CUADERNOS DE TRABAJO
ESCUELA UNIVERSITARIA DE ESTADÍSTICA

Sensitivity to hyperprior parameters in Gaussian Bayesian networks.

M.A. Gómez-Villegas
P. Main
H. Navarro
R. Susi

Cuaderno de Trabajo número 03/2010

UCM
UNIVERSIDAD
COMPLUTENSE
MADRID

Los Cuadernos de Trabajo de la Escuela Universitaria de Estadística constituyen una apuesta por la publicación de los trabajos en curso y de los informes técnicos desarrollados desde la Escuela para servir de apoyo tanto a la docencia como a la investigación.

Los Cuadernos de Trabajo se pueden descargar de la página de la Biblioteca de la Escuela www.ucm.es/BUCM/est/ y en la sección de investigación de la página del centro www.ucm.es/centros/webs/eest/

CONTACTO: Biblioteca de la E. U. de Estadística
Universidad Complutense de Madrid
Av. Puerta de Hierro, S/N
28040 Madrid
Tlf. 913944035
buc_est@buc.ucm.es

ISSN: 1989-0567

Sensitivity to hyperprior parameters in Gaussian Bayesian networks

M.A. Gómez-Villegas[a], P. Main[a], H. Navarro[b] and R. Susi[c]

[a] *Departamento de Estadística e Investigación Operativa, Universidad Complutense de Madrid, 28040 Madrid, Spain*

[b] *Departamento de Estadística, Investigación Operativa y Cálculo numérico, UNED, 28040 Madrid, Spain*

[c] *Departamento de Estadística e Investigación Operativa III, Universidad Complutense de Madrid, 28040 Madrid, Spain*

Abstract

Our focus is on learning Gaussian Bayesian networks (GBNs) from data. In GBNs the multivariate normal joint distribution can be alternatively specified by the normal regression models of each variable given its parents in the DAG (directed acyclic graph). In the latter representation the parameters are the mean vector, the regression coefficients and the corresponding conditional variances. The problem of Bayesian learning in this context has been handled with different approximations, all of them concerning the use of different priors for the parameters considered. We work with the most usual prior given by the normal/inverse gamma form. In this setting we are interested in evaluating the effect of prior hyperparameters choice on posterior distribution. The Kullback-Leibler divergence measure is used as a tool to define local sensitivity comparing the prior and posterior deviations. This method can be useful to decide the values to be chosen for the hyperparameters.

Key words: Gaussian Bayesian networks, Kullback-Leibler divergence, Bayesian linear regression

Introduction

Bayesian networks (BNs) are graphical probabilistic models of interactions between a set of variables where the joint probability distribution can be described in graphical terms.

BNs consist of qualitative and quantitative parts. The qualitative part is given by a DAG (directed acyclic graph) useful to define dependences and independencies among variables. The DAG shows the set of variables of the model at nodes and the presence/absence of arcs represents dependence/independence between variables. In the quantitative part, it is necessary to determine the set of parameters that describes the conditional probability distribution of each variable, given its parents in the DAG, to compute the joint probability distribution of the model as a factorization.

In this work, we focus on a subclass of BNs known as Gaussian Bayesian networks (GBNs). GBNs are defined as BNs where the joint probability density of $\mathbf{X} = (X_1, X_2, ..., X_p)^T$ is a multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}$ the $p-$dimensional mean vector, $\boldsymbol{\Sigma}$ the $p \times p$ positive definite covariance matrix and the dependence structure is shown in a DAG.

As in BNs, the joint density can be factorized using the conditional probability densities for every 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 density

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

being μ_i the mean of X_i, β_{ji} the regression coefficients of X_i with respect to $X_j \in pa(X_i)$, and v_i 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 possible to determine the parameters of the joint distribution. The means μ_i are the elements of the $p-$dimensional mean vector $\boldsymbol{\mu}$, and the covariance matrix $\boldsymbol{\Sigma}$ can be obtained with the coefficients b_{ji} and v_i, as follows: let $\mathbf{D}$ be a diagonal matrix $\mathbf{D} = diag(v)$ with the conditional variances $v^T = (v_1, ..., v_p)$ and let $\mathbf{B}$ be a strictly upper triangular matrix with the regression coefficients b_{ji} where $j \in \{1, ..., i-1\}$. Then, $\boldsymbol{\Sigma} = [(I - \mathbf{B})^{-1}]^T \mathbf{D} (I - \mathbf{B})^{-1}$ (see [1]).

In general, building a BN is a difficult task because it requires the user to specify the quantitative and qualitative parts of the network. Experts knowledge is important to fix the dependence structure between the variables of the network and to determine a large set of parameters. In this process, it is possible to work with a database of cases, nevertheless the experience and knowledge of experts is also necessary. In GBNs the conditional specification of the model is easy for experts, because they only have to describe univariate distributions. Then, for each X_i variable (node i in the DAG), it is necessary to specify its mean, the regression coefficients between X_i and each parent $X_j \in pa(X_i)$ and the conditional variance of X_i given its parents. Moreover, with this specification each arc in the DAG can be represented with the corresponding regression coefficient and the model is specified by the normal regression model of each variable given its parents.

Our objective in this work, is to study uncertainty about the parameters of the conditional specification. With this aim, the effect of different values for the prior hyperparameters on the posterior distribution is studied.

The problem of Bayesian learning in this context has been handled with different approximations depending on the different priors for the parameters considered (see [2] and [3]). We deal with the most usual: the normal/gamma inverse prior.

The effect of hyperparameters is studied with the Kullback-Leibler divergence [4]. This measure is used to define an appropriate local sensitivity measure to compare prior and posterior deviations. Then, with the obtained results it is possible to decide the values to be chosen for the hyperparameters considered.

Some sensitivity analyses have been developed to study uncertainty about the parameters of a GBN. [5] performed a one-way sensitivity analysis investigating the impact of small changes in the network parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. [6] proposed a one-way sensitivity analysis evaluating global sensitivity measure, rather than local aspects as location and dispersion, over the network's output. Moreover, as a generalization of this one, in [7] a $n-$way sensitivity analysis is presented. The problem of perturbed structures is also studied in [8].

The paper is organized as follows. In Section 1 the problem assessment is introduced so as the distributions considered. Section 2 is devoted to the calculation of Kullback-Leibler divergence measures. A local sensitivity measure is introduced in Section 3 and finally in Sections 4 and 5 some examples and conclusions are shown.

1 Preliminary framework

As we introduced before, the interest model is given by the conditional specification of a GBN, where the parameters are $\{\boldsymbol{\mu}, \mathbf{B}, \mathbf{D}\}$ with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} 0 & \beta_{12} & \dots & \beta_{1p} \\ & \ddots & & \\ & & \ddots & \beta_{p-1p} \\ 0 & & & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{D} = \begin{pmatrix} v_1 & & 0 \\ & \ddots & \\ 0 & & v_p \end{pmatrix}$$

Let us suppose $\boldsymbol{\mu} = \mathbf{0}$. Then, the parameters to be considered are the regression coefficients and the conditional variances of each X_i given its parents in the DAG. Note that $\beta_{ji} = 0$ if X_j (for $j < i$) is not a parent of X_i.

Selecting columns of $\mathbf{B}$ matrix and denoting $\beta_i = \begin{pmatrix} \beta_{1i} \\ \beta_{2i} \\ \vdots \\ \beta_{i-1,i} \end{pmatrix}$, for $i > 1$ the parameters to be considered now are $\{v_1, \beta_i, v_i\}_{i>1}$.

In next subsections, prior distributions, likelihood functions and posterior distributions are computed for the parameters $\{v_1, \beta_i, v_i\}_{i>1}$. Furthermore, orphan nodes (node/variable without parents in the DAG) are considered different from nodes with parents in the DAG. Thus, all the distributions of interest are determined for both cases.

1.1 ***Nodes with parents***

Let us consider a general node X_i with a nonempty set of parents $pa\,(X_i) \subset \{X_1, \ldots, X_{i-1}\}$.

1.1.1 Prior Distribution

From the normal standard theory, an Inverted Wishart is used as a prior distribution for the covariance matrix then a Wishart prior for the precision matrix $\Sigma^{-1} \sim W_p(\lambda, \tau^{-1} I_p)$. It can be shown the implied prior distributions of the normal-inverse gamma form are

$\beta_i | v_i \sim N_{i-1}\,(0, \tau^{-1} v_i I_{i-1})$ with the hyperparameter $\tau > 0$.

$v_i \sim IG\left(\frac{\lambda+i-p}{2}, \frac{\tau}{2}\right)$ with the hyperparameters $\lambda > p$ and the previous $\tau > 0$.

The corresponding expressions of prior distributions are given below

$$\pi(\beta_i | v_i)_{v_i>0} = \frac{1}{(2\pi)^{\frac{i-1}{2}} \left|\frac{v_i}{\tau}\right|^{\frac{i-1}{2}}} \exp\left\{-\frac{\tau}{2v_i}\beta_i^T \beta_i\right\} \propto \frac{\exp\left\{-\frac{\tau}{2v_i}\beta_i^T \beta_i\right\}}{\left(\frac{v_i}{\tau}\right)^{\frac{i-1}{2}}} =$$
$$= \left(\frac{\tau}{v_i}\right)^{\frac{i-1}{2}} \exp\left\{-\frac{\tau}{2v_i}\beta_i^T \beta_i\right\}, \beta_i \in \mathbb{R}^{i-1}$$

$$\pi(v_i) = \frac{\left(\frac{\tau}{2}\right)^{\left(\frac{\lambda+i-p}{2}\right)}}{\Gamma\left(\frac{\lambda+i-p_i}{2}\right)} v_i^{-\left(\frac{\lambda+i-p}{2}+1\right)} \exp\left\{-\frac{\frac{\tau}{2}}{v_i}\right\} \propto \frac{\exp\left\{-\frac{\tau}{2v_i}\right\}}{v_i^{\left(\frac{\lambda+i-p}{2}+1\right)}}, v_i > 0$$

Finally, the joint prior distribution can be computed by

$$\pi(\beta_i, v_i) = \pi(\beta_i|v_i)\pi(v_i), \beta_i \in \mathbb{R}^{i-1} \text{ and } v_i > 0$$

1.1.2 Likelihood function

A random sample of size n is observed giving the next data matrix

$$\left(\begin{array}{ccc|c|cc} x_{11} & x_{12} & \dots & x_{1i} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2i} & \dots & x_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{ni} & \dots & x_{np} \end{array}\right)$$

For the variable X_i we have to consider the observations of its parents $pa\,(X_i)$

$$X_{pa_i} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1i-1} \\ x_{21} & x_{22} & \dots & x_{2i-1} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{ni-1} \end{pmatrix}$$

as well as the observations of X_i, $x_i = (x_{1i}, x_{2i}, ..., x_{ni})^T$ and the regression model

$$x_i = X_{pa_i}\beta_i + \varepsilon_i; i = 1, \dots, p$$

with $\varepsilon_i \sim N_n(0, v_i I_n)$.

Then, the likelihood function is as follows

$$L(v_i, \beta_i; x_i, X_{pa_i}) \propto \frac{1}{(v_i)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2v_i}\left[(n-(i-1))S_i^2 + (\beta_i - \hat{\beta}_i)^T X_{pa_i}^T X_{pa_i}(\beta_i - \hat{\beta}_i)\right]\right\}$$

$\beta_i \in \mathbb{R}^{i-1}, v_i > 0$

with

$$\hat{\beta}_i = \left(X_{pa_i}^T X_{pa_i}\right)^{-1} X_{pa_i}^T x_i$$

and

$$S_i^2 = \frac{\left(x_i - X_{pa_i}\hat{\beta}_i\right)^T \left(x_i - X_{pa_i}\hat{\beta}_i\right)}{n-(i-1)} = \frac{x_i^T x_i - x_i^T X_{pa_i}\left(X_{pa_i}^T X_{pa_i}\right)^{-1} X_{pa_i}^T x_i}{n-(i-1)}$$

1.1.3 Posterior distribution

The joint posterior distribution is given by

$\pi(\beta_i, v_i | x_i X_{pa_i}) = \pi(\beta_i|v_i)\pi(v_i)L(v_i, \beta_i; x_i, X_{pa_i}) \propto$

$$\frac{\tau^{\frac{i-1}{2}}}{v_i^{\frac{\lambda+(i-p)+(i-1)+n}{2}+1}} \exp\{-\frac{1}{2v_i}[\tau + (n-(i-1))S_i^2 + \underbrace{\tau\beta_i^T\beta_i + (\beta_i - \hat{\beta}_i)^T X_{pa_i}^T X_{pa_i}(\beta_i - \hat{\beta}_i)}_{(A)}]\}$$

then substituting $\hat{\beta}_i$ with its value and making some calculations it yields

$(A) = \tau\beta_i^T\beta_i + (\beta_i - \hat{\beta}_i)^T X_{pa_i}^T X_{pa_i}(\beta_i - \hat{\beta}_i) =$

$$= \tau\beta_i^T\beta_i + \beta_i^T X_{pa_i}^T X_{pa_i}\beta_i - \beta_i^T \underbrace{X_{pa_i}^T X_{pa_i}\hat{\beta}_i}_{X_{pa_i}^T x_i} - \underbrace{\hat{\beta}_i^T X_{pa_i}^T X_{pa_i}}_{x_i^T X_{pa_i}}\beta_i + \hat{\beta}_i^T X_{pa_i}^T X_{pa_i}\hat{\beta}_i =$$

$$= \beta_i^T \underbrace{(\tau I_{i-1} + X_{pa_i}^T X_{pa_i})}_{M_i}\beta_i - x_i^T X_{pa_i} M_i (M_i)^{-1}\beta_i - \beta_i^T M_i (M_i)^{-1} X_{pa_i}^T x_i + \hat{\beta}_i^T X_{pa_i}^T X_{pa_i}\hat{\beta}_i =$$

$$= \underbrace{\hat{\beta}_i^T X_{pa_i}^T X_{pa_i}\hat{\beta}_i - \tilde{\beta}_i^T M_i\tilde{\beta}_i}_{(B)} + (\beta_i - \tilde{\beta}_i)^T M_i(\beta_i - \tilde{\beta}_i)$$

where $\boxed{M_i = \tau I_{i-1} + X_{pa_i}^T X_{pa_i} \text{ and } \tilde{\beta}_i = M_i^{-1} X_{pa_i}^T x_i}$

Therefore, returning to the posterior density expression

$\pi(\beta_i, v_i | x_i X_{pa_i}) \propto$

$$\frac{\tau^{\frac{i-1}{2}}}{v_i^{\frac{\lambda+(i-p)+(i-1)+n}{2}+1}} \exp\{-\frac{1}{2v_i}[\tau + \underbrace{(n-(i-1))S_i^2 + (B)}_{(C)} + (\beta_i - \tilde{\beta}_i)^T M_i(\beta_i - \tilde{\beta}_i)]\}$$

where

$$(C) = (n-(i-1))S_i^2 + (B) = x_i^T x_i - \underbrace{x_i^T X_{pa_i}\left(X_{pa_i}^T X_{pa_i}\right)^{-1}}_{\hat{\beta}_i^T} X_{pa_i}^T x_i + (B) =$$

$$= x_i^T x_i - \hat{\beta}_i^T X_{pa_i}^T x_i \left(X_{pa_i}^T X_{pa_i}\right) \left(X_{pa_i}^T X_{pa_i}\right)^{-1} X_{pa_i}^T x_i + (B) =$$

$$= \boxed{x_i^T x_i - \tilde{\beta}_i^T M_i \tilde{\beta}_i = x_i^T x_i - x_i^T X_{pa_i} \left(M_i\right)^{-1} X_{pa_i}^T x_i = q_i}$$

Thus,

$\pi(\beta_i, v_i | x_i X_{pa_i}) \propto$

$\frac{\tau^{\frac{i-1}{2}}}{v_i^{\frac{\lambda+(i-p)+(i-1)+n}{2}+1}} \exp\left\{-\frac{1}{2v_i}\left[\tau + q_i + (\beta_i - \tilde{\beta}_i)^T M_i (\beta_i - \tilde{\beta}_i)\right]\right\}$, with $\beta_i \in \mathbb{R}^{i-1}$ and $v_i > 0$.

It follows immediately the posterior densities of the parameters in the model

$\pi(\beta_i | v_i x_i X_{pa_i}) \propto \exp\left\{-\frac{1}{2v_i}\left[(\beta_i - \tilde{\beta}_i)^T M_i (\beta_i - \tilde{\beta}_i)\right]\right\}$, $\beta_i \in \mathbb{R}^{i-1}$

then, a normal distribution $N_{i-1}\left(\tilde{\beta}_i, v_i \left(M_i\right)^{-1}\right)$

$\pi(v_i | x_i X_{pa_i}) \propto$

$$\frac{\tau^{\frac{i-1}{2}}}{v_i^{\frac{\lambda+(i-p)+(i-1)+n}{2}+1}} \exp\left\{-\frac{1}{2v_i}\left(\tau + q_i\right)\right\} \underbrace{\int_{\mathbb{R}^{i-1}} \exp\left\{-\frac{1}{2v_i}\left[(\beta_i - \tilde{\beta}_i)^T M_i (\beta_i - \tilde{\beta}_i)\right]\right\} d\beta_i}_{\det\left(v_i M_i^{-1}\right) \propto (v_i)^{i-1}}$$

so that

$\pi(v_i | x_i X_{pa_i}) \propto \frac{\tau^{\frac{i-1}{2}}}{v_i^{\frac{\lambda+(i-p)+n}{2}+1}} \exp\left\{-\frac{1}{2v_i}\left(\tau + q_i\right)\right\}$, $v_i > 0$

then, an Inverse-Gamma distribution $IG\left(\frac{\lambda+(i-p)+n}{2}, \frac{\tau+q_i}{2}\right)$.

1.2 ***Orphan nodes***

When a node X_i has no parents in the DAG, there is no arc to X_i, then $\beta_{ki} = 0$ (for every $k < i$). Then the parameter to be studied is only v_i.

1.2.1 Prior distribution, likelihood function and posterior distribution

If a node X_i has no parents, the normal distribution to be considered is the marginal $N_1(0, v_i)$ and the prior distribution has to be $\pi(v_i) \sim IG\left(\frac{\lambda+i-p}{2}, \frac{\tau}{2}\right)$.

The data are the observations of X_i

$$x_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{pmatrix}$$

then, the likelihood function is

$$L(v_i; x_i) \propto \frac{1}{(v_i)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2v_i}\left[x_i^T x_i\right]\right\}, v_i > 0$$

therefore, the posterior distribution of the parameter is given by

$$\pi(v_i|x_i) = \pi(v_i)L(v_i; x_i) \propto \frac{1}{v_i^{\frac{\lambda+(i-p)+n}{2}+1}} \exp\left\{-\frac{1}{2v_i}\left[\tau + x_i^T x_i\right]\right\}, v_i > 0.$$

2 Divergence measure

In this section we compute the Kullback-Leibler divergence to evaluate uncertainty in hyperparameters in terms of additive perturbations, $\delta \in \mathbb{R}^+$. Then, the objective is to evaluate the effect of different peturbed hyperparameters by means of the Kullback-Leibler divergence.

Throughout this work, perturbed models obtained by adding a $\delta \in \mathbb{R}^+$ perturbation to the hyperparameters, are denoted by $\pi^\delta(\cdot)$. The original model corresponds to $\delta = 0$.

Moreover, to evaluate joint distributions next result relating marginal and conditional divergences is used.

$$D_{KL}(f^\delta(x,y) \mid f(x,y)) = D_{KL}\left(f^\delta(y) \mid f(y)\right) + \int f(y) D_{KL}(f^\delta(x|y) \mid f(x|y)) dy \tag{1}$$

Given that the joint prior and posterior distributions are of the same form $\pi(\beta, v) = \pi(\beta|v)\,\pi(v)$, expression (1) can be applied both to prior and posterior distributions by comparing the original and the perturbed model.

2.1 ***Nodes with parents***

Let X_i be a general node with a nonempty set of parents $pa(X_i) \subset \{X_1, \ldots, X_{i-1}\}$.

2.1.1 *Prior hyperparameter perturbation $\lambda \to \lambda + \delta$*

The hyperparameter λ appears only in the distribution of the parameter v_i. Then, with (1) the Kullback-Leibler divergence of the joint distribution corresponds to the marginal distribution of v_i. Next expressions are the prior and posterior distributions for the original and perturbed models.

Prior distributions:

Original model $\pi(v_i) \sim IG\left(\frac{\lambda+(i-p)}{2}, \frac{\tau}{2}\right)$

Perturbed model $\pi^{\delta}(v_i) \sim IG\left(\frac{\lambda+\delta+(i-p)}{2}, \frac{\tau}{2}\right)$

Posterior distributions:

Original model $\pi(v_i|x_i X_{pa_i}) \sim IG\left(\frac{\lambda+(i-p)+n}{2}, \frac{\tau+q_i}{2}\right)$

Perturbed model $\pi^{\delta}(v_i|x_i X_{pa_i}) \sim IG\left(\frac{\lambda+\delta+(i-p)+n}{2}, \frac{\tau+q_i}{2}\right)$

Then, divergences between joint densities are

Prior distributions:

$D_{KLprior} = D_{KL}(\pi^{\delta}(\beta_i, v_i) \mid \pi(\beta_i, v_i)) = D_{KL}(\pi^{\delta}(v_i) \mid \pi(v_i))$

$$\boxed{D_{KLprior} = \ln \frac{\Gamma\left(\frac{\lambda+\delta+(i-p)}{2}\right)}{\Gamma\left(\frac{\lambda+(i-p)}{2}\right)} - \left(\frac{\delta}{2}\right) \Psi\left(\frac{\lambda+(i-p)}{2}\right)}$$

with $\Psi(x)$ the digamma function.

Posterior distributions:

$D_{KLposterior} = D_{KL}(\pi^{\delta}(\beta_i, v_i|x_i X_{pa_i}) \mid \pi(\beta_i, v_i|x_i X_{pa_i})) =$

$= D_{KL}(\pi^{\delta}(v_i|x_i X_{pa_i}) \mid \pi(v_i|x_i X_{pa_i}))$

$$\boxed{D_{KLposterior} = \ln \frac{\Gamma\left(\frac{\lambda+\delta+(i-p)+n}{2}\right)}{\Gamma\left(\frac{\lambda+(i-p)+n}{2}\right)} - \left(\frac{\delta}{2}\right) \Psi\left(\frac{\lambda+(i-p)+n}{2}\right)}$$

2.1.2 Prior hyperparameter perturbation $\tau \rightarrow \tau + \delta$

The hyperparameter τ appears in the distribution of both parameters β_i and v_i. Next expressions are the prior and posterior distributions for the original and perturbed models as well as the Kullback-Leibler divergence calculated later.

Prior distributions:

$$\text{Original model } \pi\left(\beta_i|v_i\right) \sim N_{i-1}\left(0, \tau^{-1} v_i I_{i-1}\right)$$

$$\text{Perturbed model } \pi^{\delta}\left(\beta_i|v_i\right) \sim N_{i-1}\left(0, (\tau+\delta)^{-1} v_i I_{i-1}\right)$$

and

$$\text{Original model } \pi(v_i) \sim IG\left(\tfrac{\lambda+i-p}{2}, \tfrac{\tau}{2}\right)$$

$$\text{Perturbed model } \pi^{\delta}(v_i) \sim IG\left(\tfrac{\lambda+i-p}{2}, \tfrac{\tau+\delta}{2}\right)$$

Posterior distributions:

$$\text{Original model } \pi(\beta_i|v_i x_i X_{pa_i}) \sim N_{i-1}\left(\tilde{\beta}_i, v_i \left(M_i\right)^{-1}\right)$$

$$\text{Perturbed model } \pi^{\delta}(\beta_i|v_i x_i X_{pa_i}) \sim N_{i-1}\left(\tilde{\beta}_i^{\delta}, v_i \left(M_i^{\delta}\right)^{-1}\right)$$

and

$$\text{Original model } \pi(v_i|x_i X_{pa_i}) \sim IG\left(\tfrac{\lambda+(i-p)+n}{2}, \tfrac{\tau+q_i}{2}\right)$$

$$\text{Perturbed model } \pi^{\delta}(v_i|x_i X_{pa_i}) \sim IG\left(\tfrac{\lambda+(i-p)+n}{2}, \tfrac{\tau+\delta+q_i^{\delta}}{2}\right)$$

with $q_i^{\delta} = x_i^T x_i - x_i^T X_{pa_i} \left(M_i^{\delta}\right)^{-1} X_{pa_i}^T x_i$

Therefore, divergences between joint densities are

Prior distributions:

$$D_{KLprior} = D_{KL}(\pi^{\delta}(\beta_i, v_i) \mid \pi(\beta_i, v_i)) =$$

$$= D_{KL}(\pi^{\delta}(v_i) \mid \pi(v_i)) + \textstyle\int \pi(v_i) \underbrace{D_{KL}(\pi^{\delta}\left(\beta_i|v_i\right) \mid \pi\left(\beta_i|v_i\right)}_{\text{it does not depend on } v_i} dv_i =$$

$$= \tfrac{(i-1)}{2}\left[\left(\tfrac{\delta}{\tau}\right) - \ln\left(1+\tfrac{\delta}{\tau}\right)\right] + \tfrac{\lambda+(i-p)}{2}\left[\left(\tfrac{\delta}{\tau}\right) - \ln\left(1+\tfrac{\delta}{\tau}\right)\right]$$

$$\boxed{D_{KLprior}=\frac{\lambda+(i-p)+(i-1)}{2}\left[\left(\frac{\delta}{\tau}\right)-\ln\left(1+\frac{\delta}{\tau}\right)\right]}$$

Posterior distributions:

$D_{KLposterior} = D_{KL}(\pi^{\delta}(\beta_i, v_i|x_i X_{pa_i}) \mid \pi(\beta_i, v_i|x_i X_{pa_i})) =$

$= \int \pi(v_i|x_i X_{pa_i}) D_{KL}(\pi^{\delta}\left(\beta_i|v_i, x_i X_{pa_i}\right) \mid \pi\left(\beta_i|v_i, x_i X_{pa_i}\right) dv_i +$

$+D_{KL}(\pi^{\delta}(v_i|x_i X_{pa_i}) \mid \pi(v_i|x_i X_{pa_i})) = (1) + (2)$

$$(1) = \frac{1}{2}\left[\begin{array}{c} \underbrace{\ln\frac{|M_i|}{\left|M_i^{\delta}\right|}}_{(i)} + \underbrace{tr\left(M_i^{\delta} M_i^{-1}\right)}_{(ii)} - (i-1)+ \\ +\underbrace{(\tilde{\beta}_i - \tilde{\beta}_i^{\delta})^T M_i^{\delta}(\tilde{\beta}_i - \tilde{\beta}_i^{\delta})}_{(iii)} \underbrace{\int \frac{1}{v_i}\frac{\left(\frac{\tau+q_i}{2}\right)^{\frac{\lambda+(i-p)+n}{2}}}{\Gamma\left(\frac{\lambda+(i-p)+n}{2}\right)} v_i^{-\left(\frac{\lambda+(i-p)+n}{2}+1\right)} \exp\left\{-\frac{1}{2v_i}\left(\tau+q_i\right)\right\} dv_i}_{\frac{\lambda+(i-p)+n}{\tau+q_i}} \end{array}\right]$$

with some calculations

$$\left\{\begin{array}{c} \text{(i) } M_i^{\delta} = M_i + \delta I_{i-1} \rightarrow \left\{\begin{array}{c} M_i^{\delta} M_i^{-1} = I_{i-1} + \delta M_i^{-1} \\ M_i^{-1} = \left(M_i^{\delta}\right)^{-1}\left(I_{i-1} + \delta M_i^{-1}\right) \end{array}\right. \rightarrow \\ \rightarrow \left\{ M_i^{-1} - \left(M_i^{\delta}\right)^{-1} = \delta\left(M_i^{\delta}\right)^{-1} M_i^{-1} \right. \\ \text{(ii) } tr\left(M_i^{\delta} M_i^{-1}\right) = (i-1) + \delta tr\left(M_i^{-1}\right) \\ \text{(iii) } (\tilde{\beta}_i - \tilde{\beta}_i^{\delta})^T M_i^{\delta}(\tilde{\beta}_i - \tilde{\beta}_i^{\delta}) = \\ = x_i^T X_{pa_i}\left(M_i^{-1} - \left(M_i^{\delta}\right)^{-1}\right)^T M_i^{\delta}\left(M_i^{-1} - \left(M_i^{\delta}\right)^{-1}\right) X_{pa_i}^T x_i = \\ = \delta^2 \tilde{\beta}_i^T \left(M_i^{\delta}\right)^{-1} \tilde{\beta}_i \end{array}\right\}$$

it yields

$$(1) = \boxed{\frac{1}{2}\left[\ln\frac{|M_i|}{\left|M_i^{\delta}\right|} + \delta tr\left(M_i^{-1}\right) + \frac{\lambda+(i-p)+n}{\tau+q_i}\delta^2\tilde{\beta}_i^T\left(M_i^{\delta}\right)^{-1}\tilde{\beta}_i\right]}$$

$$(2) = \boxed{\frac{\lambda+(i-p)+n}{2}\left[-\ln\left(1+\frac{\delta+\left(q_i^{\delta}-q_i\right)}{\tau+q_i}\right) + \frac{\delta+\left(q_i^{\delta}-q_i\right)}{\tau+q_i}\right]}.$$

Adding these last equations we obtain the divergence measure between the original and perturbed posterior distributions.

2.2 Orphan nodes

Previous calculations are used for evaluating differences between distributions in this case.

2.2.1 Prior hyperparameter perturbation $\lambda \to \lambda + \delta$

The results are the same as for nodes with parents.

2.2.2 Prior hyperparameter perturbation $\tau \to \tau + \delta$

Now the divergence **between prior distributions** is the first summand of the expression for nodes with parents

$$\boxed{D_{KLprior} = D_{KL}(\pi^{\delta}(v_i) \mid \pi(v_i)) = \frac{\lambda+(i-p)}{2}\left[\left(\frac{\delta}{\tau}\right) - \ln\left(1+\frac{\delta}{\tau}\right)\right]}$$

and **between posterior distributions** the Kullback-Leibler divergence is

$$\boxed{D_{KLposterior} = D_{KL}(\pi^{\delta}(v_i|x_i) \mid \pi(v_i|x_i)) = \frac{\lambda+(i-p)+n}{2}\left[\frac{\delta}{\tau+x_i^T x_i} - \ln\left(1+\frac{\delta}{\tau+x_i^T x_i}\right)\right]}$$

3 Sensitivity measure

To asses the sensitivity of the posterior to prior variations given by small perturbations in the hyperprior parameters, we introduce a local sensitivity measure given by

$$\boxed{Sens = \lim_{\delta\to 0}\frac{D_{KLposterior}}{D_{KLprior}} = \lim_{\delta\to 0}\frac{D_{KL}(\pi^{\delta}(\beta_i,v_i|x_iX_{pa_i})|\pi(\beta_i,v_i|x_iX_{pa_i}))}{D_{KL}(\pi^{\delta}(\beta_i,v_i)|\pi(\beta_i,v_i))}}$$

3.1 Nodes with parents

3.1.1 Hyperparameter perturbation $\lambda \to \lambda + \delta$

In this case

$$Sens\left(\lambda\right)=\lim_{\delta\to 0}\frac{D_{KL}(\pi^{\delta}(\beta_i,v_i|x_iX_{pa_i})|\pi(\beta_i,v_i|x_iX_{pa_i}))}{D_{KL}(\pi^{\delta}(\beta_i,v_i)|\pi(\beta_i,v_i))}=$$

$$=\lim_{\delta\to 0}\frac{D_{KL}(\pi^{\delta}(v_i|x_iX_{pa_i})|\pi(v_i|x_iX_{pa_i}))}{D_{KL}(\pi^{\delta}(v_i)|\pi(v_i))}=$$

$$=\lim_{\delta\to 0}\frac{\ln\frac{\Gamma\left(\frac{\lambda+\delta+(i-p)+n}{2}\right)}{\Gamma\left(\frac{\lambda+(i-p)+n}{2}\right)}-\left(\frac{\delta}{2}\right)\Psi\left(\frac{\lambda+(i-p)+n}{2}\right)}{\ln\frac{\Gamma\left(\frac{\lambda+\delta+(i-p)}{2}\right)}{\Gamma\left(\frac{\lambda+(i-p)}{2}\right)}-\left(\frac{\delta}{2}\right)\Psi\left(\frac{\lambda+(i-p)}{2}\right)}=\lim_{\delta\to 0}\frac{\frac{d}{d\delta}\Psi\left(\frac{\lambda+(i-p)+n+\delta}{2}\right)}{\frac{d}{d\delta}\Psi\left(\frac{\lambda+(i-p)+\delta}{2}\right)}$$

$$\boxed{Sens\left(\lambda\right)=\frac{\Psi'\left(\frac{\lambda+(i-p)+n}{2}\right)}{\Psi'\left(\frac{\lambda+(i-p)}{2}\right)}<1}$$

with Ψ' the trigamma function.

Note that it is always less than one because the trigamma function $\Psi'(x)$ is monotone decreasing as well it is monotonically dominated when the node index increases.

3.1.2 Hyperparameter perturbation $\tau\to\tau+\delta$

First, it can be considered

$$Sens\left(\tau\right)=\lim_{\delta\to 0}\frac{D_{KL}(\pi^{\delta}(\beta_i,v_i|x_iX_{pa_i})|\pi(\beta_i,v_i|x_iX_{pa_i}))}{D_{KL}(\pi^{\delta}(\beta_i,v_i)|\pi(\beta_i,v_i))}=$$

$$=\lim_{\delta\to 0}\frac{(1)+(2)}{D_{KL}(\pi^{\delta}(\beta_i,v_i)|\pi(\beta_i,v_i))}=(1^*)+(2^*)$$

By calculating separately the two summands we obtain the limit.

$$\boxed{\boxed{(1^*)}}$$

$$(1^*)=\lim_{\delta\to 0}\frac{\frac{1}{2}\left[\ln\frac{|M_i|}{\left|M_i^{\delta}\right|}+\delta tr\left(M_i^{-1}\right)+\frac{\lambda+(i-p)+n}{\tau+q_i}\delta^2\tilde{\beta}_i^T\left(M_i^{\delta}\right)^{-1}\tilde{\beta}_i\right]}{\frac{\lambda+(i-p)+(i-1)}{2}\left[\left(\frac{\delta}{\tau}\right)-\ln\left(1+\frac{\delta}{\tau}\right)\right]}\underset{\boxed{\text{L'Hospital's rule}}}{\overset{\uparrow}{=}}$$

$$=\lim_{\delta\to 0}\frac{-\frac{d}{d\delta}\ln\left|M_i^{\delta}\right|+tr\left(M_i^{-1}\right)+\frac{\lambda+(i-p)+n}{\tau+q_i}\frac{d}{d\delta}\left(\delta^2\tilde{\beta}_i^T\left(M_i^{\delta}\right)^{-1}\tilde{\beta}_i\right)}{(\lambda+(i-p)+(i-1))\frac{\delta}{\tau(\tau+\delta)}}$$

Let $\{\lambda_k,e_k\}_{k=1,\dots,i-1}$ be the eigenvalues and eigenvectors of the $X_{pa_i}^TX_{pa_i}$ matrix, then $\{\lambda_k+\tau,e_k\}_{k=1,\dots,i-1}$ are the corresponding ones of M_i and $\{\lambda_k+\tau+\delta,e_k\}_{k=1,\dots,i-1}$ of M_i^{δ}. Therefore an eigen analysis of the $X_{pa_i}^TX_{pa_i}$ matrix allows us to find the limit in terms of these elements.

$(1^*) =$

$$= \lim_{\delta\to 0} \frac{-\frac{d}{d\delta}\ln\prod_{k=1}^{i-1}(\lambda_k+\tau+\delta)+\sum_{k=1}^{i-1}\frac{1}{\lambda_k+\tau}+\frac{\lambda+(i-p)+n}{\tau+q_i}\frac{d}{d\delta}\{\delta^2\tilde{\beta}_i^T P\begin{pmatrix}\frac{1}{\lambda_1+\tau+\delta} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \frac{1}{\lambda_{i-1}+\tau+\delta}\end{pmatrix}P^T\tilde{\beta}_i\}}{(\lambda+(i-p)+(i-1))\frac{\delta}{\tau(\tau+\delta)}}$$

with $P = \left(e_1 \vdots \ldots \vdots e_{i-1}\right)$ the eigenvectors orthogonal matrix, then

$$\frac{d}{d\delta}\left(\delta^2\tilde{\beta}_i^T P\begin{pmatrix}\frac{1}{\lambda_1+\tau+\delta} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \frac{1}{\lambda_{i-1}+\tau+\delta}\end{pmatrix}P^T\tilde{\beta}_i\right) =$$

$$= \frac{d}{d\delta}\left(\delta^2 z_i^T\begin{pmatrix}\frac{1}{\lambda_1+\tau+\delta} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \frac{1}{\lambda_{i-1}+\tau+\delta}\end{pmatrix}z_i\right) =$$

$$= \frac{d}{d\delta}\sum_{k=1}^{i-1}\frac{z_{i\ k}^2\delta^2}{\lambda_k+\tau+\delta} = \sum_{k=1}^{i-1} z_{i\ k}^2\frac{\delta^2+2\delta(\lambda_k+\tau)}{\lambda_k+\tau+\delta},$$

with $z_i = P^T\tilde{\beta}_i = \begin{pmatrix} z_{i\ 1}\\ \vdots \\ z_{i\ i-1}\end{pmatrix}$.

Therefore

$$(1^*) = \lim_{\delta\to 0}\frac{\tau(\tau+\delta)\left[-\sum_{k=1}^{i-1}\frac{1}{\lambda_k+\tau+\delta}+\sum_{k=1}^{i-1}\frac{1}{\lambda_k+\tau}+\frac{\lambda+(i-p)+n}{\tau+q_i}\sum_{k=1}^{i-1}z_{i\ k}^2\frac{\delta^2+2\delta(\lambda_k+\tau)}{\lambda_k+\tau+\delta}\right]}{\delta(\lambda+(i-p)+(i-1))} =$$

$$= \boxed{\frac{\tau^2}{(\lambda+(i-p)+(i-1))}\left[\sum_{k=1}^{i-1}\frac{1}{(\lambda_k+\tau)^2}+\frac{\lambda+(i-p)+n}{\tau+q_i}2\tilde{\beta}_i^T M_i^{-1}\tilde{\beta}_i\right]}$$

$$\boxed{\boxed{(2^*)}}$$

$$(2^*) = \lim_{\delta\to 0}\frac{\frac{\lambda+(i-p)+n}{2}\left[-\ln\left(1+\frac{\delta+\left(q_i^\delta-q_i\right)}{\tau+q_i}\right)+\frac{\delta+\left(q_i^\delta-q_i\right)}{\tau+q_i}\right]}{\frac{\lambda+(i-p)+(i-1)}{2}\left[-\ln\left(1+\frac{\delta}{\tau}\right)+\left(\frac{\delta}{\tau}\right)\right]}$$

The previous limit can be obtained using next general result with $\lim_{x\to 0} h(x) = 0$

$$\lim_{x\to 0} \frac{-\ln\left(1+\frac{x+h(x)}{c_2}\right)+\frac{x+h(x)}{c_2}}{-\ln\left(1+\frac{x}{c_1}\right)+\left(\frac{x}{c_1}\right)} \underset{\substack{\uparrow \\ \boxed{\text{L'Hospital's rule}}}}{=} \frac{c_1^2}{c_2^2} \lim_{x\to 0} \left(1+\frac{d}{dx}h(x)\right)^2$$

then,

$$(2^*) = \frac{\lambda+(i-p)+n}{\lambda+(i-p)+(i-1)} \frac{\tau^2}{(\tau+q_i)^2} \lim_{\delta\to 0} \left(1+\frac{d}{d\delta}q_i^\delta\right)^2.$$

Now we determine $\frac{d}{d\delta}q_i^\delta$

$$q_i^\delta = x_i^T x_i - x_i^T X_{pa_i} \left(M_i^\delta\right)^{-1} X_{pa_i}^T x_i$$

and with an eigen analysis of the $X_{pa_i}^T X_{pa_i}$ matrix and P as above, it follows

$$x_i^T X_{pa_i} \left(M_i^\delta\right)^{-1} X_{pa_i}^T x_i = x_i^T X_{pa_i} P \begin{pmatrix} \frac{1}{\lambda_1+\tau+\delta} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\lambda_{i-1}+\tau+\delta} \end{pmatrix} P^T X_{pa_i}^T x_i =$$

$$= \sum_{k=1}^{i-1} \frac{w_{i\,k}^2}{\lambda_k+\tau+\delta} \left(\to_{\delta\to 0} w_i^T \begin{pmatrix} \frac{1}{\lambda_1+\tau} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\lambda_{i-1}+\tau} \end{pmatrix} w_i = x_i^T x_i - q_i, \text{ effectively} \right), \text{with}$$

$$w_i = P^T X_{pa_i}^T x_i = \begin{pmatrix} w_{i\,1} \\ \vdots \\ w_{i\,i-1} \end{pmatrix}.$$

Thus

$$\frac{d}{d\delta}\left(x_i^T X_{pa_i} \left(M_i^\delta\right)^{-1} X_{pa_i}^T x_i\right) = \sum_{k=1}^{i-1} \frac{-w_{i\,k}^2}{(\lambda_k+\tau+\delta)^2} \to_{\delta\to 0} \sum_{k=1}^{i-1} \frac{-w_{i\,k}^2}{(\lambda_k+\tau)^2} = -\tilde{\beta}_i^T \tilde{\beta}_i$$

and

$$\lim_{\delta\to 0} \left(1+\frac{d}{d\delta}q_i^\delta\right)^2 = \left(1+\tilde{\beta}_i^T \tilde{\beta}_i\right)^2$$

yielding

$$(2^*) = \boxed{\frac{\lambda+(i-p)+n}{\lambda+(i-p)+(i-1)} \frac{\tau^2}{(\tau+q_i)^2} \left(1+\tilde{\beta}_i^T \tilde{\beta}_i\right)^2}.$$

As a final result

$\lim_{\delta\to 0} \frac{D_{KL}(\pi^{\delta}(\beta_i,v_i|x_i X_{pa_i})|\pi(\beta_i,v_i|x_i X_{pa_i}))}{D_{KL}(\pi^{\delta}(\beta_i,v_i)|\pi(\beta_i,v_i))} =$

$$\boxed{\begin{array}{c} Sens\,(\tau) = \frac{\tau^2}{(\lambda+(i-p)+(i-1))}\left[\sum_{k=1}^{i-1}\frac{1}{(\lambda_k+\tau)^2}+\frac{\lambda+(i-p)+n}{\tau+q_i}2\tilde{\beta}_i^T M_i^{-1}\tilde{\beta}_i\right]+ \\ +\frac{\lambda+(i-p)+n}{\lambda+(i-p)+(i-1)}\frac{\tau^2}{(\tau+q_i)^2}\left(1+\tilde{\beta}_i^T\tilde{\beta}_i\right)^2 \end{array}}$$

3.2 Orphan nodes

The only perturbation to be analyzed corresponds to the hyperparameter τ because the same results of nodes with parents can be applied to orphan nodes if λ is considered.

3.2.1 Hyperparameter perturbation $\tau \to \tau+\delta$

$Sens\,(\tau) = \lim_{\delta\to 0} \frac{D_{KL}(\pi^{\delta}(v_i|x_i)|\pi(v_i|x_i))}{D_{KL}(\pi^{\delta}(v_i)|\pi(v_i))} =$

$= \lim_{\delta\to 0} \frac{\lambda+(i-p)+n}{\lambda+(i-p)} \frac{-\ln\left(1+\frac{\delta}{\tau+x_i^T x_i}\right)+\frac{\delta}{\tau+x_i^T x_i}}{-\ln\left(1+\frac{\delta}{\tau}\right)+\left(\frac{\delta}{\tau}\right)}$

$$\boxed{Sens\,(\tau) = \frac{\lambda+(i-p)+n}{\lambda+(i-p)}\frac{\tau^2}{\left(\tau+x_i^T x_i\right)^2}}$$

4 Experiments

Let us consider a GBN with parameters β_{ji} and v_i being $j < i$ and a dependence structure given by the DAG in Figure 1 (see [7]).

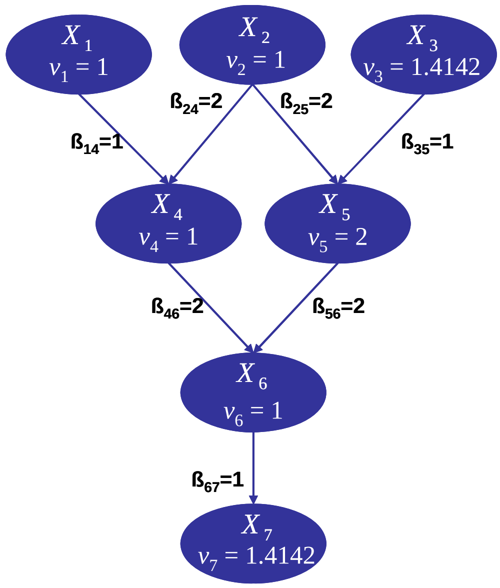

Figure 1. DAG representation of the GBN of interest

An artificial sample of size $n = 1000$ is simulated.

With the sensitivity measure introduced in Section 3, next results are obtained for both kind of perturbations

Sensitivity measure when the hyperparameter perturbation is $\lambda \to \lambda + \delta$

$\lambda \backslash X_i$	X_1	X_2	X_3	X_4	X_5	X_6	X_7
8	0.002	0.003	0.004	0.004	0.005	0.006	0.007
15	0.008	0.009	0.010	0.011	0.012	0.013	0.014
25	0.018	0.019	0.020	0.021	0.022	0.023	0.024
50	0.0428	0.043	0.044	0.044	0.045	0.046	0.047
150	0.125	0.126	0.127	0.128	0.128	0.129	0.129
500	0.330	0.331	0.331	0.332	0.332	0.333	0.333
1000	**0.498**	**0.499**	**0.499**	**0.499**	**0.499**	**0.500**	**0.500**
10000	0.909	0.909	0.909	0.909	0.909	0.909	0.909

Table 1. Sensitivity measure for different values of perturbed λ

As it can be seen, the sensitivity measure of each variable is very similar for all the nodes. Moreover, the measure increases with the values of λ but in all cases is less than 1.

Sensitivity measure when the hyperparameter perturbation is $\tau \to \tau + \delta$

Figure 2 shows the sensitivity measure obtained for $\tau > 0$ with different color lines for each variable visualizing the node numbers in the circles.

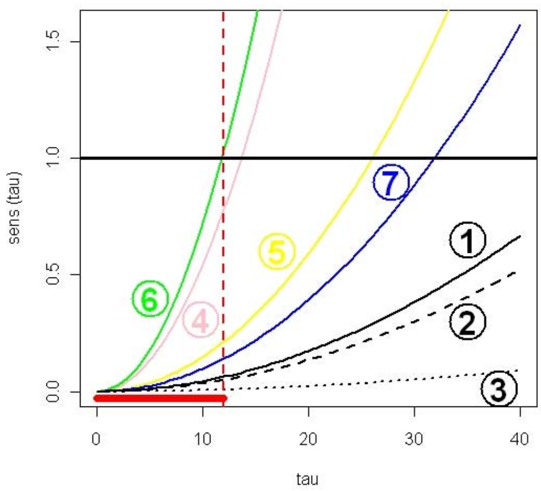

Figure 2. Sensitivity measure for τ hyperparameter with the simulated sample of the DAG discussed above

When $Sens(\tau) < 1$, posterior Kullback-Leibler divergence is smaller than prior one for infinitely small perturbations. Therefore recommended values of τ can be those with $Sens(\tau) < 1$. In Figure 2, it can be seen that X_6 is the most sensitive node for all the values of τ, then if its sensitivity measure is restricted to be less than one, the rest of the nodes will be controlled. The red zone of recommended values corresponds to $\tau < 12.130363$.

5 Conclusions

In this work a sensitivity analysis to evaluate the effect of unknown prior hyperparameters in GBN is developed. The Kullback-Leibler divergence is used to determine deviations of perturbed models from the original ones, both in prior and posterior distributions. A local sensitivity measure to compare posterior and prior behavior to hyperparameters perturbations is proposed. From a robust Bayesian perspective, a range of values for the hyperparameters satisfying our sensitivity measure less than one is desirable in order to get a posterior effect to hyperparameter perturbations smaller than prior. It is shown that this condition is always satisfied for the hyperparameter λ, whereas the hyperparameter τ needs a particular analysis for each network.

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

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

[2] Dobra A, Hans C, Jones B, Nevins JR, Yao G and West M. Sparse graphical models for exploring gene expression data. *Journal of Multivariate Analysis* (2004); 91:196-212.

[3] Geiger, D. and Heckerman, D. Learning Parameters Priors for Directed Acyclic Graphical Models and the Characterization of Several Probability Distributions. *The Annals of Statistics* (2002); 30:1412-1440.

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

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

[6] Gómez-Villegas M.A, 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.

[7] Gómez-Villegas M.A, Main P, Susi R. Sensitivity of Gaussian Bayesian Networks to Inaccuracies in Their Parameters. *Proceedings of the 4^{th} European Workshop on Probabilistic Graphical Models*. Hirtshals, Denmark: (2008): 145-152.

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

Cuadernos de Trabajo
Escuela Universitaria de Estadística

CT03/2010 **Sensitivity to hyperprior parameters in Gaussian Bayesian networks.**
M.A. Gómez-Villegas, P. Main, H. Navarro y R. Susi

CT02/2010 **Las políticas de conciliación de la vida familiar y laboral desde la perspectiva del empleador. Problemas y ventajas para la empresa.**
R. Albert, L. Escot, J.A. Fernández Cornejo y M.T. Palomo

CT01/2010 **Propiedades exóticas de los determinantes**
Venancio Tomeo Perucha

CT05/2009 **La predisposición de las estudiantes universitarias de la Comunidad de Madrid a auto-limitarse profesionalmente en el futuro por razones de conciliación**
R. Albert, L. Escot y J.A. Fernández Cornejo

CT04/2009 **A Probabilistic Position Value**
A. Ghintran, E. González–Arangüena y C. Manuel

CT03/2009 **Didáctica de la Estadística y la Probabilidad en Secundaria: Experimentos motivadores**
A. Pajares García y V. Tomeo Perucha

CT02/2009 **La disposición entre los hombres españoles a tomarse el permiso por nacimiento. ¿Influyen en ello las estrategias de conciliación de las empresas?**
L. Escot, J.A. Fernández-Cornejo, C. Lafuente y C. Poza

CT01/2009 **Perturbing the structure in Gaussian Bayesian networks**
R. Susi, H. Navarro, P. Main y M.A. Gómez-Villegas

CT09/2008 **Un experimento de campo para analizar la discriminación contra la mujer en los procesos de selección de personal**
L. Escot, J.A. Fernández Cornejo, R. Albert y M.O. Samamed

CT08/2008 **Laboratorio de Programación. Manual de Mooshak para el alumno**
D. I. de Basilio y Vildósola, M. González Cuñado y C. Pareja Flores

CT07/2008 **Factores de protección y riesgo de infidelidad en la banca comercial**
J. Mª Santiago Merino

CT06/2008 **Multinationals and foreign direct investment: Main theoretical strands and empirical effects**
María C. Latorre

CT05/2008 **On the Asymptotic Distribution of Cook's distance in Logistic Regression Models**
Nirian Martín y and Leandro Pardo

CT04/2008 **La innovación tecnológica desde el marco del capital intelectual**
Miriam Delgado Verde, José Emilio Navas López, Gregorio Martín de Castro y Pedro López Sáez

CT03/2008 **Análisis del comportamiento de los indecisos en procesos electorales: propuesta de investigación funcional predictivo-normativa**
J. Mª Santiago Merino

CT02/2008 **Inaccurate parameters in Gaussian Bayesian networks**
Miguel A. Gómez-Villegas, Paloma Main and Rosario Susi

CT01/2008 **A Value for Directed Communication Situations.**
E. González–Arangüena, C. Manuel, D. Gómez, R. van den Brink

UNIVERSIDAD COMPLUTENSE
MADRID