Dealing with uncertainty in Gaussian Bayesian networks from a regression perspective

Miguel A. Gómez-Villegas[a], Paloma Main[a], Hilario Navarro[b] and Rosario Susi[a]
[a]Universidad Complutense Madrid, Spain
[b]Universidad Nacional de Educación a Distancia, Spain

Abstract

Some sensitivity analyses have been developed to evaluate the impact of uncertainty about the mean vector and the covariance matrix that specify the joint distribution of the variables in the nodes of a Gaussian Bayesian network (GBN). Nevertheless, uncertainty about the alternative conditional specification of GBN based on the regression coefficients of each variable given its parents in the directed acyclic graph (DAG), has received low attention in the literature. In this line, we focus on evaluating the effect of regression coefficients misspecification by means of the Kullback-Leibler (KL) divergence.

1 Introduction

GBNs are defined as Bayesian networks (BNs) where the joint probability density of $\mathbf{X} = (X_1, X_2, ..., X_p)^T$ is a multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}$ the $p-$dimensional mean vector and $\boldsymbol{\Sigma}$ the $p \times p$ positive definite covariance matrix using a directed acyclic graph (DAG) to represent the dependence structure of the variables.

As in BNs, the joint density can be factorized using the conditional probability densities of X_i $(i = 1, ..., p)$ given its parents in the DAG, $pa(X_i) \subset \{X_1, ..., X_{i-1}\}$. These are univariate normal distributions with densities

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

being μ_i the marginal mean of X_i, b_{ji} the regression coefficients of X_i given $X_j \in pa(X_i)$, and v_i the conditional variance of X_i given its parents in the DAG. Note that if $b_{ji} = 0$ then X_j is not a parent of X_i.

The parameters of the joint distribution can be obtained from the previous conditional specification. More concretely, the means $\{\mu_i\}$ are obviously the elements of the $p-$dimensional mean vector $\boldsymbol{\mu}$ and

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

(Shachter and Kenley, 1989) where $\mathbf{D}$ is a diagonal matrix $\mathbf{D} = diag(\mathbf{v})$ with the conditional variances $\mathbf{v}^T = (v_1, ..., v_p)$ and $\mathbf{B}$ a strictly upper triangular matrix with the regression coefficients $b_{ji}, j \in \{1, ..., i-1\}$.

The specification based on $(\boldsymbol{\mu}, \mathbf{B}, \mathbf{D})$ is more manageable for experts because they only have to describe univariate distributions. Moreover, the DAG can be improved by adding the numerical values of the regression coefficient and conditional variance to the corresponding arc and node, respectively.

Nevertheless, there can still be considerably uncertainty about parameters. Sensitivity analysis is an important phase of any modelling procedure. (Castillo and Kjærulff, 2003) performed a one-way sensitivity analysis investigating the impact of small changes in the network parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Alternatively, (Gómez-Villegas, Main and Susi, 2007) proposed a one-way global sensitivity analysis instead of considering local aspects as location and dispersion, over the network's output. Also, in (Gómez-Villegas, Main and Susi, 2008) a $n-$way sensitivity analysis is presented as a generalization of

[1] *E-mail addresses:* ma.gv@mat.ucm.es (M.A. Gómez-Villegas), pmain@mat.ucm.es (P. Main), hnavarro@ccia.uned.es (H. Navarro), rsusi@estad.ucm.es (R. Susi)

the previous one both using the KL divergence to evaluate the impact of perturbations.

It is well known the KL divergence is an non-symmetric measure that evaluates the amount of information available to discriminate between two probability distributions. We have chosen it because we want to compare the global behaviors of two probability distributions.

The directed KL divergence between the probability densities $f(w)$ and $f'(w)$, defined over the same domain is

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

The expression for multivariate normal distributions is given by

$$\begin{aligned} &D_{KL}(f' \mid f) = \\ &= \frac{1}{2}\left[\ln \frac{|\mathbf{\Sigma}'|}{|\mathbf{\Sigma}|} + tr\left(\mathbf{\Sigma}\left(\mathbf{\Sigma}'\right)^{-1} - \mathbf{I}_p\right)\right] + \\ &+ \frac{1}{2}\left[\left(\boldsymbol{\mu}' - \boldsymbol{\mu}\right)^T \left(\mathbf{\Sigma}'\right)^{-1} \left(\boldsymbol{\mu}' - \boldsymbol{\mu}\right)\right], \end{aligned}$$

where f and f' are densities of normal distributions $N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$ and $N_p(\boldsymbol{\mu}', \mathbf{\Sigma}')$ respectively.

In general, if $\mathbf{X} = (X_1, X_2, ..., X_p)^T$ is a random vector normally distributed with parameters $(\mathbf{0}, \mathbf{\Sigma})$ where the covariance matrix is inaccurately specified, the effect of a perturbation $\mathbf{\Lambda}_{p\times p}$ measured in terms of a directed Kullback-Leibler divergence can be expressed as follows

$$D_{KL}(f|f^{\mathbf{\Sigma}}) = \frac{1}{2}[\ln \frac{|\mathbf{\Sigma}|}{|\mathbf{\Sigma}+\mathbf{\Lambda}|} + tr(\mathbf{\Sigma}^{-1}(\mathbf{\Sigma}+\mathbf{\Lambda}))]$$

$$= -\frac{1}{2}[\ln |\mathbf{I}_p + \mathbf{\Lambda}\mathbf{\Sigma}^{-1}| - tr(\mathbf{\Lambda}\mathbf{\Sigma}^{-1})] \leq$$

$$\leq \frac{1}{4} tr(\mathbf{\Sigma}^{-1}\mathbf{\Lambda})^2 = \frac{1}{4}||\mathbf{\Sigma}^{-1}\mathbf{\Lambda}||_F^2 =$$

$$= \frac{1}{4}\sum_{i=1}^{p} \boldsymbol{\lambda}_{\mathbf{i}}^{\mathbf{T}} (\mathbf{\Sigma}^{-1})^{\mathbf{2}} \boldsymbol{\lambda}_{\mathbf{i}}$$

being $f^{\mathbf{\Sigma}}$ the density function with the perturbed covariance matrix $\mathbf{\Sigma}+\mathbf{\Lambda}$, $\boldsymbol{\lambda}_{\mathbf{i}}$ $(i = 1, ..., p)$ each of the columns of $\mathbf{\Lambda}$, with the necessary restrictions to get symmetric and positive definite matrices $\mathbf{\Sigma}+\mathbf{\Lambda}$ and $\mathbf{\Lambda}\mathbf{\Sigma}^{-1}$, $||\cdot||_F$ the Frobenius matrix norm and $tr(\cdot)$ the trace function.

However, as the directed KL divergence $D_{KL}(f' \mid f)$ can be interpreted as the information lost when f' is used to approximate f, in the following sensitivity analyses f has to be the original model and f' the perturbed one opposite to previously used divergence.

Herein we focus on the repercussion of a misspecified $\mathbf{B}_{p\times p}$ matrix, while the rest of the conditional parameters are known. To evaluate perturbation effects, the proper directed KL divergence is used in all the studied cases.

The paper is organized as follows. In Section 2 the problem is stated and analyzed for constant errors. In Section 3 random perturbations are considered; some examples illustrate the behavior of the proposed measure for both local and global uncertainty.

2 Fixed misspecification of B

Let f be the density of a multivariate normal distribution $N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$ with conditional parameters $\boldsymbol{\mu}, \mathbf{B}$ and $\mathbf{D}$. Denoting by $\mathbf{\Delta}_B$ the matrix with additive perturbations on $\mathbf{B}$ and $f^{\mathbf{B}}$ the corresponding perturbed density $N_p\left(\boldsymbol{\mu}, \mathbf{\Sigma}^{\mathbf{\Delta_B}}\right)$, where

$$\mathbf{\Sigma}^{\mathbf{\Delta_B}} \equiv \left[(\mathbf{I}_p - \mathbf{B} - \mathbf{\Delta_B})^{-1}\right]^T \mathbf{D}(\mathbf{I}_p - \mathbf{B} - \mathbf{\Delta_B})^{-1}$$

(see (Susi, Navarro, Main and Gómez-Villegas, 2009)), the KL divergence for comparing two covariance matrices when the means are equal is

$$D_{KL}^{\mathbf{B}}\left(f^{\mathbf{B}} \mid f\right) = \frac{1}{2}\left[trace\left(\mathbf{\Sigma}\left(\mathbf{\Sigma}^{\mathbf{\Delta_B}}\right)^{-1}\right) - p\right]. \tag{1}$$

It should be noted that as $\mathbf{I}_p - \mathbf{B}$ and $\mathbf{I}_p - \mathbf{B} - \mathbf{\Delta_B}$ are upper triangular matrices with diagonal entries equal to one then $\ln \frac{|\mathbf{\Sigma}^{\mathbf{\Delta_B}}|}{|\mathbf{\Sigma}|} = 0$.

Now, given that

$$\begin{aligned} \mathbf{\Sigma}\left(\mathbf{\Sigma}^{\mathbf{\Delta_B}}\right)^{-1} &= \mathbf{I}_p - \left[(\mathbf{I}_p - \mathbf{B})^{-1}\right]^T \mathbf{\Delta}_{\mathbf{B}}^{\mathbf{T}} + \\ &- \mathbf{\Sigma}\mathbf{\Delta_B}(\mathbf{I}_p - \mathbf{B})^{-1}\mathbf{\Sigma}^{-1} + \mathbf{\Sigma}\mathbf{\Delta_B}\mathbf{D}^{-1}\mathbf{\Delta}_{\mathbf{B}}^{\mathbf{T}} \end{aligned}$$

and

$$\begin{aligned} &trace\left(\left[(\mathbf{I}_p - \mathbf{B})^{-1}\right]^T \mathbf{\Delta}_{\mathbf{B}}^{\mathbf{T}}\right) = \\ &= trace\left(\mathbf{\Delta_B}(\mathbf{I}_p - \mathbf{B})^{-1}\right), \end{aligned}$$

$$trace\left(\mathbf{\Sigma\Delta_B}(\mathbf{I}_p-\mathbf{B})^{-1}\mathbf{\Sigma}^{-1}\right)=$$
$$=trace\left(\mathbf{\Sigma}^{-1}\mathbf{\Sigma\Delta_B}(\mathbf{I}_p-\mathbf{B})^{-1}\right)$$

the divergence in (1) can be restored as

$$\begin{aligned}
&D_{KL}^{\mathbf{B}}\left(f^{\mathbf{B}}\mid f\right)= \\
&=\frac{1}{2}[p-2trace\left(\mathbf{\Delta_B}(\mathbf{I}_p-\mathbf{B})^{-1}\right)+ \\
&+trace\left(\mathbf{\Sigma\Delta_B D}^{-1}\mathbf{\Delta_B^T}\right)-p]= \\
&=\frac{1}{2}\left[tr\left(\mathbf{\Sigma\Delta_B D}^{-1}\mathbf{\Delta_B^T}\right)\right]= \\
&=\frac{1}{2}\left[tr\left(\mathbf{\Delta_B^T\Sigma\Delta_B D}^{-1}\right)\right]. \qquad (2)
\end{aligned}$$

Note that $(\mathbf{I}_p-\mathbf{B})^{-1}$ is an upper triangular matrix with diagonal entries equal to one and $\mathbf{\Delta_B}(\mathbf{I}_p-\mathbf{B})^{-1}$ is an upper triangular matrix with diagonal entries zero.

Under the same assumptions, if $\boldsymbol{\delta}_{(i)}$ denotes an $(i-1)$-dimensional vector of *local* errors in node i —produced by an erroneous estimation or elicitation of the node i with its parents relationships, the perturbation matrix on $\mathbf{B}$, with only this error source is

$$\mathbf{\Delta_{B,i}}=\begin{pmatrix}
 & & & \boxed{\mathbf{i}} & & \\
0 & \cdots & 0 & \boldsymbol{\delta}_{(i)1} & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & 0 & \boldsymbol{\delta}_{(i)i-1} & \cdots & 0 \\
0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 0
\end{pmatrix}.$$

Thus, denoting $f^{\mathbf{B},i}$ as the density with coefficients matrix $\mathbf{B}+\mathbf{\Delta_{B,i}}$, the effect on the joint distribution can be expressed by

$$D_{KL}^{\mathbf{B},i}\left(f^{\mathbf{B},i}\mid f\right)=\frac{1}{2}\left[tr\left(\mathbf{\Delta_{B,i}^T\Sigma\Delta_{B,i}D}^{-1}\right)\right]=$$
$$=\frac{1}{2v_i}\sum_{k=1}^{i-1}\boldsymbol{\delta}_{(i)k}tr\left(\boldsymbol{\sigma}_{(i-1)k}\left(\boldsymbol{\delta}_{(i)k}\right)^T\right), \qquad (3)$$

$i=2,....,p$, being $\boldsymbol{\sigma}_{(i-1)k}$ the *k-th* column of the submatrix $\mathbf{\Sigma}_{(i-1)}$ built with the first $i-1$ rows and columns of $\mathbf{\Sigma}$. The local perturbation divergence in (3) also may be written in the form:

$$\begin{aligned}
D_{KL}^{\mathbf{B},i}\left(f^{\mathbf{B},i}\mid f\right)&=\frac{1}{2v_i}\sum_{k=1}^{i-1}\boldsymbol{\delta}_{(i)k}\left\langle\boldsymbol{\sigma}_{(i-1)k},\boldsymbol{\delta}_{(i)}\right\rangle= \\
&=\frac{1}{2v_i}\boldsymbol{\delta}_{(i)}^T\mathbf{\Sigma}_{(i-1)}\boldsymbol{\delta}_{(i)}= \\
&=\frac{1}{2v_i}\left\|\mathbf{U}_{(i-1)}\boldsymbol{\delta}_{(i)}\right\|_2^2
\end{aligned}$$

with $\mathbf{U}=(\mathbf{I}_p-\mathbf{B})^{-1}$ and the submatrix $\mathbf{U}_{(i-1)}$ determined by the first $i-1$ rows and columns of $\mathbf{U}$.

Returning to the measure of interest, $D_{KL}^{\mathbf{B}}\left(f^{\mathbf{B}}\mid f\right)$ in (2), it can be immediately obtained that

$$D_{KL}^{\mathbf{B}}\left(f^{\mathbf{B}}\mid f\right)=\sum_{i=2}^{p}D_{KL}^{\mathbf{B},i}\left(f^{\mathbf{B},i}\mid f\right)=$$
$$=\frac{1}{2}\sum_{i=2}^{p}\frac{1}{v_i}\boldsymbol{\delta}_{(i)}^T\mathbf{\Sigma}_{(i-1)}\boldsymbol{\delta}_{(i)}. \qquad (4)$$

Then, the total effect can be expressed as the sum of individual effects and consequently, the global sensitivity analysis can be performed through local analyses of nodes. Some direct results may be useful in applications:

- If all the components in $\boldsymbol{\delta}_{(i)}$ are equal to τ, it follows
$$D_{KL}^{\mathbf{B},i}\left(f^{\mathbf{B},i}\mid f\right)=\frac{\tau^2}{2v_i}\sum_{k=1}^{i-1}\sum_{t=1}^{i-1}\sigma_{tk}$$
- The possibly erroneous arc deletion from node j to node i having $b_{ji}\neq 0$, would yield
$$D_{KL}^{\mathbf{B},i}\left(f^{\mathbf{B},i}\mid f\right)=\frac{1}{2v_i}b_{ji}^2\sigma_{jj}$$
- The effect of arc inclusion from node j to node i introducing b_{ji}^* is also
$$D_{KL}^{\mathbf{B},i}\left(f^{\mathbf{B},i}\mid f\right)=\frac{1}{2v_i}b_{ji}^{*2}\sigma_{jj}$$

These last two cases describe the impact of the conditional and marginal variances for an uncertain knowledge of the exact model giving the distance between DAGs obtained by adding or removing arcs.

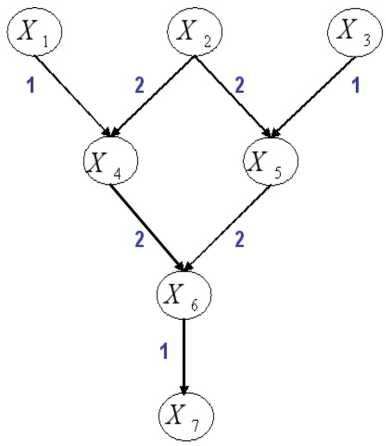

Figure 1: DAG with regression coefficients on the arcs

Example 1. Let us consider the GBN in Figure1 with parameters:

$$\boldsymbol{\mu} = \mathbf{0}, \mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{D} = diag\,(1,1,2,1,4,1,2)\,.$$

Using that

$$\boldsymbol{\Sigma} = [(\mathbf{I}_7 - \mathbf{B})^{-1}]^T \mathbf{D} (\mathbf{I}_7 - \mathbf{B})^{-1}$$

the covariance matrix is

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

The divergences reflecting the effect of each arc removal are shown in Figure 2. It is observed that divergence increases as the depth of the node grows with a maximum in the arc from X_6 to X_7. Therefore, it gives us information about

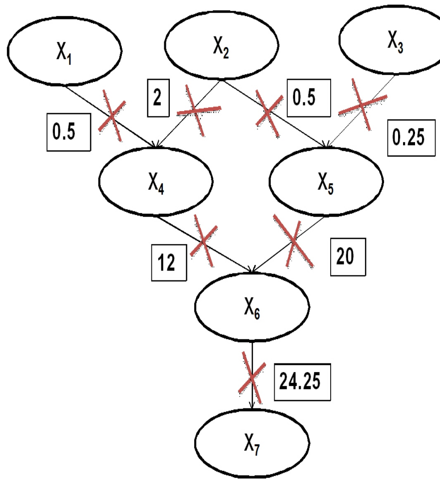

Figure 2: Measure of deviation for each arc removal

the difference between the original GBN and the particular networks obtained by means of the cancellation of some regression coefficients.

3 The random case

The expression (4) relates easily, global effect to local errors effects due to no interaction. Now, we are going to replace the hypothesis that $\boldsymbol{\Delta}_{\mathbf{B}}$ is known by the assumption that it is a random matrix. The main aim is using the relation (4) to evaluate the impact of uncertainty in $\mathbf{B}$ for a GBN. Both this measure and its value for different nodes can be useful to point the most sensitive nodes so as to compare structures with respect to uncertainty sensitivity.

If we suppose each vector $\boldsymbol{\delta}_{(i)}$ is distributed independent from the remaining errors as $N_{i-1}\left(\mathbf{0}, \mathbf{E}_i\right), i = 2, ..., p$, while the common value $v_1 = v_2 = \cdots = v_p = v$ is known, each random variable $\boldsymbol{\delta}_{(i)}^T \boldsymbol{\Sigma}_{(i-1)} \boldsymbol{\delta}_{(i)}$ is a quadratic form in normal variables with a chi-squared distribution with $i-1$ degrees of freedom $\chi^2_{(i-1)}$, if and only if $\mathbf{E}_i \boldsymbol{\Sigma}_{(i-1)}$ is a symmetric idempotent matrix of rank $i-1$ (Rencher, 2000).

Also, if $\mathbf{Y}$ is a random vector with mean vector $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$ and $\mathbf{M}_{p \times p}$ is a non-random matrix, then (Rencher, 2000)

$$E(\mathbf{Y}^T \mathbf{M} \mathbf{Y}) = \boldsymbol{\mu}^T \mathbf{M} \boldsymbol{\mu} + tr(\mathbf{M}\boldsymbol{\Sigma})$$

$$Var(\mathbf{Y}^T\mathbf{M}\mathbf{Y}) = 4\boldsymbol{\mu}^T\mathbf{M}\boldsymbol{\Sigma}\boldsymbol{\mu}+2tr((\mathbf{M}\boldsymbol{\Sigma})^2). \quad (5)$$

Thus, it results that the mean of the random variable $D^{\mathbf{B}}_{KL}\left(f^{\mathbf{B}} \mid f\right)$ is

$$E\left[D^{\mathbf{B}}_{KL}\left(f^{\mathbf{B}} \mid f\right)\right] = \frac{1}{2v}\sum_{i=2}^{p} tr\left(\Sigma_{(i-1)}\mathbf{E}_i\right).$$

If we express the covariance matrix $\mathbf{E}_i$ in terms of the difference with $\Sigma^{-1}_{(i-1)}$, that is

$$\mathbf{E}_i = \Sigma^{-1}_{(i-1)} + \mathbf{A}_i \ ,$$

then

$$tr\left(\Sigma_{(i-1)}\mathbf{E}_i\right) = tr\left(I_{(i-1)} + \Sigma_{(i-1)}\mathbf{A}_i\right) = \\ = i-1+tr\left(\Sigma_{(i-1)}\mathbf{A}_i\right).$$

It follows immediately that for a positive semidefinite matrix $\mathbf{A}_i$ —it could be denoted by $\mathbf{E}_i$ "larger than" $\Sigma^{-1}_{(i-1)}$— the minimum mean impact is obtained when $\mathbf{A}_i = \mathbf{0}$ being $k = \frac{1}{2v}\frac{p}{2}(p-1)$ a lower bound for the mean impact of all the matrices $\mathbf{E}_i$ larger than $\Sigma^{-1}_{(i-1)}$. Moreover, imposing $\mathbf{E}_i = \Sigma^{-1}_{(i-1)}$ we could assure each summand is distributed as a $\frac{1}{2v}\chi^2_{(i-1)}$ because $\mathbf{E}_i\boldsymbol{\Sigma}_{(i-1)} = \mathbf{I}_{i-1}$ is a symmetric idempotent matrix of rank $i-1$. The lower bound k can be considered to define the mean relative sensitivity by

$$\frac{E\left[D^{\mathbf{B}}_{KL}\left(f^{\mathbf{B}} \mid f\right)\right]}{k} = \frac{\sum_{i=2}^{p} tr\left(\Sigma_{(i-1)}\mathbf{E}_i\right)}{\frac{p}{2}(p-1)},$$

whenever $\mathbf{E}_i$ "larger than" $\Sigma^{-1}_{(i-1)}$ could be assumed.

3.1 Independent errors

When the hypothesis of independent errors with common variance ϵ^2 can be accepted, that is

$$\mathbf{E}_i = \epsilon^2\mathbf{I}_{(i-1)}, \ i = 2, ..., p \ ,$$

using (5), the quadratic forms of each component have first and second order moments given by:

- $E\left[\boldsymbol{\delta}^T_{(i)}\boldsymbol{\Sigma}_{(i-1)}\boldsymbol{\delta}_{(i)}\right] = \epsilon^2 tr\left(\boldsymbol{\Sigma}_{(i-1)}\right)$
- $Var\left[\boldsymbol{\delta}^T_{(i)}\boldsymbol{\Sigma}_{(i-1)}\boldsymbol{\delta}_{(i)}\right] = 2\epsilon^4 tr[(\boldsymbol{\Sigma}_{(i-1)})^2]$

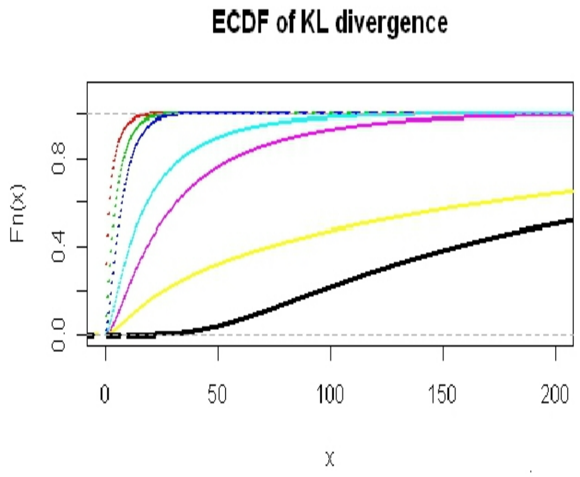

Figure 3: Node divergence from node 2 *(red)* to node 7 *(yellow)* and global divergence *(black)*

Consequently, under the stated conditions, we obtain an increasing average effect according to node depth in the network, independently of the network specification. Figure 3 illustrates the random behavior of Kullback-Leibler divergences displaying the empirical cumulative distribution function (ECDF) for samples from

$$\frac{1}{2v}\boldsymbol{\delta}^T_{(i)}\boldsymbol{\Sigma}_{(i-1)}\boldsymbol{\delta}_{(i)} \ , \ i = 2, ..., 7 \ ,$$

as well as the random global effect for the example discussed above. We have used a simulated sample of size 100, 000, with $v = 1$ and $\epsilon^2 = 5$, for each case.

In this setting, a reasonable procedure to evaluate sensitivity to uncertainty in $\mathbf{B}$ is to analyze the normalized ratio

$$S^{\mathbf{B}}(f) \equiv D^{\mathbf{B}}_{KL}\left(f^{\mathbf{B}} \mid f\right)/\epsilon^2 \ ,$$

that can be interpreted as the distribution variation in terms of the uncertainty variation. Obviously, the random divergence $D^{\mathbf{B}}_{KL}\left(f^{\mathbf{B}} \mid f\right)$ changes with ϵ^2; Figure 4 shows the ECDFs behavior for some ϵ^2 values in Example 1, exhibiting an apparent dominance relation. Nevertheless, the mean as well as the variance of $S^{\mathbf{B}}(f)$ do not depend on ϵ^2; more concretely

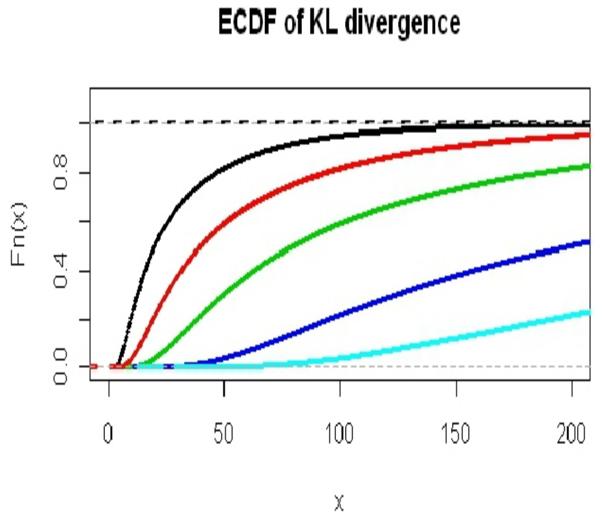

Figure 4: Empirical cumulative distribution function of Kullback-Leibler divergences for GBN in Example 1 with different independent errors: $\varepsilon^2 = 0.5$ *(black)*, 1 *(red)*, 2 *(green)*, 5 *(dark blue)*, 10 *(light blue)*

- $E\left[S^{\mathbf{B}}(f)\right] = \frac{1}{2v}\sum_{i=2}^{p} tr\left(\Sigma_{(i-1)}\right)$
- $Var\left[S^{\mathbf{B}}(f)\right] = \frac{1}{2v^2}\sum_{i=2}^{p} tr\left(\Sigma_{(i-1)}^2\right)$

Relying on the moments invariance we propose to evaluate the GBN sensitivity to uncertainty in **B** by

$$E\left[S^{\mathbf{B}}(f)\right] = \frac{1}{2v}\sum_{i=2}^{p} tr\left(\Sigma_{(i-1)}\right) = \\ = \frac{1}{2v}\sum_{i=1}^{p-1} \sigma_{ii}(p-i),$$

where σ_{ii} denotes the variance of X_i. Then, the relative contribution of each node to the total sensitivity measure will be given by

$$\frac{tr\left(\Sigma_{(i-1)}\right)}{\sum_{i=1}^{p-1}\sigma_{ii}(p-i)}. \quad (6)$$

This result enables us to classify nodes according to greatest contribution to global sensitivity.

Example 2. For the GBN in Example 1 it is obtained

$$E\left[S^{\mathbf{B}}(f)\right] = 63$$

Using (6), the numerical results by nodes are

(2) 0.008, (3) 0.016,
(4) 0.024, (5) 0.071,
(6) 0.12, (7) 0.76

Here the most important values are the contribution of the nodes to the global mean normalized divergence, resulting a significantly large influence for node (7) compared to the rest of nodes in the DAG. Therefore, independent random errors in the regression coefficients of nodes (1) to (5) do not describe very different joint models, however, that is not the case for node (7) and some effort has to be made to bring some additional information to asses the correct regression coefficient value.

4 Conclusions

The factorization of the joint distribution in GBN leads us to an additive decomposition of the Kullback-Leibler divergence. Then, for misspecified regression coefficients, the weight each node has in the global deviation of the initial structure can be determined. Modelling uncertainty with independent random errors provides a highly simplified analysis to achieve an uncertainty sensitivity measure definition that can be easily handled.

Acknowledgments

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

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

M.A. Gómez-Villegas, P. Main, R. Susi. 2007. Sensitivity Analysis in Gaussian Bayesian Networks

Using a Divergence Measure. *Communications in Statistics: Theory and Methods*, 36(3):523–539.

M.A. Gómez-Villegas, P. Main and R. Susi. 2008. Sensitivity of Gaussian Bayesian Networks to Inaccuracies in Their Parameters. *Proceedings of the 4th EuropeanWorkshop on Probabilistic Graphical Models*, pages 145–152.

A. Rencher. 2000. *Linear Models in Statistics*. New York : Wiley.

R. Shachter and C. Kenley. 1989. Gaussian influence diagrams. *Management Science*, 35:527–550.

R. Susi, H. Navarro, P. Main and M.A. Gómez-Villegas. Perturbing the structure in Gaussian Bayesian networks. 2009. *Cuadernos de Trabajo de la E.U. de Estadística*, CT01/2009: 1–21. http://www.ucm.es/info/eue/pagina/cuadernos _trabajo/CT01_2009.pdf.