

# Sensitivity Analysis in Discrete Bayesian Networks

Enrique Castillo\*, José Manuel Gutiérrez\* and Ali S. Hadi\*\*

\* *Department of Applied Mathematics and Computational Sciences,  
University of Cantabria, SPAIN*

\*\* *Department of Social Statistics, Cornell University, USA*

## ABSTRACT

The paper presents an efficient computational method for performing sensitivity analysis in discrete Bayesian networks. The method exploits the structure of conditional probabilities of a target node given the evidence. First, the set of parameters which are relevant to the calculation of the conditional probabilities of the target node is identified. Next, this set is reduced by removing those combinations of the parameters which either contradict the available evidence or are incompatible. Finally, using the canonical components associated with the resulting subset of parameters, the desired conditional probabilities are obtained. In this way, an important saving in the calculations is achieved. The proposed method can also be used to compute exact upper and lower bounds for the conditional probabilities, hence a sensitivity analysis can be easily performed. Examples are used to illustrate the proposed methodology.

**Key Words:** Propagation of uncertainty, Symbolic probabilistic inference, Canonical components, Efficient computations.

## 1 Introduction

Evidence propagation in Bayesian networks has been an active area of research during the last two decades. Consequently, several exact and approximate propagation methods have been proposed in the literature; see, for example, Pearl (1986,1988), Lauritzen and Spiegelhalter (1988), and Castillo, Gutiérrez and Hadi (1996). These methods, however, require that the joint probabilities of the nodes be specified numerically.

One aim of the analysis of discrete Bayesian networks is often to compute the conditional probabilities of a target node in the network. A question that usually arises in this context is that of sensitivity analysis, that is, how sensitive are these conditional probabilities to small changes in the parameters and/or evidence values?

One way of performing sensitivity analysis is to change the parameters values and then, using an evidence propagation method, monitor the effects of these changes on the conditional probabilities. Clearly, this brute force method is computationally intensive.

Another way of performing sensitivity analysis is suggested by Laskey (1995) who measures the impact of a small changes in one parameter on a target probability of interest. This is done using the partial derivative of output probabilities with respect to parameter being varied.

Sensitivity analysis can also be performed using symbolic probabilistic inference (SPI). For example, Li and D'Ambrosio (1994) and Chang and Fung (1995) give a goal directed algorithms which perform only those calculations that are required to respond to queries.

Castillo, Gutiérrez and Hadi (1995) perform symbolic calculations by first replacing the values of the initial probabilities by symbolic parameters, then using computer packages with symbolic computational capabilities (such as, Mathematica and Maple) to propagate uncertainty. This leads to probabilities which are expressed as functions of the parameters instead of actual numbers. Thus, the answers to specific sensitivity analysis queries can then be obtained directly without the need to redo the computations. This method, however, is suitable for Bayesian networks of a small number of variables, but is inefficient for larger networks due to the need for using symbolic packages. Nevertheless, the symbolic representation of the initial probabilities was useful in determining the algebraic structure of probabilities as a function of the parameters and/or evidence values. This algebraic structure leads to the following conclusions:

1. The conditional probabilities are ratios of polynomial functions of parameters and evidences, and
2. Numerical methods can be used to calculate the coefficients of the polynomials using the so called canonical components.

In this paper we further examine the algebraic and dependency structures of probabilities. We found that not all the terms of the general polynomial functions actually contribute to the conditional probabilities. Important implications of this finding include:

1. Substantial computational savings can be achieved by identifying and using only the relevant parameters in the polynomials.
2. The symbolic expressions of conditional probabilities can also be used to obtain lower and upper bounds for the marginal probabilities. These bounds can provide valuable information for performing sensitivity analysis of a Bayesian network.
3. An important advantage of the proposed method is that it can be performed using the currently available numeric propagation methods, thus making both symbolic computations and sensitivity analysis feasible even for large networks.

Section 2 gives the necessary notation. Section 3 reviews some recent results about the algebraic structure of conditional probabilities. Section 4 gives algorithms for efficient computations of the desired conditional probabilities. In Section 5 we illustrate the method described in Section 4 by an example. Section 6 shows how to obtain lower and upper bounds for the conditional probabilities. Finally, Section 7 gives some conclusions.

## 2 Notation

Let  $X = \{X_1, X_2, \dots, X_n\}$  be a set of  $n$  discrete variables, each can take values in the set  $\{0, 1, \dots, r_i - 1\}$ , where  $r_i$  is the cardinality (number of states) of variable  $X_i$ . A Bayesian network over  $X$  is a pair  $(D, P)$ , where the graph  $D$  is a directed acyclic graph (DAG) with one node for each variable in  $X$  and  $P = \{P_1(X_1|\Pi_1), \dots, P_n(X_n|\Pi_n)\}$  is a set of  $n$  conditional probabilities, one for each variable. Note that  $P_i(X_i|\Pi_i)$  gives the probabilities

of  $X_i$ , given the values of the variables in its parent set  $\Pi_i$ . Using the chain rule, the joint probability density of  $X$  can be written as the product of the above conditional probabilities, that is,

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P_i(X_i|\Pi_i). \quad (1)$$

Some of the conditional probability distributions in (1) can be specified numerically and others symbolically, that is,  $P_i(X_i|\Pi_i)$  can be a parametric family. When  $P_i(X_i|\Pi_i)$  is a parametric family, we refer to the node  $X_i$  as a chance node. A convenient choice of the parameters in this case is given by

$$\theta_{ij\pi} = P_i(X_i = j|\Pi_i = \pi), \quad j \in \{0, \dots, r_i - 1\}, \quad (2)$$

where  $\pi$  is any possible instantiation of the parents of  $X_i$ . Thus, the first subscript in  $\theta_{ij\pi}$  refers to the node number, the second subscript refers to the state of the node, and the remaining subscripts refer to the parents' instantiations. Since  $\sum_{j=0}^{r_i-1} \theta_{ij\pi} = 1$ , for all  $i$  and  $\pi$ , any one of the parameters can be written as one minus the sum of all others. For example,  $\theta_{i0\pi}$  is

$$\theta_{i0\pi} = 1 - \sum_{j=1}^{r_i-1} \theta_{ij\pi}. \quad (3)$$

To simplify the notation in cases where a variable  $X_i$  does not have parents, we use  $\theta_{ij}$  to denote  $P_i(X_i = j)$ ,  $j \in \{0, \dots, r_i - 1\}$ . We illustrate this notation using the following example.

**Example 1** Consider a discrete Bayesian network consisting of three variables  $X = \{X_1, X_2, X_3\}$  whose corresponding DAG  $D$  is given in Figure 1. The structure of  $D$  implies that the joint probability of the set of nodes can be written, in the form of (1), as:

$$P(X_1, X_2, X_3) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2). \quad (4)$$

For simplicity, but without loss of generality, assume that all nodes represent binary variables with values in the set  $\{0, 1\}$ . This and the structure of the probability distribution in (4) imply that the joint probability distribution of the three variables depends on 14 parameters  $\Theta = \{\theta_{ij\pi}\}$ . These parameters are given in Table 1. Note, however, that only 7 of the parameters are free (because the probabilities in each conditional distribution must add up to unity). These 7 parameters are given in Table 1 under either the column labeled  $X_i = 0$  or the column labeled  $X_i = 1$ .

The symbolic method of Castillo et al. (1995) can be used to calculate the conditional probabilities of single nodes when the parameters are given in symbolic form as is the case here. For example, suppose that the target node is  $X_3$ . Using the symbolic method, the probabilities  $P(X_3 = 0|evidence)$  for different evidences are computed and displayed in Table 2. In this paper we show how these symbolic expressions for the conditional probabilities can be computed efficiently by exploiting the algebraic and the dependency structures of the parameters.

Node		Parameters	
$X_i$	Parents	$X_i = 0$	$X_i = 1$
$X_1$	None	$\theta_{10} = P(X_1 = 0)$	$\theta_{11} = P(X_1 = 1)$
$X_2$	$X_1$	$\theta_{200} = P(X_2 = 0 X_1 = 0)$ $\theta_{201} = P(X_2 = 0 X_1 = 1)$	$\theta_{210} = P(X_2 = 1 X_1 = 0)$ $\theta_{211} = P(X_2 = 1 X_1 = 1)$
$X_3$	$X_1, X_2$	$\theta_{3000} = P(X_3 = 0 X_1 = 0, X_2 = 0)$ $\theta_{3001} = P(X_3 = 0 X_1 = 0, X_2 = 1)$ $\theta_{3010} = P(X_3 = 0 X_1 = 1, X_2 = 0)$ $\theta_{3011} = P(X_3 = 0 X_1 = 1, X_2 = 1)$	$\theta_{3100} = P(X_3 = 1 X_1 = 0, X_2 = 0)$ $\theta_{3101} = P(X_3 = 1 X_1 = 0, X_2 = 1)$ $\theta_{3110} = P(X_3 = 1 X_1 = 1, X_2 = 0)$ $\theta_{3111} = P(X_3 = 1 X_1 = 1, X_2 = 1)$

Table 1: Conditional probability tables associated with the network in Figure 1.

Evidence	$P(X_3 = 0 evidence)$
None	$\theta_{10}\theta_{200}\theta_{3000} + \theta_{10}\theta_{3001} - \theta_{10}\theta_{200}\theta_{3001} + \theta_{201}\theta_{3010} -$ $-\theta_{10}\theta_{201}\theta_{3010} + \theta_{3011} - \theta_{10}\theta_{3011} - \theta_{201}\theta_{3011} + \theta_{10}\theta_{201}\theta_{3011}$
$X_1 = 0$	$\theta_{200}\theta_{3000} + \theta_{3001} - \theta_{200}\theta_{3001}$
$X_2 = 0$	$\frac{\theta_{10}\theta_{200}\theta_{3000} + \theta_{201}\theta_{3010} - \theta_{10}\theta_{201}\theta_{3010}}{\theta_{10}\theta_{200} + \theta_{201} - \theta_{10}\theta_{201}}$

Table 2: Symbolic expressions for the probability  $P(X_3 = 0|evidence)$  for several evidence cases for the network in Example 1.

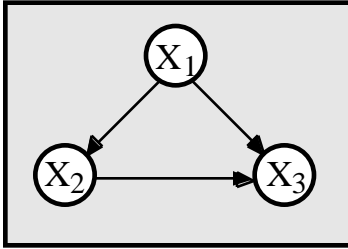


Figure 1: An example of a three-node Bayesian Network.

### 3 Algebraic Structure of Conditional Probabilities

Castillo et al. (1995) give the following theorems which characterize the algebraic structure of conditional probabilities of single nodes.

**Theorem 1** *The prior marginal probability of any set of nodes  $Y$  is a polynomial in the parameters of degree less than or equal to the minimum of the number of parameters or nodes. However, it is a first degree polynomial in each parameter.*

For example, as can be seen in the first row of Table 2, the prior marginal probability of node  $X_3$  given no evidence is a polynomial of first degree in each of the symbolic parameters.

**Theorem 2** *The posterior marginal probability of any set of nodes  $Y$ , i.e., the conditional of the set  $Y$  given some evidence  $\mathcal{E}$ , is a ratio of two polynomial functions of the parameters. Furthermore, the denominator polynomial is the same for all nodes.*

For example, the last two rows in Table 2 show that the posterior distribution of node  $X_3$  given some evidence values, is a ratio of two polynomials (note that the first of these two cases is a polynomial function of the parameters, but this is only because the denominator in this case is equal to 1).

The second part of Theorem 2 states that the denominator polynomial is the same for all nodes. For example, the denominators of the rational functions  $P(X_1 = i|X_2 = 0)$  and  $P(X_3 = j|X_2 = 0)$ , for all values of  $i$  and  $j$ , are the same. This implies that the denominator is a normalizing constant and need not be explicitly computed in every case.

Theorems 1 and 2 guarantee that the conditional probabilities of a target node given some evidence is either a polynomial or a ratio of two polynomials. The general form of these polynomials is:

$$\sum_{m_r \in \mathcal{M}} c_r m_r, \tag{5}$$

where  $c_r$  is the numerical coefficient associated with the monomial  $m_r$ . The set of monomials  $\mathcal{M}$  is formed by taking a cartesian product of the subsets of the parameters. Note that the representation of the joint probability  $P(X)$  in (1), implies that parameters with the same index  $i$  (e.g.  $\theta_{ij\pi}$  and  $\theta_{ik\pi}$ ) cannot appear in the same monomial. For example,  $\theta_{200}$  and  $\theta_{201}$ , in Example 1. For this reason the monomials are constructed by taking a cartesian product, rather all possible combinations of the parameters.

In the next section we develop a method for computing these polynomials, and hence  $P(X_i|\mathcal{E})$ , in an efficient way.

## 4 Efficient Computations of Conditional Probabilities

The proposed method consists of three steps:

1. Identify the minimal subset of the parameters which contains sufficient information to compute the conditional probabilities,
2. Construct the monomials by taking the cartesian product of the subsets of sufficient parameters, then eliminate the monomials which contain infeasible combinations of the parameters, and
3. Compute the polynomial coefficients required to compute the desired conditional probabilities.

These steps are presented in details below.

### 4.1 Identifying the Set of Relevant Nodes

The conditional probability  $P(X_i|\mathcal{E})$  does not necessarily involve all nodes. Thus, the computations of  $P(X_i|\mathcal{E})$  can be simplified by identifying only the set of nodes that are relevant to the calculation of  $P(X_i|\mathcal{E})$ . This set of relevant nodes can be obtained using either one of the two algorithms given in Geiger et al. (1990) and Shachter (1990). The first of these algorithms is given below.

**Algorithm 1** (*Identifies the Set of Relevant Nodes*)

- **Input:** A Bayesian network  $(D, P)$  and two sets of nodes: a target set  $Y$  and an evidential set  $\mathcal{E}$  (possibly empty).
- **Output:** The set of relevant nodes  $V$  needed to compute  $P(Y|\mathcal{E})$ .
- **Step 1:** Construct a DAG  $D'$  by augmenting  $D$  with a dummy node  $V_i$  and adding a link  $V_i \rightarrow X_i$  for every chance node  $X_i$  in  $D$ .
- **Step 2:** Identify the set  $V$  of dummy nodes in  $D'$  not  $d$ -separated from  $Y$  by  $\mathcal{E}$ .

The node  $V_i$  represents the parameters,  $\Theta_i$ , of node  $X_i$ . Step 2 of Algorithm 1 can be carried out in linear time using an algorithm provided by Geiger et al. (1990). Using this algorithm one can significantly reduce the set of parameters to be considered in the analysis.

We now illustrate Algorithm 1 using the Bayesian network of Example 1. We identify the relevant set of nodes needed to calculate the conditional probability  $P(X_3|evidence)$  in three different cases:

1. Case 1: No evidence.
2. Case 2: Evidence  $X_1 = 0$ .
3. Case 3: Evidence  $X_2 = 0$ .

The first step of Algorithm 1 is common for all three cases:

- **Step 1:** In this example, all the nodes are chance nodes because the corresponding probability tables are given symbolically. We construct a new DAG  $D'$  by adding the dummy nodes  $\{V_1, V_2, V_3\}$  and the corresponding links, as shown in Figure 2. From Table 1, the sets of parameters corresponding to the dummy nodes are:

$$\begin{aligned} \text{Node } V_1 : \Theta_1 &= \{\theta_{10}, \theta_{11}\}, \\ \text{Node } V_2 : \Theta_2 &= \{\theta_{200}, \theta_{201}, \theta_{210}, \theta_{211}\}, \\ \text{Node } V_3 : \Theta_3 &= \{\theta_{3000}, \theta_{3001}, \theta_{3010}, \theta_{3011}, \theta_{3100}, \theta_{3101}, \theta_{3110}, \theta_{3111}\}. \end{aligned}$$

Note that we are dealing with all possible parameters associated with the nodes, without considering the relationships among them (see Equation (3)). Dealing with all parameters will facilitate finding the coefficients of the polynomials in an efficient way as we shall see in Section 4.4.

- **Step 2:** Figure 3 shows the moralized ancestral graph associated with node  $X_3$  for the above three cases. From these graphs we conclude the following:
  - Case 1: No evidence. All nodes  $V_i$  are not  $d$ -separated from the target node  $X_3$  as can be seen in Figure 3(a). Thus,  $V = \{V_1, V_2, V_3\}$ .
  - Case 2: Evidence  $X_1 = 0$ . Figure 3(b) shows that only node  $V_1$  is  $d$ -separated from  $X_3$  by  $X_1$ . Thus,  $V = \{V_2, V_3\}$ .
  - Case 3: Evidence  $X_2 = 0$ . Figure 3(c) shows that none of the dummy nodes is  $d$ -separated from  $X_3$  by  $X_2$ . Then,  $V = \{V_1, V_2, V_3\}$ .

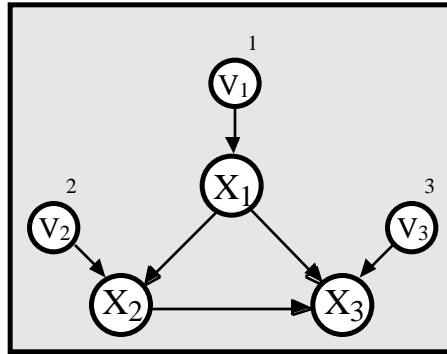
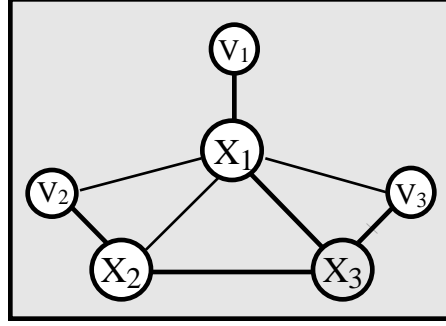


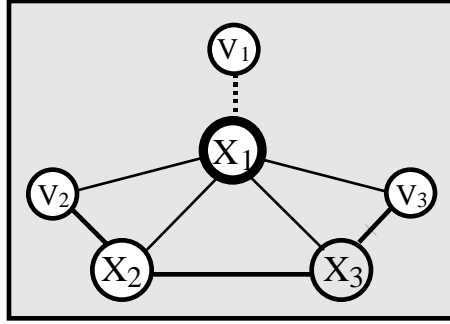
Figure 2: Augmented graph obtained by adding a dummy node  $V_i$  and a link  $V_i \rightarrow X_i$ , for every chance node  $X_i$ .

## 4.2 Identifying the Set of Sufficient Parameters

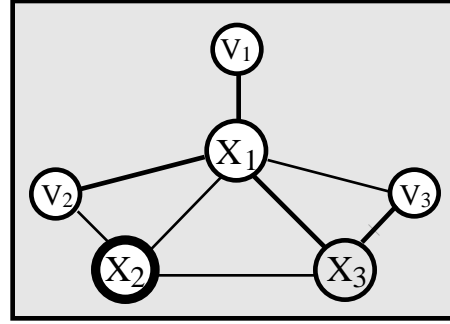
The set of relevant nodes  $V$  is identified by Algorithm 1. Let  $\Theta$  be the set of all the parameters associated with the dummy nodes  $V_i$  that are included in  $V$ . Note that the set  $\Theta$  contains all



(a) No evidence



(b) Evidence  $X_1 = 0$



(c) Evidence  $X_2 = 0$

Figure 3: Identifying relevant nodes for three different evidence cases.

the parameters that appear in the polynomial expression needed to compute  $P(X_i|\mathcal{E})$ . When identifying the set of relevant nodes (and hence the set of sufficient parameters  $\Theta$ ), Algorithm 1 takes into consideration only the set of evidence variables, but it does not make use of their values. By considering the values of the evidence variables, the set of sufficient parameters  $\Theta$  can be reduced even further by identifying and eliminating the set of parameters which are in contradiction with the evidence. These parameters are identified using the following two rules:

- **Rule 1:** Eliminate the parameters  $\theta_{ij\pi}$  if  $x_i \neq j$  for  $X_i \in \mathcal{E}$ .
- **Rule 2:** Eliminate the parameters  $\theta_{ij\pi}$  if parents' instantiations  $\pi$  are incompatible with the evidence.

The resultant set  $\Theta$  now contains the minimal sufficient subset of parameters. The following algorithm identifies such a subset:

**Algorithm 2** (*Identifies Minimal Subset of Sufficient Parameters*)

- **Input:** A Bayesian network  $(D, P)$  and two sets of nodes: a target set  $Y$  and an evidential set  $\mathcal{E}$  (possibly empty).
- **Output:** The minimum set of parameters  $\Theta$  that contains sufficient information to compute  $P(Y|\mathcal{E})$ .



- **Step 1:** Use Algorithm 1 to calculate the set of relevant nodes  $V$  and the associated set of parameters  $\Theta$  that contains sufficient information to compute  $P(Y|\mathcal{E})$ .
- **Step 2:** If there is evidence, remove from  $\Theta$  the parameters  $\theta_{ij\pi}$  if  $x_i \neq j$  for  $X_i \in \mathcal{E}$  (Rule 1).
- **Step 3:** If there is evidence, remove from  $\Theta$  the parameters  $\theta_{ij\pi}$  if the values of parents' instantiations  $\pi$  are incompatible with the evidence (Rule 2).

We illustrate Algorithm 2 using the Bayesian network in Figure 1 and the three cases mentioned above.

- **Step 1:** The results of this step are given in Step 2 of Algorithm 1. Therefore, the sets of sufficient parameters associated with the three cases are:

$$\begin{aligned} \text{Case 1 (no evidence)} : \quad & \Theta = \{\Theta_1, \Theta_2, \Theta_3\}, \\ \text{Case 2 } (X_1 = 0) : \quad & \Theta = \{\Theta_2, \Theta_3\}, \\ \text{Case 3 } (X_2 = 0) : \quad & \Theta = \{\Theta_1, \Theta_2, \Theta_3\}, \end{aligned}$$

- **Step 2:** The results of this step are given for each case below:
  - Case 1: No Evidence. Since there is no evidence, Step 2 does not apply here. Thus, no reduction of  $\Theta$  is possible at this step.
  - Case 2: Evidence  $X_1 = 0$ . The set  $\Theta = \{\Theta_2, \Theta_3\}$  does not contain parameters associated with the evidence node  $X_1$ . Therefore, no parameters are removed from  $\Theta$  at this step.
  - Case 3: Evidence  $X_2 = 0$ . The parameters  $\theta_{210}$ , and  $\theta_{211}$  are removed from  $\Theta$  because they do not match the evidence  $X_2 = 0$  (they indicate that  $X_2 = 1$ ).
- **Step 3:** The results of this step are given for each case below:
  - Case 1: No evidence. Step 3 does not apply because there is no evidence. Thus,  $\Theta = \{\Theta_1, \Theta_2, \Theta_3\}$  is the minimal set of sufficient parameters needed to calculate  $P(X_3)$ .
  - Case 2: Evidence  $X_1 = 0$ . The instantiations of the parents associated with parameters  $\theta_{201}, \theta_{211}$  do not match the evidence  $X_1 = 0$ . The same is true for the parameters  $\theta_{3j10}$  and  $\theta_{3j11}$ , for all values of  $j$ . Thus, we remove these parameters from  $\Theta$  and obtain

$$\Theta = \{\{\theta_{201}, \theta_{211}\}; \{\theta_{3000}, \theta_{3001}, \theta_{3100}, \theta_{3101}\}\},$$

which is the minimal subset of parameters needed to calculate  $P(X_3|X_1 = 0)$ .

Note that the number of parameters is reduced from 14 to 6 parameters (or from 7 to 3 free parameters).

- Case 3: Evidence  $X_2 = 0$ . The parameters  $\theta_{3j01}$  and  $\theta_{3j11}$ , for all values of  $j$  contradict the evidence  $X_2 = 0$ , hence they are removed from  $\Theta$ . The resultant minimal sufficient subset of parameters is

$$\Theta = \{\{\theta_{10}, \theta_{11}\}; \{\theta_{200}, \theta_{201}\}; \{\theta_{3000}, \theta_{3010}, \theta_{3100}, \theta_{3110}\}\}.$$

The final results of applying Algorithms 1 and 2 to the Bayesian network of Example 1 are summarized in Table 3. We make the following remarks:

1. In Case 1 of no evidence, Algorithms 1 and 2 did not decrease the number of initial parameters, 14, because (1) there is no evidence and (2) there is no independency structure in the Bayesian network of Example 1. When there is no evidence but the structure of the network is not highly dependent, Algorithm 1 can still produce a substantial reduction in number of initial parameters, as we shall see in Section 5.
2. When evidence is available (as in Cases 2 and 3), Algorithm 2 produces a more substantial reduction in the number of parameters than Algorithm 1, as would be expected. For example, in Case 2, the two algorithms reduced the number of parameters by 2 and 6, respectively.
3. By comparing the expressions for the probability  $P(X_3|\mathcal{E})$ , written in symbolic form as given in Table 2, with the parameters in Table 3, we see that the results in the two tables agree. For example,  $P(X_3 = 0|X_1 = 0)$  does not depend of the parameters in  $\Theta_1$ , whereas  $P(X_3 = 0|X_2 = 0)$  does depend on all parameters. Note that Table 2 shows the probabilities as function of only the free parameters.

### 4.3 Identifying Feasible Monomials

Once the minimal sufficient subsets of parameters has been identified, they are combined to obtain the final polynomial required to compute the conditional probabilities. As stated in Section 3, the monomials are obtained by taking the cartesian product of the minimal sufficient subsets of parameters. The set of all monomials obtained by the cartesian product can be reduced further by eliminating the set of all infeasible combinations of the parameters. This reduction can be done using the following rule:

- **Rule 3:** Parameters associated with contradicting conditioning instantiations cannot appear in the same monomial. For example, in Example 1,  $\theta_{200}$  (which conditions on  $X_1 = 0$ ) and  $\theta_{3010}$  (which conditions on  $X_1 = 1$ ) cannot occur simultaneously.

Combining Algorithm 2 with the above rule, we obtain the following algorithm:

**Algorithm 3** (*Identifies Feasible Monomials*)

- **Input:** A Bayesian network  $(D, P)$  and two sets of nodes: a target set  $Y$  and an evidential set  $\mathcal{E}$  (possibly empty).

Case	Initially	
	Parameters	Number
No evidence	$\{\Theta_1, \Theta_2, \Theta_3\}$	14
$X_1 = 0$	$\{\Theta_1, \Theta_2, \Theta_3\}$	14
$X_2 = 0$	$\{\Theta_1, \Theta_2, \Theta_3\}$	14
Case	After Algorithm 1	
	Parameters	Number
No evidence	$\{\Theta_1, \Theta_2, \Theta_3\}$	14
$X_1 = 0$	$\{\Theta_2, \Theta_3\}$	12
$X_2 = 0$	$\{\Theta_1, \Theta_2, \Theta_3\}$	14
Case	After Algorithm 2	
	Parameters	Number
No evidence	$\{\Theta_1, \Theta_2, \Theta_3\}$	14
$X_1 = 0$	$\{\theta_{200}, \theta_{210}, \theta_{3000}, \theta_{3001}, \theta_{3100}, \theta_{3101}\}$	6
$X_2 = 0$	$\{\theta_{10}, \theta_{11}, \theta_{200}, \theta_{201}, \theta_{3000}, \theta_{3010}, \theta_{3100}, \theta_{3110}\}$	8

Table 3: Set of relevant parameters needed to calculate  $P(X_3|evidence)$ , for three different evidence cases before and after applying Algorithms 1, 2.

- **Output:** *The minimum set of monomials  $\mathcal{M}$  which forms the polynomial expression needed to compute the probability  $P(Y|\mathcal{E})$ .*
- **Step 1:** *Using Algorithm 2, identify the set  $\Theta$  of minimal sufficient parameters.*
- **Step 2:** *Obtain the set of monomials  $\mathcal{M}$  by taking the cartesian product of the subsets of parameters in  $\Theta$ .*
- **Step 3:** *Using Rule 3, remove from  $\mathcal{M}$  those monomials which contain a set of incompatible parameters.*

Table 4 shows the set of minimum monomials obtained initially, and after applying Algorithms 2 and 3, to the three evidence cases mentioned above. As an illustrative example, we apply Algorithm 3 to obtain the feasible monomials in Case 2: Evidence  $X_1 = 0$ .

- **Step 1:** The minimal sufficient set of parameters obtained by Algorithm 2 is:

$$\Theta = \{\{\theta_{200}, \theta_{210}\}; \{\theta_{3000}, \theta_{3001}, \theta_{3100}, \theta_{3101}\}\},$$

as shown in Table 3.

- **Step 2:** The set of monomials obtained by taking the cartesian product is:

$$\begin{aligned} &\theta_{200}\theta_{3000}, \quad \theta_{200}\theta_{3001}, \quad \theta_{200}\theta_{3100}, \quad \theta_{200}\theta_{3101} \\ &\theta_{210}\theta_{3000}, \quad \theta_{210}\theta_{3001}, \quad \theta_{210}\theta_{3100}, \quad \theta_{210}\theta_{3101}. \end{aligned}$$

Note that, at this step, the set  $\mathcal{M}$  has been reduced from 64 to 8 candidate monomials.

Case	Initially	
	Monomials	Number
No evidence	$\Theta_1 * \Theta_2 * \Theta_3$	64
$X_1 = 0$	$\Theta_1 * \Theta_2 * \Theta_3$	64
$X_2 = 0$	$\Theta_1 * \Theta_2 * \Theta_3$	64
Case	After Algorithm 2	
	Monomials	Number
No evidence	$\Theta_1 * \Theta_2 * \Theta_3$	64
$X_1 = 0$	$\{\theta_{200}, \theta_{210}\} * \{\theta_{3000}, \theta_{3001}, \theta_{3100}, \theta_{3101}\}$	8
$X_2 = 0$	$\{\theta_{10}, \theta_{11}\} * \{\theta_{200}, \theta_{201}\} * \{\theta_{3000}, \theta_{3010}, \theta_{3100}, \theta_{3110}\}$	16
Case	After Algorithm 3	
	Monomials	Number
No evidence	$\{\{\theta_{10}\} * \{\theta_{200}\} * \{\theta_{3000}, \theta_{3100}\}; \{\theta_{10}\} * \{\theta_{210}\} * \{\theta_{3001}, \theta_{3101}\};$ $\{\theta_{11}\} * \{\theta_{201}\} * \{\theta_{3010}, \theta_{3110}\}; \{\theta_{11}\} * \{\theta_{211}\} * \{\theta_{3011}, \theta_{3111}\}\}$	8
$X_1 = 0$	$\{\{\theta_{200}\} * \{\theta_{3000}, \theta_{3100}\}; \{\theta_{210}\} * \{\theta_{3001}, \theta_{3101}\}\}$	4
$X_2 = 0$	$\{\theta_{10}\theta_{200}\theta_{3000}, \theta_{11}\theta_{201}\theta_{3110}\}$	2

Table 4: Set of monomials needed to calculate  $P(X_3|evidence)$ , for three different evidence cases.

- **Step 3:** The parameters  $\theta_{3001}$ , and  $\theta_{3101}$  indicate that  $X_2 = 1$ . By Rule 3, they can not appear in the same monomial with parameter  $\theta_{200}$ , which indicates that  $X_2 = 0$ . The same is true for parameters  $\theta_{210}$ ,  $\theta_{3000}$ , and  $\theta_{3100}$ . Thus, four monomials are eliminated and the set is reduced to:

$$\begin{aligned} &\theta_{200}\theta_{3000}, \quad \theta_{200}\theta_{3110} \\ &\theta_{210}\theta_{3001}, \quad \theta_{210}\theta_{3101}. \end{aligned}$$

Thus, the number of monomials is reduced to 4.

As can be seen from Table 4, the number of candidate monomials has been reduced to a minimum after applying Algorithm 3.

## 4.4 Computing the Polynomial Coefficients

The set of monomials  $\mathcal{M}$  constructed by Algorithm 3 contains all the monomials needed to compute  $P(X_i = j|\mathcal{E})$  for  $j = 0, \dots, r_i - 1$ . This set can be divided into  $r_i$  subsets where the  $j$ -th subset  $\mathcal{M}_j$  contains the set of monomials needed to compute  $P(X_i = j|\mathcal{E})$  for one value of  $j$ . Let  $n_j$  be the number of monomials in  $\mathcal{M}_j$  and  $m_{jk}$  be the  $k$ -th monomial in the subset  $\mathcal{M}_j$ . Note that the monomials are products of certain subsets of the parameters  $\Theta$ . From (5), the polynomial needed to compute  $P(X_i = j|\mathcal{E})$  is of the form

$$p_j(\Theta) = \sum_{m_{jk} \in \mathcal{M}_j} c_{jk} m_{jk} \propto P(X_i = j|\mathcal{E}), \quad j = 0, \dots, r_i - 1. \quad (6)$$

Thus,  $P(X_i = j|\mathcal{E})$  can be written as a linear convex combination of the monomials in  $\mathcal{M}_j$ . Our objective now is to compute the coefficients  $c_{jk}$ .

If the parameters  $\Theta$  are assigned numerical values, say  $\theta$ , then  $p_j(\theta)$  can be obtained by replacing  $\Theta$  by  $\theta$  and using any numeric propagation method to compute  $P(X_i = j|\mathcal{E}, \Theta = \theta)$ . Thus, we have

$$P(X_i = j|\mathcal{E}, \Theta = \theta) \propto p_j(\theta) = \sum_{m_{jk} \in \mathcal{M}_j} c_{jk} m_{jk}. \quad (7)$$

The term  $p_j(\theta)$  represents the unnormalized probability  $P(X_i = j|\mathcal{E}, \Theta = \theta)$ . Note that in (7) all the monomials and  $p_j(\theta)$  are known numbers and the only unknowns are the coefficients  $c_{jk}$ ,  $k = 1, \dots, n_j$ . To compute these coefficients, we need to construct any set of  $n_j$  independent equations each is of the form (7). These equations can be obtained using  $n_j$  sets of distinct values of  $\Theta$ . Let these values be denoted by  $\theta_1, \dots, \theta_{n_j}$ . Let  $T_j$  be the  $n_j \times n_j$  non-singular matrix whose  $ik$ -th element is the values of the monomial  $m_{jk}$  obtained by replacing  $\Theta$  by  $\theta_i$ , the  $i$ -th set of numeric values of  $\Theta$ . Let

$$c_j = \begin{pmatrix} c_{j1} \\ \vdots \\ c_{jn_j} \end{pmatrix}, \quad \text{and} \quad p_j = \begin{pmatrix} P(X_i = j|\mathcal{E}, \Theta = \theta_1) \\ \vdots \\ P(X_i = j|\mathcal{E}, \Theta = \theta_{n_j}) \end{pmatrix}.$$

From (7) the  $n_j$  independent linear equations can be written as

$$T_j c_j = p_j,$$

which implies that the coefficients  $c_{jk}$  are given by

$$c_j = T_j^{-1} p_j.$$

The values of the coefficients  $c_{jk}$  can then be substituted in (6) and the unnormalized probability  $p_j(\theta)$  is expressed as a function of  $\Theta$ .

The above calculations are summarized in the following algorithm.

**Algorithm 4** (*Computes Polynomial Coefficients*)

- **Input:** A Bayesian network  $(D, P)$ , a target node  $X_i$  and an evidential set  $\mathcal{E}$  (possibly empty).
- **Output:** The polynomial coefficients  $c_{jk}$  in (6).
- **Step 1:** Use Algorithm 3 to identify the minimum set of monomials  $\mathcal{M}$  needed to calculate the probability  $P(X_i|\mathcal{E})$ .
- **Step 2:** For each possible state  $j$  of node  $X_i$ :  $j = 0, \dots, (r_i - 1)$ . Build the subset  $\mathcal{M}_j$  by considering those monomials in  $\mathcal{M}$  containing some parameter of the form  $\theta_{ij\pi}$ , for some  $\pi$ . Note that this process divide the set  $\mathcal{M}$  in  $r_i - 1$  different sets of monomials.
- **Step 3:** For each possible state  $j$  of node  $X_i$ , calculate the coefficients  $c_{jk}$ ,  $k = 1, \dots, n_j$ , as follows:
  1. Construct the  $n_j \times n_j$  nonsingular matrix  $T_j$  such that  $T_j c_j = p_j$ .
  2. Use any numeric propagation method to compute the corresponding vector  $p_j$ .

$\mathcal{M}_0$	$\mathcal{M}_1$
$\theta_{200}\theta_{3000}$	$\theta_{210}\theta_{3001}$
$\theta_{200}\theta_{3100}$	$\theta_{210}\theta_{3101}$

Table 5: Required monomials to determine the indicated probabilities.

3. Compute  $c_j = T_j^{-1}p_j$ .

Note that the matrix  $T_j$  in Step 3 is not unique. One can take advantage of this fact and choose the values of  $\Theta$  which produce a simple matrix  $T_j$ . The use of the extreme values 0 or 1 for the parameters in  $\Theta$  usually produces a simple form of  $T_j$ . In this case the matrix  $T_j$  contains the so called canonical components. Algorithm 4, including this process of constructing  $T_j$ , is illustrated using the network in Example 1 and Case 2: Evidence  $X_1 = 0$ .

- **Step 1:** In Section 4.3 we applied Algorithm 3 and found the minimal set of feasible polynomials needed to calculate  $P(X_3|X_1 = 0)$ . These monomials are shown in Table 5.
- **Step 2:** Table 5 also shows the subsets of monomials  $\mathcal{M}_0$ ,  $\mathcal{M}_1$ , needed to calculate  $P(X_3 = 0|X_1 = 0)$ , and  $P(X_3 = 1|X_1 = 0)$ , respectively.
- **Step 3:** For  $j = 0$  we need to construct  $T_0$  using

$$\begin{aligned}
p_0(\Theta) &= c_{01}m_{01} + c_{02}m_{02} \\
&= c_{01}\theta_{200}\theta_{3000} + c_{02}\theta_{200}\theta_{3100}.
\end{aligned} \tag{8}$$

Since we have two coefficients, we need two independent equations which are obtained by specifying two distinct sets of values of the parameters

$$\Theta = \{\theta_{200}, \theta_{210}, \theta_{3000}, \theta_{3100}, \theta_{3001}, \theta_{3101}\}.$$

A simple way of selecting values of  $\Theta$  is as follows. To obtain the  $i$ -th set  $\theta_i$  we set all the parameters in  $m_{0i}$  equal to one and all other free parameters equal to zero. Thus, the first set is obtained by setting  $(\theta_{200}, \theta_{3000}) = (1, 1)$  and all other free parameters equal to zero. The second set is obtained by setting  $(\theta_{200}, \theta_{3100}) = (1, 1)$  and all other free parameters equal to zero. This yields the two sets:

$$\begin{aligned}
\theta_1 &= (1, 0, 1, 0, 1, 0) \\
\theta_2 &= (1, 0, 0, 1, 1, 0).
\end{aligned}$$

Note that both cases are obtained by setting the free parameter  $\theta_{3101}$  equal to zero (using Equation (3)). Thus, the two equations are:

$$\begin{aligned}
p_0(\theta_1) &= c_{01} \times 1 \times 1 + c_{02} \times 1 \times 0 = c_{01}, \\
p_0(\theta_2) &= c_{01} \times 1 \times 0 + c_{02} \times 1 \times 1 = c_{02}.
\end{aligned}$$

This implies that

$$T_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and the coefficients are given by

$$c_0 = T_1^{-1}p_0 = p_0,$$

where

$$c_0 = p_0 = \begin{pmatrix} p_0(\theta_1) \\ p_0(\theta_2) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (9)$$

Note that  $p_0(\theta_1)$  and  $p_0(\theta_2)$  are obtained by performing two numerical propagations, one using  $\Theta = \theta_1$  and the other using  $\Theta = \theta_2$ .

We repeat this process for  $j = 1$ . The polynomial equation is

$$\begin{aligned} p_1(\Theta) &= c_{11}m_{11} + c_{12}m_{12} \\ &= c_{11}\theta_{210}\theta_{3001} + c_{12}\theta_{210}\theta_{3101}. \end{aligned} \quad (10)$$

We need two sets of values of  $\Theta$ . The first set is obtained by setting  $(\theta_{210}, \theta_{3001}) = (1, 1)$  and all other free parameters equal to zero. The second set is obtained by setting  $(\theta_{210}, \theta_{3101}) = (1, 1)$  and all other free parameters equal to zero. This yields the two sets:

$$\begin{aligned} \theta_1 &= (0, 1, 1, 0, 1, 0) \\ \theta_2 &= (0, 1, 1, 0, 0, 1) \end{aligned}$$

and the two equations are:

$$\begin{aligned} p_1(\theta_1) &= c_{11} \times 1 \times 1 + c_{12} \times 1 \times 0 = c_{11} \\ p_1(\theta_2) &= c_{11} \times 1 \times 0 + c_{12} \times 1 \times 1 = c_{12}, \end{aligned}$$

which implies that

$$T_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We use a numerical propagation method to compute  $p_1(\theta_1)$  and  $p_1(\theta_2)$  and obtain the coefficients.

$$c_1 = \begin{pmatrix} p_1(\theta_1) \\ p_1(\theta_2) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad (11)$$

and Algorithm 4 concludes.

Note that the conditional probabilities can be obtained by substituting the values of the coefficients in the corresponding equation. For example, for  $j = 0$ , we obtain the conditional probability by substituting the values in (9) in (8):

$$P(X_3 = 0 | X_1 = 0) \propto \theta_{200}\theta_{3000} + \theta_{200}\theta_{3100},$$

which agrees with the probability  $P(X_3 = 0|X_1 = 0)$  in Table 2 which were obtained by symbolic propagation (note that  $\theta_{3101} = 1 - \theta_{3000}$ ).

It is interesting to note here that all coefficients  $c_{ij}$  in (9) and (11) are found to be 1. This is, in fact, not a coincidence in this case because it can be easily shown that if all the nodes of a network are chance nodes (as is the case here), then all coefficients are equal to 1 and there no need to execute Algorithm 4 in this case.

## 4.5 The Proposed Algorithm

Algorithm 4 gives the polynomial coefficients required to compute the unnormalized probabilities given in (6). The required conditional probabilities  $P(X_i = j|\mathcal{E})$  can then be obtained by normalizing the unnormalized probabilities. We, therefore, propose the following algorithm for computing  $P(X_i = j|\mathcal{E})$ . This algorithm is obtained by combining Algorithms 1-4 with the final normalizing step.

**Algorithm 5** (*Computes  $P(X_i|\mathcal{E})$* )

- **Input:** A Bayesian network  $(D, P)$ , a target node  $X_i$  and an evidential set  $\mathcal{E}$  (possibly empty).
- **Output:** The probabilities  $P(X_i|\mathcal{E})$ .
- **Step 1:** Construct a DAG  $D'$  by augmenting  $D$  with a dummy node  $V_i$  and adding a link  $V_i \rightarrow X_i$  for every chance node  $X_i$  in  $D$ . The node  $V_i$  represents the parameters,  $\Theta_i$ , of node  $X_i$ .
- **Step 2:** Identify the set  $V$  of dummy nodes in  $D'$  not  $d$ -separated from  $Y$  by  $\mathcal{E}$ , and let  $\Theta$  be the set of all the parameters associated with the dummy nodes  $V_i$  that are included in  $V$ .
- **Step 3:** If there is evidence, remove from  $\Theta$  the parameters  $\theta_{ij\pi}$  if  $x_i \neq j$  for  $X_i \in \mathcal{E}$  (Rule 1).
- **Step 4:** If there is evidence, remove from  $\Theta$  the parameters  $\theta_{ij\pi}$  if the set of values of parents' instantiations  $\pi$  are incompatible with the evidence (Rule 2).
- **Step 5:** Obtain the set of monomials  $\mathcal{M}$  by taking the cartesian product of the subsets of parameters in  $\Theta$ .
- **Step 6:** Using Rule 3, remove from  $\mathcal{M}$  those monomials which contain a set of incompatible parameters.
- **Step 7:** For each possible state  $j$  of node  $X_i$ :  $j = 0, \dots, (r_i - 1)$ . Build the subset  $\mathcal{M}_j$  by considering those monomials in  $\mathcal{M}$  containing some parameter of the form  $\theta_{ij\pi}$ , for some  $\pi$ . Note that this process divide the set  $\mathcal{M}$  in  $r_i - 1$  different sets of monomials.
- **Step 8:** For each possible state  $j$  of node  $X_i$ , calculate the coefficients  $c_{jk}$ ,  $k = 1, \dots, n_j$ , as follows:
  1. Construct the  $n_j \times n_j$  nonsingular matrix  $T_j$  such that  $T_j c_j = p_j$ .



2. Use any numeric propagation method to compute the corresponding vector  $p_j$ .
  3. Compute  $c_j = T_j^{-1}p_j$ .
- **Step 9:** Calculate the unnormalized probabilities  $p_j(\Theta)$ ,  $j = 0, \dots, r_i - 1$  and the conditional probabilities  $P(X_i = j|\mathcal{E}) = p_j(\Theta)/N$ , where

$$N = \sum_{j=0}^{r_i-1} p_j(\Theta)$$

is the normalizing constant.

## 5 An Illustrative Example

To illustrate the proposed Algorithm 5 we use the following example. Suppose we have a discrete Bayesian network consisting of seven variables  $X = \{X_1, X_2, \dots, X_7\}$  with the corresponding DAG  $D$  as given in Figure 4. The structure of  $D$  implies that the joint probability of the set of nodes can be written as:

$$P(X) = P(X_1)P(X_2|X_1)P(X_3|X_1)P(X_4|X_2, X_3)P(X_5|X_3)P(X_6|X_4)P(X_7|X_4). \quad (12)$$

For simplicity, but without loss of generality, assume that all nodes represent binary variables with values in the set  $\{0, 1\}$ . This and the structure of the network in Figure 4 imply that the joint probability distribution of the seven variables depends on 30 parameters. However, only 15 of the parameters are free (because the probabilities in each conditional distribution must add up to unity). These 15 parameters are given in Table 6. Note that six of the free parameters (those associated with nodes  $X_2$  and  $X_4$ ) are assigned fixed numerical values and the remaining nine are given symbolically. Thus, the chance nodes in this case are  $\{X_1, X_3, X_5, X_6, X_7\}$ .

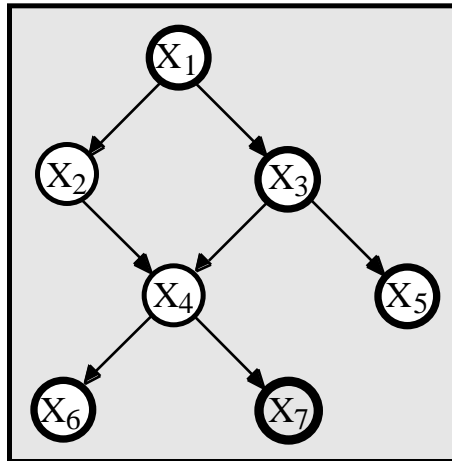


Figure 4: An example of a six-node Bayesian Network.

For illustrative purposes, suppose now that the target node is  $X_7$  and that we wish to compute the conditional probabilities  $P(X_7|X_1 = 1)$ . Then, using Algorithm 5, we do the following:

Node		Parameters	
$X_i$	Parents	$X_i = 0$	$X_i = 1$
$X_1$	None	$\theta_{10} = P(X_1 = 0)$	$\theta_{11} = P(X_1 = 1)$
$X_2$	$X_1$	$\theta_{200} = P(X_2 = 0 X_1 = 0) = 0.2$ $\theta_{201} = P(X_2 = 0 X_1 = 1) = 0.5$	$\theta_{210} = P(X_2 = 1 X_1 = 0) = 0.8$ $\theta_{211} = P(X_2 = 1 X_1 = 1) = 0.5$
$X_3$	$X_1$	$\theta_{300} = P(X_3 = 0 X_1 = 0)$ $\theta_{301} = P(X_3 = 0 X_1 = 1)$	$\theta_{310} = P(X_3 = 1 X_1 = 0)$ $\theta_{311} = P(X_3 = 1 X_1 = 1)$
$X_4$	$X_2, X_3$	$\theta_{4000} = P(X_4 = 0 X_2 = 0, X_3 = 0) = 0.1$ $\theta_{4001} = P(X_4 = 0 X_2 = 0, X_3 = 1) = 0.2$ $\theta_{4010} = P(X_4 = 0 X_2 = 1, X_3 = 0) = 0.3$ $\theta_{4011} = P(X_4 = 0 X_2 = 1, X_3 = 1) = 0.4$	$\theta_{4100} = P(X_4 = 1 X_2 = 0, X_3 = 0) = 0.9$ $\theta_{4101} = P(X_4 = 1 X_2 = 0, X_3 = 1) = 0.8$ $\theta_{4110} = P(X_4 = 1 X_2 = 1, X_3 = 0) = 0.7$ $\theta_{4111} = P(X_4 = 1 X_2 = 1, X_3 = 1) = 0.6$
$X_5$	$X_3$	$\theta_{500} = P(X_5 = 0 X_3 = 0)$ $\theta_{501} = P(X_5 = 0 X_3 = 1)$	$\theta_{510} = P(X_5 = 1 X_3 = 0)$ $\theta_{511} = P(X_5 = 1 X_3 = 1)$
$X_6$	$X_4$	$\theta_{600} = P(X_6 = 0 X_4 = 0)$ $\theta_{601} = P(X_6 = 0 X_4 = 1)$	$\theta_{610} = P(X_6 = 1 X_4 = 0)$ $\theta_{611} = P(X_6 = 1 X_4 = 1)$
$X_7$	$X_4$	$\theta_{700} = P(X_7 = 0 X_4 = 0)$ $\theta_{701} = P(X_7 = 0 X_4 = 1)$	$\theta_{710} = P(X_7 = 1 X_4 = 0)$ $\theta_{711} = P(X_7 = 1 X_4 = 1)$

Table 6: Numeric and symbolic conditional probability tables associated with the network in Figure 4.

- **Step 1:** We need to add to the initial graph  $D$  shown in Figure 4 the nodes  $V_1, V_3, V_5, V_6, V_7$ , whose corresponding parameters sets are:

$$\begin{aligned}
 \text{Node } V_1 : \Theta_1 &= \{\theta_{10}, \theta_{11}\}, \\
 \text{Node } V_3 : \Theta_3 &= \{\theta_{300}, \theta_{301}, \theta_{310}, \theta_{311}\}, \\
 \text{Node } V_5 : \Theta_5 &= \{\theta_{500}, \theta_{501}, \theta_{510}, \theta_{511}\}, \\
 \text{Node } V_6 : \Theta_6 &= \{\theta_{600}, \theta_{601}, \theta_{610}, \theta_{611}\}, \\
 \text{Node } V_7 : \Theta_7 &= \{\theta_{700}, \theta_{701}, \theta_{710}, \theta_{711}\}.
 \end{aligned}$$

The result is shown in Figure 5.

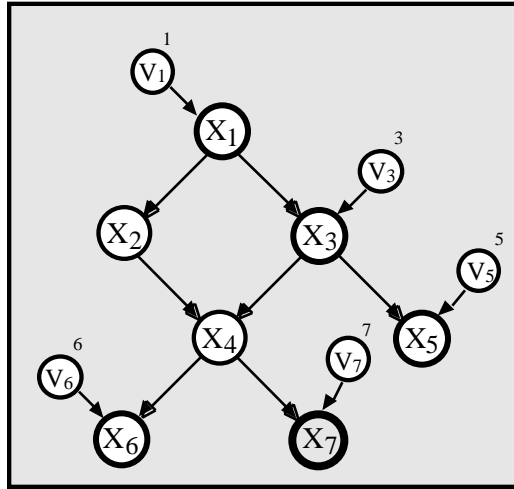


Figure 5: Augmented graph after adding a dummy node  $V_i$  for every chance node  $X_i$ .

$\mathcal{M}_0$	$\mathcal{M}_1$
$\theta_{301}\theta_{700}$	$\theta_{301}\theta_{710}$
$\theta_{301}\theta_{701}$	$\theta_{301}\theta_{711}$
$\theta_{311}\theta_{700}$	$\theta_{311}\theta_{710}$
$\theta_{311}\theta_{701}$	$\theta_{311}\theta_{711}$

Table 7: Required monomials to determine the indicated probabilities.

- **Step 2:** The set  $V$  of dummy nodes not  $d$ -separated from  $X_7$  by  $X_1$  is found to be  $V = \{V_3, V_7\}$ . Thus, the set of all parameters associated with the dummy nodes that are included in  $V$  is

$$\Theta = \{\{\theta_{300}, \theta_{301}, \theta_{310}, \theta_{311}\}; \{\theta_{700}, \theta_{701}, \theta_{710}, \theta_{711}\}\}.$$

Note that at this step we have reduced the number of parameters from 18 to 8 (or the number of free parameters from 9 to 4).

- **Step 3:** The set  $\Theta$  does not contain parameters associated with the evidential node  $X_1$ . Thus, no reduction is possible applying Rule 1.
- **Step 4:** Since  $\theta_{300}$  and  $\theta_{310}$  are not compatible with the evidence, we can remove from  $\Theta$  these parameters obtaining the minimum set of sufficient parameters:

$$\Theta = \{\{\theta_{301}, \theta_{311}\}; \{\theta_{700}, \theta_{701}, \theta_{710}, \theta_{711}\}\}.$$

- **Step 5:** The initial set of candidate monomials is given by taking the cartesian product of the minimal sufficient subsets, that is,  $\mathcal{M} = \{\theta_{301}, \theta_{311}\} * \{\theta_{700}, \theta_{701}, \theta_{710}, \theta_{711}\}$ . Thus, the candidate monomials are shown in Table 7.
- **Step 6:** The parents of nodes  $X_3$  and  $X_7$  do not have common elements, hence all monomials shown in Table 7 are feasible monomials.
- **Step 7:** The sets of monomials  $\mathcal{M}_0$  and  $\mathcal{M}_1$  needed to calculate  $P(X_7 = 0|X_1 = 1)$  and  $P(X_7 = 1|X_1 = 1)$ , respectively, are shown in the Table 7.
- **Step 8:** For  $j = 0$  we have the following polynomial equation:

$$\begin{aligned} p_0(\Theta) &= c_{01}m_{01} + c_{02}m_{02} + c_{03}m_{03} + c_{04}m_{04} \\ &= c_{01}\theta_{301}\theta_{700} + c_{02}\theta_{301}\theta_{701} + c_{03}\theta_{311}\theta_{700} + c_{04}\theta_{311}\theta_{701}. \end{aligned} \quad (13)$$

Thus, taking the canonical components

$$\{\theta_1, \theta_2, \theta_3, \theta_4\} = \{(1, 0, 1, 0, 1, 0), (1, 0, 0, 1, 1, 0), (0, 1, 1, 0, 1, 0), (0, 1, 0, 1, 1, 0)\},$$

for the set of sufficient parameters  $\Theta = \{\theta_{301}, \theta_{311}, \theta_{700}, \theta_{701}, \theta_{710}, \theta_{711}\}$ , we get the following system of equations:

$X_7 = 0$			
$(\theta_{301}, \theta_{311}, \theta_{700}, \theta_{701}, \theta_{710}, \theta_{711})$	$p_0(\theta)$	monomials	Coefficients
$(1, 0, 1, 0, 1, 0)$	0.15	$\theta_{301}\theta_{700}$	$c_{01} = 0.15$
$(1, 0, 0, 1, 1, 0)$	0.85	$\theta_{301}\theta_{701}$	$c_{02} = 0.85$
$(0, 1, 1, 0, 1, 0)$	0.35	$\theta_{311}\theta_{700}$	$c_{03} = 0.35$
$(0, 1, 0, 1, 1, 0)$	0.65	$\theta_{311}\theta_{701}$	$c_{04} = 0.65$
$X_7 = 1$			
$(\theta_{301}, \theta_{311}, \theta_{700}, \theta_{701}, \theta_{710}, \theta_{711})$	$p_1(\theta)$	monomials	Coefficients
$(1, 0, 1, 0, 1, 0)$	0.15	$\theta_{301}\theta_{710}$	$c_{11} = 0.15$
$(1, 0, 0, 1, 1, 0)$	0.85	$\theta_{301}\theta_{711}$	$c_{12} = 0.85$
$(0, 1, 1, 0, 1, 0)$	0.35	$\theta_{301}\theta_{710}$	$c_{13} = 0.35$
$(0, 1, 0, 1, 1, 0)$	0.65	$\theta_{311}\theta_{711}$	$c_{14} = 0.65$

Table 8: Monomial coefficients and their corresponding values of  $p_j(\theta)$ .

$$c_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} p_0(\theta_1) \\ p_0(\theta_2) \\ p_0(\theta_3) \\ p_0(\theta_4) \end{pmatrix} = \begin{pmatrix} 0.15 \\ 0.85 \\ 0.35 \\ 0.65 \end{pmatrix}. \quad (14)$$

Similarly, for  $j = 1$  we get

$$c_1 = \begin{pmatrix} p_1(\theta_1) \\ p_1(\theta_2) \\ p_1(\theta_3) \\ p_1(\theta_4) \end{pmatrix} = \begin{pmatrix} 0.15 \\ 0.85 \\ 0.35 \\ 0.65 \end{pmatrix}. \quad (15)$$

Table 8 shows the results of calculating the numerical probabilities needed in above expressions.

- **Step 9:** Finally, combining (13) and (14) we get the final polynomial expressions.

$$P(X_7 = 0 | X_1 = 1) \propto 0.15\theta_{301}\theta_{700} + 0.85\theta_{301}\theta_{701} + 0.35\theta_{311}\theta_{700} + 0.65\theta_{311}\theta_{701}. \quad (16)$$

Similarly, for  $X_7 = 1$  we get

$$P(X_7 = 1 | X_1 = 1) \propto 0.15\theta_{301}\theta_{710} + 0.85\theta_{301}\theta_{711} + 0.35\theta_{311}\theta_{710} + 0.65\theta_{311}\theta_{711}. \quad (17)$$

Now, we can apply the relationships among the parameters in (3) to simplify above expressions. In this case, we consider:  $\theta_{311} = 1 - \theta_{301}$ . Thus, we get:

$$\begin{aligned} P(X_7 = 0 | X_1 = 1) &\propto 0.15\theta_{301}\theta_{700} + 0.85\theta_{301}\theta_{701} + (1 - \theta_{301})(0.35\theta_{700} + 0.65\theta_{701}) \\ &= 0.35\theta_{700} - 0.2\theta_{301}\theta_{700} + 0.65\theta_{701} + 0.2\theta_{301}\theta_{701}. \end{aligned} \quad (18)$$

Similarly,

$$P(X_7 = 1 | X_1 = 1) \propto 1 - 0.35\theta_{700} + 0.2\theta_{301}\theta_{700} - 0.65\theta_{701} - 0.2\theta_{301}\theta_{701}. \quad (19)$$

Finally, adding the unnormalized probabilities in (18) and (19) we get the normalizing constant. In this case, the normalizing constant is 1. Thus, the probabilities  $P(X_7 = j | X_1 = 1)$  are given in (18) and (19).

## 6 Lower and Upper Bounds for Probabilities

The symbolic expressions of conditional probabilities obtained by Algorithm 5 can also be used to obtain lower and upper bounds for the marginal probabilities. These bounds can provide valuable information for performing sensitivity analysis of a Bayesian network. To compute these bounds, we first need the following result.

**Theorem 3 (Bela Martos, 1964)** *If the linear fractional functional of a vector  $u$ ,*

$$\frac{c * u - c_0}{d * u - d_0}, \quad (20)$$

*where  $c$  and  $d$  are vector coefficients and  $c_0$  and  $d_0$  are real constants, is defined in the convex polyhedral  $Au \leq a_0, u \geq 0$ , where  $A$  is a constant matrix and  $a_0$  is a constant vector, and the denominator in (20) does not vanish in the polyhedral, then the functional reaches the maximum at least in one of the vertices of the polyhedron.*

It can be seen from Theorem 3 that lower and upper bounds are attained at one of the canonical components (vertices of the feasible convex parameter set). Thus, from Theorem 3, the lower and upper bounds for the ratio of polynomial probabilities  $P(X_i = j | \mathcal{E})$  are given by the minimum and maximum, respectively, of the numerical values attained by this probability over all the possible canonical components associated with the parameters contained in  $\Theta$ , i.e. for all possible combinations of extreme values of the parameters (the vertices of the parameters set). As an example we compute the lower and upper bounds associated with all the variables in the Bayesian network in Section 5, first for the case of no evidence and second for the case of evidence  $X_2 = 0$ . For comparison purposes, we reduce the number of symbolic parameters from 9 to 5 (by replacing the parameters of variable  $X_3$  and  $X_6$  by numeric values, that is,  $\theta_{300} = 0.3$ ,  $\theta_{301} = 0.4$ ,  $\theta_{600} = 0.5$ ,  $\theta_{601} = 0.3$ ), and then compute the bounds and compare them with those obtained in the 9-parameter cases. Tables 9 and 10 show the lower and upper bounds for the four different cases.

Several remarks can be made here:

1. The range (the difference between lower and upper bounds) of probabilities is non-decreasing in the number of symbolic parameters. For example, the ranges for the 5-parameter case are no larger than those for the 9-parameter case (e.g., in Table 10, the range of  $X_6$  reduces from 1 to 0.004). These results are expected, because less symbolic parameters means less uncertainty.

		Case 1: 9 parameters			Case 2: 5 parameters		
Node	State	Lower	Upper	Range	Lower	Upper	Range
$X_1$	0	0.000	1.000	1.000	0.000	1.000	1.000
	1	0.000	1.000	1.000	0.000	1.000	1.000
$X_2$	0	0.200	0.500	0.300	0.200	0.500	0.300
	1	0.500	0.800	0.300	0.500	0.800	0.300
$X_3$	0	0.000	1.000	1.000	0.300	0.400	0.100
	1	0.000	1.000	1.000	0.600	0.700	0.100
$X_4$	0	0.150	0.380	0.230	0.270	0.320	0.050
	1	0.620	0.850	0.230	0.680	0.730	0.050
$X_5$	0	0.000	1.000	1.000	0.000	1.000	1.000
	1	0.000	1.000	1.000	0.000	1.000	1.000
$X_6$	0	0.000	1.000	1.000	0.354	0.364	0.010
	1	0.000	1.000	1.000	0.636	0.646	0.010
$X_7$	0	0.000	1.000	1.000	0.000	1.000	1.000
	1	0.000	1.000	1.000	0.000	1.000	1.000

Table 9: Lower and upper bounds for the initial marginal probabilities (no evidence).

- By comparison with the bounds in Table 9, we see that the ranges in Table 10 are generally smaller than those in Table 9. Again, these results are expected because more evidence means less uncertainty.

## 7 Conclusions

The symbolic structure of prior and posterior probabilities of Bayesian networks are characterized as either polynomials or ratios of two polynomial functions of the parameters, respectively. Not all terms in the polynomials, however, are relevant to the computations of the probabilities of a target node. We present methods for identifying the set of relevant parameters. This leads to substantial computational savings. In addition, an important advantage of the proposed method is that it can be performed using the currently available numeric propagation methods, thus making both symbolic computations and sensitivity analysis feasible even for large networks.

## Acknowledgments

This paper was written while J. M. Gutiérrez was visiting Cornell University. We thank the University of Cantabria and Dirección General de Investigación Científica y Técnica (DGICYT) (project PB94-1056), for support of this work.

		Case 1: 9 parameters			Case 2: 5 parameters		
Node	State	Lower	Upper	Range	Lower	Upper	Range
$X_1$	0	0.000	1.000	1.000	0.000	1.000	1.000
	1	0.000	1.000	1.000	0.000	1.000	1.000
$X_2$	0	1.000	1.000	0.000	1.000	1.000	0.000
	1	0.000	0.000	0.000	0.000	0.000	0.000
$X_3$	0	0.000	1.000	1.000	0.300	0.400	0.100
	1	0.000	1.000	1.000	0.600	0.700	0.100
$X_4$	0	0.100	0.300	0.200	0.220	0.240	0.020
	1	0.700	0.900	0.200	0.760	0.780	0.020
$X_5$	0	0.000	1.000	1.000	0.000	1.000	1.000
	1	0.000	1.000	1.000	0.000	1.000	1.000
$X_6$	0	0.000	1.000	1.000	0.344	0.348	0.004
	1	0.000	1.000	1.000	0.652	0.656	0.004
$X_7$	0	0.000	1.000	1.000	0.000	1.000	1.000
	1	0.000	1.000	1.000	0.000	1.000	1.000

Table 10: Lower and upper bounds for the conditional probabilities  $P(X_i|X_2 = 0)$ .

## References

- Castillo, E., Gutiérrez, J. M., and Hadi, A. S. (1995), “Parametric Structure of Probabilities in Bayesian Networks,” in Lectures Notes in Artificial Intelligence: Symbolic and Quantitative Approaches to Reasoning and Uncertainty (C. Froidevaux and J. Kohlas, eds), New York: Springer-Verlag, 946, 89–98.
- Castillo, E., Gutiérrez, J. M., and Hadi, A. S. (1996), Expert Systems and Probabilistic Network Models, New York: Springer-Verlag.
- Chang, K-C. and Fung, R. (1995), “Symbolic Probabilistic Inference with Both Discrete and Continuous Variables,” IEEE Transactions on Systems, Man and Cybernetics, 25, 6, 910–916.
- Li, Z., and D’Ambrosio, B. (1994), “Efficient Inference in Bayes Nets as a Combinatorial Optimization Problem,” International Journal of Approximate Reasoning, 11, 1, 55–81.
- Geiger, D., Verma, T., and Pearl, J. (1990), “Identifying Independence in Bayesian Networks,” Networks, 20, 507–534.
- Laskey, K. B. (1995), “Sensitivity Analysis for Probability Assessments in Bayesian Networks,” IEEE Transactions on Systems, Man and Cybernetics, 25, 6, 901–909.

- Lauritzen, S. L., and Spiegelhalter, D. J. (1988), “Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems,” *Journal of the Royal Statistical Society (B)*, 50, 157–224.
- Martos, B. (1964), “Hyperbolic Programming,” *Naval Research Logistic Quarterly*, 11, 135–156.
- Pearl, J. (1986), “Fusion, Propagation and Structuring in Belief Networks,” *Artificial Intelligence*, 29, 241–288.
- Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo, CA.
- Shachter, R. D. (1990), “An Ordered Examination of Influence Diagrams,” *Networks*, 20, 535–563.
- Shachter, R. D., Andersen, S. K., and Szolovits, P. (1994), “Global Conditioning for Probabilistic Inference in Belief Networks,” in *Proceedings of Uncertainty in Artificial Intelligence*, Morgan Kaufmann Publishers, 514–522.