Evaluating the difference between graph structures in Gaussian Bayesian networks

Miguel Ángel Gómez-Villegas [a] Paloma Main [a,*]
Hilario Navarro [b] Rosario Susi [c]

[a] *Dpto. Estadística e I.O., Fac. Ciencias Matemáticas, Univ. Complutense de Madrid, 28040-Madrid, Spain*

[b] *Dpto. de Estadística, I.O. y Cálc. Numérico, Fac. Ciencias, UNED, 28040-Madrid, Spain*

[c] *Dpto. Estadística e I.O. III, E.U. Estadística, Univ. Complutense de Madrid, 28040-Madrid, Spain*

Abstract

In this work, we evaluate the sensitivity of Gaussian Bayesian networks to perturbations or uncertainties in the regression coefficients of the network arcs and the conditional distributions of the variables. The Kullback-Leibler divergence measure is used to compare the original network to its perturbation. By setting the regression coefficients to zero or non-zero values, the proposed method can remove or add arcs, making it possible to compare different network structures. The methodology is implemented with some case studies.

Key words: Gaussian Bayesian networks, Conditional representation, Sensitivity analysis, Kullback-Leibler divergence measure.

Introduction

A Bayesian network (BN) is a probabilistic model of causal interactions between a set of variables, where the joint probability distribution is described in graphical terms. Probabilistic networks have become an increasingly popular

* Corresponding author.
Email address: pmain@mat.ucm.es (Paloma Main).

paradigm for reasoning through uncertain, complex models in a variety of situations, including AI, medical diagnosis and data mining (Kjærulff, Madsen and Anders 2008).

This model consists of two parts: one qualitative and the other quantitative. Its qualitative aspect is a directed, acyclic graph (DAG), with nodes and arcs that represent a set of variables and their relationships respectively. Based on the dependence structure depicted in the graph, the joint distribution of the variables can be factorized in terms of the univariate conditional distributions of each variable given its parents. These distributions constitute the quantitative portion of the model.

Building a BN is a difficult task, because all of the individual distributions and relationships between variables need to be correctly specified. Expert knowledge is essential to fix the dependence structure among variables of the network and to determine a large set of parameters. Databases can aid the process, but provide incomplete data and only partial knowledge of the domain. Thus, any assessments obtained using only databases are inevitably inaccurate (van der Gaag, Renooij and Coupé 2007).

The present research is restricted to a subclass of BNs known as Gaussian Bayesian networks (GBNs). The quantitative portion of a GBN consists of a univariate normal distribution for each variable given its parents in the DAG. Also, the joint probability distribution of the model is constrained to be a multivariate normal distribution.

For each variable X_i, the experts have to provide its mean, the regression coefficients between X_i and each parent $X_j \in Pa(X_i) \subset \{X_1, \ldots, X_{i-1}\}$, and the conditional variance of X_i given its parents in the DAG. This specification is easy for experts, because they only have to describe univariate distributions. Moreover, the arcs in the DAG can be expressed in terms of the regression coefficients.

Our interest in this paper is the sensitivity of GBNs defined by these parameters. This subject has not been frequently treated in the literature, because sensitivity analyses usually perturb the the joint parameters instead of the conditional parameters. However, it is easy to model the presence or absence of arcs by adopting regression coefficients different from or equal to zero. Thus, it is also possible to study the effect of changes in the qualitative part of the network. An objective evaluation of this effect may also reveal that a simpler DAG structure yields equivalent results.

In Section 1 we define GBNs, present some general concepts, and introduce a working example. In Section 2 we describe the methodology used to study the sensitivity of the GBN and calculate the sensitivity of our example. In Section 3 we vary the network structure and present a metrology example: the calibration of an electronic level using a sine table. The paper ends with some conclusions and suggestions for further research.

1 General Concepts

Throughout this work, random variables will be denoted by capital letters. Moreover, in the multidimensional case, boldface characters will be used.

A BN is defined by a pair $(\mathcal{G}, \mathcal{P})$, where $\mathcal{G}$ is the DAG with nodes representing variables and arcs showing the dependence structure. $\mathcal{P}$ is a set of conditional probability distributions P, each representing the distribution of one random variable X_i given all of its parents in the DAG. That is, $\mathcal{P} \equiv P(X_i \mid pa(X_i))$ $\forall\ X_i$, where $Pa(X_i) \subset \{X_1, \ldots, X_{i-1}\}$. When X_i has no incoming arcs (no parents), $P(X_i \mid pa(X_i))$ stands for the marginal $P(X_i)$.

The joint probability distribution of a BN can be defined in terms of the elements of $\mathcal{P}$, as the product of the conditional and marginal probability distributions:

$$P(\mathbf{X}) = \prod_{i=1}^{n} P(X_i \mid pa(X_i)). \tag{1}$$

BNs have been studied by authors such as Pearl (1988), Lauritzen (1996) and Jensen and (2007) among others.

It is common to consider BNs with discrete variables. Nevertheless, it is possible to work with some continuous distributions. For example, a GBN is a BN where all the variables of the model are Gaussian. Specifically, in a GBN the joint probability density of $\mathbf{X} = \{X_1, \ldots, X_n\}$ is a multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$f(\mathbf{x}) = (2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}\ . \tag{2}$$

Here $\boldsymbol{\mu}$ is the n-dimensional mean vector and $\boldsymbol{\Sigma}$ the $n \times n$ positive definite covariance matrix.

Alternatively, the joint density can be factorized using the conditional probability densities for every X_i $(i = 1, .., n)$ given its parents. These are univariate normal distributions, and can be obtained from the joint density as

$$f(x_i \mid pa(x_i)) \sim N(x_i \mid \mu_i + \sum_{j=1}^{i-1} \beta_{ij}(x_j - \mu_j), v_i)$$

where μ_i is the mean of X_i, β_{ji} are the regression coefficients of X_i with respect to $X_j \in Pa(X_i)$, and v_i is the conditional variance of X_i given its parents. Note that $\beta_{ji} = 0$ if and only if there is no link from X_j to X_i.

From the conditional specification it is also possible to get the parameters of the joint distribution. The means μ_i are the elements of the n-dimensional mean vector $\boldsymbol{\mu}$, and the covariance matrix $\mathbf{\Sigma}$ can be obtained with $\{v_i\}$ and $\left\{\beta_{ji}\right\}$ as follows. Let $\mathbf{D}$ be a diagonal matrix with the conditional variances $\mathbf{v}^T = (v_1, \ldots, v_n)$, $\mathbf{D} = diag(\mathbf{v})$. Let $\mathbf{B}$ be a strictly upper triangular matrix with the regression coefficients β_{ji} as columns. Then the covariance matrix $\mathbf{\Sigma}$ can be computed as

$$\mathbf{\Sigma} = [(\mathbf{I} - \mathbf{B})^{-1}]^T \mathbf{D} (\mathbf{I} - \mathbf{B})^{-1} \tag{3}$$

For details, see Shachter and Kenley (1989).

As we remarked in Section 1, the conditional parameters, $\{v_i\}$ and $\left\{\beta_{ji}\right\}$ may be easier to specify by experts than $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$. However, the two representations are completely equivalent when defining a GBN.

Now we introduce a working example of a GBN.

Example 1 *Our sample problem concerns the amount of time that a machine will work before failing. The machine is made up of 7 elements, each with its own random time to failure X_i $(i = 1, ..., 7)$. The elements are connected as shown in Figure 1; the regression coefficients between variables are written next to the arcs.*

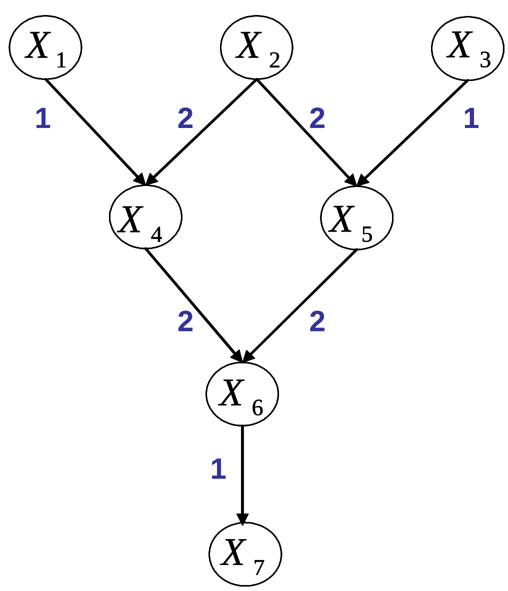

Figure 1 DAG of the GBN in Example 1

The time that each element continues working is given by a normal distribution. The joint probability distribution of $\mathbf{X} = \{X_1, X_2, ..., X_7\}$ is a multivariate normal distribution.

Experts with the machine give the following parameters:

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 3 \\ 2 \\ 1 \\ 4 \\ 5 \\ 8 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0\ 0\ 0\ 1\ 0\ 0\ 0 \\ 0\ 0\ 0\ 2\ 2\ 0\ 0 \\ 0\ 0\ 0\ 0\ 1\ 0\ 0 \\ 0\ 0\ 0\ 0\ 0\ 2\ 0 \\ 0\ 0\ 0\ 0\ 0\ 2\ 0 \\ 0\ 0\ 0\ 0\ 0\ 0\ 1 \\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \end{pmatrix}$$

and

$$\mathbf{D} = \begin{pmatrix} 1\ 0\ 0\ 0\ 0\ 0\ 0 \\ 0\ 1\ 0\ 0\ 0\ 0\ 0 \\ 0\ 0\ 2\ 0\ 0\ 0\ 0 \\ 0\ 0\ 0\ 1\ 0\ 0\ 0 \\ 0\ 0\ 0\ 0\ 4\ 0\ 0 \\ 0\ 0\ 0\ 0\ 0\ 1\ 0 \\ 0\ 0\ 0\ 0\ 0\ 0\ 2 \end{pmatrix}$$

where $\mathbf{B}$ is the strictly upper triangular $(j < i)$ matrix of regression coefficients β_{ji} given in the DAG, and $\mathbf{D}$ is the diagonal matrix of conditional variances $\mathbf{v}^T = (v_1, \ldots, v_n)$, $\mathbf{D} = diag(\mathbf{v})$.

Computing the joint parameters with (3) yields the equation $\mathbf{X} \sim N(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = \begin{pmatrix} 1 \\ 3 \\ 2 \\ 1 \\ 4 \\ 5 \\ 8 \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 2 & 2 \\ 0 & 1 & 0 & 2 & 2 & 8 & 8 \\ 0 & 0 & 2 & 0 & 2 & 4 & 4 \\ 1 & 2 & 0 & 6 & 4 & 20 & 20 \\ 0 & 2 & 2 & 4 & 10 & 28 & 28 \\ 2 & 8 & 4 & 20 & 28 & 97 & 97 \\ 2 & 8 & 4 & 20 & 28 & 97 & 99 \end{pmatrix}$$

We now have both representations of the GBN to work with.

2 Sensitivity in GBN

To build a BN is a difficult task, and expert knowledge is necessary to define the model. Furthermore, as discussed in the Introduction, even the parameters offered by experts may be inaccurate.

Sensitivity analysis is a general technique for evaluating the effects of inaccuracies in model parameters on the conclusions.

In a BN, the desired output is the marginal distribution of an interesting variable. This function is computed from the quantitative parameters that specify the BN. The result will be sensitive to inaccuracy in any parameter, but not all parameters require the same level of accuracy. Typically, some variables have more impact on the network's output than others (van der Gaag, Renooij and Coupé, 2007).

The model's sensitivity can be measured by varying one parameter while keeping the others fixed. This method is known as *one-way sensitivity analysis*. The output can also be studied while simultaneously changing a set of parameters, known as *n-way sensitivity analysis* (van der Gaag, Renooij and Coupé 2007).

In recent years, a few authors have published sensitivity analyses of BNs (Laskey, 1995; Coupé, van der Gaag and Habbema, 2000; Kjærulff and van der Gaag, 2000; Bednarsky, Cholewa and Frid, 2004; Chan and Darwiche, 2005). All of these works, while useful, deal exclusively with discrete BNs.

The sensitivity of GBNs has been studied only by Castillo and Kjaerulff (2003) and Gómez-Villegas, Main and Susi (2007, 2008). Castillo and Kjaerulff proposed a one-way sensitivity analysis (based on Laskey, 1995) investigating the impact of small changes in the network parameters.

Our first paper on this topic (Gómez-Villegas, Main and Susi 2007) is also a one-way sensitivity analysis. It differs from that of Castillo and Kjærulff (2003) in that we evaluated a global sensitivity measure rather than local aspects of the variable distributions such as location and dispersion. Specifically, our method was based on computing a divergence measure similar in spirit to that Chan and Darwiche (2005) but different in detail due to the variables considered. In our 2008 paper we extended this work to an n-way sensitivity analysis, studying the joint effect of a set of uncertain parameters on the network's output.

In the present paper, we shall undertake a model sensitivity analysis using the conditional specification of the network. Until now, all sensitivity analyses proposed for GBNs have studied variations in the elements of $\boldsymbol{\mu}$ and Σ. This research studies variations in a $\mathbf{D}$ and $\mathbf{B}$, that is, the arcs in the network and their regression coefficients.

Arcs can be eliminated or included by appropriate perturbations of the coefficients in $\mathbf{B}$. In Section 3, we shall study cases where one arc is removed from or added to the model. With the proposed n-way analysis, it will be possible to determine the sensitivity of each node to changes in the structure of the network.

To evaluate the difference between the original GBN and a perturbed networks, we compute the Kullback-Leibler (KL) divergence measure. In the next subsection, we present and justify our use of the KL divergence. In the following sections, we shall describe our methodology and the results of the n-way sensitivity analysis on our example GBN.

2.1 The Kullback-Leibler divergence

The KL divergence is an non-symmetric measure that evaluates the amount of information available to discriminate between two probability distributions. It is used because we wish to compare the global behaviors of two probability distributions (for more details, see Kullback and Leibler, 1951).

The KL divergence between the original probability density $f(w)$ and the perturbed density $f'(w)$, defined over the same domain, is given by

$$KL(f'(w) \mid f(w)) = \int_{-\infty}^{\infty} f(w) \ln \frac{f(w)}{f'(w)} dw \ . \tag{4}$$

For multivariate normal distributions, it is evaluated as follows:

$$KL(f, f') = \frac{1}{2}\left[\ln \frac{|\boldsymbol{\Sigma}'|}{|\boldsymbol{\Sigma}|} + tr\left(\boldsymbol{\Sigma}\boldsymbol{\Sigma}'^{-1}\right) - \dim(\mathbf{X})\right] + \frac{1}{2}\left[(\boldsymbol{\mu}' - \boldsymbol{\mu})^T \boldsymbol{\Sigma}'^{-1} (\boldsymbol{\mu}' - \boldsymbol{\mu})\right],$$

where f is the joint probability density of $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and f' is the joint probability density of $\mathbf{X}' \sim N(\boldsymbol{\mu}', \boldsymbol{\Sigma}')$.

2.2 Methodology and results

As mentioned above, sensitivity analysis compares the original model to a perturbed model.

The original model consists of parameters based on expert knowledge: the mean vector $\boldsymbol{\mu}$ and the matrices $\mathbf{B}$ and $\mathbf{D}$.

The perturbed model represents uncertainties in the parameters using a vector $\boldsymbol{\delta}$ and the matrices $\boldsymbol{\Delta}_{\mathbf{B}}$ and $\boldsymbol{\Delta}_{\mathbf{D}}$. Their elements are additive perturbations on the corresponding elements of $\boldsymbol{\mu}$, $\mathbf{B}$ and $\mathbf{D}$ respectively. Thus, $\boldsymbol{\Delta}_{\mathbf{B}}$ is an upper triangular matrix and $\boldsymbol{\Delta}_{\mathbf{D}}$ is a diagonal matrix.

In the sensitivity analysis, we shall consider three different perturbed models: one using $\boldsymbol{\delta}$, one using $\boldsymbol{\Delta}_{\mathbf{B}}$, and the last using $\boldsymbol{\Delta}_{\mathbf{D}}$. A separate KL divergence will be calculated for each model.

Let $(\mathcal{G}, \mathcal{P})$ be a GBN with parameters $\boldsymbol{\mu}$, $\mathbf{B}$ and $\mathbf{D}$, where $\boldsymbol{\mu}$ is the vector variables means and $\mathbf{B}$, $\mathbf{D}$ are matrices of the regression coefficients and conditional variances respectively. The joint distribution is $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

For a set of variations in the parameters $\boldsymbol{\mu}$, $\mathbf{B}$ or $\mathbf{D}$, the following result holds:

(1) When a perturbation $\boldsymbol{\delta}$ is added to the mean vector, we compare the original model $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with the perturbed model $N\left(\boldsymbol{\mu}^{\boldsymbol{\delta}}, \boldsymbol{\Sigma}\right)$, where $\boldsymbol{\mu}^{\boldsymbol{\delta}} \equiv \boldsymbol{\mu} + \boldsymbol{\delta}$. The KL divergence is

$$KL^{\boldsymbol{\mu}}(f^{\boldsymbol{\mu}} \mid f) = \frac{1}{2}\left[\boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta}\right] . \tag{5}$$

(2) When the perturbation $\boldsymbol{\Delta}_{\mathbf{B}}$ is added to $\mathbf{B}$, we compare the original model $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with the perturbed model $N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{\boldsymbol{\Delta}_{\mathbf{B}}}\right)$, where $\Sigma^{\boldsymbol{\Delta}_{\mathbf{B}}} \equiv \left[(\mathbf{I} - \mathbf{B} - \boldsymbol{\Delta}_{\mathbf{B}})^{-1}\right]^T \mathbf{D} (\mathbf{I} - \mathbf{B} - \boldsymbol{\Delta}_{\mathbf{B}})^{-1}$. The KL divergence is

$$KL^{\mathbf{B}}(f^{\mathbf{B}} \mid f) = \frac{1}{2}\left[trace\left(\boldsymbol{\Sigma} \boldsymbol{\Delta}_{\mathbf{B}} \mathbf{D}^{-1} \boldsymbol{\Delta}_{\mathbf{B}}^T\right)\right] . \tag{6}$$

(3) When the perturbation $\boldsymbol{\Delta}_{\mathbf{D}}$ is added to $\mathbf{D}$, we compare the original model $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with the perturbed model $N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}^{\boldsymbol{\Delta}_{\mathbf{D}}}\right)$, where $\Sigma^{\boldsymbol{\Delta}_{\mathbf{D}}} \equiv \left[(\mathbf{I} - \mathbf{B})^{-1}\right]^T (\mathbf{D} + \boldsymbol{\Delta}_{\mathbf{D}}) (\mathbf{I} - \mathbf{B})^{-1}$. The KL divergence is

$$KL^{\mathbf{D}}(f^{\mathbf{D}} \mid f) = \frac{1}{2}\left[\ln \frac{|\mathbf{D} + \boldsymbol{\Delta}_{\mathbf{D}}|}{|\mathbf{D}|} + trace((\mathbf{D} + \boldsymbol{\Delta}_{\mathbf{D}})^{-1} \mathbf{D}) - n\right] . \tag{7}$$

Equation (7) can be expressed in terms of the conditional variances v_i as

$$KL^{\mathbf{D}}(f^{\mathbf{D}} \mid f) = -\frac{1}{2}\left[\sum_{i=1}^{n}\left(\ln\left(1 - \frac{\delta_i}{v_i + \delta_i}\right) + \frac{\delta_i}{v_i + \delta_i}\right)\right] ,$$

where δ_i is an individual perturbation to the conditional variances of X_i.

Now we shall manipulate Equation (6) to arrive at a more useful expression. The following identities can be proven through standard linear algebra techniques:

$$(\boldsymbol{\Sigma})^{-1} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\left[(\mathbf{I} - \mathbf{B})^{-1}\right]^{T}\boldsymbol{\Delta}_{\mathbf{B}}^{T} - \boldsymbol{\Delta}_{\mathbf{B}}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Delta}_{\mathbf{B}}\mathbf{D}^{-1}\boldsymbol{\Delta}_{\mathbf{B}}^{\mathbf{T}}$$

and

$$\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}^{\boldsymbol{\Delta}_{\mathbf{B}}}\right)^{-1} = \mathbf{I}_n - \left[(\mathbf{I} - \mathbf{B})^{-1}\right]^{T}\boldsymbol{\Delta}_{\mathbf{B}}^{\mathbf{T}} - \boldsymbol{\Sigma}\boldsymbol{\Delta}_{\mathbf{B}}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}\boldsymbol{\Delta}_{\mathbf{B}}\mathbf{D}^{-1}\boldsymbol{\Delta}_{\mathbf{B}}^{\mathbf{T}} .$$

Operating with the traces, it follows that

$$\frac{1}{2}\left[trace\left(\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}^{\boldsymbol{\Delta}_{\mathbf{B}}}\right)^{-1}\right) - n\right] = \frac{1}{2}\left[trace\left(\boldsymbol{\Sigma}\boldsymbol{\Delta}_{\mathbf{B}}\mathbf{D}^{-1}\boldsymbol{\Delta}_{\mathbf{B}}^{\mathbf{T}}\right)\right] .$$

If there exist some inaccuracies in the conditional parameters describing a GBN, a sensitivity analysis can be undertaken using these expressions. By calculating the KL divergences for a range of perturbations, one can determine which parameters are most crucial to describing the network accurately. If the KL divergence is close to zero, for example, it is possible to conclude that the network is not sensitive to the proposed variation.

Our methodology evaluates the effects of simultaneous perturbations in a set of parameters. This tool opens up new lines of research, making it possible to trace the evolution of perturbations through different levels of the graph, or infer the behavior of an unknown variable from the values of evidential variables.

The next examples illustrate the procedure described in this section. We introduce two cases with different assumptions about the uncertain parameters.

Example 2 *Working with the GBN given in Example 1, let us say that experts disagree on the values of certain parameters: the mean of X_6 could be either 4 or 5, and the mean of X_7 could be either 7 or 8. Opinions also differ regarding the regression coefficients between X_4 and its parent X_2, and between X_5 and its parent X_2. Finally, the conditional variances of X_2, X_4 and X_5 might be different from those intially given. The perturbations $\boldsymbol{\delta}$, $\boldsymbol{\Delta}_{\mathbf{D}}$ and $\boldsymbol{\Delta}_{\mathbf{B}}$ are set* *to*

$$\boldsymbol{\delta}=\begin{pmatrix}0\\0\\0\\0\\0\\-1\\-1\end{pmatrix}\;\boldsymbol{\Delta}_{\mathbf{B}}=\begin{pmatrix}0&0&0&0&0&0&0\\0&0&0&-1&-1&0&0\\0&0&0&0&0&0&0\\0&0&0&0&0&0&0\\0&0&0&0&0&0&0\\0&0&0&0&0&0&0\\0&0&0&0&0&0&0\end{pmatrix}\;\boldsymbol{\Delta}_{\mathbf{D}}=\begin{pmatrix}0&0&0&0&0&0&0\\0&1&0&0&0&0&0\\0&0&0&0&0&0&0\\0&0&0&1&0&0&0\\0&0&0&0&1&0&0\\0&0&0&0&0&0&0\\0&0&0&0&0&0&0\end{pmatrix}$$

The perturbations are added to $\boldsymbol{\mu}$, $\mathbf{B}$ and $\mathbf{D}$ respectively, yielding three different perturbed models. The KL divergences between the original model and each perturbed model are given below.

$KL^{\boldsymbol{\mu}}(f^{\boldsymbol{\mu}} \mid f) = 0.5$

$KL^{\mathbf{B}}(f^{\mathbf{B}} \mid f) = 0.625$

$KL^{\mathbf{D}}(f^{\mathbf{D}} \mid f) = 0.204$

The values obtained are rather small because the KL divergence can be infinity but we can fairly say that the network is less sensitive to perturbations in the variances than in the means or regression coefficients.

The second example shows the effect of perturbing more parameters.

Example 3 *For the GBN in Example 1, the experts now disagree on the values of more parameters. Using* (i) *and* (i, j) *to denote specific elements of a vector and matrix respectively, we add the following perturbations to the* $\boldsymbol{\delta}$, $\boldsymbol{\Delta}_{\mathbf{D}}$ *and* $\boldsymbol{\Delta}_{\mathbf{B}}$ *of Example 2.*

$$\boldsymbol{\delta}(2) = \boldsymbol{\delta}(3) = 1; \boldsymbol{\Delta}_{\mathbf{B}}^{(4,6)} = \boldsymbol{\Delta}_{\mathbf{B}}^{(5,6)} = -1; \boldsymbol{\Delta}_{\mathbf{D}}^{(6,6)} = 2, \boldsymbol{\Delta}_{\mathbf{D}}^{(7,7)} = 3$$

The KL divergences are

$KL^{\boldsymbol{\mu}}(f^{\boldsymbol{\mu}} \mid f) = 4.375$

$KL^{\mathbf{B}}(f^{\mathbf{B}} \mid f) = 12.625$

$KL^{\mathbf{D}}(f^{\mathbf{D}} \mid f) = 0.579$

These values are significantly larger than those obtained in Example 2. We can conclude that the network is more sensitive to the new perturbations $\boldsymbol{\mu}$ and $\mathbf{B}$. However, the KL divergence is still rather small with respect to perturbation

$\mathbf{\Delta_D}$. Thus, the conditional variances do not require the same level of accuracy as $\boldsymbol{\mu}$ and $\mathbf{B}$.

Some other considerations concerning KL measure calibration will be dealt with in the case study defined below.

3 Perturbing the structure of a GBN

The regression coefficients $\left\{\beta_{ji}\right\}$ show the degree of association between X_i and its parents j. If $\beta_{ji} = 0$, there is no arc in the DAG from X_j to X_i. Therefore, it is possible to study variations of the structure of the DAG by perturbing the elements of the $\mathbf{B}$ matrix. When we change a coefficient β_{ji} to zero, the corresponding arc from X_j to X_i is removed. Likewise, changing a coefficient from zero to some other value introduces a new arc. Sometimes experts do not agree on the qualitative part of the model, in which case this form of analysis is necessary.

A word of caution, however: introducing a new arc can lead to cycles in the graph that are impossible to work with in the GBN framework.

By computing expression (6) for networks modified in this manner, we can also determine which relationships are essential to the original network. With this analysis we can take away relatively unimportant arcs until we obtain the simplest structure similar to the original model. This is the subject of the next example.

Example 4 *Removing only one arc, we want to find the resulting dependence structure closest to the GBN introduced in Example 1.*

We consider 4 different graphs, shown in panels (a), (b), (c) and (d) of Figure 2.

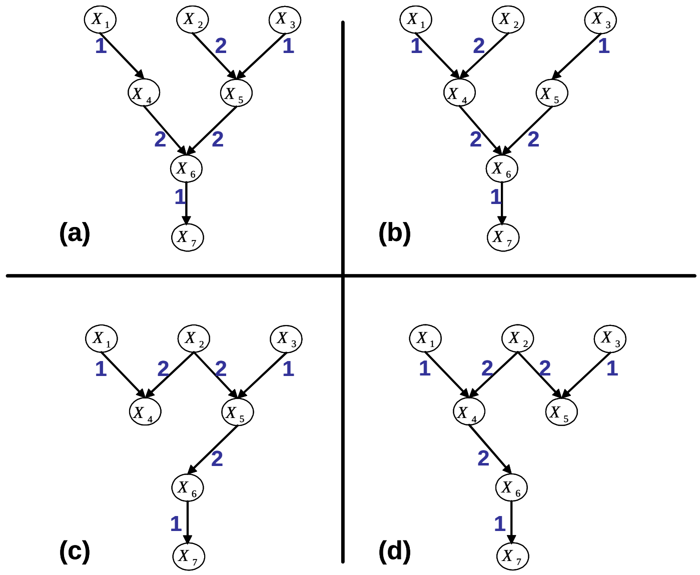

Figure 2 Four different one arc removed situations

The resulting KL divergences are given in Table 1.

	DKL	*from*		*to*
(a)	2	X_2	$\rightarrow$	X_4
(b)	0.5	X_2	$\rightarrow$	X_5
(c)	12	X_4	$\rightarrow$	X_6
(d)	20	X_5	$\rightarrow$	X_6

Table 1 KL divergences one arc removed

According to this result, only the model (b) could replace the original model given in Example 2. Thus, the dependence between X_2 and X_5 can be removed.

The analysis confirms that it is not possible to remove one of the arcs between X_6 and its parents, X_4 and X_5; the perturbed models (c) and (d) are very different from the original model because two of the variables no longer have any influence on X_7. The arc between X_2 and X_4 could be removed, but the KL divergence for this case is larger.

4 Case study

In metrology, uncertainties are essential when comparing measurements to each other or with reference values in a specification or standard. The most widely used method in this field is established in the book *Guide to the expression of uncertainty in measurement* (ISO GUM), which establishes general rules for evaluating and expressing uncertainty. In this setting, we now consider the effects of uncertainty in the calibration of small-angle measurement instruments (Piratelli-Filho and Di Giacomo 2003). Specifically, we examine the calibration of an electronic level using a sine table, as represented schematically in Figure 3.

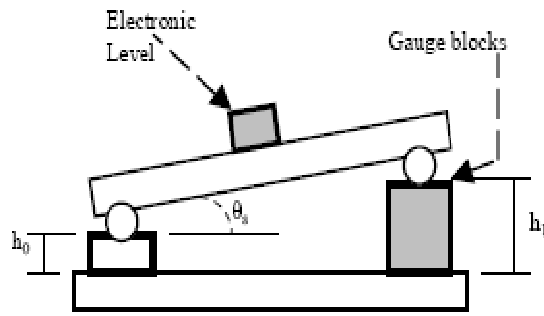

Figure 3 Experimental assembly at sine bar

Following the ISO GUM methodology, we determine the variables influencing the measurement and their relations. The measurement is then represented by a GBN. This is an approximation to the way the model is usually handled, allowing the possibility of determining the relative influences of the variables.

The variables involved in the measurement procedure are defined below.

ΔT_{20}: deviation of the temperature in the measuring room from the reference temperature 20 °C with the standard uncertainty $u_{\Delta T_{20}} = 0.41\,°\text{C}$

ΔT_{dif}: difference between the temperatures of the table and gauge block, with the standard uncertainty $u_{\Delta T_{dif}} = 0.029\,°\text{C}$

h_1: difference of the gauge block height from the standard, with the standard uncertainty $u_{h_1} = 0.025\,\mu\text{m}$

h_0: difference of the initial height on the sine table from the standard, with the standard uncertainty $u_{h_0} = 0.025\,\mu\text{m}$

L: difference in the length of sine table (distance between cylinders) from the standard, with the standard uncertainty $u_L = 0.52\,\mu\text{m}$

C: deviations of the sine bar cylinders from perfect roundness, with the standard uncertainty $u_C = 0.06\,\mu\text{m}$

R: bias associated with the instrument resolution, with the standard uncertainty $u_R = 0.29\,\mu\text{m}$

θ: difference between the measured angle from the standard, with the standard uncertainty $u_\theta = 0.033\,°\text{C}$.

Gaussian distributions are assumed for all variables, with the standard uncertainties interpreted as conditional standard deviations for each variable given its parents. The relationships between the variables are expressed by the DAG in Figure 4. The marginal distributions of the variables are computed using the software package Hugin Lite 7.2, a limited version of Hugin Developer / Hugin Researcher®. This program can be downloaded from www.hugin.com.

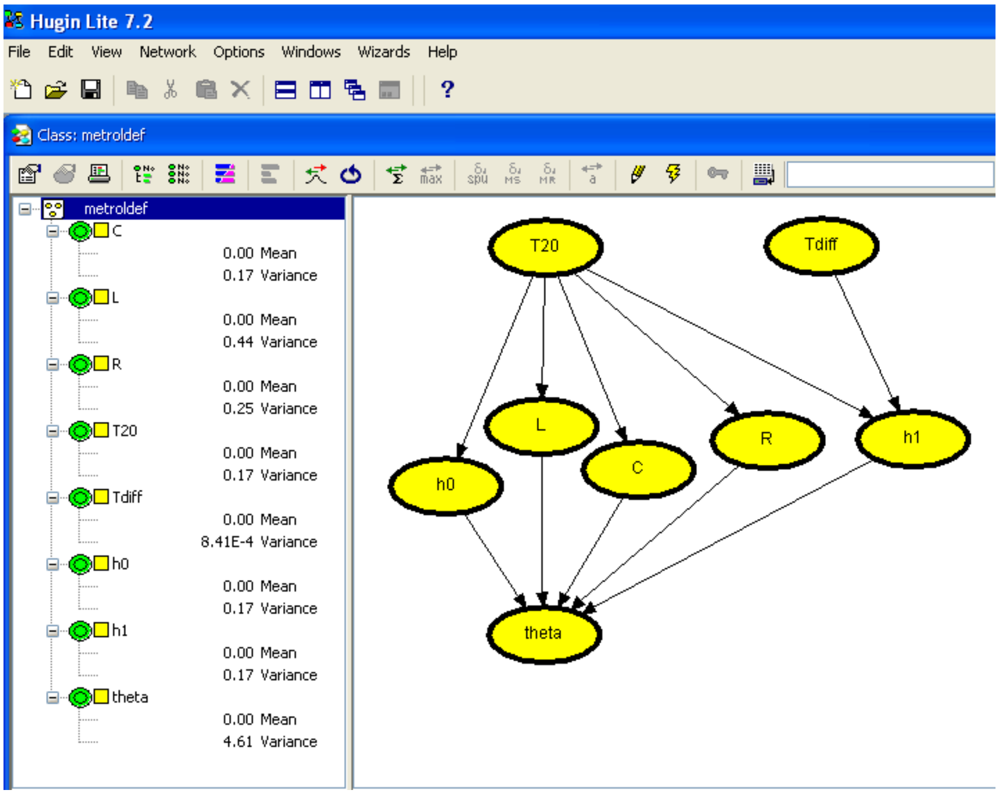

Figure 4 DAG of the small angle measurement case study

With these specifications and using Equation (3), the covariance matrix is given by

$$\Sigma = \begin{array}{c|l} \begin{pmatrix} 0.17 & 0 & 0.17 & 0.17 & 0.17 & 0.17 & 0.17 & 0.85 \\ 0 & 0.000841 & 0.000841 & 0 & 0 & 0 & 0 & 0.000841 \\ 0.17 & 0.000841 & 0.171466 & 0.17 & 0.17 & 0.17 & 0.17 & 0.851466 \\ 0.17 & 0 & 0.17 & 0.170625 & 0.17 & 0.17 & 0.17 & 0.850625 \\ 0.17 & 0 & 0.17 & 0.17 & 0.4404 & 0.17 & 0.17 & 1.1204 \\ 0.17 & 0 & 0.17 & 0.17 & 0.17 & 0.1736 & 0.17 & 0.8536 \\ 0.17 & 0 & 0.17 & 0.17 & 0.17 & 0.17 & 0.2541 & 0.9341 \\ 0.85 & 0.000841 & 0.851466 & 0.850625 & 1.1204 & 0.8536 & 0.9341 & 4.611280 \end{pmatrix} & \begin{matrix} \Delta T_{20} \\ \Delta T_{dif} \\ h_1 \\ h_0 \\ L \\ C \\ R \\ \theta \end{matrix} \end{array} .$$

Eliminating one arc at a time and computing the KL divergence with respect to the original model, we obtain the results in Table 2.

DKL	from		to
136	ΔT_{20}	$\rightarrow$	h_1
0.31435	ΔT_{20}	$\rightarrow$	L
23.6111	ΔT_{20}	$\rightarrow$	C
1.0107	ΔT_{20}	$\rightarrow$	R
136	ΔT_{20}	$\rightarrow$	h_0
0.6728	ΔT_{dif}	$\rightarrow$	h_0
78.34022	h_0	$\rightarrow$	θ
202.20386	L	$\rightarrow$	θ
79.70615	C	$\rightarrow$	θ
116.66667	R	$\rightarrow$	θ
78.72635	h_1	$\rightarrow$	θ

Table 2 KL divergences one arc removed

The most influential relationship is that between the length of the sine table and the measured angle. We can get a sense for the relative importance of this perturbation by calculating the maximum KL divergence obtained by removing more than one arc. The values and removed arcs are given in Table

3.

# of arcs	DKL	arcs removed
2	474.977	$L \to \theta \quad R \to \theta$
3	866.896	$L \to \theta \quad R \to \theta \quad C \to \theta$
4	1413.942	$L \to \theta \quad R \to \theta \quad C \to \theta \quad h_1 \to \theta$
5	2116.708	$L \to \theta \quad R \to \theta \quad C \to \theta \quad h_1 \to \theta \quad h_0 \to \theta$
6	2252.708	$L \to \theta \quad R \to \theta \quad C \to \theta \quad h_1 \to \theta$ $h_0 \to \theta \quad \Delta T_{20} \to h_1(h_0)$
7	2388.708	$L \to \theta \quad R \to \theta \quad C \to \theta \quad h_1 \to \theta$ $h_0 \to \theta \quad \Delta T_{20} \to h_0 \quad \Delta T_{20} \to h_1$
8	2412.32	$L \to \theta \quad R \to \theta \quad C \to \theta \quad h_1 \to \theta$ $h_0 \to \theta \quad \Delta T_{20} \to h_1 \quad \Delta T_{20} \to h_0 \quad \Delta T_{20} \to C$
9	2413.33	$L \to \theta \quad R \to \theta \quad C \to \theta \quad h_1 \to \theta$ $h_0 \to \theta \quad \Delta T_{20} \to h_1 \quad \Delta T_{20} \to h_0 \quad \Delta T_{20} \to C$ $\Delta T_{20} \to R$
10	2414	$L \to \theta \quad R \to \theta \quad C \to \theta \quad h_1 \to \theta$ $h_0 \to \theta \quad \Delta T_{20} \to h_1 \quad \Delta T_{20} \to h_0 \quad \Delta T_{20} \to C$ $\Delta T_{20} \to R \quad \Delta T_{dif} \to h_1$

Table 3 maximum KL divergences 2,...,10 arcs removed

The case of complete independence between the variables (that is, removing all eleven arcs) diverges from the initial network by 2414.317. Figure 5 plots the KL divergence in Table 3 against the number of arcs removed, normalizing to this value.

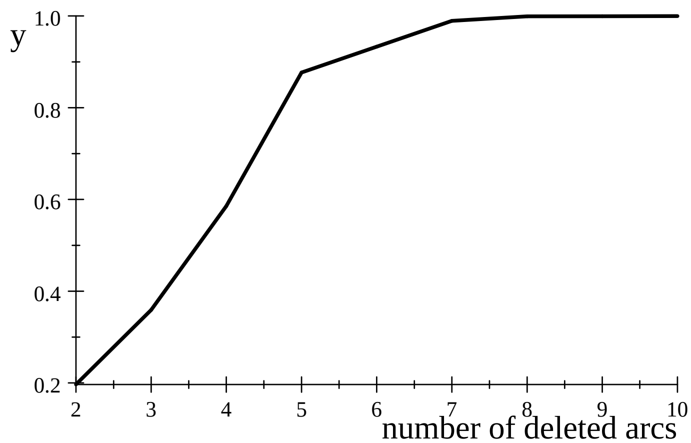

Figure 5 relative maximum DKL with the number of deleted arcs

Of course, it is intuitively obvious that the original network is nearer to a completely connected network than to an independent network. Another useful benchmark is the KL divergence between the original network and a completely connected graph with the same eight nodes and all arcs having coefficients equal to 1. This result is 588.4907.

5 Conclusions

This paper contributes to the problem of sensitivity analysis in GBNs in three ways. First, it describes how to characterize uncertainty in the conditional specifications of a GBN. Second, it explains how to analyze the network's sensitivity to perturbations in the network parameters (means, conditional variances, and the regression coefficients between a variable and its parents in the DAG). Third, it proposes a method of n-way sensitivity analysis that provides a global vision of the difference between the original network and its perturbation. We evaluate and discuss the proposed sensitivity analysis with an example GBN and several cases of uncertainty.

An important use of the model is evaluating the network's sensitivity to structural variations. By replacing regression coefficients with zero or non-zero values, the new method can remove or add arcs in the DAG. The results are applied to a metrology case study: the calibration of an electronic level using a sine table.

Further research will focus on applying the previous results to establish the sensitivity of a network to specific nodes in the DAG. Another interesting extension to the model is including prior evidence on some of the variables; by this means we can evaluate the effect of perturbations on evidence propagation.

6 Acknowledgements

This research has been supported by the Spanish *Ministerio de Ciencia e Innovación*, Grant MTM 2008–03282, and partially by GR58/08-A, 910395 - *Métodos Bayesianos* by the BSCH-UCM from Spain.

References

[1] Bednarski M, Cholewa W, Frid W. Identification of sensitivities in Bayesian networks. Engineering Applications of Artificial Intelligence 2004;17:327-335.

[2] Castillo E, Kjærulff U. Sensitivity analysis in Gaussian Bayesian networks using a symbolic-numerical technique. Reliability Engineering and System Safety 2003;79:139-148.

[3] Chan H, Darwiche A. A distance Measure for Bounding Probabilistic Belief Change. International Journal of Approximate Reasoning 2005;38(2):149-174.

[4] Coupé VMH, van der Gaag LC, Habbema JDF. Sensitivity analysis: an aid for belief-network quantification. The Knowledge Engineering Review 2000;15(3):215-232.

[5] Gómez-Villegas MA, Main P, Susi R. Sensitivity Analysis in Gaussian Bayesian Networks Using a Divergence Measure. Communications in Statistics: Theory and Methods 2007;36(3):523-539.

[6] Gómez-Villegas MA, Main P, Susi R. Sensitivity of Gaussian Bayesian Networks to Inaccuracies in Their Parameters. Proceedings of the 4th European Workshop on Probabilistic Graphical Models. Hirtshals, Denmark: 2008: 145-152.

[7] Hugin, http://www.hugin.com/

[8] Jensen FV, Lauritzen SL, Olensen KG. Bayesian updating in causal probabilistic networks by local computation. Computational Statistics Quarterly 4 (1990) 269-282.

[9] Kjærulff U, van der Gaag LC. Making Sensitivity Analysis Computationally Efficient. In Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence. San Francisco, CA, USA: 2000:315-325.

[10] Kjaerulff, Uffe B., Madsen, Anders L. Bayesian Networks and Influence Diagrams. A Guide to Construction and Analysis. New York: Springer; 2008.

[11] Kullback S, Leibler RA. On Information and Sufficiency. Annals of Mathematical Statistics 1951;22:79-86.

[12] Laskey KB. Sensitivity Analysis for Probability Assessments in Bayesian Networks. IEEE Transactions on Systems, Man and Cybernetics 1995;25:901-909.

[13] Lauritzen SL, Spiegelhalter DJ. Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems. Journal of the Royal Statistical Society, Series B 1988;50(2):157-224.

[14] Lauritzen SL. Graphical Models. Oxford: Clarendon Press; 1996.

[15] Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Mateo, CA: Morgan Kaufmann; 1988.

[16] Piratelli-Filho, A. Di Giacomo, B. Uncertainty evaluation in small angle calibration using ISO GUM approach and Monte Carlo method. XVII IMEKO World Congress Metrology in the 3^{rd} MIllenium 2003.

[17] Shachter R, Kenley C. Gaussian influence diagrams. Management Science 1989; 35:527-550.

[18] van der Gaag LC, Renooij S, Coupé VMH. Sensitivity analysis of probabilistic networks. StuddFuzz 2007; 213: 103-124.