Sensitivity Analysis in Normal Bayesian Networks

Enrique Castillo[1], **Uffe Kjaerulff**[2] **and Linda C. van der Gaag**[3]

[1] *Department of Applied Mathematics and Computational Sciences, University of Cantabria, Avda. Castros s/n, 39005 Santander, Spain.* (castie@ccaix3.unican.es)

[2] *Department of Computer Science, Aalborg University, Fredrik Bajers Vej 7, DK-9220, Aalborg, Denmark.* (uk@cs.auc.dk)

[3] *Computer Science Department, Utrecht University, P.O. Box 80.089, 3508 TB Utrecht, The Netherlands.* (linda@cs.uu.nl)

Abstract

The paper discusses the problem of sensitivity analysis in normal Bayesian networks. The algebraic structure of the conditional means and variances, as linear and quadratic functions of the parameters, are used to simplify the sensitivity analysis. In particular the probabilities of conditional variables exceeding given values and related probabilities are analyzed, and full expressions for the partial derivatives obtained. An example of application is used to illustrate all the concepts and methods.

Key Words: Sensitivity, Normal models, Bayesian networks.

1 Introduction

Sensitivity analysis is becoming an important and popular area of work. When solving practical problems, applied scientists are not satisfied enough with getting results coming from models, but they require a sensitivity analysis, indicating how sensitive the resulting numbers are to changes in the parameter values, to be performed (see Castillo, Gutiérrez and Hadi [5, 4], Castillo, Gutiérrez, Hadi and Solares [5], Castillo, Solares, and Gómez.[6, 7, 8]).

In some cases, the parameter selection has an extreme importance in the final results. For example, it is well known how sensitive are the distributional asumptions and parameter values to tail distributions (see Galambos [13] or Castillo [2]). If this influence is neglected, the consequences can be desastrous. Thus, the relevance of sensitivity analysis.

Laskey [16] seems to be the first to address the complexity of sensitivity analysis of Bayesian networks, byintroducing a method for computing the partial derivative of a posterior marginal probability with respect to a given parameter. Castillo, Gutiérrez and Hadi[5, 4] show that the function expressing the posterior probability is a quotient of linear functions in the parameters and the evidence values in the discrete case, and of the means, variances and evidence values, but covariances can appear squared. This discovery allows simplifying sensitivity analysis and making it computationally efficient (see, for example, Kjaerulff and van der Gaag [15], or Darwiche [12]).

In this paper we address the problem of sensitivity analysis in normal Bayesian networks and show how changes in the parameter and evidence values influence marginal and conditional probabilities given the evidence.

Normal Bayesian networks are specially useful when dealing with reliability analysis. It is well known that in this area, the normal procedure consists of transforming the initial set of variables into a set of normal variables, using the Rosenblatt transformation [19], that transforms a set of continuous variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ into a set of normal variables. Rosenblatt converts first the original multidimensional variable $\mathbf{X}$ into a set of independent standard uniform random variables $\mathbf{U}$, using the transformation:

$$\begin{array}{rcl} U_1 & = & F_1(X_1) \\ U_2 & = & F_2(X_2|X_1) \\ \vdots & \vdots & \vdots \\ U_n & = & F_n(X_n|X_1, X_2, \ldots, X_n) \end{array} \tag{1}$$

where, $F_1(\cdot), F_2(\cdot|\cdot), \ldots, F_n(\cdot|\cdot)$ are the conditional cdfs of the set $\{X_1, X_2, \ldots, X_n\}$.

Next, using the transformation

$$Z_i = \Phi^{-1}(U_i);\ i = 1, 2, \ldots, n, \tag{2}$$

where $\Phi(x)$is the cdf of the standard $N(0, 1)$ random variable, the random vector $\mathbf{Z}$ becomes normal $N(\mathbf{0}, \mathbf{I})$.

It is interesting to realize that the Rosenblatt transformation transforms $\mathbf{X}$ to non-dimensional variables $\mathbf{U}$.

This paper is structured as follows. In Section 2 we remind the reader about normal Bayesian networks and introduce our working example. In Section 3 we discuss how to perform an exact propagation in normal Bayesian networks. Section 4 is devoted to symbolic propagation. Section 5 analyses the sensitivity problem. Finally, Section 6 gives some conclusions.

2 Normal Bayesian Network Models

In this section we introduce Bayesian network models, but we first remind the reader the definition of Bayesian network.

Definition 1 (Bayesian network) *A Bayesian network is a pair (D, P), where D is a directed acyclic graph (DAG), $P = \{p(x_1|\pi_1), \ldots, p(x_n|\pi_n)\}$ is a set of n conditional probability densities (CPD), one for each variable, and Π_i is the set of parents of node X_i in D. The set P defines the associated joint probability density as*

$$p(\mathbf{x}) = \prod_{i=1}^{n} p(x_i|\pi_i). \tag{3}$$

The DAG D is a minimal directed I-map of $p(\mathbf{x})$.

The main two advantages of Bayesian networks are: (a) the factorization implied by (3), and (b) the fact that conditionally independence relations can be inferred directly from the graph D.

Definition 2 (Gaussian Bayesian network) *A Bayesian network is said to be a Gaussian Bayesian network if and only if the JPD associated with its variables X is a multivariate normal distribution, $N(\mu, \Sigma)$, i.e., with joint probability density function:*

$$f(x) = (2\pi)^{-n/2}|\Sigma|^{-1/2} \exp\left\{-1/2(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}, \tag{4}$$

where μ is the n-dimensional mean vector, Σ is the $n \times n$ covariance matrix, $|\Sigma|$ is the determinant of Σ, and μ^T denotes the transpose of μ.

Gaussian Bayesian networks have been treated, among others, by Kenley [14], Shachter and Kenley [20]), and Castillo, Gutiérrez and Hadi [3]. The JPD of the variables in a Gaussian Bayesian network can be specified as in (3) by the product of a set of CPDs whose joint probability density function is given by

$$f(x_i|\pi_i) \sim N\left(\mu_i + \sum_{j=1}^{i-1}\beta_{ij}(x_j - \mu_j),\ v_i\right), \tag{5}$$

where β_{ij} is the regression coefficient of X_j in the regression of X_i on the parents of X_i, Π_i, and

$$v_i = \Sigma_i - \Sigma_{i\Pi_i}\Sigma_{\Pi_i}^{-1}\Sigma_{i\Pi_i}^T$$

is the conditional variance of X_i, given $\Pi_i = \pi_i$, where Σ_i is the unconditional variance of X_i, $\Sigma_{i\Pi_i}$ is the covariances between X_i and the variables in Π_i, and Σ_{Π_i} is the covariance matrix of Π_i. Note that β_{ij} measures the strength of the relationship between X_i and X_j. If $\beta_{ij} = 0$, then X_j is not a parent of X_i.

Note that while the conditional mean $\mu_{x_i|\pi_i}$ depends on the values of the parents π_i, the conditional variance does not depend on these values. Thus, the natural set of CPDs defining a normal Bayesian network is given by a collection of parameters $\{\mu_1, \ldots, \mu_n\}$, $\{v_1, \ldots, v_n\}$, and $\{\beta_{ij} \,|\, j < i\}$, as shown in (5).

Alternatively, we can define a normal JPD function by giving its mean μ vector and its precision matrix $W = \Sigma^{-1}$. Shachter and Kenley [20] describe the general transformation from $\{v_1, \ldots, v_n\}$ and $\{\beta_{ij} : j < i\}$ to W. They use the following recursive formula, in which $W(i)$ denotes the $i \times i$ upper left submatrix of W and β_i denotes the column vector $\{\beta_{ij} : j < i\}$:

$$W(i+1) = \begin{pmatrix} W(i) + \dfrac{\beta_{i+1}\beta_{i+1}^T}{v_{i+1}} & \dfrac{-\beta_{i+1}}{v_{i+1}} \\ \dfrac{-\beta_{i+1}^T}{v_{i+1}} & \dfrac{1}{v_{i+1}} \end{pmatrix}, \tag{6}$$

with $W(1) = 1/v_1$.

Thus, we have two alternative representations of the JPD of a normal Bayesian network. The following is an illustrative example of a normal Bayesian network.

Example 1 (Normal Bayesian network) Assume that we are studying the river in Figure 1(a), where we have indicated the four cross sections A, B, C and D, where the water discharges are measured. The mean time of the water going from A to B and from B to D is one day, and the mean time from C to D is two days. Thus, we register the set (A, B, C, D) with the corresponding delays. Assume that the joint water discharges can be assumed to be normal distributions and that we are interested in predicting B and D, one and two days later, respectively, from the observations of A and C.

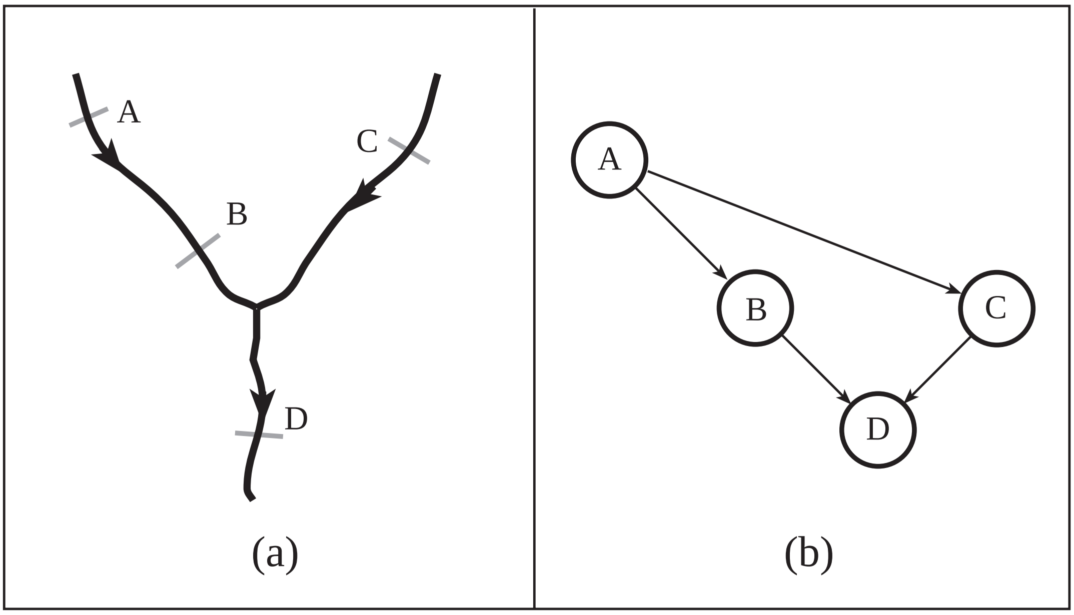

Figure 1: (a) the river in Example 1 and the selected cross sections, and (b) the Bayesian network used to solve the problem.

In Figure 1(b) we have shown the graph associated with a Bayesian network that shows the dependence structure of the variables involved.

Suppose that the random variable (A, B, C, D) is normally distributed, i.e., $\{A, B, C, D\} \sim N(\mu, \Sigma)$. A normal Bayesian network is defined by specifying the set of CPDs appearing in the factorization (3), which gives

$$f(a, b, c, d) = f(a)f(b|a)f(c|a)f(d|b, c), \tag{7}$$

where

$$\begin{aligned} f(a) &\sim N(\mu_A, v_A), \\ f(b|a) &\sim N(\mu_B + \beta_{BA}(a - \mu_A), v_B), \\ f(c|a) &\sim N(\mu_C + \beta_{CA}(a - \mu_A), v_C), \\ f(d|b, c) &\sim N(\mu_D + \beta_{DB}(b - \mu_B) + \beta_{DC}(c - \mu_C), v_D). \end{aligned} \tag{8}$$

This set of CPDs constitutes one of two equivalent representations of the normal Bayesian network. The parameters involved in this representation are $\{\mu_A, \mu_B, \mu_C, \mu_D\}$, $\{v_A, v_B, v_C, v_D\}$, and $\{\beta_{BA}, \beta_{CB}, \beta_{DB}, \beta_{DC}\}$.

An alternative representation can be obtained using (6). In this case, after four iterations, we finally obtain the matrix

$$W = \begin{pmatrix} \dfrac{1}{v_A} + \dfrac{\beta_{BA}^2}{v_B} + \dfrac{\beta_{CA}^2}{v_C} & -\dfrac{\beta_{BA}}{v_B} & -\dfrac{\beta_{CA}}{v_C} & 0 \\ -\dfrac{\beta_{BA}}{v_B} & \dfrac{1}{v_B} + \dfrac{\beta_{DB}^2}{v_D} & \dfrac{\beta_{DB}\beta_{DC}}{v_D} & -\dfrac{\beta_{DB}}{v_D} \\ -\dfrac{\beta_{CA}}{v_C} & \dfrac{\beta_{DB}\beta_{DC}}{v_D} & \dfrac{1}{v_C} + \dfrac{\beta_{DC}^2}{v_D} & -\dfrac{\beta_{DC}}{v_D} \\ 0 & -\dfrac{\beta_{DB}}{v_D} & -\dfrac{\beta_{DC}}{v_D} & \dfrac{1}{v_D} \end{pmatrix}.$$

Note that so far, all parameters have been considered in symbolic form. Thus, we can specify a Bayesian model by assigning numerical values to the parameters above. For example, for

$$\mu_A = 3, \;\; \mu_B = 4, \;\; \mu_C = 9, \;\; \mu_D = 14.$$

$$v_A = 4; \; v_B = 1; \; v_C = 4; \; v_D = 1, \; \beta_{BA} = 1, \; \beta_{CA} = 2, \; \beta_{DB} = 1, \; \beta_{DC} = 1.$$

we get

$$\mu = \begin{pmatrix} 3 \\ 4 \\ 9 \\ 14 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 4 & 4 & 8 & 12 \\ 4 & 5 & 8 & 13 \\ 8 & 8 & 20 & 28 \\ 12 & 13 & 28 & 42 \end{pmatrix}.$$

■

3 Exact Propagation in Gaussian Networks

Several algorithms have been proposed in the literature to solve the problems of evidence propagation in these models. Some of them have originated from the methods for discrete models. For example, Normand and Tritchler [18] introduce an algorithm for evidence propagation in Gaussian network models using the same idea of the polytrees algorithm. Lauritzen [17] suggests a modification of the join tree algorithm to propagate evidence in mixed models.

Several algorithms use the structure provided by (3) and (5) for evidence propagation (see Xu and Pearl [21], and Chang and Fung [11]). In this section we present a conceptually simple and efficient algorithm that uses the covariance matrix representation. An incremental implementation of the algorithm allows updating probabilities, as soon as a single piece of evidence is observed. The main result is given in the following theorem, which characterizes the CPDs obtained from a Gaussian JPD (see, for example, Anderson [1]).

Theorem 1 Conditionals of a Gaussian distribution. *Let Y and Z be two sets of random variables having a joint multivariate Gaussian distribution with mean vector and covariance matrix given by*

$$\mu = \begin{pmatrix} \mu^Y \\ \mu^Z \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma^{YY} & \Sigma^{YZ} \\ \Sigma^{ZY} & \Sigma^{ZZ} \end{pmatrix},$$

where μ^Y and Σ^{YY} are the mean vector and covariance matrix of Y, μ^Z and Σ^{ZZ} are the mean vector and covariance matrix of Z, and Σ^{YZ} is the covariance of Y and Z. Then the CPD of Y given $Z = z$ is multivariate Gaussian with mean vector $\mu^{Y|Z=z}$ and covariance matrix $\Sigma^{Y|Z=z}$ that are given by

$$\mu^{Y|Z=z} = \mu^Y + \Sigma^{YZ}\Sigma^{ZZ^{-1}}(z - \mu^Z), \tag{9}$$

$$\Sigma^{Y|Z=z} = \Sigma^{YY} - \Sigma^{YZ}\Sigma^{ZZ^{-1}}\Sigma^{ZY}. \tag{10}$$

Note that the conditional mean $\mu^{Y|Z=z}$ depends on z but the conditional variance $\Sigma^{Y|Z=z}$ does not.

Theorem 1 suggests an obvious procedure to obtain the means and variances of any subset of variables $Y \subset X$, given a set of evidential nodes $E \subset X$ whose values are known to be $E = e$. Replacing Z in (9) and (10) by E, we obtain the mean vector and covariance matrix of the conditional distribution of the nodes in Y. Note that considering $Y = X \setminus E$ we get the joint distribution of the remaining nodes, and then we can answer questions involving the joint distribution of nodes instead of the usual information that refers only to individual nodes.

The above introduced methods for evidence propagation in Gaussian Bayesian network models use the same idea, but perform local computations by taking advantage of the factorization of the JPD as a product of CPDs.

In order to simplify the computations, it is more convenient to use an incremental method, updating one evidential node at a time (taking elements one by one from E). In this case we do not need to calculate the inverse of a matrix because it degenerates to a scalar. Moreover, μ^Y and Σ^{YZ} are column vectors, and Σ^{ZZ} is also a scalar. Then the number of calculations needed to update the probability distribution of the nonevidential variables given a single piece of evidence is linear in the number of variables in X. Thus, this algorithm provides a simple and efficient method for evidence propagation in Gaussian Bayesian network models.

Due to the simplicity of this incremental algorithm, the implementation of this propagation method in the inference engine of an expert system is an easy task. The algorithm gives the CPD of the nonevidential nodes Y given the evidence $E = e$. The performance of this algorithm is illustrated in the following example.

Example 2 Propagation in Gaussian Bayesian network models. Consider the Gaussian Bayesian network given in Figure 1. Suppose we have the evidence $\{A = 7, C = 17, B = 8\}$.

If we apply expressions (9) and (10) to propagate evidence, we obtain the following:

After evidence $A = 7$**:** In the first iteration step, we consider the first evidential node $A = 7$. We obtain the following mean vector and covariance matrix for the rest of the nodes $Y = \{B, C, D\}$:

$$\mu^{Y|A=7} = \begin{pmatrix} 8 \\ 17 \\ 26 \end{pmatrix}; \ \Sigma^{YY|A=7} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 4 & 4 \\ 1 & 4 & 6 \end{pmatrix}. \qquad (11)$$

After evidence $A = 7, C = 17$**:** The second step of the algorithm adds evidence $C = 17$; we obtain the following mean vector and covariance matrix for the rest of the nodes $Y = \{B, D\}$:

$$\mu^{Y|A=7,C=17} = \begin{pmatrix} 8 \\ 26 \end{pmatrix}; \ \Sigma^{YY|A=7,C=17} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}. \qquad (12)$$

After evidence $A = 7, C = 17, B = 8$**:** Finally, after considering evidence $B = 8$ we get the conditional mean and variance of D, which are given by $\mu^{D|A=7,C=17,B=8} = 26$, $\sigma^{DD|A=7,C=17,B=8} = 1$.

∎

4 Symbolic Propagation in Gaussian Bayesian Networks

In Section 3 we presented several methods for exact propagation in Gaussian Bayesian networks. Some of these methods have been extended for symbolic computation (see, for example, Chang and Fung [11] and Lauritzen [17]). In this section we illustrate symbolic propagation in Gaussian Bayesian networks using the conceptually simple method given in Section 3. When dealing with symbolic computations, all the required operations must be performed by a program with symbolic manipulation capabilities. Figure 2 shows the *Mathematica* code for the symbolic implementation of the method given in Section 3. The code calculates the mean and variance of all nodes given the evidence in the evidence list.

Example 3 Consider the set of variables $X = \{A, B, C, D\}$ with mean vector and covariance matrix

$$\mu = \begin{pmatrix} p \\ 4 \\ 9 \\ q \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} a & 4 & d & f \\ 4 & 5 & 8 & c \\ d & 8 & 20 & 28 \\ f & c & 28 & b \end{pmatrix}. \qquad (13)$$

Note that some means and variances are specified in symbolic form, and that we have

$$\sigma^{YY} = \begin{pmatrix} 5 & c \\ c & b \end{pmatrix}, \ \sigma^{ZZ} = \begin{pmatrix} a & d \\ d & 20 \end{pmatrix}, \ \sigma^{YZ} = \begin{pmatrix} 4 & 8 \\ f & 28 \end{pmatrix}. \qquad (14)$$

```
(* Definition of the JPD *)
M={p,4,9,q};
V={{a, 4, d, f},
   {4, 5, 8, c},
   {d, 8, 20, 28},
   {f, c, 28, b}};
(* Nodes and evidence *)
X={A,B,C,D};
Ev={A,C};
ev={x1,x3};
(* Incremental updating of M and V *)
NewM=Transpose[List[M]];
NewV=V;
For[k=1, k<=Length[Ev], k++,
(* Position of the ith element of E[[k]] in X *)
    i=Position[X,Ev[[k]]][[1,1]];
    My=Delete[NewM,i];
    Mz=NewM[[i,1]];
    Vy=Transpose[Delete[Transpose[Delete[NewV,i]],i]];
    Vz=NewV[[i,i]];
    Vyz=Transpose[List[Delete[NewV[[i]],i]]];
    NewM=My+(1/Vz)*(ev[[k]]-Mz)*Vyz;
    NewV=Vy-(1/Vz)*Vyz.Transpose[Vyz];
(* Delete ith element *)
    X=Delete[X,i];
(* Printing results *)
    Print["Iteration step = ",k];
    Print["Remaining nodes = ",X];
    Print["M = ",Together[NewM]];
    Print["V = ",Together[NewV]];
    Print["--------------------"];
]
```

Figure 2: *Mathematica* code for symbolic propagation of evidence in a Gaussian Bayesian network model.

We use the *Mathematica* code in Figure 2 to calculate the conditional means and variances of all nodes. The first part of the code defines the mean vector and covariance matrix of the Bayesian network. Table 1 shows the initial marginal probabilities of the nodes (no evidence) and the conditional probabilities of the nodes given each of the evidences $\{A = x_1\}$ and $\{A = x_1, C = x_3\}$. An examination of the results in Table 1 shows that the conditional means and variances are rational expressions, that is, ratios of polynomials in the parameters. Note, for example, that for the case of evidence $\{A = x_1, C = x_3\}$, the polynomials are first-degree in p, q, a, b, x_1, and x_3, that is, in the mean and variance parameters and in the evidence variables, and second-degree in d, f, i.e., the covariance parameters. Note also the common denominator for the rational functions giving the conditional means and variances. ■

The fact that the mean and variances of the conditional probability distributions of the nodes are rational functions of polynomials is given by the following theorem (see Castillo, Gutiérrez, Hadi, and Solares [5]).

Theorem 2 *Consider a Gaussian Bayesian network over a set of variables $X = \{X_1, \ldots, X_n\}$ with mean vector μ and covariance matrix Σ. Partition X, μ, and Σ as $X = \{Y, Z\}$,*

$$\mu = \begin{pmatrix} \mu^Y \\ \mu^Z \end{pmatrix}, \qquad \text{and} \qquad \Sigma = \begin{pmatrix} \Sigma^{YY} & \Sigma^{YZ} \\ \Sigma^{ZY} & \Sigma^{ZZ} \end{pmatrix},$$

where μ^Y and Σ^{YY} are the mean vector and covariance matrix of Y, μ^Z and Σ^{ZZ} are the mean vector and covariance matrix of Z, and Σ^{YZ} is the covariance of Y and Z. Suppose that Z is the set of evidential nodes. Then the conditional probability distribution of any variable $X_i \in Y$ given Z is normal, with mean and variance that are ratios of polynomial functions in the evidential variables and the related parameters in μ and Σ. The polynomials involved are at most of degree one in the conditioning variables and in the mean and variance parameters and are of degree two in the covariance parameters involving at least one Z (evidential) variable. Finally, the polynomial in the denominator is the same for all nodes.

Proof: From Theorem 1,

$$\mu^{Y|Z=z} = \mu^Y + \Sigma^{YZ}\Sigma^{ZZ^{-1}}(z - \mu^Z). \tag{15}$$

Note that $\Sigma^{YZ}\Sigma^{ZZ^{-1}}(z - \mu^Z)$ is a rational function because it can be written as the quotient of the polynomials $\Sigma^{YZ} adj(\Sigma^{ZZ})(z - \mu^Z)$ and $det(\Sigma^{ZZ})$, where $adj(\Sigma^{ZZ})$ is the adjoint matrix of Σ^{ZZ} and $det(\Sigma^{ZZ})$ is the determinant of Σ^{ZZ}. Therefore, the conditional expectation $\mu^{Y|Z=z}$ in (15) is μ^Y plus a rational function, which implies that $\mu^{Y|Z=z}$ is a rational function with polynomial denominator $det(\Sigma^{ZZ})$. Note also that each parameter appears in only one of the three factors above, which implies linearity in each parameter.

Similarly, from Theorem 1 the conditional variance is

$$\Sigma^{YY|Z=z} = \Sigma^{YY} - \Sigma^{YZ}\Sigma^{ZZ^{-1}}\Sigma^{ZY}, \tag{16}$$

No Evidence		
Node	Mean	Variance
A	p	a
B	4	5
C	9	20
D	q	b
Evidence $A = x_1$		
Node	Mean $\mu^{Y\|A=x_1}$	Variance $\sigma^{YY\|A=x_1}$
A	x_1	0
B	$\frac{4(a-p+x_1)}{a}$	$\frac{5a-16}{a}$
C	$\frac{9a-dp+dx_1}{a}$	$\frac{20a-d^2}{a}$
D	$\frac{-fp+aq+fx_1}{a}$	$\frac{ab-f^2}{a}$
Evidence $A = x_1$ and $C = x_3$		
Node	Mean $\mu^{Y\|A=x_1,C=x_3}$	Variance $\sigma^{YY\|A=x_1,C=x_3}$
A	x_1	0
B	$\frac{4(2a+(9-d)d+(2d-20)p+(20-2d)x_1+(2a-d)x_3}{20a-d^2}$	$\frac{36a+64d-5d^2-320}{20a-d^2}$
C	x_3	0
D	$\frac{-252a+9df+(28d-20f)p+(20a-d^2)q}{20a-d^2} + \frac{(20f-28d)x_1+(28a-df)x_3}{20a-d^2}$	$\frac{(20ab-bd^2+56df-20f^2-784a}{20a-d^2}$

Table 1: Means and variances of the marginal probability distributions of nodes, initially and after evidence.

which is Σ^{YY} minus the rational function $\Sigma^{YZ}\Sigma^{ZZ^{-1}}\Sigma^{ZY}$. This implies that $\Sigma^{Y|Z=z}$ is a rational function with polynomial denominator $|\Sigma^{ZZ}|$. Note also that all parameters except the covariances in Σ^{YZ} and Σ^{ZZ} appear in only one of the factors, which implies linearity in these parameters. On the contrary, the covariance parameters in Σ^{YZ} and Σ^{ZZ} appear in two factors, and hence they generate second-degree terms in the polynomials.

Finally, the denominator polynomial is of second degree in the covariance parameters of Σ^{ZZ}, because of the symmetry of the covariance matrix. ■

Note that because the denominator polynomial is identical for all nodes, for implementation purposes it is more convenient to calculate and store all the numerator polynomials for each node and calculate and store the common denominator polynomial separately.

5 Sensitivity Analysis

When dealing with Normal Bayesian networks, one is normally involved in calculating probabilities of the form:

$$\begin{array}{rcl} P(X_i > a|\mathbf{e}) & = & 1 - F_{X_i|\mathbf{e}}(a), \\ P(X_i \le a|\mathbf{e}) & = & F_{X_i|\mathbf{e}}(a) \\ P(a < X_i \le b|\mathbf{e}) & = & F_{X_i|\mathbf{e}}(b) - F_{X_i|\mathbf{e}}(a) \end{array} \tag{17}$$

and one is required to perform a sensibility analysis on this probabilities with respect to a given parameter θ or evidence value e. Thus, it becomes important to know the partial derivatives

$$\frac{\partial F_{X_i|\mathbf{e}}(a;\mu(\theta),\sigma(\theta))}{\partial \theta} \text{ and } \frac{\partial F_{X_i|\mathbf{e}}(a;\mu(\theta),\sigma(\theta))}{\partial e}.$$

We can write

$$\frac{\partial F_{X_i|\mathbf{e}}(a;\mu(\theta),\sigma(\theta))}{\partial \theta} = \frac{\partial F_{X_i|\mathbf{e}}(a;\mu(\theta),\sigma(\theta))}{\partial \mu}\frac{\partial \mu(\theta)}{\partial \theta} + \frac{\partial F_{X_i|\mathbf{e}}(a;\mu(\theta),\sigma(\theta))}{\partial \sigma}\frac{\partial \sigma(\theta)}{\partial \theta} \tag{18}$$

where θ is one of the parameters.

Since

$$F_{X_i|\mathbf{e}}(a;\mu(\theta),\sigma(\theta)) = \Phi(\frac{a-\mu(\theta)}{\sigma(\theta)}) \tag{19}$$

we have

$$\frac{\partial F_{X_i|\mathbf{e}}(a;\mu(\theta),\sigma(\theta))}{\partial \mu} = f_{N(0,1)}(\frac{a-\mu(\theta)}{\sigma(\theta)})\left(\frac{-1}{\sigma(\theta)}\right) \tag{20}$$

and

$$\frac{\partial F_{X_i|\mathbf{e}}(a;\mu(\theta),\sigma(\theta))}{\partial \sigma} = f_{N(0,1)}(\frac{a-\mu(\theta)}{\sigma(\theta)})\left(\frac{\mu(\theta)-a}{\sigma(\theta)^2}\right) \tag{21}$$

and then (18) becomes

$$\frac{\partial F_{X_i|\mathbf{e}}(a;\mu(\theta),\sigma(\theta))}{\partial \theta} = f_{N(0,1)}(\frac{a-\mu(\theta)}{\sigma(\theta)})\left[\left(\frac{-1}{\sigma(\theta)}\right)\frac{\partial \mu(\theta)}{\partial \theta} + \left(\frac{\mu(\theta)-a}{\sigma(\theta)^2}\right)\frac{\partial \sigma(\theta)}{\partial \theta}\right] \tag{22}$$

Thus, the partial derivatives $\dfrac{\partial F_{X_i|\mathbf{e}}(a;\mu(\theta),\sigma(\theta))}{\partial\theta}$ can be obtained by a single evaluation of $\mu(\theta)$ and $\sigma(\theta)$, and determining the partial derivatives $\dfrac{\partial\mu(\theta)}{\partial\theta}$ and $\dfrac{\partial\sigma(\theta)}{\partial\theta}$ with respect to all the parameters being considered. Thus, the calculus of these partial derivatives becomes crucial.

There are two ways of calculating these partial derivatives:

1. Using the algebraic structure of the conditional means and variances.

2. Direct differentiations of the formulas (15) and (16).

We use both method below.

5.1 Sensitivity based on the algebraic structure of conditional means and variances

To calculate $\dfrac{\partial\mu_N(\theta)}{\partial\theta}$ and $\dfrac{\partial\sigma_N(\theta)}{\partial\theta}$ for node N we need to know the dependence of $\mu_N(\theta)$ and $\sigma_N(\theta)$ on the parameter θ. This can be done with the help of Theorem 2. To illustrate we use the previous example.

From Theorem 2 we can write

$$\mu_N^{Y|A=x_1,C=x_3}(a) = \frac{\alpha_1 a+\beta_1}{\gamma a+\delta} \tag{23}$$

$$\sigma_N^{Y|A=x_1,C=x_3}(a) = \frac{\alpha_2 a+\beta_2}{\gamma a+\delta}, \tag{24}$$

where N is B or D, and since we have only 6 unknowns, calculation of $\mu_N^{Y|A=x_1,C=x_3}$ and $\sigma_N^{Y|A=x_1,C=x_3}$ for three different values of a allows determining the constant coefficients $\alpha_1,\alpha_2,\beta_1,\beta_2,\gamma$ and δ. Then, the partial derivatives with respect to a becomes

$$\frac{\partial\mu_N^{Y|A=x_1,C=x_3}(a)}{\partial a} = \frac{\alpha_1\delta-\beta_1\gamma}{(\gamma a+\delta)^2} \tag{25}$$

$$\frac{\partial\sigma_N^{Y|A=x_1,C=x_3}(a)}{\partial a} = \frac{\alpha_2\delta-\beta_2\gamma}{(\gamma a+\delta)^2}. \tag{26}$$

Similarly, from Theorem 2 we can write

$$\mu_N^{Y|A=x_1,C=x_3}(f) = \frac{\alpha_3 f+\beta_3}{\gamma_1} \tag{27}$$

$$\sigma_N^{Y|A=x_1,C=x_3}(f) = \frac{\alpha_4 f+\beta_4}{\gamma_1} \tag{28}$$

$$\tag{29}$$

and since we have only 5 unknowns, calculation of a total of 5 values of $\mu_N^{Y|A=x_1,C=x_3}(f)$ and $\sigma_N^{Y|A=x_1,C=x_3}(f)$ for different values of f allows determining the constant coefficients $\alpha_3, \alpha_4, \beta_3, \beta_4$ and γ_1. Then, the partial derivatives with respect to f becomes

$$\frac{\partial \mu_N^{Y|A=x_1,C=x_3}(f)}{\partial f} = \frac{\alpha_3}{\gamma_1} \tag{30}$$

$$\frac{\partial \sigma_N^{Y|A=x_1,C=x_3}(f)}{\partial f} = \frac{\alpha_4}{\gamma a_1}. \tag{31}$$

It is worthwhile mentioning that if $N = B$, then $\alpha_3 = \alpha_4 = \beta_3 = \beta_4 = 0$, and we need no calculations.

Finally, we can also obtain the partial derivatives with respect to evidence values. From Theorem 2 we can write

$$\mu_N^{Y|A=x_1,C=x_3}(x_1) = \alpha_5 x_1 + \beta_5 \tag{32}$$

$$\sigma_N^{Y|A=x_1,C=x_3}(x_1) = \gamma_2 \tag{33}$$

and since we have only 3 unknowns, calculation of a total of 3 values of $\mu_N^{Y|A=x_1,C=x_3}(x_1)$ and $\sigma_N^{Y|A=x_1,C=x_3}(x_1)$ for different values of f allows determining the constant coefficients α_5, β_5 and γ_2. Then, the partial derivatives with respect to f becomes

$$\frac{\partial \mu_N^{Y|A=x_1,C=x_3}(x_1)}{\partial x_1} = \alpha_5 \tag{34}$$

$$\frac{\partial \sigma_N^{Y|A=x_1,C=x_3}(x_1)}{\partial x_1} = 0. \tag{35}$$

It is worthwhile mentioning that if partial derivatives with respect to several parameters are to be calculated, the number of calculations reduces even more because some of them are common.

5.2 Sensitivity based on direct differentiation

To calculate the partial derivatives $\frac{\partial \mu(\theta)}{\partial \theta}$ and $\frac{\partial \sigma(\theta)}{\partial \theta}$ we need to develop the expressions for $\mu^{Y|Z=z}$ and $\Sigma^{Y|Z=z}$ in terms of the parameters. We consider only the case where the set Y has a unique component, that we denote y.

Equations (15) and (16) for a single variable become

$$\mu^{Y|Z=z}(\theta) = \mu^Y(\theta) + \Sigma^{YZ}(\theta)\Sigma^{ZZ^{-1}}(\theta)(z - \mu^Z(\theta)). \tag{36}$$

and

$$\sigma^{YY|Z=z}(\theta) = \Sigma^{YY}(\theta) - \Sigma^{YZ}(\theta)\Sigma^{ZZ^{-1}}(\theta)\left(\Sigma^{YZ}(\theta)\right)^T, \tag{37}$$

where θ is the single parameter under consideration.

It must be noted that with the exception of $\Sigma^{YZ}(\theta)$, only one of the functions in (36) and (37), can be dependent on θ.

$$
\begin{aligned}
\mu_y^{Y|Z=z} &= \mu_y^Y + \sum_z \sigma_{yz}^{YZ} \sum_k \sigma_{zk}^{ZZ^{-1}} \left(z_k - \mu_k^Z \right) \\
&= \begin{cases}
\mu_y^Y + \sum_z \sigma_{yz}^{YZ} / \sigma_{zz}^{ZZ} \left(z_z - \mu_z^Z \right) & if \quad |Z| = 1 \\
\mu_y^Y + \dfrac{\sum_{\bar{z}} \sigma_{y\bar{z}}^{YZ} \sum_{\bar{k}} (-1)^{z+k} \sigma_{zk}^{ZZ} \left(z_{\bar{k}} - \mu_{\bar{k}}^Z \right)}{\sum_k (-1)^{z+k} \sigma_{zk}^{ZZ} \sigma_{\bar{z}\bar{k}}^{ZZ}} & if \quad |Z| = 2 \\
\mu_y^Y + \dfrac{\sum_r \sigma_{yr}^{YZ} \sum_s \sum_{k \neq s} (-1)^{z+k+r+s} \sigma_{zk}^{ZZ} \left| \sigma_{(rz)(sk)}^{ZZ} \right| \left(z_s - \mu_s^Z \right)}{\sum_k (-1)^{z+k} \sigma_{zk}^{ZZ} \left| \sigma_{(z)(k)}^{ZZ} \right|} & if \quad |Z| > 2
\end{cases}
\end{aligned}
\tag{38}
$$

$$
\begin{aligned}
\sigma_{yy}^{YY|Z=z} &= \sigma_{yy}^{YY} - \sum_z \sigma_{yz}^{YZ} \sum_k \sigma_{zk}^{ZZ^{-1}} \sigma_{ky}^{ZY} = \sigma_{yy}^{YY} - \sum_z \sigma_{yz}^{YZ} \sum_k \sigma_{zk}^{ZZ^{-1}} \sigma_{yk}^{YZ} \\
&= \begin{cases}
\sigma_{yy}^{YY} - \sum_z \sigma_{yz}^{YZ} / \sigma_{zz}^{ZZ} \sigma_{yz}^{YZ} & if \quad |Z| = 1 \\
\sigma_{yy}^{YY} - \dfrac{\sum_{\bar{z}} \sigma_{y\bar{z}}^{YZ} \sum_{\bar{k}} (-1)^{z+k} \sigma_{zk}^{ZZ} \sigma_{y\bar{k}}^{YZ}}{\sum_k (-1)^{z+k} \sigma_{zk}^{ZZ} \sigma_{\bar{z}\bar{k}}^{ZZ}} & if \quad |Z| = 2 \\
\sigma_{yy}^{YY} - \dfrac{\sum_r \sigma_{yr}^{YZ} \sum_s \sum_{k \neq s} (-1)^{z+k+r+s} \sigma_{zk}^{ZZ} \left| \sigma_{(rz)(sk)}^{ZZ} \right| \sigma_{ys}^{YZ}}{\sum_k (-1)^{z+k} \sigma_{zk}^{ZZ} \left| \sigma_{(z)(k)}^{ZZ} \right|} & if \quad |Z| > 2
\end{cases}
\end{aligned}
\tag{39}
$$

where r refers to an arbitrary row of Σ^{ZZ} different from row z, and $\sigma_{(zr)(ks)}^{ZZ^{-1}}$ is the matrix $\sigma^{ZZ^{-1}}$ after removing rows z and r and columns k and s, and $\bar{r}$ means the row or column different from r.Note that for $|Z| = 1$, σ_{zk} becomes σ_{zz}.

From (38) and (39) we have only the following cases:

Case 1: $\theta = \mu_r^Y$: then, we have

$$
\frac{\partial \mu_y^{Y|Z=z}}{\partial \mu_r^Y} = \delta_{yr}, \quad \frac{\partial \sigma_{yy}^{Y|Z=z}}{\partial \mu_r^Y} = 0. \tag{40}
$$

Case 2: $\theta = \mu_k^Z$: then, we have

$$
\frac{\partial \mu_y^{Y|Z=z}}{\partial \mu_k^Z} = -\sum_z \sigma_{yz}^{YZ} \sigma_{zk}^{ZZ^{-1}}, \quad \frac{\partial \sigma_{yy}^{Y|Z=z}}{\partial \mu_k^Z} = 0. \tag{41}
$$

Case 3: $\theta = \sigma_{rs}^{ZZ}$: then, the partial derivative $\dfrac{\partial \mu_y^{Y|Z=z}}{\partial \sigma_{zk}^{ZZ}}$ becomes

$$\begin{cases} \dfrac{-\sigma_{yz}^{YZ}\left(z_z - \mu_z^Z\right)}{(\sigma_{zz}^{ZZ})^2} & if \quad |Z| = 1 \\[2ex] \dfrac{\sigma_{y\bar{z}}^{YZ}(-1)^{z+k}\left(z_{\bar{k}} - \mu_{\bar{k}}^Z\right)\left|\Sigma^{ZZ}\right| - (-1)^{z+k}\sigma_{\bar{z}\bar{k}}^{ZZ}\left(\sum_{\bar{z}}\sigma_{y\bar{z}}^{YZ}\sum_{\bar{k}}(-1)^{z+k}\sigma_{zk}^{ZZ}\left(z_{\bar{k}} - \mu_{\bar{k}}^Z\right)\right)}{\left|\Sigma^{ZZ}\right|^2} & if \quad |Z| = 2 \\[2ex] \dfrac{\sum_r \sigma_{yr}^{YZ}\sum_{s\neq k}(-1)^{z+k+r+s}\left|\sigma_{(rz)(sk)}^{ZZ}\right|\left(z_s - \mu_s^Z\right)\left|\Sigma^{ZZ}\right|}{\left|\Sigma^{ZZ}\right|^2} \\ -\dfrac{(-1)^{z+k}\left|\sigma_{(z)(k)}^{ZZ}\right|\left(\sum_r \sigma_{yr}^{YZ}\sum_s\sum_{k\neq s}(-1)^{z+k+r+s}\sigma_{zk}^{ZZ}\left|\sigma_{(rz)(sk)}^{ZZ}\right|\left(z_s - \mu_s^Z\right)\right)}{\left|\Sigma^{ZZ}\right|^2} & if \quad |Z| > 2 \end{cases} \tag{42}$$

Note that for $|Z| = 1$ we have $z = r = s$.

Similarly, the partial derivative $\dfrac{\partial \sigma_{yy}^{Y|Z=z}}{\partial \sigma_{rs}^{ZZ}}$ becomes

$$\begin{cases} \dfrac{-\sigma_{yz}^{YZ}\sigma_{yz}^{YZ}}{(\sigma_{zz}^{ZZ})^2} & if \quad |Z| = 1 \\[2ex] \dfrac{\sigma_{y\bar{z}}^{YZ}(-1)^{z+k}\left(z_{\bar{k}} - \mu_{\bar{k}}^Z\right)\left|\Sigma^{ZZ}\right| - (-1)^{z+k}\sigma_{\bar{z}\bar{k}}^{ZZ}\left(\sum_{\bar{z}}\sigma_{y\bar{z}}^{YZ}\sum_{\bar{k}}(-1)^{z+k}\sigma_{zk}^{ZZ}\sigma_{y\bar{k}}^{YZ}\right)}{\left|\Sigma^{ZZ}\right|^2} & if \quad |Z| = 2 \\[2ex] \dfrac{\sum_r \sigma_{yr}^{YZ}\sum_{s\neq k}(-1)^{z+k+r+s}\left|\sigma_{(rz)(sk)}^{ZZ}\right|\sigma_{ys}^{YZ}\left|\Sigma^{ZZ}\right|}{\left|\Sigma^{ZZ}\right|^2} \\ -\dfrac{(-1)^{z+k}\left|\sigma_{(z)(k)}^{ZZ}\right|\left(\sum_r \sigma_{yr}^{YZ}\sum_s\sum_{k\neq s}(-1)^{z+k+r+s}\sigma_{zk}^{ZZ}\left|\sigma_{(rz)(sk)}^{ZZ}\right|\sigma_{ys}^{YZ}\right)}{\left|\Sigma^{ZZ}\right|^2} & if \quad |Z| > 2 \end{cases} \tag{43}$$

Case 4: $\theta = \sigma_{rs}^{YY}$: then, we have

$$\frac{\partial \mu_y^{Y|Z=z}}{\partial \sigma_{rs}^{YY}} = 0, \quad \frac{\partial \sigma_{yy}^{Y|Z=z}}{\partial \sigma_{rs}^{YY}} = \delta_{rs}, \tag{44}$$

where δ_{rs} are the Kronecker deltas.

Table 2: Normalized partial derivatives with respect of all parameters.

t	$\dfrac{t\ \partial P(B>11 \mid A=7, C=17)}{\partial t}$	$\dfrac{t\ \partial P(D>30 \mid A=7, C=17)}{\partial t}$
p	-0.01330	-0.01550
q	0	0.07233
a	0.00886	0
b	0	0.2170
d	0	0.04134
f	0	-0.0620
x_1	0.03102	0.03617
x_3	0	0.08783

Case 5: $\theta = \sigma_{yz}^{YZ}$: then, we have

$$\frac{\partial \mu_y^{Y|Z=z}}{\partial \sigma_{yz}^{YZ}} = \sum_k \sigma_{zk}^{ZZ^{-1}} \left(z_k - \mu_k^Z \right), \quad \frac{\partial \sigma_{yy}^{Y|Z=z}}{\partial \sigma_{yz}^{YZ}} = -2 \sum_k \sigma_{zk}^{ZZ^{-1}} \sigma_{yk}^{YZ}. \tag{45}$$

Case 6: If we look for the sensitivity with respect to observed values we need to consider z_k:

$$\frac{\partial \mu_y^{Y|Z=z}}{\partial z_k} = \sum_z \sigma_{yz}^{YZ} \sigma_{zk}^{ZZ^{-1}}, \quad \frac{\partial \sigma_{yy}^{Y|Z=z}}{\partial z_k} = 0. \tag{46}$$

Example 4 We continue with the previous example and calculate now the probabilities of B exceeding the critical value 11, and D exceeding the critical value 30, because they have been determined as those producing important damages in the associated areas B and D, respectively. Using Expressions (17), (19) and (12) we get

$$P(B > 11|A = 7, C = 17) = 1 - F_{B|A=7,C=17}(11) = 1 - \Phi(\frac{11-8}{1}) = 0.00135$$

$$P(D > 30|A = 7, C = 17) = 1 - F_{D|A=7,C=17}(30) = 1 - \Phi(\frac{30-26}{\sqrt{2}}) = 0.00234$$

In Table 2 we have calculated all the normalized partial derivatives of the failure probabilities. We have used the parameter values for the normalization. ■

6 Conclusions

Sensitivity analysis in normal Bayesian networks is gratedly simplified due to the knowledge of the algebraic structure of conditional mean and variances. The fact that conditional means and variances are quotients of linear or quadratic functions of the parameters and evidence values, and which of them appear as linear or quadratic terms, allows an efficient evaluation. Closed expressions for the partial derivatives of probabilities of the form $P(X_i > a|\mathbf{e})$, $P(X_i \leq a|\mathbf{e})$ and $P(a < X_i \leq b|\mathbf{e})$ with respect to the parameters can be obtained.

References

[1] Anderson, T. W. (1984), *An introduction to multivariate statistical analysis, 2nd edition.* John Wiley and Sons, New York.

[2] Castillo, E. *Extreme value theory in engineering.* Academic Press, New York, 1988.

[3] Castillo, E., Gutiérrez, J. M. and Hadi, A. S. *Expert systems and probabilistic network models.* Springer Verlag, New York, 1997.

[4] Castillo, E., Gutiérrez, J. M. and Hadi, A. S. Sensitivity analysis in discrete Bayesian networks. *IEEE Transactions on Systems, Man and Cybernetics*, Vol 26, N. 7, 412–423, 1997.

[5] Castillo, E., Gutiérrez, J. M., Hadi, A. S. and Solares, C. Symbolic propagation and sensitivity analysis in Gaussian Bayesian networks with application to damage assessment,. *Artificial Intelligence in Engineering*, 11:173–181, 1997.

[6] Castillo, E., Solares, C., and Gómez, P. Tail sensitivity analysis in Bayesian networks. In Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence (UAI'96), Portland (Oregon), Morgan Kaufmann Publishers, San Francisco, California, 133–140, 1996.

[7] Castillo, E., Solares, C., and Gómez, P. Estimating extreme probabilities using tail simulated data. *International Journal of Approximate Reasoning*, 1996.

[8] Castillo, E., Solares, C., and Gómez, P. High probability one-sided confidence intervals in reliability Models. *Nuclear Science and Engineering*, Vol. 126, 158–167, 1997.

[9] Castillo, E., Solares, C., and Gómez, P. Tail uncertainty analysis in complex systems. *Artificial Intelligence* 96(2), 395-419, 1997.

[10] Castillo, E., Sarabia, J. M., Solares, C., and Gómez, P. Uncertainty analyses in fault trees and Bayesian networks using FORM/SORM methods. *Reliability Engineering and System Safety (in press)*, 1998.

[11] Chang, K. C. and Fung, R. (1991), Symbolic probabilistic inference with continuous variables. In *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence.* Morgan Kaufmann Publishers, San Mateo, CA, 77–85.

[12] Darwiche, A. A differential approach to inference in Bayesian networks. UAI2000.

[13] Galambos, J. *The asymptotic theory of extreme order statistics.* Robert E. Krieger Publishing Company. Malabar, Florida, 1987.

[14] Kenley, C. R. (1986), *Influence diagram models with continuous variables.* Ph.D. Thesis, Stanford, Stanford University.

[15] Kjaerulff, U. and van der Gaag, L. C. Making sensitivity analysis computationally efficient. UAI2000.

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

[17] Lauritzen, S. L. (1992), Propagation of probabilities, means, and variances in mixed graphical association models. *Journal of the American Statistical Association*, 87:1098–1108.

[18] Normand, S.-L. and Tritchler, D. (1992), Parameter updating in Bayes network. *Journal of the American Statistical Association*, 87:1109–1115.

[19] M. Rosenblatt. Remarks on a multivariate transformation. *Ann. Math. Stat. 23(3)*,470–472, 1952.

[20] Shachter, R. and Kenley, C. (1989), Gaussian influence diagrams. *Management Science*, 35(5):527–550.

[21] Xu, L. and Pearl, J. (1989), Structuring causal tree models with continuous variables. In Kanal, L. N., Levitt, T. S., and Lemmer, J. F., editors, *Uncertainty in Artificial Intelligence 3.* North Holland, Amsterdam, 209–219.