

A New Method for Efficient Symbolic Propagation 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 Statistics, Cornell University, USA*

ABSTRACT

The paper presents a new efficient method for uncertainty propagation in discrete Bayesian networks in symbolic, as opposed to numeric, form, when considering some of the probabilities of the Bayesian network as parameters. The algebraic structure of the conditional probabilities of any set of nodes, given some evidence, is characterized as ratios of linear polynomials in the parameters. We use this result to carry out these symbolic expressions efficiently by calculating the coefficients of the polynomials involved, using standard numerical algorithms. The numeric canonical components method is proposed as an alternative to symbolic computations, gaining in speed and simplicity. It is also shown how to avoid redundancy when calculating the numeric canonical components probabilities using standard message-passing methods. The canonical components can also be used to obtain lower and upper bounds for the symbolic expression associated with the probabilities. Finally, we analyze the problem of symbolic evidence, which allows answering multiple queries regarding a given set of evidential nodes. In this case, the algebraic structure of the symbolic expressions obtained for the probabilities are shown to be ratios of non-linear polynomial expressions. Then we can perform symbolic inference with only a small set of symbolic evidential nodes. The methodology is illustrated by examples.

Key Words: Symbolic calculus, Sensitivity analysis, Propagation of uncertainty, Canonical components.

1 Introduction

Bayesian networks are powerful tools both for graphically representing the relationships among a set of variables and for dealing with uncertainties in expert systems. A key problem in Bayesian networks is evidence propagation, that is, obtaining the posterior distributions of variables when some evidence is observed. Several efficient methods for propagation of evidence in Bayesian networks have been proposed in recent years. Exact methods exploit the independence structure contained in the network to efficiently propagate uncertainty (see, for example, Kim and Pearl (1983), Lauritzen and Spiegelhalter (1988), Jensen, Olesen, and Andersen (1990), Pearl (1988), and Shachter, Andersen, and Szolovits (1994)). Stochastic simulation

constitute an interesting alternative in highly connected networks, where exact algorithms may become inefficient (Pearl (1986), Henrion (1988), Shachter and Peot (1990a), Fung and Chang (1990), Bouckaert, Castillo, and Gutiérrez (1996)). Recently, search-based approximation algorithms, which search for high probability configurations through a space of possible values, have emerged as an alternative to the above methods in special cases as, for example, in Bayesian networks with extreme probabilities (Poole (1993), Santos and Shimony (1994), Li and D'Ambrosio (1995)).

However, all exact and approximate methods require that the joint probabilities of the nodes be specified numerically, that is, all the parameters must be assigned numeric values. In practice, exact numeric specification of these parameters may not be available or it may happen that the subject matter specialists can specify only ranges of values for the parameters rather than their exact values. In such cases, there is a need for symbolic methods which are able to deal with the parameters themselves, without assigning them numeric values.

Symbolic propagation leads to solutions which are expressed as functions of the parameters in symbolic form. Thus, the answers to general queries can be given symbolically in terms of the parameters and the answers to specific queries can then be obtained by plugging the values of the parameters in the solution which is given in symbolic form, without need to redo the propagation. Furthermore, symbolic propagation allows one to study the sensitivity of the results to changes in parameter values with little additional computational effort.

Recently, two main approaches have been proposed for symbolic inference in Bayesian networks. The symbolic probabilistic inference algorithm (SPI) (Shachter, D'Ambrosio, and DelFabero (1990b), Li and D'Ambrosio (1994)) is a goal directed method which performs only those calculations that are required to respond to queries. Symbolic expressions can be obtained by postponing evaluation of expressions, maintaining them in symbolic form. On the other hand, Castillo, Gutiérrez and Hadi (1995a, 1995b, 1996) perform symbolic calculations using slightly modified versions of standard numerical propagation algorithms 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. As opposed to SPI algorithm, this method is not goal oriented, but allows us to obtain symbolic expressions for all the nodes in the network.

However, both methods suffer from the same problem: they need to use special programs, or extra computational efforts implementing the necessary code, to carry out the symbolic computations. Furthermore, computing and simplifying symbolic expressions is a computationally expensive task, and it becomes increasingly inefficient when dealing with large networks, or large numbers of symbolic parameters. In this paper we present an efficient approach to symbolic propagation that takes advantage of the polynomial structure of the probabilities of the nodes to avoid symbolic computations. The main idea of the method is obtaining the symbolic expressions

through a numerical algorithm to compute the coefficients of the associated polynomials. Then, all the computations are carried out numerically, avoiding the use of the computationally expensive symbolic manipulations. The main findings of this paper are the following:

- The algebraic structure of initial or updated conditional probabilities of events, given the evidence, when considering as parameters several probabilities in the Bayesian network has been characterized as ratios of linear polynomials in each one of the parameters.
- Taking advantage of this structure of the probabilities, a new method which obtains symbolic expression of the parameters by performing only numerical computations is introduced. Each symbolic query can be answered by performing numeric computations (numeric canonical components) using any of the mentioned above standard methods. Furthermore we show that, when using message-passing algorithms to obtain the numerical values associated with the canonical components, some of the messages are common to several components. Consequently, important savings in computation time can be obtained by avoiding redundant calculations.
- The upper and lower bounds for the symbolic expressions for the probabilities are attained at canonical cases. Thus, calculating these values does not require extra computation effort, because it reduces to the simple operation of finding the maximum and minimum values and can be done during the process of calculating these expressions. These bounds provide a useful information about the sensitivity of certain parameters in the probability of the nodes in the network.
- When introducing symbolic evidence in the network, the symbolic expressions for the conditional probabilities are now ratios of non-linear polynomials on the symbolic probabilities and symbolic evidential parameters. Then, symbolic propagation of symbolic evidence can only be efficiently performed in cases with a small number of symbolic evidential nodes.

The rest of the paper is organized as follows. Section 2 gives the notation and the basic framework. In Section 3 we present some theoretical results characterizing the algebraic structure of the probabilities when dealing with symbolic parameters. In Section 4 we take advantage of this algebraic structure to introduce the numeric canonical components method for efficient symbolic propagation of uncertainty. The problem of finding lower and upper bounds for the symbolic expressions corresponding to the probabilities of the nodes is analyzed in Section 5. Section 6 introduces symbolic evidence and illustrates how to obtain symbolic expression in this situation using the previous method.

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 $\{0, 1, \dots, r_i\}$. A Bayesian network over X is a pair (G, P) , where the graph G 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 probability distributions (CPD), one for each variable, where Π_i is the set of parents of node X_i in G . Using the chain rule, the joint probability distribution (JPD) 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 CPD 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 *symbolic node*. A convenient choice of the parameters in this case is given by

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

where π 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 $\sum_{j=0}^{r_i} \theta_{ij\pi} = 1$, for all i and π , any one of the parameters can be written as one minus the sum of all others. For example, $\theta_{ir_i\pi}$ is

$$\theta_{ir_i\pi} = 1 - \sum_{j=0}^{r_i-1} \theta_{ij\pi}. \quad (2)$$

To simplify the notation in cases where a variable X_i does not have parents, we use θ_{ij} to denote $P_i(X_i = j)$, $j \in \{0, \dots, r_i\}$.

For illustrative purpose, we shall use the following example.

Example 1 Consider a discrete Bayesian network consisting of the set of variables $X = \{X_1, \dots, X_8\}$ with the relationships represented in the DAG G , given in Figure 1. The structure of G implies that the JPD, $P(x)$, of the set of nodes can be written, in the form of (1), as:

$$P(x) = P(x_1)P(x_2|x_1)P(x_3|x_1)P(x_4|x_2, x_3)P(x_5|x_3)P(x_6|x_4)P(x_7|x_4)P(x_8|x_5). \quad (3)$$

For simplicity, but without loss of generality, we assume that all nodes represent binary variables with values in the set $\{0, 1\}$. This and the structure of the probability distribution in (3) imply that the JPD of the eight variables depends on 34 parameters $\Theta = \{\theta_{ij\pi}\}$. These parameters are given in Table 1. Note, however, that only 4 of the

parameters are given in symbolic form; the rest are assigned fixed numerical values. Therefore, in this example, the only symbolic nodes are X_3 and X_6 . Moreover, only two of these 4 symbolic parameters are free (because the probabilities in each CPD must add up to one).

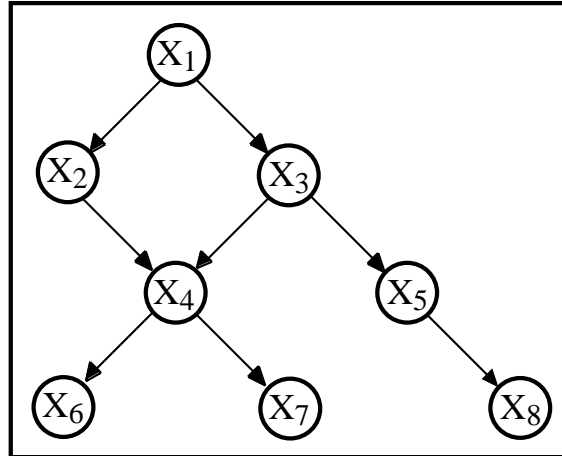


Figure 1: Example of a Bayesian Network.

In this situation, the symbolic propagation algorithms mentioned above allow us to obtain the marginal, or conditional, probabilities of the nodes in the network as functions of the free parameters θ_{300} and θ_{600} . For illustrative purposes Table 2 shows the symbolic expressions corresponding to the initial probabilities and Table 3 shows the conditional probabilities, given the evidence $\{X_2 = 1, X_5 = 1\}$.

Tables 2 and 3 can then be used to answer several queries regarding initial or evidential probabilities associated with the network in Figure 1 simply by plugging in specific values for the parameters. For example, in Figure 2 we show the resulting conditional probabilities for the nodes in the network, given the evidence $\{X_2 = 1, X_5 = 1\}$, for the values of the parameters $\theta_{300} = 0.4$ and $\theta_{600} = 0.8$. These probabilities can be obtained from the information in Table 3 by simple substitution. All this results have been obtained using the symbolic package Mathematica, with the code generated by the program *X-pert SymbolicTM* (Castillo et al. (1996)), an inference symbolic code generator for Mathematica and Maple programs. However, the same results can be carried out using the SPI algorithm (Shachter et al. 1990b).

3 Algebraic Structure of Probabilities

An examination of the probabilities given in Tables 2 and 3 reveals the fact that there is a common structure underlying the symbolic expressions associated with the probabilities of the nodes. In this section, we characterize the algebraic structure of

X	Parameters	
	$X = 0$	$X = 1$
X_1	$\theta_{10} = P(X_1 = 0) = 0.2$	$\theta_{11} = P(X_1 = 1) = 0.8$
X_2	$\theta_{200} = P(X_2 = 0 X_1 = 0) = 0.3$ $\theta_{201} = P(X_2 = 0 X_1 = 1) = 0.5$	$\theta_{210} = P(X_2 = 1 X_1 = 0) = 0.7$ $\theta_{211} = P(X_2 = 1 X_1 = 1) = 0.5$
X_3	$\theta_{300} = P(X_3 = 0 X_1 = 0)$ $\theta_{301} = P(X_3 = 0 X_1 = 1) = 0.5$	$\theta_{310} = P(X_3 = 1 X_1 = 0)$ $\theta_{311} = P(X_3 = 1 X_1 = 1) = 0.5$
X_3	$\theta_{4000} = P(X_4 = 0 X_2 = 0, X_3 = 0) = 0.1$ $\theta_{4001} = P(X_4 = 0 X_2 = 0, X_3 = 1) = 0.3$ $\theta_{4010} = P(X_4 = 0 X_2 = 1, X_3 = 0) = 0.8$ $\theta_{4011} = P(X_4 = 0 X_2 = 1, X_3 = 1) = 0.4$	$\theta_{4100} = P(X_4 = 1 X_2 = 0, X_3 = 0) = 0.9$ $\theta_{4101} = P(X_4 = 1 X_2 = 0, X_3 = 1) = 0.7$ $\theta_{4110} = P(X_4 = 1 X_2 = 1, X_3 = 0) = 0.2$ $\theta_{4111} = P(X_4 = 1 X_2 = 1, X_3 = 1) = 0.6$
X_5	$\theta_{500} = P(X_5 = 0 X_3 = 0) = 0.3$ $\theta_{501} = P(X_5 = 0 X_3 = 1) = 0.1$	$\theta_{510} = P(X_5 = 1 X_3 = 0) = 0.7$ $\theta_{511} = P(X_5 = 1 X_3 = 1) = 0.9$
X_6	$\theta_{600} = P(X_6 = 0 X_4 = 0)$ $\theta_{601} = P(X_6 = 0 X_4 = 1) = 0.9$	$\theta_{610} = P(X_6 = 1 X_4 = 0)$ $\theta_{611} = P(X_6 = 1 X_4 = 1) = 0.1$
X_7	$\theta_{700} = P(X_7 = 0 X_4 = 0) = 0.3$ $\theta_{701} = P(X_7 = 0 X_4 = 1) = 0.6$	$\theta_{710} = P(X_7 = 1 X_4 = 0) = 0.7$ $\theta_{711} = P(X_7 = 1 X_4 = 1) = 0.4$
X_8	$\theta_{800} = P(X_8 = 0 X_5 = 0) = 0.2$ $\theta_{801} = P(X_8 = 0 X_5 = 1) = 0.4$	$\theta_{810} = P(X_8 = 1 X_5 = 0) = 0.8$ $\theta_{811} = P(X_8 = 1 X_5 = 1) = 0.6$

Table 1: The CPD required to define the JPD of X_1, \dots, X_8 with four symbolic parameters θ_{300} , θ_{310} , θ_{600} , and θ_{610} .

X_i	$P(X_i = 0)$
X_1	0.2
X_2	0.46
X_3	$0.4 + 0.2 \theta_{300}$
X_4	$0.424 - 0.056 \theta_{300}$
X_5	$0.18 + 0.04 \theta_{300}$
X_6	$0.5184 + 0.0504 \theta_{300} + 0.424 \theta_{600} - 0.056 \theta_{300} \theta_{600}$
X_7	$0.4728 + 0.0168 \theta_{300}$
X_8	$0.364 - 0.008 \theta_{300}$

Table 2: Initial marginal probability distributions of the nodes.

X_i	$P(X_i = 0 X_2 = 1, X_5 = 1)$
X_1	$(0.126 - 0.028 \theta_{300}) / (0.446 - 0.028 \theta_{300})$
X_2	0
X_3	$(0.140 + 0.098 \theta_{300}) / (0.446 - 0.028 \theta_{300})$
X_4	$(0.164 - 0.021 \theta_{300}) / (0.446 - 0.028 \theta_{300})$
X_5	0
X_6	$(0.253 - 0.006 \theta_{300} + 0.164 \theta_{600} - 0.021 \theta_{300} \theta_{600}) / (0.446 - 0.028 \theta_{300})$
X_7	$(0.218 - 0.010 \theta_{300}) / (0.446 - 0.028 \theta_{300})$
X_8	$(0.178 - 0.011 \theta_{300}) / (0.446 - 0.028 \theta_{300})$

Table 3: Conditional probabilities of the nodes, given $e = \{X_2 = 1, X_5 = 1\}$.

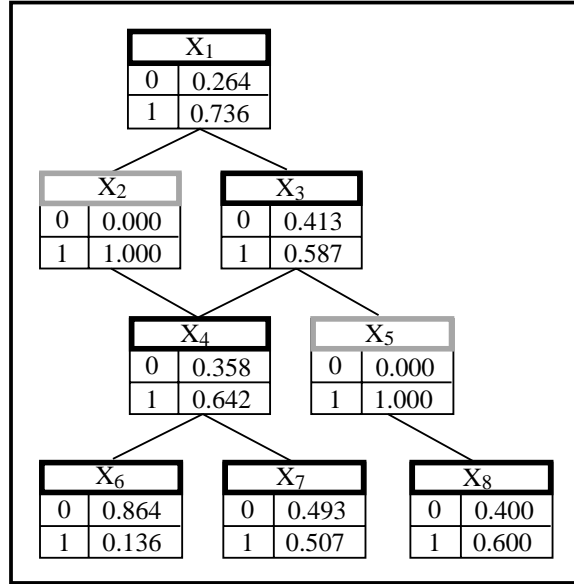


Figure 2: Marginal posterior probability distributions of the nodes given the evidence $e = \{X_2 = 1, X_5 = 1\}$ and the symbolic parameter's values $\theta_{300} = 0.4, \theta_{600} = 0.8$. Light rectangles show the evidence nodes.

these expressions as functions of the symbolic parameters. We start with the case of marginal probabilities and later we analyze the case of conditional probabilities.

Theorem 1 *The prior probability of any instantiation, (x_1, \dots, x_n) , of the nodes in the network is a polynomial in the symbolic parameters of degree less than or equal to the number of symbolic nodes. However, it is a first degree polynomial in each parameter.*

Proof: According to (1) the probability of an instantiation (x_1, \dots, x_n) is

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | \pi_i) = \prod_{i=1}^n \theta_{ix_i \pi_i}.$$

Note that all the parameters appearing in the above product are associated with different variables, and some of them may be specified numerically. Thus, $P(x_1, \dots, x_n)$ is a monomial of degree less than or equal the number of symbolic nodes. Note that $P(x_1, \dots, x_n)$ may become a polynomial when considering only the set of free parameters (see (2)). This simply requires replacing the parameters $\theta_{ix_i \pi_i}$ by

$$1 - \prod_{j=0}^{r_i-1} \theta_{ij \pi_i}.$$

Therefore, we create as many different monomials as the cardinality of X_i , but each of the resulting monomials is still first degree in each parameter. ■

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

Proof: For simplicity, assume $Y = \{X_1, \dots, X_r\}$. Then $P(y)$ is the sum of the probabilities of a subset of instantiations:

$$P(y) = P(x_1, \dots, x_r) = \sum_{x_{r+1}, \dots, x_n} P(x_1, \dots, x_r, x_{r+1}, \dots, x_n) = \sum_{x_{r+1}, \dots, x_n} \prod_{i=1}^n \theta_{ix_i \pi_i}$$

Therefore, the prior marginals of any node are also polynomials of the symbolic parameters of first degree in each parameter. ■

For example, as can be seen in Table 2, the prior marginal probability of all nodes in the Bayesian network of Example 1 are polynomials of first degree in each of the free symbolic parameters θ_{300} and θ_{600} .

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

Proof: We have:

$$P(y|e) = \frac{P(y, e)}{P(e)}.$$

Using Corollary 1, both the numerator and the denominator are first degree polynomials on the symbolic parameters, since they are marginal probabilities of a subset of nodes of the network. ■

Note that, in the above expressions, the denominator polynomial is the same for any conditional probability $P(y|e)$, for a given evidence set $E = e$. Then, in practical situations and for implementation purposes, it is more convenient to calculate and store only the numerator polynomials for each node and calculate the common denominator polynomial by normalization.

For example, the probabilities in Table 3 show that the posterior distribution of the nodes, given the evidence $\{X_2 = 1, X_5 = 1\}$, is a ratio of two polynomials, and that the denominator polynomial is the same for all nodes.

Once we know the structure of the marginal and conditional probabilities we can exploit it to obtain symbolic results in an efficient way.

4 Efficient Symbolic Inference

Suppose that we are dealing with a set of symbolic nodes $\{X_{i_1}, \dots, X_{i_s}\} \subset X$. Let $\Theta = \{\Theta_1, \dots, \Theta_s\}$ be the set of symbolic parameters included in the network, where Θ_k stands for the symbolic parameters associated with the symbolic node X_{i_k} , with $k = 1, \dots, s$. Corollaries 1 and 2 guarantee that the conditional probabilities of a typical node X_i , given some evidence $E = e$, $P(X_i = j|E = e), j = 0, \dots, r_i$, is either a polynomial or a ratio of two polynomials. Theorem 1 guarantees that each monomial forming these polynomials contains at most one parameter of Θ_k , for each $k = 1, \dots, s$. Therefore, we build the set of feasible monomials, M , by taking the cartesian product of the sets of symbols corresponding to the different symbolic nodes forming Θ . Then, we have

$$M = \Theta_1 \times \dots \times \Theta_s. \quad (4)$$

Thus, the general form of these polynomials is:

$$\sum_{m_r \in M} c_r m_r, \quad (5)$$

where c_r is the numerical coefficient associated with the monomial $m_r \in M$.

In this section, we develop a new method, which we refer to as the *numeric canonical components*, for computing the coefficients c_r by instantiating the symbolic parameters, Θ , and computing the resulting numerical probabilities. Then, once the coefficients have been obtained, the polynomials and, hence, the probabilities $P(x_i|e)$ can be easily obtained.

We will also show that there are some analogies between the canonical components symbolic method and the conditioning algorithms used to propagate uncertainty in multiply-connected networks (see, for example, Pearl (1988)). Both methods perform several numerical propagations associated with different instantiations of certain parameters in the network to achieve the desired solution.

4.1 Numeric Canonical Components Method

Let M be the set of monomials needed to compute $P(X_i = j|E = e)$ for $j = 0, \dots, r_i$. Let m be the number of monomials in M . From (5), the polynomial needed to compute $P(X_i = j|E = e)$ is of the form:

$$P(X_i = j|E = e) \propto \sum_{m_k \in M} c_k^{ij} m_k = p_{ij}(\Theta), \quad j = 0, \dots, r_i. \quad (6)$$

The term $p_{ij}(\Theta)$ represents the unnormalized probability $P(X_i = j|E = e)$. Thus, $p_{ij}(\Theta)$ can be written as a linear combination of the monomials in M . Our objective now is to compute the coefficients c_k^{ij} .

If the parameters Θ are assigned numerical values, say θ , then $p_{ij}(\theta)$ can be obtained by replacing Θ by θ and using any numeric propagation method to compute $P(X_i = j|E = e, \Theta = \theta)$. Similarly, the monomials m_k takes a numerical value, the product of the parameters involved in m_k . Thus, we have

$$\sum_{m_k \in M} c_k^{ij} m_k = p_{ij}(\theta). \quad (7)$$

Note that in (7) all the monomials m_k , and the unnormalized probability $p_{ij}(\theta)$ are known numbers, and the only unknowns are the coefficients c_k^{ij} , $k = 1, \dots, m$. To compute these coefficients, we need to construct any set of m independent equations each is of the form (7). These equations can be obtained using m sets of distinct instantiations Θ . Let these values be denoted by $C = \{\theta_1, \dots, \theta_m\}$. We refer to the elements in C as the *canonical components*. Let \mathbf{T}_{ij} be the $m \times m$ non-singular matrix, whose ik -th element is the value of the monomial m_k obtained by replacing Θ by θ_i , the i -th instantiation of Θ . We refer to the matrix \mathbf{T}_{ij} as the *canonical matrix* associated with the set of canonical components C . Let

$$\mathbf{c}^{ij} = \begin{pmatrix} c_1^{ij} \\ \vdots \\ c_m^{ij} \end{pmatrix}, \quad \text{and} \quad \mathbf{p}_{ij} = \begin{pmatrix} p_{ij}(\theta_1) \\ \vdots \\ p_{ij}(\theta_m) \end{pmatrix}. \quad (8)$$

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

$$\mathbf{T}_{ij} \mathbf{c}_{ij} = \mathbf{p}_{ij}, \quad (9)$$

which implies that the vector of coefficients \mathbf{c}_{ij} is given by

$$\mathbf{c}_{ij} = \mathbf{T}_{ij}^{-1} \mathbf{p}_{ij}. \quad (10)$$

The values of the coefficients in \mathbf{c}_{ij} can then be substituted in (6) and the unnormalized probability $p_{ij}(\Theta)$ is expressed as a function of Θ .

Therefore, Equations (6), (8) and (9) provide an efficient algorithm for symbolic propagation that does not require any symbolic computation. We refer to this algorithm as the *numeric canonical components method* (NCCM). This algorithm is summarized as follows:

Algorithm 1 (Numeric Canonical Components):

1. Construct m sets of instantiations of Θ : $\theta_1, \dots, \theta_m$, providing m independent linear equations in \mathbf{c}_{ij} , when substituting the values of θ_i in (7).
2. Calculate the $m \times m$ non-singular matrix \mathbf{T}_{ij} whose ik -th element is the value of the monomial m_k obtained by replacing Θ by θ_i , the i -th instantiation of Θ .
3. Compute the vector of probabilities \mathbf{p}_{ij} in (8) using any standard numerical propagation method.
4. Solve the linear system of equations (9) to obtain the desired coefficients \mathbf{c}_{ij} .
5. Substitute the obtained values of \mathbf{c}_{ij} in (6) and normalize to obtain the symbolic expression for the probabilities $P(X_i = j | E = e)$.

Note that, Step 3 of Algorithm 1 requires the use of a numeric propagation method to propagate uncertainty as many times as the number of possible combinations of the symbolic parameters. This means that the number of numerical propagations increases combinatorially with the number of symbolic parameters. This problem is also present in other propagation algorithms. For example, conditioning algorithms suffer from this problem with respect the number of nodes in the cutset. Therefore, the role of symbolic nodes in the canonical components symbolic method is similar to the role of conditioning nodes in the conditioning algorithms.

Algorithm 1 requires calculating and solving a linear system of equations. In the following, we show that by imposing certain conditions in the symbolic parameters, it is always possible to find a set of canonical components whose corresponding \mathbf{T}_{ij} matrix is the identity matrix. Thus, the symbolic expressions associated with the probabilities can be obtained directly without the need for solving the linear system of equations (9).

Consider a typical symbolic node X_i with the associated parameters $\theta_{ij\pi}$. Some of these parameters may be specified numerically, and some may be given in symbolic form. Assume that a subset of the parameters $\theta_{ij\pi}$ is given in symbolic form for a given

instantiation π of Π_i , and the parameters for all other instantiations are numeric. For example, the Bayesian network of Example 1 satisfies this assumption because the symbolic parameters corresponding to each of the symbolic nodes are associated with the same instantiation of the set of parents (see Table 2). For example, X_1 is the only parent of the symbolic node X_3 , and both the symbolic parameters θ_{300} and θ_{310} are associated with the same instantiation of X_1 : $X_1 = 0$.

In this situation, the canonical components resulting from considering extreme values for the symbolic parameters produce an identity canonical matrix \mathbf{T}_{ij} . The next theorem states this fact.

Theorem 2 *Given a set of symbolic nodes $\{X_{i_1}, \dots, X_{i_s}\}$ with associated symbolic parameters $\Theta = \{\Theta_1, \dots, \Theta_s\}$, where $\Theta_k = \{\theta_{i_k j \pi_k}, j = 0, \dots, r_{i_k}\}$, and π_k is a given instantiation of Π_{i_k} , then, the canonical matrix associated with the set of canonical components C defined by the cartesian product $C = C_1 \times \dots \times C_n$, where*

$$C_k = \{\{\theta_{i_k 0 \pi_k} = 1, \theta_{i_k 1 \pi_k} = 0, \dots, \theta_{i_k r_{i_k} \pi_k} = 0\}, \dots, \{\theta_{i_k 0 \pi_k} = 0, \theta_{i_k 1 \pi_k} = 0, \dots, \theta_{i_k r_{i_k} \pi_k} = 1\}\}$$

is the $m \times m$ identity matrix.

Proof: From (4), the set of monomials m_k is given by: $M = \Theta_1 \times \dots \times \Theta_s$. In this case, using the assumption $\Theta_k = \{\theta_{i_k j \pi_k}, j = 0, \dots, r_{i_k}\}$, we have:

$$M = \{\{\theta_{i_1 0 \pi_1}, \dots, \theta_{i_1 r_{i_1} \pi_1}\} \times \dots \times \{\theta_{i_s 0 \pi_s}, \dots, \theta_{i_s r_{i_s} \pi_s}\}\}.$$

Therefore, any instantiation of the symbolic parameters in C annihilates all the monomials in M but one. Let $\theta \in C$ be a typical canonical component. Then, all the parameters in Θ_k are zero, but one, $\theta_{i_k j_k \pi_k}$, for $k = 1, \dots, m$. Thus, substituting this numerical values in the monomials in M , the monomial $\theta_{i_1 j_1 \pi_1} \dots \theta_{i_s j_s \pi_s}$ takes the value 1; the rest of monomials vanish. Therefore, every row in the matrix \mathbf{T}_{ij} contains one element with the value 1 and all the others are 0. Finally, the process of the construction of M and C guarantees that the matrix \mathbf{T}_{ij} is the identity matrix of order $m \times m$. ■

Then, the solution \mathbf{c}_{ij} of the system of linear equations (9) becomes:

$$\mathbf{c}_{ij} = \mathbf{p}_{ij}. \quad (11)$$

Therefore, in this situation, the Algorithm 1 can be simplified as follows:

Algorithm 2 (Modified Numeric Canonical Components):

1. Construct m sets of instantiations of Θ in the form indicated in Theorem 2, $C = \{\theta_1, \dots, \theta_m\}$.
2. Compute the vector of probabilities $\mathbf{p}_{ij} = (p_{ij}(\theta_1), \dots, p_{ij}(\theta_m))$, using any standard propagation numeric method.

3. Set $\mathbf{c}_{ij} = \mathbf{p}_{ij}$, substitute \mathbf{c}_{ij} in (6), and normalize to obtain the symbolic expression for the probabilities $P(X_i = j|E = e)$.

Note that with the assumption considered in Theorem 2, the canonical components allow us to perform symbolic propagation in a straightforward and efficient way. If the assumption given in Theorem 2 is not satisfied, then the resulting canonical matrix associated with the canonical set of parameters C may be different from the identity matrix, and it will be necessary to solve the system of equations (10) to obtain the polynomial coefficients.

The performance of this algorithm is illustrated in the following section with an example.

4.2 Numeric Canonical Components by Example

Consider the network in Figure 1 and suppose that we are given the evidence $e = \{X_2 = 1, X_5 = 1\}$. We wish to assess the influence of the symbolic parameters on the conditional probabilities of the remaining nodes. In this example the set of symbolic nodes is $\{X_3, X_6\}$ and the set of parameters is $\Theta = \{\Theta_3, \Theta_6\} = \{\{\theta_{300}, \theta_{310}\}, \{\theta_{600}, \theta_{610}\}\}$ (see Table 1). Then, the set of feasible monomials is given by

$$\begin{aligned} M &= \Theta_3 \times \Theta_6 \\ &= \{\theta_{300}\theta_{600}, \theta_{300}\theta_{610}, \theta_{310}\theta_{600}, \theta_{310}\theta_{610}\} \\ &= \{m_1, m_2, m_3, m_4\}. \end{aligned}$$

Then, for any node X_i in the network, we know that the unnormalized conditional probability $P(X_j = j|e)$ is a polynomial function of the form

$$P(X_j = j|e) \propto \sum_{k=1}^4 c_k^{ij} m_k = p_{ij}(\Theta). \quad (12)$$

Thus, our aim consists of obtaining the coefficients $\{c_k^{ij} ; k = 1, \dots, 4\}$ for each node X_i and each possible value j . To this aim, we consider the canonical components associated with the symbolic set of parameters Θ . In this case, given that we are dealing with binary variables, there are only two possible canonical combinations of the parameters in Θ_i , $\{1, 0\}$ and $\{0, 1\}$. Then, we have the following set of canonical components:

$$\begin{aligned} C &= \{\{1, 0\}, \{0, 1\}\} \times \{\{1, 0\}, \{0, 1\}\} \\ &= \{\{1, 0; 1, 0\}, \{1, 0; 0, 1\}, \{0, 1; 1, 0\}, \{0, 1; 0, 1\}\} \\ &= \{c_1, c_2, c_3, c_4\}. \end{aligned}$$

Then, by instantiating the symbolic parameters to the corresponding values given in its canonical components all the monomials appearing in (12) become either 0 or 1. Then, we obtain an expression that only depends on the coefficients c^{ij} :

$$\begin{aligned} p_{ij}(\Theta = c_1) &= c_1^{ij}, & p_{ij}(\Theta = c_2) &= c_2^{ij}, \\ p_{ij}(\Theta = c_3) &= c_3^{ij}, & p_{ij}(\Theta = c_4) &= c_4^{ij}. \end{aligned}$$

Thus, in this case, the matrix \mathbf{T}_{ij} is the identity matrix because all the symbolic parameters of the symbolic nodes are associated with the same instantiation of the set of parents. Then, we have:

$$\begin{pmatrix} c_1^{ij} \\ c_2^{ij} \\ c_3^{ij} \\ c_4^{ij} \end{pmatrix} = \begin{pmatrix} p_{ij}(c_1) \\ p_{ij}(c_2) \\ p_{ij}(c_3) \\ p_{ij}(c_4) \end{pmatrix}. \quad (13)$$

It is interesting to point out here that the feasible set (the set generated by all feasible parameter values) for the probabilities of any set of nodes is the convex hull generated by the canonical probabilities.

In Figure 3 we show the unnormalized conditional probabilities $p_{ij}(c_k)$ of all nodes, given the evidence $e = \{X_2 = 1, X_5 = 1\}$, associated with the four possible canonical components. Using these values we can obtain all the rational functions in Table 3. For instance, from Figure 3 we get the following values for the node X_6 :

$$\begin{pmatrix} c_1^{60} \\ c_2^{60} \\ c_3^{60} \\ c_4^{60} \end{pmatrix} = \begin{pmatrix} 0.390 \\ 0.247 \\ 0.418 \\ 0.254 \end{pmatrix}, \quad (14)$$

and

$$\begin{pmatrix} c_1^{61} \\ c_2^{61} \\ c_3^{61} \\ c_4^{61} \end{pmatrix} = \begin{pmatrix} 0.028 \\ 0.171 \\ 0.029 \\ 0.193 \end{pmatrix}, \quad (15)$$

that is, the coefficients of the numerator polynomials for $X_6 = 0$ and $X_6 = 1$, respectively. Then, substituting these values in (12) we obtain:

$$\begin{aligned} P(X_6 = 0|e) &\propto 0.390 \theta_{300} \theta_{600} + 0.247 \theta_{300} \theta_{610} + 0.418 \theta_{310} \theta_{600} + 0.254 \theta_{310} \theta_{610}. \\ P(X_6 = 1|e) &\propto 0.028 \theta_{300} \theta_{600} + 0.171 \theta_{300} \theta_{610} + 0.029 \theta_{310} \theta_{600} + 0.193 \theta_{310} \theta_{610}. \end{aligned}$$

Adding both polynomials we obtain the denominator normalizing polynomial, that is,

$$\begin{pmatrix} c_1^{60} \\ c_2^{60} \\ c_3^{60} \\ c_4^{60} \end{pmatrix} + \begin{pmatrix} c_1^{61} \\ c_2^{61} \\ c_3^{61} \\ c_4^{61} \end{pmatrix} = \begin{pmatrix} 0.418 \\ 0.418 \\ 0.447 \\ 0.447 \end{pmatrix}.$$

Thus, we have:

$$\begin{aligned} P(X_6 = 0|e) &= (0.390 \theta_{300} \theta_{600} + 0.247 \theta_{300} \theta_{610} + 0.418 \theta_{310} \theta_{600} + 0.254 \theta_{310} \theta_{610})/d, \\ P(X_6 = 1|e) &= (0.028 \theta_{300} \theta_{600} + 0.171 \theta_{300} \theta_{610} + 0.029 \theta_{310} \theta_{600} + 0.193 \theta_{310} \theta_{610})/d, \end{aligned}$$

where $d = 0.418 \theta_{300} \theta_{600} + 0.418 \theta_{300} \theta_{610} + 0.447 \theta_{310} \theta_{600} + 0.447 \theta_{310} \theta_{610}$.

Finally, eliminating the dependent parameters we get the expression shown in Table 3. Note that the only symbolic operation realized in this process is simplifying the final expression to eliminate the dependent parameters. However, this is an optional operation and, in some cases, it is more convenient to keep the expression with all the parameters, and simplify the numerical results after plugging in some specific values for the parameters.

We should mention here that, although we are using only exact propagation methods for illustrative purposes, the methodology continues to work when using approximate or simulation propagation methods, such as those described in the introduction, or even combine exact and approximate methods.

4.3 Efficient Computation of Canonical Components

The proposed symbolic inference method requires several applications of a numeric approximate or exact propagation algorithm to calculate the numerical probabilities associated with each one of the canonical components. Therefore, the number of propagations increases combinatorially with the number of symbolic parameters. However, when propagating uncertainty in the canonical cases using some of the message-passing algorithms (see, e.g., Kim and Pearl (1983), Jensen et al. (1990) or Shachter et al. (1994)), we can save many calculations because some messages are common to several canonical components. In this section, we illustrate this fact in two of the main message passing algorithms: the algorithm for polytrees (see, for example, Kim and Pearl (1983), and Pearl (1988)), and the clustering algorithm (Shachter et al. (1994)).

Figure 4 illustrates the message-passing process corresponding to a typical node X_i in a polytree when applying Kim-Pearl algorithm. In this figure, Θ^k stands for the set of parameters contained in the connected component associated with node X_k when dropping the link $X_k - X_i$, while Θ_k is the set of parameters associated with node X_k . Note that the message from node X_k to node X_i depends only on those parameters,

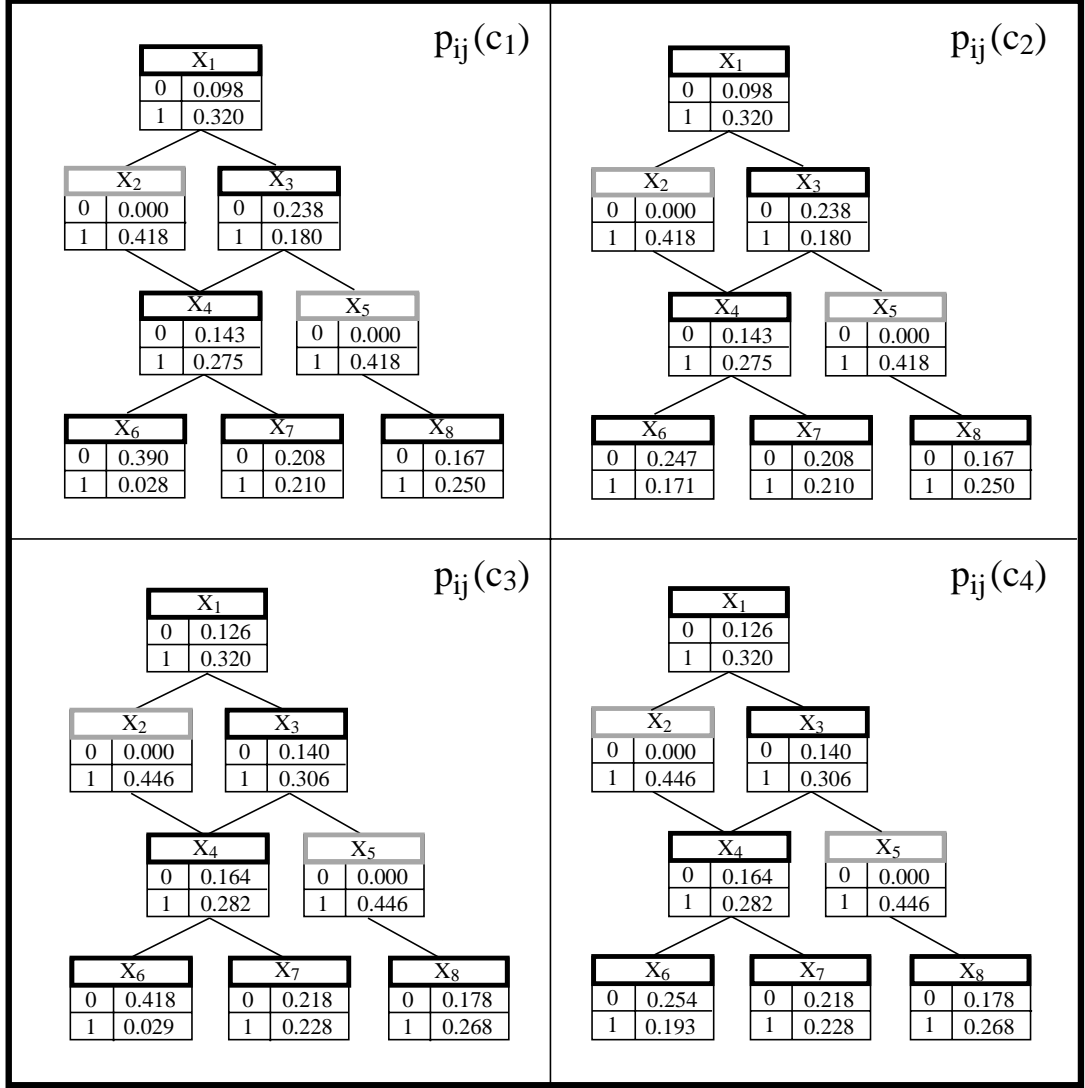


Figure 3: The four elemental canonical cases. The first column in the table for each node X_i is the state of X_i and the second column is the unnormalized marginal probabilities of X_i , given the evidence $e = \{X_2 = 1, X_5 = 1\}$. Light rectangles show the evidence nodes.

while the message passed from X_i to X_k depends on the remaining parameters in the Bayesian network. Note also that if Θ^k does not contain any symbolic parameter, then all the messages coming from this region of the graph to X_i need to be calculated only once, because they have the same value for every canonical component.

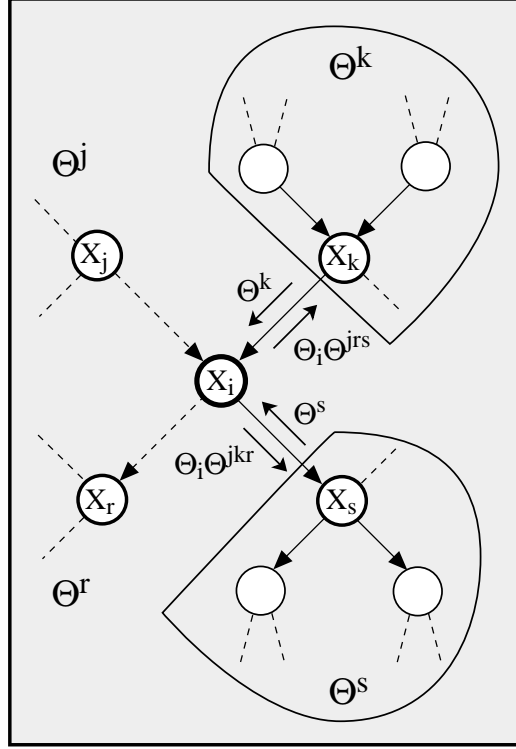


Figure 4: Parametric dependence of the messages in a general node in a polytree.

The same computational savings are obtained when applying clustering message-passing algorithms. In this case, the situation is the same except that we deal with a tree of clusters (sets of nodes) instead of a tree of nodes.

For instance, suppose we are given the multiply-connected graph in Figure 1 with the numeric and symbolic probabilities in Table 1. Figure 5 shows all the messages needed to propagate evidence in a family tree associated with the Bayesian network of Figure 1, using this method. In Figure 5, the cluster messages are indicated by arrows. We can distinguish two types of messages:

1. Messages with no index. These are common messages for all the canonical components. Thus need to be calculated only once.
2. Messages with one or more indices such as Θ_3 , Θ_6 , or $\Theta_3\Theta_6$. These messages depend on these parameters and then we must calculate as many different messages as the number of monomials associated with them.

Thus, in this example we can build the rational function associated with the node marginals by performing only half of the computations.

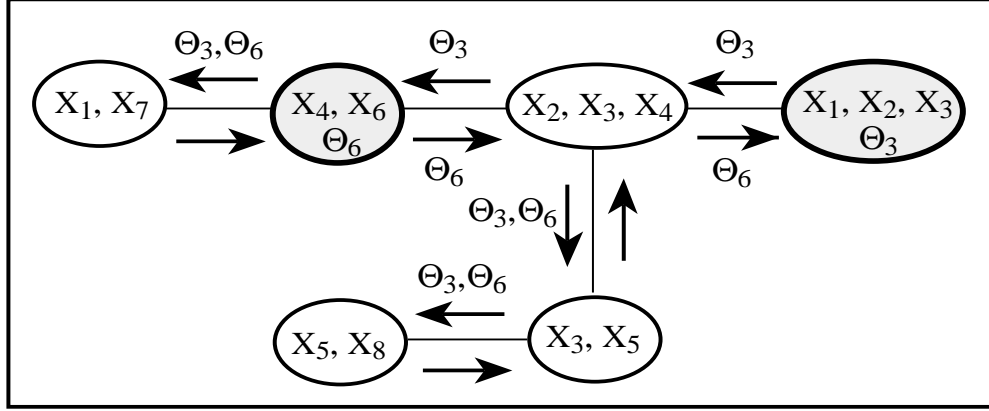


Figure 5: Family tree and messages affected by the relevant parameters. The clusters involving potential functions depending on parameters Θ_3 or Θ_6 are outlined.

5 Upper and Lower Bounds for Probabilities

The symbolic expressions obtained for the probabilities, such as those in Tables 2 and 3, can also be used to obtain upper and lower bounds for the marginal probabilities which is a useful information for performing sensitivity analysis (Castillo, Gutiérrez, and Hadi (1995c)). In this section we show that the bounds of the obtained ratios of polynomials are attained at one of the canonical components (vertices of the feasible convex parameter set). Thus, it is very easy to obtain the lower and upper bounds of the expression associated with the probabilities when using the canonical components method.

We use the following theorem given by Bela-Martos (1964).

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

$$\frac{\mathbf{c} * \mathbf{u} - c_0}{\mathbf{d} * \mathbf{u} - d_0}, \quad (16)$$

where \mathbf{c} and \mathbf{d} are vector coefficients and c_0 and d_0 are real constants, is defined in the convex polyhedral $A\mathbf{u} \leq a_0, \mathbf{u} \geq 0$, where A is a constant matrix and a_0 is a constant vector, and the denominator in (16) does not vanish in the polyhedral, then the functional reaches the maximum at least in one of the vertices of the polyhedron.

In our case, \mathbf{u} is the set of symbolic parameters and the fractional functions (16) are the symbolic expressions associated with the probabilities. In this case, the

		$P(X_i = x_i)$			$P(X_i = x_i X_2 = 1, X_5 = 1)$		
X_i	x_i	Lower	Upper	Range	Lower	Upper	Range
X_1	0	0.200	0.200	0.000	0.234	0.282	0.048
	1	0.800	0.800	0.000	0.718	0.766	0.048
X_2	0	0.460	0.460	0.000	0.000	0.000	0.000
	1	0.540	0.540	0.000	1.000	1.000	0.000
X_3	0	0.400	0.600	0.200	0.315	0.569	0.255
	1	0.400	0.600	0.200	0.431	0.685	0.255
X_4	0	0.368	0.424	0.056	0.343	0.369	0.026
	1	0.576	0.632	0.056	0.631	0.657	0.026
X_5	0	0.180	0.220	0.040	0.000	0.000	0.000
	1	0.780	0.820	0.040	1.000	1.000	0.000
X_6	0	0.518	0.942	0.424	0.568	0.936	0.369
	1	0.058	0.482	0.424	0.064	0.432	0.369
X_7	0	0.473	0.490	0.017	0.489	0.497	0.008
	1	0.510	0.527	0.017	0.503	0.511	0.008
X_7	0	0.356	0.364	0.008	0.400	0.400	0.000
	1	0.636	0.644	0.008	0.600	0.600	0.000

Table 4: Lower and upper bounds for the initial marginal probabilities $P(X_i = x_i)$ (no evidence), and the conditional probabilities $P(X_i = x_i | X_2 = 1, X_5 = 1)$.

convex polyhedral is defined by $\mathbf{u} \leq 1, \mathbf{u} \geq 0$, that is, A is the identity matrix and $a_0 = 1$. Then, using Theorem 3, we know that the lower and upper bounds of the symbolic expressions associated with the probabilities are attained at the vertices of this polyhedron, that is, at some of the canonical components associated with the symbolic set of parameters.

As an example, Figure 3 shows the unnormalized conditional probabilities for the nodes. For the event $X_6 = 0$, normalizing the conditional probabilities corresponding to the four canonical cases, we obtain the set of probabilities $P(X_6 = 0 | X_2 = 1, X_5 = 1)$ in the four canonical cases: $\{0.936, 0.568, 0.934, 0.591\}$. Thus, the maximum possible value for the probability of this event is 0.936 and the minimum is 0.568. Table 4 shows the lower bound, the upper bound, and the range (the difference between lower and upper bounds) for all the probabilities of the nodes in the Bayesian network given in Example 1, in two different cases: no evidence, and evidence $e = \{X_2 = 1, X_5 = 1\}$.

6 Symbolic Treatment of Random Evidence

In the previous sections, we have dealt with deterministic evidence, i.e. the calculation of some conditional probability $P(X_i|E = e)$, where the evidence nodes in the set $E = \{E_1, \dots, E_m\}$ take known values $e = \{E_1 = e_1, \dots, E_m = e_m\}$. In this section we deal with symbolic evidence. Suppose that a probability $q(e_1, \dots, e_m)$ over E , with $\sum_{e_1, \dots, e_m} q(e_1, \dots, e_m) = 1$, is given. Then the conditional probability with this random evidence is

$$P_q(X_i|E = e) = P(X_i|(E_1, \dots, E_m) = (e_1, \dots, e_m) \text{ with probability } q(e_1, \dots, e_m)).$$

In this case, the conditional probability becomes

$$P_q(X_i|E = e) = \sum_{e_1, \dots, e_m} q(e_1, \dots, e_m) A_{e_1, \dots, e_m}(X_i), \quad (17)$$

where $A_{e_1, \dots, e_m}(x_i) = P(X_i = x_i|(E_1, \dots, E_m) = (e_1, \dots, e_m))$. Note that this expression can be obtained by applying the symbolic canonical components methods for each combination of values of the evidence variables.

It is important to note that $P_q(x_i|E = e)$ is also a rational function because it is a linear convex combination of the rational functions $A_{e_1, \dots, e_m}(x_i)$. However, in this case the parameters of the conditional probability distributions can appear with exponents larger than one, which implies polynomial of order larger than one in each of the parameters. This result is stated in the following theorem.

Theorem 4 *The probabilities of the nodes given a random evidence are rational functions where the degree in each parameter of the polynomials involved is at most equal to the sum of the cardinalities of the random evidential nodes.*

Proof: The polynomial denominators of the rational functions $Q_{e_1, \dots, e_k}(x_i)$ are in general different for different combinations of the evidence set (e_1, \dots, e_k) . Thus, the common denominator is the product of different rational functions. The number of these rational functions, and hence the degree of the polynomial, cannot exceed the sum of the cardinalities of the evidential nodes. ■

Example 2 Suppose that we know the deterministic evidence $X_2 = 1$, and we consider the event $X_5 = x_5$ in symbolic form, that is, the event $X_5 = 0$ is assigned a probability p and the event $X_5 = 1$ is assigned a probability $1 - p$. With this information, we wish to compute $P(X_6 = x_6|X_2 = 1, X_5 = x_5)$. In this case we have:

$$\begin{aligned} P(X_6 = 0|X_2 = 1, X_5 = 0) &= \frac{0.056 + 0.019 \theta_{300} + 0.032 \theta_{600} + 0.007 \theta_{300} \theta_{600}}{0.094 + 0.028 \theta_{300}} \\ &= s_0, \end{aligned} \quad (18)$$

and the probability of the event $X_6 = 0$ when the evidence is $X_5 = 1$ is

$$\begin{aligned} P(X_6 = 0 | X_2 = 1, X_5 = 1) &= \frac{0.253 - 0.006 \theta_{300} + 0.164 \theta_{600} - 0.021 \theta_{300} \theta_{600}}{0.446 - 0.028 \theta_{300}} \\ &= s_1. \end{aligned} \quad (19)$$

Thus, the probability of the event $X_6 = 0$ in this case is a linear convex combination of s_0 and s_1 which is equal to: $ps_0 + (1 - p)s_1 = a/b$, where

$$\begin{aligned} a &= -30.334 - 1.522 p - 8.316 \theta_{300} - 0.492 p \theta_{300} + 0.214 \theta_{300}^2 \\ &\quad + 0.464 p \theta_{300}^2 - 19.663 \theta_{600} + 1.459 p \theta_{600} \\ &\quad - 3.339 \theta_{300} \theta_{600} + 0.5 p \theta_{300} \theta_{600} + 0.75 \theta_{300}^2 \theta_{600} - 0.5 p \theta_{300}^2 \theta_{600}, \end{aligned}$$

and

$$b = -53.474 - 12.571 \theta_{300} + \theta_{300}^2.$$

Note that the polynomials involved are second degree in θ_{300} as would be expected by Theorem 4.

Note that the symbolic expressions (18) and (19) can be obtained using the symbolic canonical components method by considering the evidences $X_5 = 0$ and $X_5 = 1$, respectively. Note also that in cases where the linear convex combinations are complicated expressions, they may be kept in the expanded form and the simplification process may be realized after considering specific numerical values.

Acknowledgments

The authors are grateful to the University of Cantabria, the Dirección General de Investigación Científica y Técnica (DGICYT) (Project PB94-1056) and Iberdrola for partial support of this work.

References

- Bouckaert, R. R., Castillo, E. and Gutiérrez, J. M. (1995), "A Modified Simulation Scheme for Inference in Bayesian Networks," *International Journal of Approximate Reasoning*, (in press).
- Castillo, E., Gutiérrez, J. M., and Hadi, A. S. (1995a), "Parametric Structure of Probabilities in Bayesian Networks," in *Lectures Notes in Artificial Intelligence*, Springer-Verlag, 946, 89–98.
- Castillo, E., Gutiérrez, J. M., and Hadi, A. S. (1995b), "Symbolic Propagation in Discrete and Continuous Bayesian Networks," in *Mathematics with Vision: Proceedings of the First International Mathematica Symposium*, (V. Keranen and P. Mitic, eds.), Computational Mechanics Publications, 77–84.

- Castillo, E., Gutiérrez, J. M., and Hadi, A. S. (1995c), "Sensitivity Analysis in Discrete Bayesian Networks," Technical Report 95-10, Statistics Center, Cornell University.
- Castillo, E., Gutiérrez, J. M., and Hadi, A. S. (1996), *Expert Systems and Probabilistic Network Models*, Springer-Verlag, New York.
- Fung, R. and Chang, K-C. (1990), "Weighing and Integrating Evidence for Stochastic Simulation in Bayesian Networks," in *Uncertainty in Artificial Intelligence 5*, Machine Intelligence and Pattern Recognition Series, 10, (Henrion et al. Eds.), North Holland, Amsterdam, 209–219.
- Henrion, M. (1988), "Propagating Uncertainty in Bayesian Networks by Probabilistic Logic Sampling," in *Uncertainty in Artificial Intelligence 2*, (J.F. Lemmer and L. N. Kanal, Eds.), North Holland, Amsterdam, 317–324.
- Jensen, F. V., Olesen, K. G., and Andersen, S. K. (1990), "An Algebra of Bayesian Belief Universes for Knowledge-Based Systems," *Networks*, 20, 637–659.
- Kim, J. H. and Pearl, J. (1983), "A Computation Model for Causal and Diagnostic Reasoning in Inference Systems," in *Proceedings of the 8th International Joint Conference on AI*, Los Angeles, 190–93.
- 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.
- 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.
- Li, Z., and D'Ambrosio, B. (1995), "A Framework for Ordering Composite Beliefs in Belief Networks," *IEEE Transactions on Systems, Man, and Cybernetics*, 25, 2, 243–255.
- Martos, B. (1964), "Hyperbolic Programming," *Naval Research Logistic Quarterly*, 32, 135–156.
- Pearl, J. (1986), "Evidential Reasoning Using Stochastic Simulation of Causal Models," *Artificial Intelligence*, 32, 245–287.
- Pearl, J. (1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo, CA.

- Pool, D. (1993), “Average-case Analysis of a Search Algorithm for Estimating Prior and Posterior Probabilities in Bayesian Networks with Extreme Probabilities”, In *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, 13, 1, 606–612.
- Santos, E., and Shimony S. E. (1994), “Belief Updating by Enumerating High-Probability Independence-Based Assignments,” in *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, 506–513. Morgan Kaufmann Publishers, San Francisco.
- Shachter, R. D. and Peot, M. A. (1990a), “Simulation Approaches to General Probabilistic Inference on Belief Networks,” in *Uncertainty in Artificial Intelligence 5*, Machine Intelligence and Pattern Recognition Series, 10 (Henrion et al. Eds.), North Holland, Amsterdam, 221-231.
- Shachter, R. D., D’Ambrosio, B., and DelFabero, B. (1990b), “Symbolic Probabilistic Inference in Belief Networks,” in *Proceedings Eighth National Conference on AI*, 126–131.
- Shachter, R. D., Andersen, S. K. and Szolovits, P. (1994), “Global Conditioning for Probabilistic Inference in Belief Networks,” in *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, 514–522. Morgan Kaufmann Publishers, San Francisco.