAP} \mid \mathit{Pulm\ emb}) \cdot p(\mathit{Shunt} \mid \mathit{Pulm\ emb}) \cdot p(\mathit{Pulm\ emb}) =$$

$$= \sum_{\{\mathit{Pulm\ emb}\}} p(\mathit{high\ PAP} \mid \mathit{Pulm\ emb}) \cdot p(\mathit{Pulm\ emb}) =$$

$$= p(\mathit{high\ PAP} \mid \mathit{pulm\ emb}) \cdot p(\mathit{pulm\ emb}) + p(\mathit{high\ PAP} \mid \mathit{no\ pulm\ emb}) \cdot p(\mathit{no\ pulm\ emb}) =$$

$$= c'$$

The previous derivations show that the denominator $\Pr(\mathit{high\ PAP})$ of the probability of interest does not depend on the conditional probability under study $p(\mathit{normal\ Shunt} \mid \mathit{pulm\ emb})$. The numerator $\Pr(\mathit{normal\ Shunt} \wedge \mathit{high\ PAP})$ relates linearly to this conditional probability. We conclude that our probability of interest relates linearly to the conditional probability under study:

$$\Pr(\mathit{normal\ Shunt} \mid \mathit{high\ PAP}) = \frac{a' \cdot x + b'}{c'} = a \cdot x + b$$

The sensitivity of the probability of interest with regard to the conditional probability under study is now uniquely determined by the values of the constants a and b. These values again are computed from the assessments for the appropriate conditional probabilities in the network, as specified in Example 4.2. We find that

$a = 0.139$
$b = 0.779$

The linear function relating the probability of interest $\Pr(\mathit{normal\ Shunt} \mid \mathit{high\ PAP})$ to the conditional probability $p(\mathit{normal\ Shunt} \mid \mathit{pulm\ emb})$, denoted by x, therefore equals

$$\Pr(\mathit{normal\ Shunt} \mid \mathit{high\ PAP}) = 0.139 \cdot x + 0.779$$

The function is depicted in Figure 9. □

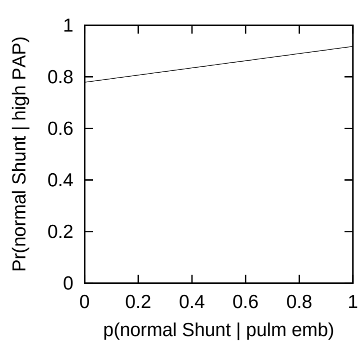

Figure 9: The function relating the probability of interest $\Pr(\mathit{normal\ Shunt} \mid \mathit{high\ PAP})$ to the conditional probability under study $p(\mathit{normal\ Shunt} \mid \mathit{pulm\ emb})$.

We would like to note that, in the special case where none of the nodes in a Bayesian belief network are observed, Proposition 4.3 implies that the network's probability of interest relates linearly to *every* conditional probability of *every* node from the sensitivity set under study.

Corollary 4.5 *Let B be a Bayesian belief network and let* Pr *be its joint probability distribution. Let V_r be the network's node of interest and let $Sen(V_r, \varnothing)$ be the sensitivity set for V_r given the empty set of observed nodes. Let $V_s \in Sen(V_r, \varnothing)$. Then, for any value v_r of V_r, we have that*

$$\Pr(v_r \mid o) = a \cdot x + b$$

for every conditional probability $x = p(v_s \mid \pi')$ of V_s, where a and b are constants that are dependent upon the values v_s of V_s and π' of $\pi_G(V_s)$.

In the foregoing, we have argued that a belief network's probability of interest relates to a conditional probability under study by a simple mathematical function. Knowledge of this function allows for considerably reducing the computational burden of a one-way sensitivity analysis of a Bayesian belief network as only the constants in the function need be known. These constants can be determined by computing the probability of interest from the network for a small number of values for a conditional probability under study and solving the resulting system of equations; systematic variation of the conditional probability is then no longer necessary. For a conditional probability that is related linearly to the probability of interest, two network computations suffice; for all other conditional probabilities, three network computations are required. The following example illustrates the basic idea.

Example 4.6 We consider again the belief network from Figure 7. As in Example 4.2, we investigate the functional relation between the probability of interest $\Pr(\textit{normal Shunt} \mid \textit{high PAP})$ and the conditional probability $x = p(\textit{high PAP} \mid \textit{no pulm emb})$. We recall that this function equals

$$\Pr(\textit{normal Shunt} \mid \textit{high PAP}) = \frac{a \cdot x + b}{c \cdot x + d}$$

In Example 4.2, we have determined the values of the constants a, b, c, and d by expressing every constant in terms of conditional probabilities from the network and subsequently filling in the appropriate assessments. The functional relation can be determined more efficiently, however, by computing the probability of interest from the network for three different values of the conditional probability under study. Note that three network computations suffice since the constant c can be eliminated from the above equation, yielding

$$\Pr(\textit{normal Shunt} \mid \textit{high PAP}) = \frac{a' \cdot x + b'}{x + c'}$$

Using the three values $x = 0.2$, $x = 0.4$, and $x = 0.6$ for the conditional probability under study and the assessments for the other conditional probabilities as specified in Example 4.2, we find by computing the probability of interest $\Pr(\textit{normal Shunt} \mid \textit{high PAP})$ from the network, the values

$$\begin{aligned} \Pr(\textit{normal Shunt} \mid \textit{high PAP})_{x=0.2} &= 0.87356 \\ \Pr(\textit{normal Shunt} \mid \textit{high PAP})_{x=0.4} &= 0.88897 \\ \Pr(\textit{normal Shunt} \mid \textit{high PAP})_{x=0.6} &= 0.89424 \end{aligned}$$

From these values, we now obtain the three linear equations

$$\frac{0.2 \cdot a' + b'}{0.2 + c'} = 0.87356 \quad \Rightarrow \quad 0.2 \cdot a' + b' - 0.87356 \cdot c' - 0.2 \cdot 0.87356 = 0$$

$$\frac{0.4 \cdot a' + b'}{0.4 + c'} = 0.88897 \quad \Rightarrow \quad 0.4 \cdot a' + b' - 0.88897 \cdot c' - 0.4 \cdot 0.88897 = 0$$

$$\frac{0.6 \cdot a' + b'}{0.6 + c'} = 0.89424 \quad \Rightarrow \quad 0.6 \cdot a' + b' - 0.89424 \cdot c' - 0.6 \cdot 0.89424 = 0$$

Solving this system of linear equations gives

$a' = 0.905$
$b' = 0.00061$
$c' = 0.00789$

It is readily verified, by dividing the values of the constants a, b, and d specified in Example 4.2 by the value of the constant c, that the mathematical function yielded coincides with the function found in Example 4.2. □

5 Experimental results

In the previous sections, we have detailed various properties that allow for reducing the computational burden of a one-way sensitivity analysis of a Bayesian belief network. In Section 3, we have argued that a belief network's probability of interest is algebraically independent of the conditional probabilities of any node that is not included in the sensitivity set under study. As sensitivity analyses with respect to these conditional probabilities are uninformative, they can be excluded from the overall analysis. In Section 4, we have argued that for any conditional probability, that pertains to a node that is included in the sensitivity set, a small number of network computations suffice to determine the sensitivity of the probability of interest with regard to a conditional probability under study. Systematic variation of conditional probabilities then is no longer necessary. Now, to gain insight into the effect of exploiting these properties, we have conducted several experiments on randomly generated Bayesian belief networks. In these experiments, we have investigated, for various different sets of networks, the number of nodes in the sensitivity set under study and the number of nodes whose conditional probabilities are related linearly to the probability of interest, as these numbers reflect the computational burden of a network's sensitivity analysis.

In each experiment, we have generated a set of one thousand connected acyclic digraphs; for details of the graph-generator used, we refer the reader to [Van der Gaag, 1994]. We have generated various sets of digraphs with fifty nodes each, comprising fifty, seventy five, one hundred, one hundred and fifty, two hundred, and two hundred and fifty arcs, respectively. As our investigations are concerned with the digraph of a Bayesian belief network only, we have refrained from quantifying the generated digraphs with conditional probabilities. For each digraph from every set, we have randomly selected a single node of interest and k observed nodes, where, for the various sets of digraphs, k is varied from zero to thirty by steps of two nodes.

To study the behaviour of our method on Bayesian belief networks that have been developed for different types of application, we have also generated various sets of digraphs for which a *diagnostic* and a *prognostic* bias, respectively, have been used in the selection of the node of interest and of the observed nodes. For diagnostic applications, we have assumed that a belief network's node of interest tends to be located in the upper part of the digraph, whereas the observed nodes are likely to be situated in its lower part. For prognostic applications, on the other hand, we have assumed that the node of interest tends to be located in the lower part of the digraph and the observed nodes in the upper part. The two biases have been realised as a two-stage selection. The selection of a node of interest in the lower part of a digraph, for example, starts with selecting a single auxiliary node in a random fashion. The node of interest is then selected from among the nodes that are assigned a lower number in a topological ordering of the digraph than the auxiliary node. For computational reasons, the maximum number of observed nodes considered with the diagnostic and prognostic biases, respectively, has been limited to sixteen nodes.

In each experiment, we have determined, for every digraph, the number of nodes in the sensitivity set for the selected node of interest given the set of observed nodes under study, and the number of nodes whose conditional probabilities are related linearly to the probability of interest. The results are summarised in Figure 10. Figure 10(a) and Figure 10(b) pertain to the digraphs for which no bias has been used in the selection of the node of interest and of the observed nodes. Figure 10(a) shows the average number of nodes in the sensitivity set, plotted against the number of observed nodes; the six curves pertain to the sets of digraphs with different numbers of arcs.

Figure 10(b) shows the average number of nodes, from the sensitivity set, whose conditional probabilities are related linearly to the probability of interest. Figure 10(c) and Figure 10(d) depict the same information for the digraphs for which a diagnostic bias has been used in the selection of the node of interest and of the observed nodes; Figure 10(e) and Figure 10(f) show the information for the digraphs for which a prognostic bias has been used.

To discuss the results obtained from our experiments, we start by considering the average number of nodes in the sensitivity set for digraphs for which the node of interest and the set of observed nodes have been selected randomly. From Figure 10(a), we see that the average number of nodes in the sensitivity set increases at first, with an increasing number of observed nodes. This property is readily explained by observing that, initially, only observations for ancestors of the node of interest, that is, only observations for nodes from the sensitivity set, allow for diminishing the set's size. For all other nodes in the digraph, an observation will either have no effect or increase the size of the sensitivity set. In the digraphs under consideration, the node of interest will, on average, be located in 'the middle' of the digraph. The number of ancestors of this node will, on average, be smaller than its number of non-ancestors. The tendency of additional observations for the ancestors of the node of interest to decrease the size of the sensitivity set will therefore be outweighed by the tendency of additional observations for its non-ancestors to increase the sensitivity set's size. Now, for a still further increasing number of observed nodes, the increase in size of the sensitivity set diminishes. In fact, when roughly eighteen observed nodes have been selected, additional observations cause the sensitivity set to decrease in size. This property is explained by observing that a new node can only be inserted into the sensitivity set if one of its descendants is selected as an observed node where it had no observed descendants before. The more observed nodes have been selected, however, the fewer nodes remain without observed descendants. On the other hand, additional observations for nodes from the sensitivity set will serve to decrease the set's size. For larger numbers of observed nodes, the sensitivity set will be quite large and the latter tendency will therefore outweigh the former, resulting in an overall decrease in the size of the sensitivity set. Figure 10(a) further reveals that a larger number of arcs in a belief network's digraph will result in a larger sensitivity set. This property is explained by observing that, in a digraph with more arcs, the node of interest is likely to have more ancestors, resulting in a larger sensitivity set to begin with. Moreover, a larger number of arcs will, on average, result in a larger number of chains between a node under consideration and the node of interest. To block the influence of this node's conditional probabilities on the probability of interest, that is, to exclude the node from the sensitivity set, on average, a larger number of observations is required. For a fixed number of observed nodes, therefore, an increase in the number of arcs leads to an increase in size of the sensitivity set.

We now consider the average number of nodes, from a sensitivity set under study, whose conditional probabilities are related linearly to the selected probability of interest. From Figure 10(b), we see that this number diminishes with an increasing number of observed nodes. This property is readily explained by observing that only the conditional probabilities of ancestors of the node of interest that do not have any observed descendants, are related linearly to the probability of interest. The more observed nodes have been selected, the fewer ancestors of the node of interest remain without observed descendants and, hence, the smaller the number of nodes whose conditional probabilities are related linearly to the probability of interest. Figure 10(b) further shows that, for a fixed number of observed nodes, the number of linearly related nodes increases with an increasing number of arcs, which conforms with the tendency of the number of ancestors of the node of interest to increase with the number of arcs.

We proceed with addressing the results from our experiments with digraphs for which a diagnostic bias has been used in the selection of the node of interest and of the set of observed nodes. From Figure 10(c) and Figure 10(d), we see that these digraphs show tendencies similar to those shown by digraphs for which no bias has been used. The initial increase in the size of the sensitivity set with an increasing number of observed nodes, however, is stronger and reaches a higher maximum for the digraphs with a diagnostic bias than for the digraphs for which no bias has been used. This property is readily explained by once more observing that, initially, only observations for ancestors of the node of interest allow for diminishing the size of the sensitivity

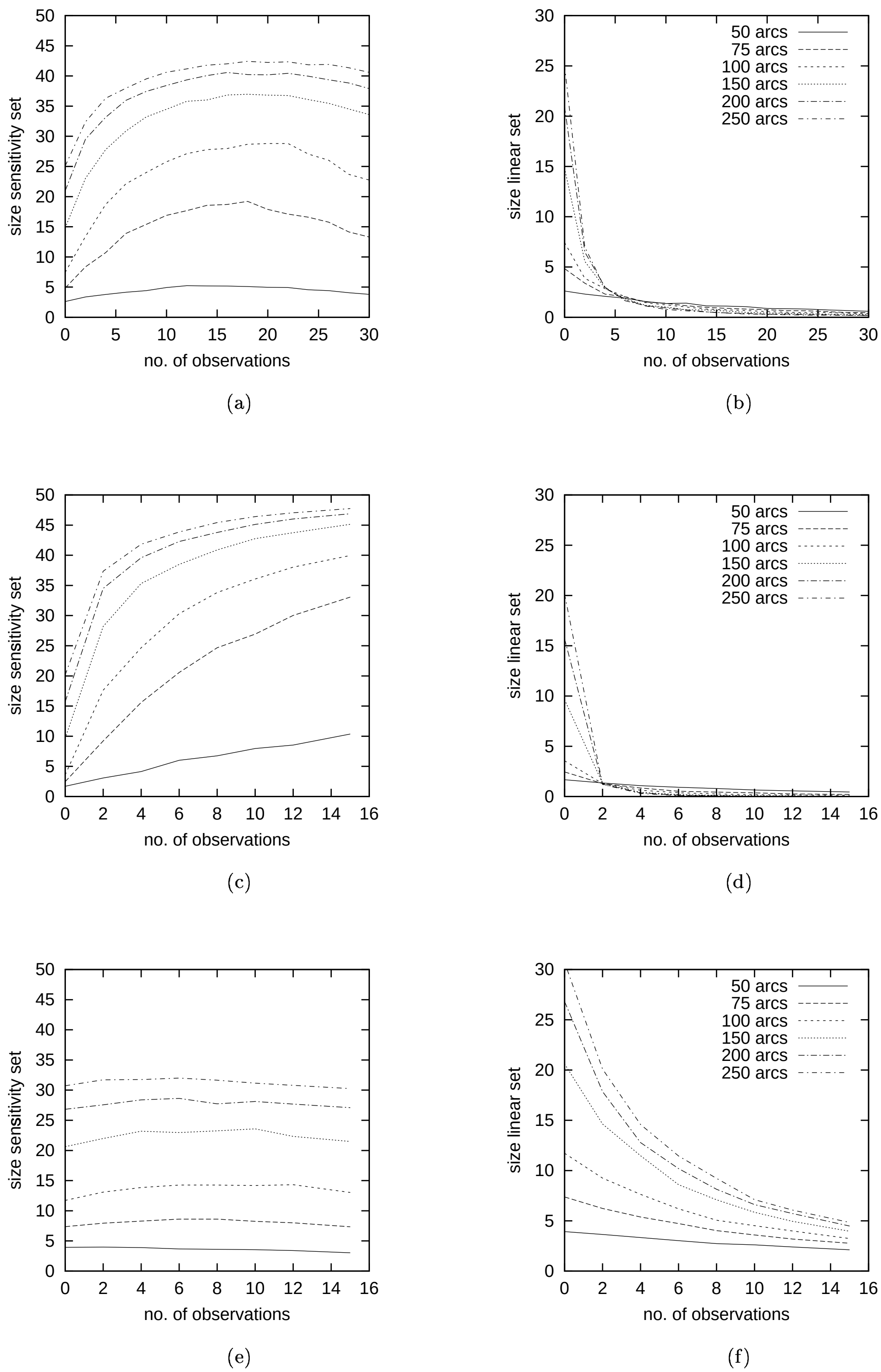

Figure 10: The average number of nodes in the sensitivity set under study and the average number of linearly related nodes, for various sets of networks without any bias (figures (a) and (b)), with a diagnostic bias (figures (c) and (d)), and with a prognostic bias (figures (e) and (f)), respectively.

set. Since the node of interest in digraphs with a diagnostic bias is, on average, situated higher in the digraph than in unbiased digraphs, its ratio of the number of ancestors to the number of non-ancestors will, on average, be smaller. As a result, the tendency of additional observations for non-ancestors to increase the size of the sensitivity set is even more dominant in digraphs for which a diagnostic bias has been used than in digraphs without any bias. The smaller number of ancestors further accounts for the stronger decrease of the number of nodes from the sensitivity set whose conditional probabilities are related linearly to the probability of interest, as revealed in Figure 10(d).

We now consider the results from our experiments with digraphs for which a prognostic bias has been used in the selection of the node of interest and of the set of observed nodes. Figure 10(e) suggests that the size of the sensitivity set for these digraphs remains reasonably constant with an increasing number of observed nodes. The tendency of additional observations for the ancestors of the node of interest to decrease the size of the sensitivity set is therefore balanced, in these digraphs, by the tendency of observations for its non-ancestors to increase the sensitivity set's size. We feel that this property is the coincidental result of the 'degree' of prognostic bias we have used. We expect that a more extreme location of the node of interest and of the observed nodes in the digraphs under study, that is, a larger ratio of the number of ancestors to the number of non-ancestors, will lead to a decrease in the size of the sensitivity set with an increasing number of observations. In fact, further experiments, using a three-stage selection for a prognostic bias, have met this expectation. Figure 10(e) further shows that the size of the sensitivity set increases with an increasing number of arcs, as we have seen before for the digraphs without any bias as well as for the digraphs for which a diagnostic bias has been used. Similar tendencies as for unbiased and for diagnostic digraphs are also seen in Figure 10(f) with respect to the number of nodes from the sensitivity set whose conditional probabilities are related linearly to the probability of interest.

6 Previous work

Sensitivity analysis is a general, well-known technique for studying the effects of the inaccuracies in the parameters of a mathematical model on the model's output; it is widely used in mathematical modelling in various different domains of application [Morgan & Henrion, 1990, Habbema *et al.*, 1990, Dippel *et al.*, 1992, Helton, 1993, Doubilet *et al.*, 1985]. As more and more Bayesian belief networks are being developed for real-life applications, interest in sensitivity analysis of belief networks is increasing. In this section, we review previous work on sensitivity analysis of belief networks. In doing so, we do not intend to give an exhaustive overview of the state of the art. We merely discuss the results from related work and compare it with the results that we have presented in the current paper.

In her work on sensitivity analysis of Bayesian belief networks, K. Blackmond Laskey has been motivated, as in fact we have been, by the observation that straightforward, systematic variation of the assessments of a network's conditional probabilities is too much time-consuming to be of practical use. She has developed an efficient method for analytically computing first-order approximations of exact analyses [Laskey, 1995]. Her method sets out by identifying, in a belief network under study, conditional probabilities that upon variation have no effect on a probability of interest. Laskey suggests two procedures for this purpose. She suggests that the assessment of every single conditional probability be varied over a small number of values, serving to reveal all conditional probabilities that have no influence on the probability of interest. For an alternative procedure, she observes that some uninfluential probabilities can be identified using graphical considerations. For this purpose, she introduces a concept similar to our sensitivity set; in fact, our notion of sensitivity set has been inspired to a large extent by her concept. Laskey's method excludes the identified uninfluential conditional probabilities from further analysis. For the remaining conditional probabilities, the effect of variation on the network's probability of interest is measured by a so-called sensitivity value. A sensitivity value is the partial derivative of the probability of interest with respect to a conditional probability under study. A sensitivity value thus provides an approximation of the effect of small deviations from the probability's assessment on the probability of interest. Laskey presents two procedures for analytically computing sensitivity values; these procedures build upon the propagation algorithm by Lauritzen and Spiegelhalter and upon Monte Carlo sampling, respectively. Compared to straightforward variation of conditional probabilities in a sensitivity analysis, Laskey's method requires considerably less computational effort.

In her method, Laskey has introduced a powerful concept upon which we have built our concept of sensitivity set. She suggests, as we do, to construct, from a belief network's digraph, an auxiliary digraph in which a predecessor X_i is added to every node V_i. She proceeds by observing that, if the auxiliary predecessor X_i of a node V_i is d-separated from the *auxiliary predecessor* X_r of the node of interest V_r, then sensitivity analyses with respect to the conditional probabilities for node V_i are uninformative as these probabilities cannot influence the probability of interest upon variation. Her observation, unfortunately, is incorrect, as it can declare several conditional probabilities to be not influential while in fact they are. For example, from her observation, we would conclude that, if no observations are available as yet, the conditional probabilities of all ancestors of the node of interest V_r are uninfluential, since $\langle\{X_i\} \mid \varnothing \mid \{X_r\}\rangle^d$ for all $V_i \in \pi^*(V_r) \setminus \{V_r\}$. To show that this conclusion is incorrect, we give an example from the ALARM-network. We are interested in the probability that the node *LV failure* takes the value *true* when no observations are available as yet. For the probability of interest, we have that

$$\Pr(\mathit{fail}) = p(\mathit{fail} \mid \mathit{history}) \cdot p(\mathit{history}) + p(\mathit{fail} \mid \mathit{no\ history}) \cdot p(\mathit{no\ history})$$

which reveals that the probability of interest is algebraically dependent upon the probabilities of the nodes *History* and *LV failure*. Since, in the absence of observations, the auxiliary predecessor of the node *History* is d-separated from the auxiliary predecessor of *LV failure*, building upon Laskey's observation would incorrectly declare the prior probabilities of the node *History* to be uninfluential. With the various lemmas presented in Section 3 of the current paper, we have shown that our concept of sensitivity set provides for correctly identifying uninfluential nodes.

As mentioned before, Laskey's method of computing sensitivity values requires considerably less computational effort than straightforward variation of probability assessments for studying sensitivity. The method, however, provides insight in the effect of small deviations from a probability's assessment only: as Laskey indicates, when larger deviations are considered, the quality of the approximation may break down rapidly. For the ALARM-network, Figure 8 illustrates how an approximation may fail to reveal the extent of the sensitivity of a probability of interest to a conditional probability under study. The figure shows the effect of variation of the assessment for the conditional probability $p(\mathit{high\ PAP} \mid \mathit{no\ pulm\ emb})$ on the probability of interest $\Pr(\mathit{normal\ Shunt} \mid \mathit{high\ PAP})$. The assessment specified for the conditional probability under consideration is 0.05. For variation of this assessment to higher values, the derivative of the sensitivity function does not change rapidly. The derivative at the specified assessment therefore provides a good approximation of the effect on the probability of interest for larger values. However, even a slight shift in the specified assessment to a smaller value has a very large effect on the derivative of the sensitivity function. The approximation therefore does not suffice. We feel that exact sensitivity analysis of a Bayesian belief network is to be preferred to approximate analysis.

We briefly review two other methods for sensitivity analysis of Bayesian belief networks that take a different approach than our method. In [Chang & Fung, 1995] and [Castillo *et al.*, 1997], the idea of symbolic propagation in belief networks is exploited for studying sensitivity. Instead of yielding a single number as the standard propagation algorithms do, a symbolic propagation algorithm yields an algebraic expression for a network's probability of interest in terms of all conditional probabilities in the network. From this expression, the sensitivity of the probability of interest to a conditional probability under study is readily computed, basically by filling in the specified assessments for all other conditional probabilities. A disadvantage of building upon symbolic propagation is that it is quite time-consuming. We therefore feel that methods for sensitivity analysis that build upon the faster standard propagation algorithms are preferred. In [Spiegelhalter, 1989], a method for sensitivity analysis of Bayesian belief networks is presented that builds upon an explicit specification of the inaccuracies in a network's conditional probabilities. As in our method, an auxiliary graph is constructed from the digraph of a belief network by adding an auxiliary predecessor to every node. The auxiliary predecessor now captures second-order distributions for the conditional probabilities of its successor. Using standard propagation algorithms, the effects of the specified inaccuracies on a probability of interest are readily computed. A disadvantage of this method is that it requires an explicit specification of the inaccuracies in a belief network's probability assessments. As second-order distributions for the specified assessments often are not available, assumptions on the nature of the inaccuracies have to be made that may not be realistic. We would like to note that with our method for studying sensitivity no assumptions with regard to the inaccuracies involved are necessary.

While in this paper we have focused on *sensitivity analysis* of Bayesian belief networks, we would like to note that the reliability of a belief network's output can in addition be studied by subjecting the network to an *uncertainty analysis*. In an uncertainty analysis of a belief network, the assessments of *all* conditional probabilities of the network's quantitative part are varied simultaneously. To this end, for each conditional probability, values are drawn from some probability distribution. Uncertainty analysis of a Bayesian belief network serves to reveal the overall reliability of the network's output. Uncertainty analysis, however, yields less insight into the effect of single conditional probabilities than sensitivity analysis does. Previous experiments with uncertainty analysis of Bayesian belief networks have led to the suggestion that belief networks are highly insensitive to inaccuracies in the assessments of their conditional probabilities [Henrion *et al.*, 1996, Pradhan *et al.*, 1996]. In these experiments, performed on belief networks for diagnostic applications, a measure of the reliability of a network's diagnosis is obtained by assuming a log-normal distribution for every conditional probability, having the initially specified assessment for its mean, and subsequently averaging over the probability of the true diagnosis for various diagnostic situations. Unfortunately, when using probability distributions to model inaccuracies in the assessments for a network's conditional probabilities, it is not the *average* of the probabilities of the true diagnosis that reflects the effects of these inaccuracies, but the *variation* in these probabilities. In addition, we would like to note that the reported results are based on experience with a single belief network only, in which the conditional probability distributions have been simplified using noisy-OR and noisy-MAX assumptions. From the results reported so far for uncertainty analysis of Bayesian belief networks, therefore, no decisive conclusions can be drawn. We feel that the sensitivity of a network's probability of interest to the various conditional probabilities involved will vary from application to application. In fact, sensitivity analysis of a Bayesian belief network for congenital heart disease has shown that a network's conditional probabilities can have a large effect on a probability of interest [Coupé *et al.*, 1999b].

7 Conclusions

The assessments obtained for the various conditional probabilities of a Bayesian belief network are inevitably inaccurate, due to incompleteness of data and partial knowledge of the problem under study. The inaccuracies in these probability assessments can severely compromise the reliability of the network's output. To gain insight into the reliability of a probability of interest computed from a belief network, the network can be subjected to a sensitivity analysis. A sensitivity analysis can be performed by systematically varying the assessments for one or more of the network's conditional probabilities simultaneously. We have argued that even for a rather small belief network such a straightforwardly performed analysis is highly time-consuming. In this paper, we have shown that, by qualitative considerations pertaining to a belief network's digraph, various conditional probabilities can be identified that upon variation cannot influence the network's probability of interest. Analyses with respect to these probabilities are uninformative and can therefore be excluded from the overall analysis. More specifically, we have shown that a sensitivity analysis of a Bayesian belief network can be restricted to the conditional probabilities of the nodes from the sensitivity set for the network's node of interest. Excluding uninformative analyses can lead to a considerable reduction in the computational burden of a sensitivity analysis, as is evidenced by the results from the experiments we have performed on randomly generated belief networks. In the paper, we have further shown that for sensitivity analyses that are informative, simple mathematical functions exist expressing the network's probability of interest in terms of the conditional probabilities under study. Knowledge of these functions allows for even further reduction of the computational burden of a sensitivity analysis, as only the constants in the functions need be determined, rendering systematic variation of conditional probabilities unnecessary.

In this paper, we have focused attention on a one-way sensitivity analysis of a Bayesian belief network in which the network's conditional probabilities are investigated one at a time. For such an analysis, we have detailed the mathematical function expressing the network's probability of interest in terms of a single conditional probability. More specifically, we have shown that, in general, the probability of interest relates as a quotient of two linear functions to a conditional probability under study. In essence, it is also possible to investigate the effect of simultaneous variation of two or more conditional probabilities. Such a higher-order sensitivity analysis can, just as a one-way analysis, be restricted to the conditional probabilities of the nodes that are included in the sensitivity set for a belief network's node of interest. Moreover, for higher-order sensitivity analyses also functional relations exist between a network's probability of interest and the conditional probabilities under study. Although not reported in this paper, we have detailed the functions that hold in a two-way sensitivity analysis in which conditional probabilities are studied pairwise. These functions comprise terms for the separate effects of each of the two conditional probabilities being investigated as well as terms for their joint effect. More specifically, the probability of interest, in general, relates as a quotient of two bi-linear functions to the probabilities under study. The more conditional probabilities of a belief network are investigated simultaneously, the more involved the mathematical functions will be. We feel that the results of higher-order sensitivity analyses in which three or more conditional probabilities are studied simultaneously will in general be very hard to interpret.

In the near future, we envision further experiments with our method of sensitivity analysis on real-life Bayesian belief networks. In these experiments, we would like to study the reliability of belief network's output in general. Also, we would like to evaluate in more detail the effect of the location of the node of interest and of the observed nodes in a network's digraph. In addition, we envision further investigation of the properties of sensitivity analysis, both from a theoretical and an experimental point of view. Our experiments so far on randomly generated belief networks and on the ALARM-network have shown considerable computational savings. Motivated by these initial results, we hope to be able to arrive at a generally applicable, practicable method for sensitivity analysis of Bayesian belief networks.

References

[Andreassen *et al.*, 1987] S. Andreassen, M. Woldbye, B. Falck, S.K. Andersen (1987). MUNIN – A causal probabilistic network for interpretation of electromyographic findings. *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pp. 366 – 372.

[Beinlich *et al.*, 1989] I.A. Beinlich, H.J. Suermondt, R.M. Chavez, G.F. Cooper (1989). The ALARM monitoring system: a case study with two probabilistic inference techniques for belief networks. In: J. Hunter, J. Cookson, J. Wyatt (Eds). *Proceedings of the Second Conference on Artificial Intelligence in Medicine*, Springer-Verlag, Berlin, pp. 247 – 256.

[Castillo *et al.*, 1997] E. Castillo, J.M. Gutiérrez, A.S. Hadi (1997). Sensitivity analysis in discrete Bayesian Networks. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 27, pp. 412 – 423.

[Chang & Fung, 1995] K. Chang, R. Fung (1995). Symbolic probabilistic inference with both discrete and continuous variables. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 25, pp. 910 – 916.

[Coupé *et al.*, 1999a] V.M.H. Coupé, L.C. van der Gaag, J.D.F. Habbema (1999). Sensitivity analysis: an aid for belief-network quantification. *Knowledge Engineering Review*, to appear.

[Coupé *et al.*, 1999b] V.M.H. Coupé, N. Peek, J. Ottenkamp, J.D.F. Habbema (1999). Using sensitivity analysis for efficient quantification of a belief-network. *Artificial Intelligence in Medicine*, to appear.

[Dippel *et al.*, 1992] D.W.J. Dippel, J.W.M. ter Berg, J.D.F. Habbema (1992). Screening for unruptured familial intracranial aneurysms. A decision analysis. *Acta Neurologica Scandinavica*, vol. 86, pp. 381 – 389.

[Doubilet *et al.*, 1985] P. Doubilet, C.B. Begg, M.C. Weinstein, P. Braun, B.J. McNeil (1985). Probabilistic sensitivity analysis using monte carlo simulation: a practical approach. *Medical Decision Making*, vol. 5, pp. 157 – 177.

[Druzdzel & Van der Gaag, 1995] M.J. Druzdzel, L.C. van der Gaag (1995). Elicitation of probabilities for belief networks: combining qualitative and quantitative information. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 141 – 148.

[Habbema *et al.*, 1990] J.D.F. Habbema, P.M.M. Bossuyt, D.J.W. Dippel (1990). Analysing clinical decision analyses. *Statistics in Medicine*, vol. 9, pp. 1229 – 1242.

[Heckerman *et al.*, 1992] D.E. Heckerman, E.J. Horvitz, B.N. Nathwani (1992). Toward normative expert systems. Part 1: The Pathfinder project. *Methods of Information in Medicine*, vol. 31, pp. 90 – 105.

[Helton, 1993] J.C. Helton (1993). Uncertainty and sensitivity analysis techniques for use in performance assessment for radoactive waste disposal. *Reliability Engineering and System Safety*, vol. 42, pp. 327 – 367.

[Henrion *et al.*, 1996] M. Henrion, M. Pradhan, B. Del Favero, K. Huang, G. Provan, P. O'Rorke (1996). Why is diagnosis using belief networks insensitive to imprecision in probabilities ? *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence*, pp. 307 – 314.

[Laskey, 1995] K.B. Laskey (1995). Sensitivity analysis for probability assessments in Bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 25, pp. 901 – 909.

[Lauritzen & Spiegelhalter, 1988] S.L. Lauritzen, D.J. Spiegelhalter (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society, Series B*, vol. 50, pp. 157 – 224.

[Morgan & Henrion, 1990] M.G. Morgan, M. Henrion (1990). *Uncertainty, a Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, Cambridge.

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

[Pradhan *et al.*, 1996] M. Pradhan, M. Henrion, G. Provan, B. Del Favero, K. Huang (1996). The sensitivity of belief networks to imprecise probabilities: an experimental investigation. *Artificial Intelligence*, vol. 85, pp. 363 – 397.

[Spiegelhalter, 1989] D. Spiegelhalter (1989). A unified approach to imprecision and sensitivity of beliefs in expert systems. In: L.N. Kanal, T.S. Levitt, J.F. Lemmer (Editors). *Uncertainty in Artificial Intelligence 3*, North-Holland, Amsterdam, pp. 199 – 208.

[Tversky *et al.*, 1982] D. Kahneman, P. Slovic, and A. Tversky (1982). *Judgement under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge.

[Van der Gaag, 1994] L.C. van der Gaag (1994). *Evidence Absorption – Experiments on Different Classes of Randomly Generated Belief Networks*, Technical Report UU-CS-94-42, Utrecht University.

[Van der Gaag & Meyer, 1998] L.C. van der Gaag, J-J.Ch. Meyer (1998). Informational independence: models and normal forms. *International Journal of Intelligent Systems*, vol. 13, pp. 83 – 109.

Appendix

In the Sections 3 and 4 of this paper we have presented various properties of sensitivity analysis of Bayesian belief networks. In Section 3, we have introduced the concept of a sensitivity set for a network's node of interest given available observations. We have shown that the conditional probabilities of the nodes that are not included in a sensitivity set under consideration upon variation cannot influence the probability of interest. For the nodes that are included in the sensitivity set, we have shown in Section 4 that the probability of interest relates to the conditional probabilities of these nodes as a quotient of two linear functions. So far, we have presented these properties with short, intuitive proofs. In this appendix, we provide full proofs for the various different properties.

In order to prove that a belief network's probability of interest for a node V_r is algebraically independent of the conditional probabilities of any node that is not included in a sensitivity set $Sen(V_r, O)$ under consideration, we have partitioned, in Definition 3.3, the set of remaining nodes into the three sets $Insen_1(V_r, O)$, $Insen_2(V_r, O)$, and $Insen_3(V_r, O)$. In the following lemma, we show that these three sets and the sensitivity set are mutually exclusive and collectively exhaustive.

Lemma A.1 *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$. Let $V_r \in V(G)$ be the network's node of interest and let $O \subseteq V(G)$ be the set of observed nodes in G. Let $Sen(V_r, O)$ be the sensitivity set for V_r given O and let $Insen_1(V_r, O)$, $Insen_2(V_r, O)$, $Insen_3(V_r, O)$, and $Sen(V_r, O)$ be defined as in Definition 3.3. Then,*

- $Insen_i(V_r, O) \cap Insen_j(V_r, O) = \varnothing$, *for all $i, j = 1, 2, 3$ with $i \neq j$;*
- $V(G) \setminus Sen(V_r, O) = \bigcup_{i=1,2,3} Insen_i(V_r, O)$.

Proof. From Definition 3.3, it is readily seen that the sets $Insen_1(V_r, O)$, $Insen_2(V_r, O)$, and $Insen_3(V_r, O)$ are mutually exclusive. In our proof, we therefore focus on the second property stated in the lemma.

To prove that $V(G) \setminus Sen(V_r, O) = \bigcup_{i=1,2,3} Insen_i(V_r, O)$, we have to show that any node that is included in one of the sets $Insen_i(V_r, O)$, is not included in $Sen(V_r, O)$, and vice versa. To show that a node V_j is not included in the set $Sen(V_r, O)$, we construct from the belief network's digraph G the auxiliary digraph G^* as defined in Definition 3.1 and show that in G^* any chain from V_j's auxiliary predecessor X_j to the node of interest V_r is blocked by O. We now begin by showing that $\bigcup_{i=1,2,3} Insen_i(V_r, O) \subseteq V(G) \setminus Sen(V_r, O)$:

- We assume that $Insen_1(V_r, O) \cup Insen_2(V_r, O) \neq \varnothing$ and consider a node $V_j \in Insen_1(V_r, O) \cup Insen_2(V_r, O)$. We observe that, in the digraph G, any chain from this node V_j to node V_r includes either a predecessor or a successor of V_j; in the auxiliary digraph G^*, therefore, any chain from V_j's auxiliary predecessor X_j to node V_r equally includes either an(other) predecessor or a successor of V_j. Now, for node V_j, we have by definition that $\langle(\{V_j\} \cup \pi_G(V_j)) \mid O \mid \{V_r\}\rangle^d_G$. From $\langle(\{V_j\} \cup \pi_G(V_j)) \mid O \mid \{V_r\}\rangle^d_G$, we have that, in the digraph G, any chain $V_j \rightarrow \cdots V_r$ from V_j to V_r that includes a successor of V_j, is blocked by O. In the auxiliary digraph G^*, therefore, any chain $X_j \rightarrow V_j \rightarrow \cdots V_r$ from X_j to V_r that includes a successor of V_j, is blocked by O. From $\langle(\{V_j\} \cup \pi_G(V_j)) \mid O \mid \{V_r\}\rangle^d_G$, we further have that, in G, any chain $V_k \cdots V_r$ from a node $V_k \in \pi_G(V_j)$ to node V_r is blocked by O. In G^*, therefore, any chain $X_j \rightarrow V_j \leftarrow V_k \cdots V_r$ from X_j to V_r that includes a node $V_k \in \pi_G(V_j)$, is blocked by O. We conclude that $\langle\{X_j\} \mid O \mid \{V_r\}\rangle^d_{G^*}$. By definition, we have that $V_j \in V(G) \setminus Sen(V_r, O)$.
- We assume that $Insen_3(V_r, O) \neq \varnothing$ and consider a node $V_j \in Insen_3(V_r, O)$. For this node V_j, we have by definition that $V_j \in V(G) \setminus \pi^*_G(V_r)$ and $\sigma^*_G(V_j) \cap O = \varnothing$. From these properties, we have that, in the digraph G, any chain from node V_j to node V_r either includes a predecessor V_k of V_j or includes a descendant $V_m \in \sigma^*_G(V_j)$ with two incoming arcs for which $\sigma^*_G(V_m) \cap O = \varnothing$. In the auxiliary digraph G^*, any chain $X_j \rightarrow V_j \leftarrow V_k \cdots V_r$ from V_j's auxiliary predecessor X_j to node V_r that includes an(other) predecessor V_k of V_j, is

blocked by O because $\sigma^*_G(V_j) \cap O = \varnothing$. Furthermore, in G^*, any chain $X_j \rightarrow V_j \rightarrow \cdots \rightarrow V_m \leftarrow \cdots V_r$ from X_j to V_r is blocked by O because $\sigma^*_G(V_m) \cap O = \varnothing$. We conclude that $\langle\{X_j\} \mid O \mid \{V_r\}\rangle^d_{G^*}$. By definition, we have that $V_j \in V(G) \setminus Sen(V_r, O)$.

From the previous observations, we conclude that $\bigcup_{i=1,2,3} Insen_i(V_r, O) \subseteq V(G) \setminus Sen(V_r, O)$; note that the property trivially holds for the case where $Insen_i(V_r, O) = \varnothing$, $i = 1, 2, 3$.

We proceed by showing that $V(G) \setminus Sen(V_r, O) \subseteq \bigcup_{i=1,2,3} Insen_i(V_r, O)$. We assume that $V(G) \setminus Sen(V_r, O) \neq \varnothing$; the property trivially holds otherwise. We now consider a node $V_j \in V(G) \setminus Sen(V_r, O)$. For this node V_j, we have by definition that $\langle\{X_j\} \mid O \mid \{V_r\}\rangle^d_{G^*}$. We distinguish between two cases, the case where $V_j \in \pi^*_G(V_r)$ and the case where $V_j \notin \pi^*_G(V_r)$:

- We assume that $V_j \in \pi^*_G(V_r)$. From $\langle\{X_j\} \mid O \mid \{V_r\}\rangle^d_{G^*}$, we have that, in the auxiliary digraph G^*, any chain $X_j \rightarrow V_j \rightarrow \cdots V_r$ from X_j to V_r that includes a successor of V_j, is blocked by O. We conclude from this observation that, in the digraph G, any chain $V_j \rightarrow \cdots V_r$ from V_j to V_r is blocked by O. Note that from $V_j \in \pi^*_G(V_r)$, we have that there exists at least one (directed) path $V_j \rightarrow \cdots \rightarrow V_r$ from V_j to V_r in G. From this path being blocked, we conclude that $\sigma^*_G(V_j) \cap O \neq \varnothing$. Now, from $\langle\{X_j\} \mid O \mid \{V_r\}\rangle^d_{G^*}$, we further observe that, in the digraph G^* any chain $X_j \rightarrow V_j \leftarrow V_k \cdots V_r$ from X_j to V_r, that includes an(other) predecessor V_k of V_j, is blocked by O. From $\sigma^*_G(V_j) \cap O \neq \varnothing$ and the previous observations, we have that, in G, any chain $V_k \cdots V_r$ from a predecessor V_k of V_j to V_r is blocked by O. We conclude that $\langle(\{V_j\} \cup \pi^*_G(V_j)) \mid O \mid \{V_r\}\rangle^*_G$ and, hence, that $V_j \in Insen_1(V_r, O)$.

- We assume that $V_j \notin \pi^*_G(V_r)$. We once more distinguish between two cases, the case where $\sigma^*_G(V_j) \cap O = \varnothing$ and the case where $\sigma^*_G(V_j) \cap O \neq \varnothing$:
  - We assume that $\sigma^*_G(V_j) \cap O = \varnothing$. By definition, we have that $V_j \in Insen_3(V_r, O)$.
  - We assume that $\sigma^*_G(V_j) \cap O \neq \varnothing$. From $\langle\{X_j\} \mid O \mid \{V_r\}\rangle^d_{G^*}$, we have that, in the auxiliary digraph G^*, any chain $X_j \rightarrow V_j \rightarrow \cdots V_r$ from X_j to V_r that includes a successor of V_j, is blocked by O. We conclude from this observation that, in the digraph G, any chain $V_j \rightarrow \cdots V_r$ from V_j to V_r is blocked by O. From $\langle\{X_j\} \mid O \mid \{V_r\}\rangle^d_{G^*}$, we further observe that, in the auxiliary digraph G^*, any chain $X_j \rightarrow V_j \leftarrow V_k \cdots V_r$ from X_j to V_r that includes an(other) predecessor V_k of V_j, is blocked by O. From $\sigma^*_G(V_j) \cap O \neq \varnothing$ and the previous observations, we have that, in the digraph G, any chain $V_k \cdots V_r$ from a predecessor V_k of V_j to V_r is blocked by O. We conclude that $\langle(\{V_j\} \cup \pi^*_G(V_j)) \mid O \mid \{V_r\}\rangle^d_G$ and, hence, that $V_j \in Insen_2(V_r, O)$.

  From the previous considerations we conclude that $V_j \in \bigcup_{i=1,2,3} Insen_i(V_r, O)$ and, hence, that $V(G) \setminus Sen(V_r, O) \subseteq \bigcup_{i=1,2,3} Insen_i(V_r, O)$.

From $V(G) \setminus Sen(V_r, O) \subseteq \bigcup_{i=1,2,3} Insen_i(V_r, O)$ and $\bigcup_{i=1,2,3} Insen_i(V_r, O) \subseteq V(G) \setminus Sen(V_r, O)$ we conclude that $V(G) \setminus Sen(V_r, O) = \bigcup_{i=1,2,3} Insen_i(V_r, O)$, as stated in the lemma. □

In Section 3, we have provided the three lemmas 3.5, 3.7, and 3.9, stating that the probability of interest of a Bayesian belief network is algebraically independent of the conditional probabilities of the nodes included in the sets $Insen_1(V_r, O)$, $Insen_2(V_r, O)$, and $Insen_3(V_r, O)$. We will provide formal proofs for these lemmas shortly. Before doing so, however, we introduce the concept of a *sensitivity ordering* of the nodes of a belief network's digraph that will be used throughout the proofs.

Definition A.2 *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$ where $V(G) = \{V_1, \ldots, V_n\}$, $n \geq 1$. Let $V_r \in V(G)$ be the network's node of interest and let $O \subseteq V(G)$ be the set of observed nodes in G. Let $Sen(V_r, O)$ be the sensitivity set for V_r given O and let $Insen_1(V_r, O)$, $Insen_2(V_r, O)$, and $Insen_3(V_r, O)$ be defined as in Definition 3.3. Let $\iota : V(G) \longleftrightarrow \{1, \ldots, n\}$ be a total ordering on $V(G)$, such that*

- *for any two nodes $V_i, V_j \in V(G)$ with $V_i \to V_j \in A(G)$, we have $\iota(V_i) < \iota(V_j)$;*
- *for any two nodes $V_i \in Insen_1(V_r, O) \cup Sen(V_r, O)$, $V_j \in Insen_2(V_r, O)$, we have $\iota(V_i) < \iota(V_j)$;*
- *for any two nodes $V_i \in Insen_2(V_r, O)$, $V_j \in Insen_3(V_r, O)$, we have $\iota(V_i) < \iota(V_j)$.*

Then, ι is a sensitivity ordering *of G with respect to V_r and O.*

For any node of interest and any set of observed nodes, there exists a sensitivity ordering of a belief network's digraph. Any such sensitivity ordering is a topological ordering of the digraph at hand.

Lemma A.3 (cf. Lemma 3.5) *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$. Let $\Pr$ be the joint probability distribution defined by B. Let $O \subseteq V(G)$ be the set of observed nodes in G and let o denote the corresponding observations. Let $V_r \in V(G)$ be the network's node of interest. Then, for any value v_r of V_r, we have that $\Pr(v_r \mid o) \not\sim p(V_i \mid \pi_G(V_i))$ for every node $V_i \in Insen_3(V_r, O)$.*

Proof. Let ι be a sensitivity ordering of G with respect to V_r and O. Without loss of generality, we assume that the nodes in G are indexed by their ordering number, that is, we assume that $\iota(V_i) = i$; we take $n \geq 1$ to be the number of nodes in G. From the rule of marginalisation, we have that the probability of interest $\Pr(v_r \mid o)$ can be written as

$$\Pr(v_r \mid o) = \frac{\displaystyle\sum_{\{V_1,\ldots,V_n\}\setminus(\{V_r\}\cup O)} \Pr((\{V_1,\ldots,V_n\} \setminus (\{V_r\} \cup O)) \wedge v_r \wedge o)}{\displaystyle\sum_{\{V_1,\ldots,V_n\}\setminus O} \Pr((\{V_1,\ldots,V_n\} \setminus O) \wedge o)}$$

In the above equation, we have used the notation $\sum_W$ to indicate summation over all possible values of the variables in the set W. In the following, we will also use the notation $|_{X=x}$; this notation is used to indicate that in the preceding formula the variables in the set X take the combination of values x. Now, using the property stated in Proposition 2.6 for the probability distribution $\Pr$ defined by the network, we find that

$$\Pr(v_r \mid o) = \frac{\displaystyle\sum_{\{V_1,\ldots,V_n\}\setminus(\{V_r\}\cup O)} \prod_{i=1,\ldots,n} p(V_i \mid \pi_G(V_i)) \Bigg|_{\substack{V_r = v_r \\ O = o}}}{\displaystyle\sum_{\{V_1,\ldots,V_n\}\setminus O} \prod_{j=1,\ldots,n} p(V_j \mid \pi_G(V_j)) \Bigg|_{O = o}}$$

From the definition of sensitivity ordering, we have that the nodes in the set $Insen_3(V_r, O)$ have the highest ordering numbers in the network's digraph; without loss of generality, we assume that $Insen_3(V_r, O)$ includes the nodes $V_{m+1}, \ldots, V_n$. Now,

$$\Pr(v_r \mid o) = \frac{\displaystyle\sum_{\{V_1,\ldots,V_n\}\setminus(\{V_r\}\cup O)} \left(\left(\prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i))\right) \cdot \left(\prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i))\right)\right) \Bigg|_{\substack{V_r = v_r \\ O = o}}}{\displaystyle\sum_{\{V_1,\ldots,V_n\}\setminus O} \left(\left(\prod_{j=m+1,\ldots,n} p(V_j \mid \pi_G(V_j))\right) \cdot \left(\prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j))\right)\right) \Bigg|_{O = o}}$$

From Definition 3.3, we know that the set $Insen_3(V_r, O)$ does not include any nodes from the set $\{V_r\} \cup O$, that is, $(\{V_r\} \cup O) \cap \{V_{m+1}, \ldots, V_n\} = \varnothing$. Since our sensitivity ordering is a topological ordering, we further know that $(\bigcup_{i=1,\ldots,m} \pi_G(V_i)) \cap \{V_{m+1}, \ldots, V_n\} = \varnothing$. Using these observations, we find that

$$\Pr(v_r \mid o) =$$

$$= \frac{\displaystyle\sum_{\{V_1,\ldots,V_m\}\setminus(\{V_r\}\cup O)} \left(\left(\sum_{\{V_{m+1},\ldots,V_n\}} \prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i)) \right) \cdot \prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r = v_r \\ O = o}}}{\displaystyle\sum_{\{V_1,\ldots,V_m\}\setminus O} \left(\left(\sum_{\{V_{m+1},\ldots,V_n\}} \prod_{j=m+1,\ldots,n} p(V_j \mid \pi_G(V_j)) \right) \cdot \prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j)) \right) \Bigg|_{O = o}}$$

The rule of marginalisation now implies that the sum terms in parentheses in the equation above equal one: for node V_n, marginalisation gives

$$\sum_{\{V_{m+1},\ldots,V_n\}} \prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i)) =$$

$$= \sum_{\{V_{m+1},\ldots,V_{n-1}\}} \left(\left(\sum_{\{V_n\}} p(V_n \mid \pi_G(V_n)) \right) \cdot \prod_{i=m+1,\ldots,n-1} p(V_i \mid \pi_G(V_i)) \right) =$$

$$= \sum_{\{V_{m+1},\ldots,V_{n-1}\}} \prod_{i=m+1,\ldots,n-1} p(V_i \mid \pi_G(V_i))$$

Recursively repeating this argument for the nodes $V_{n-1}, \ldots, V_{m+1}$ results in

$$\sum_{\{V_{m+1},\ldots,V_n\}} \prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i)) = 1$$

We conclude that

$$\Pr(v_r \mid o) = \frac{\displaystyle\sum_{\{V_1,\ldots,V_m\}\setminus(\{V_r\}\cup O)} \prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i)) \Bigg|_{\substack{V_r = v_r \\ O = o}}}{\displaystyle\sum_{\{V_1,\ldots,V_m\}\setminus O} \prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j)) \Bigg|_{O = o}}$$

which shows that the probability of interest $\Pr(v_r \mid o)$ is algebraically independent of the conditional probabilities of any node from the set $\mathit{Insen}_3(V_r, O)$, as stated in the lemma. □

So far, we have shown that a belief network's probability of interest for a node V_r given observations for nodes O is algebraically independent of the conditional probabilities of any node from the set $\mathit{Insen}_3(V_r, O)$. We now proceed by showing that this probability of interest is also algebraically independent of the conditional probabilities of the nodes from the set $\mathit{Insen}_2(V_r, O)$.

Lemma A.4 (cf. Lemma 3.7) *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$. Let* $\Pr$ *be the joint probability distribution defined by B. Let $O \subseteq V(G)$ be the set of observed nodes in G and let o denote the corresponding observations. Let $V_r \in V(G)$ be the network's node of interest. Then, for any value v_r of V_r, we have that* $\Pr(v_r \mid o) \nsim p(V_i \mid \pi_G(V_i))$ *for every node $V_i \in \mathit{Insen}_2(V_r, O)$.*

Proof. Let ι be a sensitivity ordering of G with respect to V_r and O. Without loss of generality, we assume that $\mathit{Insen}_3(V_r, O) = \varnothing$. Also without loss of generality, we assume that the nodes in G are indexed by their ordering number, that is, we assume that $\iota(V_i) = i$. We take $n \geq 1$ to be the number of nodes in G. From the definition of sensitivity ordering, we have that the nodes in $Insen_2(V_r, O)$ have the highest ordering numbers in the digraph; we assume that $Insen_2(V_r, O)$ consists of the nodes $V_{m+1}, \ldots, V_n$. For our probability of interest $\Pr(v_r \mid o)$, we find that

$$\Pr(v_r \mid o) = \frac{\sum_{\{V_1,\ldots,V_n\}\setminus(\{V_r\}\cup O)} \left(\left(\prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i))\right) \cdot \left(\prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i))\right)\right)\Bigg|_{\substack{V_r = v_r\\ O = o}}}{\sum_{\{V_1,\ldots,V_n\}\setminus O} \left(\left(\prod_{j=m+1,\ldots,n} p(V_j \mid \pi_G(V_j))\right) \cdot \left(\prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j))\right)\right)\Bigg|_{O = o}}$$

Since our sensitivity ordering is a topological ordering of G, we know that $(\bigcup_{i=1,\ldots,m} \pi_G(V_i)) \cap \{V_{m+1}, \ldots, V_n\} = \varnothing$. Since $V_r \notin Insen_2(V_r, O)$ by definition, we also have that $V_r \notin \{V_{m+1}, \ldots, V_n\}$. Using these observations, we find that

$$\Pr(v_r \mid o) =$$

$$= \frac{\sum_{\{V_1,\ldots,V_m\}\setminus(\{V_r\}\cup O)} \left(\left(\sum_{\{V_{m+1},\ldots,V_n\}\setminus O} \prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i))\right) \cdot \prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i))\right)\Bigg|_{\substack{V_r = v_r\\ O = o}}}{\sum_{\{V_1,\ldots,V_m\}\setminus O} \left(\left(\sum_{\{V_{m+1},\ldots,V_n\}\setminus O} \prod_{j=m+1,\ldots,n} p(V_j \mid \pi_G(V_j))\right) \cdot \prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j))\right)\Bigg|_{O = o}}$$

Now, from Definition 3.3, we have that the nodes $V_{m+1}, \ldots, V_n$ from the set $Insen_2(V_r, O)$ and their predecessors are d-separated from the node of interest V_r. Any predecessor of a node V_i, $i = m+1, \ldots, n$, therefore, is either included in $Insen_2(V_r, O)$ itself or is an observed node. We conclude that $(\bigcup_{i=m+1,\ldots,n} \pi_G(V_i)) \cap (\{V_1, \ldots, V_m\} \setminus (\{V_r\} \cup O)) = \varnothing$. The probability of interest can now be written as

$$\Pr(v_r \mid o) =$$

$$= \frac{\left(\sum_{\{V_{m+1},\ldots,V_n\}\setminus O} \prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i))\right) \cdot \left(\sum_{\{V_1,\ldots,V_m\}\setminus(\{V_r\}\cup O)} \prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i))\right)\Bigg|_{\substack{V_r = v_r\\ O = o}}}{\left(\sum_{\{V_{m+1},\ldots,V_n\}\setminus O} \prod_{j=m+1,\ldots,n} p(V_j \mid \pi_G(V_j))\right) \cdot \left(\sum_{\{V_1,\ldots,V_m\}\setminus O} \prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j))\right)\Bigg|_{O = o}} =$$

$$= \frac{\sum_{\{V_1,\ldots,V_m\}\setminus(\{V_r\}\cup O)} \prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i))\Bigg|_{\substack{V_r = v_r\\ O = o}}}{\sum_{\{V_1,\ldots,V_m\}\setminus O} \prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j))\Bigg|_{O = o}}$$

which shows that the probability of interest $\Pr(v_r \mid o)$ is algebraically independent of the conditional probabilities of any node from the set $Insen_2(V_r, O)$, as stated in the lemma. □

So far, we have shown that a belief network's probability of interest for a node V_r given observations for nodes O is algebraically independent of the conditional probabilities of any node from the sets $Insen_3(V_r, O)$ and $Insen_2(V_r, O)$. To conclude, we now prove that this probability of interest is also algebraically independent of the conditional probabilities of the nodes from the set $Insen_1(V_r, O)$.

Lemma A.5 (cf. Lemma 3.9) *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$. Let* Pr *be the joint probability distribution defined by B. Let $O \subseteq V(G)$ be the set of observed nodes* *in G and let o denote the corresponding observations. Let* $V_r \in V(G)$ *be the network's node of interest. Then, for any value* v_r *of* V_r*, we have that* $\Pr(v_r \mid o) \nsim p(V_i \mid \pi_G(V_i))$ *for every node* $V_i \in \mathit{Insen}_1(V_r, O)$.

Proof. Without loss of generality, we assume that $\mathit{Insen}_2(V_r, O) = \varnothing$ and $\mathit{Insen}_3(V_r, O) = \varnothing$. From these assumptions and Lemma A.1, we have that $V(G) = \mathit{Sen}(V_r, O) \cup \mathit{Insen}_1(V_r, O)$. Let the nodes from $\mathit{Sen}(V_r, O)$ be called $V_1, \ldots, V_m$ and let the nodes from $\mathit{Insen}_1(V_r, O)$ be called $V_{m+1}, \ldots, V_n$, $n \geq 1$; note that, in contrast with the proofs of the previous lemmas, the nodes are *not* indexed by their ordering number according to some sensitivity ordering of G. For our probability of interest $\Pr(v_r \mid o)$, we now find that

$$\Pr(v_r \mid o) = \frac{\left.\sum_{\{V_1,\ldots,V_n\}\setminus(\{V_r\}\cup O)} \left(\left(\prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i))\right) \cdot \left(\prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i))\right)\right)\right|_{\substack{V_r = v_r \\ O = o}}}{\left.\sum_{\{V_1,\ldots,V_n\}\setminus O} \left(\left(\prod_{j=m+1,\ldots,n} p(V_j \mid \pi_G(V_j))\right) \cdot \left(\prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j))\right)\right)\right|_{O = o}}$$

From Definition 3.3, we have that the nodes $V_{m+1}, \ldots, V_n$ from the set $\mathit{Insen}_1(V_r, O)$ and their predecessors are d-separated from the node of interest V_r. Any predecessor of a node V_i, $i = m+1, \ldots, n$, therefore, is either included in $\mathit{Insen}_1(V_r, O)$ itself or is an observed node. We conclude that $(\bigcup_{i=m+1,\ldots,n} \pi_G(V_i)) \cap (\{V_1, \ldots, V_m\} \setminus (\{V_r\} \cup O)) = \varnothing$. In addition, for every node V_i, $i = 1, \ldots, m$, from $\mathit{Sen}(V_r, O)$, we have that any predecessor that is included in the set $\mathit{Insen}_1(V_r, O)$ is an observed node. Hence, $\bigcup_{i=1,\ldots,m} \pi_G(V_i)) \cap (\{V_{m+1}, \ldots, V_n\} \setminus O) = \varnothing$. Building upon these observations, the probability of interest can be written as

$$\Pr(v_r \mid o) =$$

$$= \frac{\left.\left(\sum_{\{V_{m+1},\ldots,V_n\}\setminus O} \prod_{i=m+1,\ldots,n} p(V_i \mid \pi_G(V_i))\right) \cdot \left(\sum_{\{V_1,\ldots,V_m\}\setminus(\{V_r\}\cup O)} \prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i))\right)\right|_{\substack{V_r = v_r \\ O = o}}}{\left.\left(\sum_{\{V_{m+1},\ldots,V_n\}\setminus O} \prod_{j=m+1,\ldots,n} p(V_j \mid \pi_G(V_j))\right) \cdot \left(\sum_{\{V_1,\ldots,V_m\}\setminus O} \prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j))\right)\right|_{O = o}} =$$

$$= \frac{\left.\sum_{\{V_1,\ldots,V_m\}\setminus(\{V_r\}\cup O)} \prod_{i=1,\ldots,m} p(V_i \mid \pi_G(V_i))\right|_{\substack{V_r = v_r \\ O = o}}}{\left.\sum_{\{V_1,\ldots,V_m\}\setminus O} \prod_{j=1,\ldots,m} p(V_j \mid \pi_G(V_j))\right|_{O = o}}$$

which shows that the probability of interest $\Pr(v_r \mid o)$ is algebraically independent of the conditional probabilities of any node from the set $\mathit{Insen}_1(V_r, O)$, as stated in the lemma. □

In the foregoing, we have shown that a belief network's probability of interest is algebraically independent of the conditional probabilities of any node that is not included in the sensitivity set under consideration. We now show that the probability of interest relates to any conditional probability for a node from the sensitivity set as a quotient of two linear functions.

Proposition A.6 (cf. Proposition 4.1) *Let B be a Bayesian belief network with the digraph* $G = (V(G), A(G))$ *and let* $\Pr$ *be the joint probability distribution defined by B. Let* $O \subseteq V(G)$ *be the set of observed nodes in G and let o denote the corresponding observations. Let* V_r *be the network's node of interest and let* $\mathit{Sen}(V_r, O)$ *be the sensitivity set for* V_r *given O. Then, for any* *value v_r of V_r, we have that*

$$\Pr(v_r \mid o) = \frac{a \cdot x + b}{c \cdot x + d}$$

for every conditional probability $x = p(v_s \mid \pi')$ of every node $V_s \in Sen(V_r, O)$, where a, b, c, and d are constants that are dependent upon the values v_s of V_s and π' of $\pi_G(V_s)$.

Proof. The probability of interest $\Pr(v_r \mid o)$ for the belief network B equals

$$\Pr(v_r \mid o) = \frac{\Pr(v_r \wedge o)}{\Pr(o)}$$

Without loss of generality, we take the nodes of the belief network B to be $V_1, \ldots, V_n$, $n \geq 1$. For ease of exposition, we assume all variables in the network to be binary, taking one of the truth values *true* and *false*. We will use v_i to denote the proposition that the variable V_i takes the value *true*; $V_i = \mathit{false}$ will be denoted as $\neg v_i$. We will return to our assumption of binary variables at the end of the proof. We now consider a node V_s from the sensitivity set $Sen(V_r, O)$ under study. Without loss of generality, we investigate the sensitivity of the probability of interest with regard to the conditional probability $p(v_s \mid \pi')$ for this node, where π' is a specific combination of values for the nodes from the set $\pi_G(V_s)$. For the numerator $\Pr(v_r \wedge o)$ of the probability of interest, we find that

$$\Pr(v_r \wedge o) =$$

$$= \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_r\} \cup O)}} \prod_{i=1,\ldots,n} p(V_i \mid \pi_G(V_i)) \Bigg|_{\substack{V_r = v_r \\ O = o}} =$$

$$= \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_r\} \cup O)}} \left(p(V_s \mid \pi_G(V_s)) \cdot \prod_{\substack{i = 1,\ldots,n, \\ i \neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r = v_r \\ O = o}} =$$

$$= \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_r, V_s\} \cup \pi_G(V_s) \cup O)}} \left(p(v_s \mid \pi') \cdot \prod_{\substack{i = 1,\ldots,n, \\ i \neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r = v_r \\ O = o \\ V_s = v_s \\ \pi_G(V_s) = \pi'}} +$$

$$+ \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_r, V_s\} \cup \pi_G(V_s) \cup O)}} \left((1 - p(v_s \mid \pi')) \cdot \prod_{\substack{i = 1,\ldots,n, \\ i \neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r = v_r \\ O = o \\ V_s = \neg v_s \\ \pi_G(V_s) = \pi'}} +$$

$$+ \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_r\} \cup O), \\ \pi_G(V_s) \neq \pi'}} \left(p(V_s \mid \pi_G(V_s)) \cdot \prod_{\substack{i = 1,\ldots,n, \\ i \neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r = v_r \\ O = o}}$$

The first term in the above sum of three assembles all products that specify the conditional probability $p(v_s \mid \pi')$. The second term gathers all products specifying the complement, $p(\neg v_s \mid \pi')$, of the conditional probability under study. Note that this term, as the first one, depends on the value of $p(v_s \mid \pi')$. The third term, to conclude, collects the remaining products; these products specify for the node V_s a conditional probability that has another combination of values than π' for its conditioning part. Note that the third term does not depend on the value of the conditional probability under study. Writing x for $p(v_s \mid \pi')$, we find that

$$\Pr(v_r \wedge o) = a \cdot x + b$$

where

$$a = \sum_{\substack{\{V_1, \ldots, V_n\} \setminus \\ (\{V_r, V_s\} \cup \pi_G(V_s) \cup O)}} \left(\prod_{\substack{i = 1, \ldots, n, \\ i \neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r = v_r \\ O = o \\ V_s = v_s \\ \pi_G(V_s) = \pi'}} +$$

$$- \sum_{\substack{\{V_1, \ldots, V_n\} \setminus \\ (\{V_r, V_s\} \cup \pi_G(V_s) \cup O)}} \left(\prod_{\substack{i = 1, \ldots, n, \\ i \neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r = v_r \\ O = o \\ V_s = \neg v_s \\ \pi_G(V_s) = \pi'}}$$

and

$$b = \sum_{\substack{\{V_1, \ldots, V_n\} \setminus \\ (\{V_r, V_s\} \cup \pi_G(V_s) \cup O)}} \left(\prod_{\substack{i = 1, \ldots, n, \\ i \neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r = v_r \\ O = o \\ V_s = \neg v_s \\ \pi_G(V_s) = \pi'}} +$$

$$+ \sum_{\substack{\{V_1, \ldots, V_n\} \setminus \\ (\{V_r\} \cup O), \\ \pi_G(V_s) \neq \pi'}} \left(\prod_{i=1,\ldots,n} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{V_r = v_r \\ O = o}}$$

Note that the constants a and b are related to the conditional probability under study but are not dependent upon its value.

For the denominator $\Pr(o)$ of the probability of interest, we analogously find that

$$\Pr(o) =$$

$$= \sum_{\substack{\{V_1, \ldots, V_n\} \setminus \\ (\{V_s\} \cup \pi_G(V_s) \cup O)}} \left(p(v_s \mid \pi') \cdot \prod_{\substack{i = 1, \ldots, n, \\ i \neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{O = o \\ V_s = v_s \\ \pi_G(V_s) = \pi'}} +$$

$$+ \sum_{\substack{\{V_1, \ldots, V_n\} \setminus \\ (\{V_s\} \cup \pi_G(V_s) \cup O)}} \left((1 - p(v_s \mid \pi')) \cdot \prod_{\substack{i = 1, \ldots, n, \\ i \neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{O = o \\ V_s = \neg v_s \\ \pi_G(V_s) = \pi'}} +$$

$$+ \sum_{\substack{\{V_1, \ldots, V_n\} \setminus O, \\ \pi_G(V_s) \neq \pi'}} \left(p(V_s \mid \pi_G(V_s)) \cdot \prod_{\substack{i = 1, \ldots, n, \\ i \neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{O = o} =$$

$$= c \cdot x + d$$

once more writing x for the conditional probability under study. For the constants c and d, we have that

$$c = \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_s\}\cup\pi_G(V_s)\cup O)}} \left(\prod_{\substack{i=1,\ldots,n, \\ i\neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{O=o \\ V_s=v_s \\ \pi_G(V_s)=\pi'}} +$$

$$- \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_s\}\cup\pi_G(V_s)\cup O)}} \left(\prod_{\substack{i=1,\ldots,n, \\ i\neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{O=o \\ V_s=\neg v_s \\ \pi_G(V_s)=\pi'}}$$

and

$$d = \sum_{\substack{\{V_1,\ldots,V_n\}\setminus \\ (\{V_s\}\cup\pi_G(V_s)\cup O)}} \left(\prod_{\substack{i=1,\ldots,n, \\ i\neq s}} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{\substack{O=o \\ V_s=\neg v_s \\ \pi_G(V_s)=\pi'}} +$$

$$+ \sum_{\substack{\{V_1,\ldots,V_n\}\setminus O, \\ \pi_G(V_s)\neq\pi'}} \left(\prod_{i=1,\ldots,n} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{O=o}$$

From the previous observations, we conclude that the probability of interest $\Pr(v_r \mid o)$ equals

$$\Pr(v_r \mid o) = \frac{a \cdot x + b}{c \cdot x + d}$$

where x, a, b, c, and d are as above.

In our proof so far, we have assumed all variables in the belief network B to be binary. We would like to note that the proof can be generalised to non-binary variables, provided that for varying the value of a conditional probability $p(v_s \mid \pi')$ for a node V_s from the sensitivity set under study, the ratio of any pair of complementary probabilities $p(v_s' \mid \pi')$ and $p(v_s'' \mid \pi')$ for this node is kept fixed. □

So far, we have shown that a belief network's probability of interest relates as a quotient of two linear functions to a conditional probability under study. For a conditional probability that pertains to a node from the sensitivity set that does not have any observed descendants, this functional relation reduces to a *linear function*.

Proposition A.7 (cf. Proposition 4.3) *Let B be a Bayesian belief network with the digraph $G = (V(G), A(G))$ and let* Pr *be the joint probability distribution defined by B. Let $O \subseteq V(G)$ be the set of observed nodes in G and let o denote the corresponding observations. Let V_r be the network's node of interest and let $Sen(V_r, O)$ be the sensitivity set for V_r given O. Then, for any value v_r of V_r, we have that*

$$\Pr(v_r \mid o) = a \cdot x + b$$

for every conditional probability $x = p(v_s \mid \pi')$ for every node $V_s \in Sen(V_r, O)$ with $\sigma^*(V_s) \cap O = \varnothing$, where a and b are constants that are dependent upon the values v_s of V_s and π' of $\pi_G(V_s)$.

Proof. The probability of interest $\Pr(v_r \mid o)$ for the belief network B once more equals

$$\Pr(v_r \mid o) = \frac{\Pr(v_r \wedge o)}{\Pr(o)}$$

From the proof of Proposition A.6, we have that the numerator $\Pr(v_r \wedge o)$ in this equation relates linearly to the conditional probability x under study. More formally, we have that

$$\Pr(v_r \wedge o) = a' \cdot x + b'$$

where a' and b' are constants as specified in the proof of the proposition.

Let ι be a sensitivity ordering of G with respect to V_r and O. Without loss of generality, we assume that the nodes in G are indexed by their ordering number, that is, we assume that $\iota(V_i) = i$; we take $n \geq 1$ to be the number of nodes in G. For ease of exposition, we further assume all variables in the network to be binary, taking one of the truth values *true* and *false*. We will once more use v_i to denote the proposition that the variable V_i takes the value *true*; $V_i = \mathit{false}$ will be denoted as $\neg v_i$. Our proof can be generalised to non-binary variables as indicated in the proof of Proposition A.6. We now consider a node V_s from the sensitivity set $Sen(V_r, O)$ under study. Without loss of generality, we assume that the set $\sigma^*(V_s)$ consists of the nodes $V_s, \ldots, V_n$. We investigate the sensitivity of the probability of interest with regard to the conditional probability $p(v_s \mid \pi')$ for the node V_s, where π' is a specific combination of values for the nodes from $\pi_G(V_s)$. For the denominator $\Pr(o)$ of the probability of interest, we find that

$$\begin{aligned}\Pr(o) &= \sum_{\{V_1,\ldots,V_n\}\setminus O} \prod_{i=1,\ldots,n} p(V_i \mid \pi_G(V_i)) \Bigg|_{O=o} = \\ &= \sum_{\{V_1,\ldots,V_n\}\setminus O} \left(\left(\prod_{i=s,\ldots,n} p(V_i \mid \pi_G(V_i)) \right) \cdot \left(\prod_{i=1,\ldots,s-1} p(V_i \mid \pi_G(V_i)) \right) \right) \Bigg|_{O=o}\end{aligned}$$

Since our sensitivity ordering ι is a topological ordering, we know that $(\bigcup_{i=1,\ldots,s-1} \pi_G(V_i)) \cap \{V_s, \ldots, V_n\} = \varnothing$. In addition, we have that $\sigma^*(V_s) \cap O = \varnothing$ and, hence, that $\{V_s, \ldots, V_n\} \cap O = \varnothing$. Building upon these observations, we find that

$$\Pr(o) = \sum_{\{V_1,\ldots,V_{s-1}\}\setminus O} \left(\left(\sum_{\{V_s,\ldots,V_n\}} \prod_{i=s,\ldots,n} p(V_i \mid \pi_G(V_i)) \right) \cdot \prod_{i=1,\ldots,s-1} p(V_i \mid \pi_G(V_i)) \right) \Bigg|_{O=o}$$

The rule of marginalisation now implies that the sum term in parentheses in the equation above equals one. We conclude that

$$\begin{aligned}\Pr(o) &= \sum_{\{V_1,\ldots,V_{s-1}\}\setminus O} \prod_{i=1,\ldots,s-1} p(V_i \mid \pi_G(V_i)) \Bigg|_{O=o} = \\ &= c'\end{aligned}$$

From this derivation, we have that $\Pr(o)$ is a constant with respect to the conditional probability under study x. For our probability of interest, we now find that

$$\begin{aligned}\Pr(v_r \mid o) &= \frac{a' \cdot x + b'}{c'} \\ &= a \cdot x + b\end{aligned}$$

where $a = \frac{a'}{c'}$ and $b = \frac{b'}{c'}$. □