Journal of Multivariate Analysis 124 (2014) 214–225

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Sensitivity to hyperprior parameters in Gaussian Bayesian networks

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

[a] Departamento de Estadística e Investigación Operativa I, Universidad Complutense de Madrid, Plaza de Ciencias 3, 28040 Madrid, Spain

[b] Departamento de Estadística, Investigación Operativa y Cálculo numérico, UNED, Paseo Senda del Rey 9, 28040 Madrid, Spain

[c] Departamento de Estadística e Investigación Operativa III, Universidad Complutense de Madrid, Avda. Puerta de Hierro s/n, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 24 April 2013
Available online 11 November 2013

AMS subject classifications:
62F15
62F35

Keywords:
Gaussian Bayesian network
Kullback–Leibler divergence
Bayesian linear regression

ABSTRACT

Bayesian networks (BNs) have become an essential tool for reasoning under uncertainty in complex models. In particular, the subclass of Gaussian Bayesian networks (GBNs) can be used to model continuous variables with Gaussian distributions. Here we focus on the task of learning GBNs from data. Factorization of the multivariate Gaussian joint density according to a directed acyclic graph (DAG) provides an alternative and interchangeable representation of a GBN by using the Gaussian conditional univariate densities of each variable given its parents in the DAG. With this latter conditional specification of a GBN, the learning process involves determination of the mean vector, regression coefficients and conditional variances parameters. Some approaches have been proposed to learn these parameters from a Bayesian perspective using different priors, and therefore some hyperparameter values are tuned. Our goal is to deal with the usual prior distributions given by the normal/inverse gamma form and to evaluate the effect of prior hyperparameter choice on the posterior distribution. As usual in Bayesian robustness, a large class of priors expressed by many hyperparameter values should lead to a small collection of posteriors. From this perspective and using Kullback–Leibler divergence to measure prior and posterior deviations, a local sensitivity measure is proposed to make comparisons. If a robust Bayesian analysis is developed by studying the sensitivity of Bayesian answers to uncertain inputs, this method will also be useful for selecting robust hyperparameter values.

1. Introduction

Bayesian networks (BNs) are graphical probabilistic models of interactions between a set of variables for which the joint probability distribution can be described in graphical terms. BNs consist of qualitative and quantitative parts $(\mathcal{G}, \mathcal{P})$. The qualitative part, $\mathcal{G}$, comprises a directed acyclic graph (DAG) useful for defining dependence and independence among variables $\mathbf{X} = \{X_1, \ldots, X_p\}$. The DAG shows the set of variables of the model at nodes, and the presence of arcs represents the dependence between variables. In the quantitative part, $\mathcal{P}$, it is necessary to determine the set of

* Corresponding author.
E-mail addresses: ma.gv@mat.ucm.es, villegas@ucm.es (M.A. Gómez-Villegas).

0047-259X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmva.2013.10.022

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. Then, the set $\mathcal{P}$ defines the associated joint probability distribution

$$P(\mathbf{X}) = \prod_{i=1}^{p} P(X_i|pa(X_i))$$

with $\mathcal{P} = \{P(X_1|pa(X_1)), \ldots, P(X_p|pa(X_p))\}$.

Among others, BNs have been studied by Pearl [15], Lauritzen [14], Cowell et al. [3] and Jensen et al. [12].

In this work, we focus on a subclass of BNs known as Gaussian Bayesian networks (GBNs). GBNs have been treated by authors like Shachter, et al. [17], Castillo, et al. [1,2], Dobra, et al. [6] and Cowell, et al. [3].

GBNs are defined as BNs for which the probability density of $\mathbf{X} = (X_1, \ldots, X_p)'$ is a multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ is the p-dimensional mean vector and $\boldsymbol{\Sigma}$ is a $p \times p$ positive definite covariance matrix for which the dependence structure is shown in a DAG. Then the joint density can be factorized using the conditional probability densities for every X_i ($i = 1, \ldots, p$) given its parents in the DAG, $pa(X_i) \subset \{X_1, \ldots, X_{i-1}\}$. These are univariate normal distributions with density

$$f(x_i|pa(X_i)) \sim N_1\left(x_i|\mu_i + \sum_{j=1}^{i-1} \beta_{ji}(x_j - \mu_j), v_i\right),$$

where μ_i 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. Remark that the presence of arcs represents the dependence between variables, therefore $\beta_{ji} = 0$ if and only if there is no link from X_j to X_i, with $j < i$.

The conditional and joint specifications of a GBN are interchangeable and we can work equivalently with both parameterizations considering $\boldsymbol{\Sigma} = [(I - \mathbf{B})^{-1}]'\mathbf{D}(I - \mathbf{B})^{-1}$ [17], where $\mathbf{D}$ is the diagonal matrix $\mathbf{D} = diag(\mathbf{v})$ with the conditional variances $\mathbf{v}' = (v_1, \ldots, v_p)$ and $\mathbf{B}$ is a strictly upper triangular matrix with the regression coefficients β_{ji} with $j = 1, \ldots, i - 1$.

The problems of learning both sets of parameters are also equivalent if some particular prior distributions are used.

In general, building a BN is a difficult task because it requires the user to specify the quantitative and qualitative parts of the network. Expert knowledge is important for fixing the dependence structure between the variables of the network and for determining a large set of parameters. In this process, it is possible to work with a database of cases, but the experience and knowledge of experts is also necessary. In GBNs the conditional specification of the model is manageable 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 the 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.

Literature about sensitivity analysis in GBNs is not extensive. Authors like Castillo & Kjærulff [2] or Gómez-Villegas et al. [8–10], have studied the problem of uncertainty in parameters assignments in GBNs. Castillo & Kjærulff [2] performed a one-way sensitivity analysis to investigate the impact of small changes in the network parameters, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, by computing partial derivatives of output probability of interest with respect to inaccurate parameters. A local sensitivity analysis is developed to evaluate small changes in the parameters. Gómez-Villegas et al. [8] proposed a one-way sensitivity analysis to evaluate the impact of small and large changes in the parameters over the network's output. Then, a global sensitivity measure is proposed to study the discrepancy of the output distribution of interest between two models, the initial and a perturbed model. Both analyses deal with variations in one parameter at a time holding the others fixed. Then, both are one-way sensitivity analyses.

As a generalization of the latter approach, Gómez-Villegas et al. [10] presented an n-way sensitivity analysis to evaluate uncertainty about a set of parameters.

Our objective here is to investigate uncertainty about the parameters of the conditional specification. To achieve this, we study the effect of different values for the prior hyperparameters on the posterior distribution.

The problem of Bayesian learning in this context has been handled with different approximations [6,7]. We work with the most commonly used, the normal/inverse gamma prior.

We study the effect of hyperparameter selection using Kullback–Leibler (KL) divergence [13]. This measure is used to define an appropriate local sensitivity measure to compare small prior and posterior deviations. From the results obtained it is possible to decide the values to chose for the hyperparameters considered.

The remainder of the paper is organized as follows. Section 2 introduces the problem assessment and the distributions considered. Section 3 is devoted to calculation of KL divergence measures. A local sensitivity measure is introduced in Section 4. Section 5 includes some examples and conclusions are drawn in Section 6.

2. Preliminary framework

The model of interest is the conditional specification of a GBN with parameters $\{\boldsymbol{\mu}, \mathbf{B}, \mathbf{D}\}$, where

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

being $\boldsymbol{\mu}$ the p-dimensional mean vector, $\mathbf{B}$ a strictly upper triangular matrix with the regression coefficients β_{ji} with $j = 1, \dots, i-1$ and $\mathbf{D}$ a diagonal matrix with the conditional variances v_i where $i = 1, \dots, p$.

Without loss of generality and to simplify further developments, we 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.

It can be pointed out that if $\beta_{ji} = 0$, there is no link from X_j to X_i, then X_j is not a parent of X_i (for $j < i$). Therefore, for each variable X_i and their parents $pa(X_i) \subset \{X_1, \dots, X_{i-1}\}$ we have a vector of dimension $i-1$ with the regression coefficients for the parents and zeros for the nodes in $\{X_1, \dots, X_{i-1}\}$ not connected to X_i.

Denoting the columns of $\mathbf{B}$ matrix by $\beta_i = (\beta_{1i}, \dots, \beta_{i-1i})'$ for $i > 1$, the conditional specification is now given by $\{v_1, \beta_i, v_i\}_{i>1}$, where v_1 is the marginal variance of X_1.

In the next subsections, we compute the prior distributions, likelihood functions and posterior distributions for the parameters $\{v_1, \beta_i, v_i\}_{i>1}$. Orphan nodes (nodes or variables 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.

2.1. Nodes with parents

Consider a general node X_i with a nonempty set of parents $pa\,(X_i) \subset \{X_1, \dots, X_{i-1}\}$. Then we can establish the following distributions.

2.1.1. Prior distribution

From normal standard theory, an inverted Wishart is used as a prior distribution for the covariance matrix, and then a Wishart prior for the precision matrix $\boldsymbol{\Sigma}^{-1} \sim W_p(\lambda, \tau^{-1}I_p)$, where I_p is the identity matrix. It is well known that the normal-Wishart distribution is a conjugate family for multivariate-Normal sampling [4]. Also it has to be used as the only prior for GBNs if global parameters independence is assumed [7]. It can be shown that the implied prior distributions of the normal/inverse gamma form are $\beta_i \mid v_i \sim N_{i-1}\left(0, \tau^{-1}v_i I_{i-1}\right)$ with the hyperparameter $\tau > 0$ and $v_i \sim IG\left(\frac{\lambda+i-p}{2}, \frac{\tau}{2}\right)$, i.e., an inverse gamma with hyperparameters $\lambda > p$ and the previous $\tau > 0$. Then the joint prior distribution can be computed as

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

The corresponding prior distributions are

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

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

In Section 3 we propose a divergence measure to evaluate uncertainty about the hyperparameters λ and τ in terms of additive perturbations $\delta \in \mathbb{R}^+$, where $\lambda + \delta$ and $\tau + \delta$ are the perturbed hyperparameters. A symmetric study can be developed for negative perturbations with the corresponding restrictions.

If we are perturbing the first hyperparameter of the inverse gamma distribution, then λ is perturbed by adding δ. If we are perturbing the second hyperparameter of the inverse gamma distribution, which also appears in the variability of the normal distribution, then τ hyperparameter is perturbed by adding δ. Hereafter, λ and τ denote the first and second hyperparameters, respectively, of the inverse gamma distribution.

2.1.2. Likelihood function

Suppose that we observe a random sample of size n giving the data matrix

$$\left(\begin{array}{cc|c|cc} x_{11} & \dots & x_{1i} & \dots & x_{1p} \\ \vdots & & \vdots & & \vdots \\ x_{n1} & \dots & x_{ni} & \dots & x_{np} \end{array}\right).$$

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

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

and of X_i, $x_i = (x_{1i}, \dots, x_{ni})'$ and the regression model $x_i = X_{pa_i}\beta_i + \varepsilon_i$ $(i = 2, \dots, p)$ with $\varepsilon_i \sim N_n(0, v_i I_n)$. Then the likelihood function is

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

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

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

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

2.1.3. Posterior distribution

The joint posterior distribution is [6]

$$\pi(\beta_i, v_i \mid x_i, X_{pa_i}) \propto c_i \exp\left\{-\frac{1}{2v_i}\left[\tau + q_i + (\beta_i - \tilde{\beta}_i)' M_i (\beta_i - \tilde{\beta}_i)\right]\right\}, \quad \beta_i \in \mathbb{R}^{i-1},\ v_i > 0$$

with $c_i = \frac{\tau^{\frac{i-1}{2}}}{v_i^{\frac{\lambda+(i-p)+(i-1)+n}{2}+1}}$, $q_i = x_i' x_i - x_i' X_{pa_i}(M_i)^{-1} X'_{pa_i} x_i$, $M_i = \tau I_{i-1} + X'_{pa_i} X_{pa_i}$ and $\tilde{\beta}_i = M_i^{-1} X'_{pa_i} x_i$.

It immediately follows that the posterior densities of the parameters in the model, being $\pi(\beta_i \mid v_i, x_i, X_{pa_i})$ a Normal distribution $N_{i-1}\left(\tilde{\beta}_i, v_i (M_i)^{-1}\right)$ and $\pi(v_i \mid x_i, X_{pa_i})$ an Inverse Gamma distribution $IG\left(\frac{\lambda+(i-p)+n}{2}, \frac{\tau+q_i}{2}\right)$.

2.2. Orphan nodes

When a node X_i has no parents in the DAG, the parameter to be studied is only v_i.

2.2.1. 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 given by $x_i = (x_{1i}, \dots, x_{ni})'$. Then the likelihood function is

$$L(v_i; x_i) \propto v_i^{-\frac{n}{2}} \exp\left\{-\frac{1}{2v_i}\left(x_i' x_i\right)\right\}, \quad v_i > 0.$$

Therefore, the posterior distribution of the parameter to be considered is

$$\pi(v_i \mid x_i) \propto v_i^{-\frac{\lambda+(i-p)+n}{2}+1} \exp\left\{-\frac{1}{2v_i}\left(\tau + x_i' x_i\right)\right\}, \quad v_i > 0.$$

3. Divergence measure

In this section we compute the KL divergence to evaluate uncertainty in hyperparameters in terms of additive perturbations, $\delta \in \mathbb{R}^+$. The objective is to evaluate the effect of different perturbed hyperparameters (λ and τ) by means of the KL 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 $\delta = 0$.

To evaluate joint distributions, the next result relating marginal and conditional divergences is used:

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

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

Table 1
Prior and posterior distributions for the original and perturbed models.

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

Table 2
Prior and posterior distributions for the original and perturbed models.

Original model	Perturbed model
Prior distribution	
$\pi(v_i) \sim IG\left(\frac{\lambda+i-p}{2}, \frac{\tau}{2}\right)$ $\pi(\beta_i \mid v_i) \sim N_{i-1}\left(0, \tau^{-1} v_i I_{i-1}\right)$	$\pi^{\delta}(v_i) \sim IG\left(\frac{\lambda+i-p}{2}, \frac{\tau+\delta}{2}\right)$ $\pi^{\delta}(\beta_i \mid v_i) \sim N_{i-1}\left(0, (\tau+\delta)^{-1} v_i I_{i-1}\right)$
Posterior distribution	
$\pi(v_i \mid x_i, X_{pa_i}) \sim IG\left(\frac{\lambda+(i-p)+n}{2}, \frac{\tau+q_i}{2}\right)$	$\pi^{\delta}(v_i \mid x_i, X_{pa_i}) \sim IG\left(\frac{\lambda+(i-p)+n}{2}, \frac{\tau+\delta+q_i^{\delta}}{2}\right)$ with $q_i^{\delta} = x_i'x_i - x_i'X_{pa_i}\left(M_i^{\delta}\right)^{-1} X_{pa_i}'x_i$ and $M_i^{\delta} = (\tau+\delta)I_{i-1} + X_{pa_i}'X_{pa_i}$
$\pi(\beta_i \mid v_i, x_i, X_{pa_i}) \sim N_{i-1}\left(\tilde{\beta}_i, v_i (M_i)^{-1}\right)$	$\pi^{\delta}(\beta_i \mid v_i, x_i, X_{pa_i}) \sim N_{i-1}\left(\tilde{\beta}_i^{\delta}, v_i \left(M_i^{\delta}\right)^{-1}\right)$ with $\tilde{\beta}_i^{\delta} = (M_i^{\delta})^{-1} X_{pa_i}' x_i$

3.1. *Nodes with parents*

Let X_i be a general node with a nonempty set of parents $pa(X_i) \subset \{X_1, \ldots, X_{i-1}\}$. To compute the prior and posterior KL divergence between joint distributions of the original and perturbed models, we consider different perturbed models depending on the hyperparameter to be perturbed.

3.1.1. *Perturbed first hyperparameter of the inverse gamma distribution*

In this case, the perturbed model is obtained by adding δ to the hyperparameter λ, which only appears in the distribution of the parameter v_i. Then, using (1), the KL divergence of the joint distribution corresponds to the marginal distribution of v_i. Prior and posterior distributions for the original and perturbed models are shown in Table 1.

The *KL divergence between prior densities* is computed as

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

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

where $\Psi(x)$ is the digamma function.

The *KL divergence between posterior densities* is

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

$$D_{KLposterior} = \log \frac{\Gamma\left(\frac{\lambda+\delta+(i-p)+n}{2}\right)}{\Gamma\left(\frac{\lambda+(i-p)+n}{2}\right)} - \left(\frac{\delta}{2}\right)\Psi\left(\frac{\lambda+(i-p)+n}{2}\right). \tag{3}$$

3.1.2. *Perturbed second hyperparameter of the inverse gamma distribution*

The perturbed model is obtained by adding δ to the hyperparameter τ. This hyperparameter appears in the distribution of parameters β_i and v_i.

The prior and posterior distributions for the original and perturbed models are shown in Table 2 and then we calculate the KL divergence.

Therefore, the *KL divergence between joint prior densities* is given by

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

$$D_{KLprior} = \left(i + \frac{\lambda-(p+1)}{2}\right)\left[\left(\frac{\delta}{\tau}\right) - \log\left(1+\frac{\delta}{\tau}\right)\right]. \tag{4}$$

The expression obtained for the *KL divergence between posterior densities* is

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

$$D_{KLposterior} = \frac{1}{2}\left[\ln\frac{|M_i|}{|M_i^{\delta}|} + \delta tr\left(M_i^{-1}\right) + \frac{\lambda+(i-p)+n}{\tau+q_i}\delta^2\tilde{\beta}_i^T\left(M_i^{\delta}\right)^{-1}\tilde{\beta}_i\right] + \frac{\lambda+(i-p)+n}{2}\left[\frac{\delta+\left(q_i^{\delta}-q_i\right)}{\tau+q_i} - \log\left(1+\frac{\delta+\left(q_i^{\delta}-q_i\right)}{\tau+q_i}\right)\right]. \quad (5)$$

Details on the calculations can be found in Appendix A.

3.2. Orphan nodes

The previous calculations are used to evaluate differences between distributions in the case in which the set of parents of orphan nodes is empty.

Although the only parameter to consider in this case is υ_i, again we have to consider two different perturbed models, depending on the perturbed hyperparameter, λ or τ.

3.2.1. Perturbed first hyperparameter of the inverse gamma distribution

When uncertainty is about hyperparameter λ, the results are the same as for nodes with parents.

3.2.2. Perturbing second hyperparameter of the inverse gamma distribution

Finally, when uncertainty is about τ and the perturbed model is obtained by adding δ to the hyperparameter τ, the *KL divergence between prior distributions* is the first summand of the expression for nodes with parents:

$$D_{KLprior} = D_{KL}(\pi^{\delta}(\upsilon_i) \mid \pi(\upsilon_i))$$

$$D_{KLprior} = \frac{\lambda+(i-p)}{2}\left[\left(\frac{\delta}{\tau}\right) - \log\left(1+\frac{\delta}{\tau}\right)\right]. \quad (6)$$

The *KL divergence between posterior distributions* is given by

$$D_{KLposterior} = D_{KL}(\pi^{\delta}(\upsilon_i|x_i) \mid \pi(\upsilon_i|x_i))$$

$$D_{KLposterior} = \frac{\lambda+(i-p)+n}{2}\left[\frac{\delta}{\tau+x_i'x_i} - \log\left(1+\frac{\delta}{\tau+x_i'x_i}\right)\right]. \quad (7)$$

4. Sensitivity measure

To assess the sensitivity of the posterior to prior variations given by small perturbations in the hyperprior parameters, we consider a local sensitivity measure under KL divergence [5,11], given by

$$Sens = \lim_{\delta\to 0}\frac{D_{KLposterior}}{D_{KLprior}} = \lim_{\delta\to 0}\frac{D_{KL}(\pi^{\delta}(\beta_i, \upsilon_i \mid x_i, X_{pa_i}) \mid \pi(\beta_i, \upsilon_i \mid x_i, X_{pa_i}))}{D_{KL}(\pi^{\delta}(\beta_i, \upsilon_i) \mid \pi(\beta_i, \upsilon_i))}.$$

This local sensitivity measure is defined to compare prior and posterior deviations. With this measure it is possible to establish a range of values for the hyperparameters to achieve a sensitivity measure of less than one. This is desirable to obtain a posterior effect for hyperparameter perturbations smaller than the prior. As shown in this subsection, this condition is always satisfied for the hyperparameter λ, whereas the hyperparameter τ needs a particular analysis for each case.

4.1. Nodes with parents

For node X_i with a nonempty set of parents, the sensitivity measures obtained for different perturbed models are described below.

4.1.1. Perturbed first hyperparameter of the inverse gamma distribution

The next result is obtained by computing the sensitivity measure when uncertainty is about λ to compare prior (2) and posterior (3) deviations:

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

where Ψ' is the trigamma function.

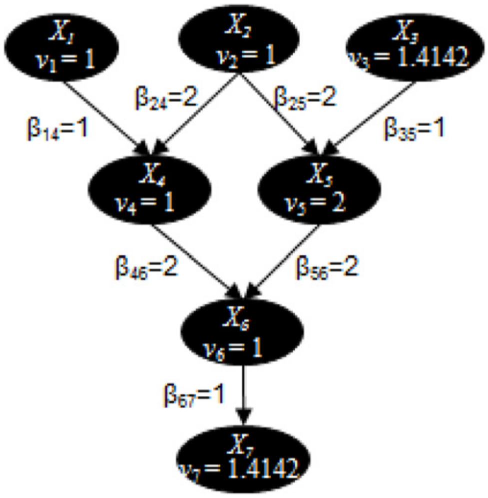

Fig. 1. Directed acyclic graph representation of the Gaussian Bayesian network of interest.

This is always less than one because the trigamma function $\Psi'(x)$ is monotone decreasing and is also monotonically dominated when the node index increases. As can be seen, the discrepancy between posterior distributions is the numerator of this expression and similarly the discrepancy between prior distributions is the denominator. This quotient less than one, is a very interesting result from a robust Bayesian perspective, because final distributions are more similar than initial distributions and therefore, the posterior effect for hyperparameter perturbation is smaller than the prior. If we have a value larger than one, some problems of sensitivity of conclusions to assumptions may occur. For a discussion about this concepts from a Bayesian perspective see [18]. In this case of uncertainty about first hyperparameter of inverse gamma distribution in nodes with parents, the condition is always satisfied for any hyperparameter λ (for details, see Appendix B).

4.1.2. Perturbed second hyperparameter of the inverse gamma distribution

When there is uncertainty about τ, the sensitivity measure that compares the divergence of prior (4) and posterior (5) is

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

where $\{\lambda_k\}_{k=1,\ldots,i-1}$ are the eigenvalues of the $X'_{pa_i}X_{pa_i}$ matrix and $\{\lambda_k + \tau\}_{k=1,\ldots,i-1}$ are those for M_i.

Optimal values of τ for a sensitivity measure of less than one can be analyzed for each GBN. The calculations are presented in Appendix B.

4.2. Orphan nodes

When node X_i has no parents in the DAG, the only perturbation to be analyzed corresponds to the hyperparameter τ because the same results for nodes with parents can be applied to orphan nodes if λ is considered.

4.2.1. Perturbed second hyperparameter of the inverse gamma distribution

The sensitivity measure computed when there is uncertainty about the hyperparameter τ to compare the divergence for prior (6) and posterior (7) is given by

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

For details of the calculations, see Appendix C.

5. Experiments

Consider a GBN with parameters β_{ji} and $v_i, j < i$, and a dependence structure given by the DAG in Fig. 1.

Table 3
Sensitivity measure for different values of perturbed λ.

λ	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
1,000	**0.498**	**0.499**	**0.499**	**0.499**	**0.499**	**0.500**	**0.500**
10,000	0.909	0.909	0.909	0.909	0.909	0.909	0.909

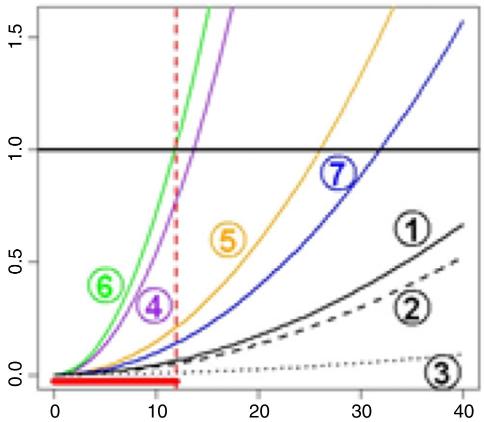

Fig. 2. Directed acyclic graph representation of the Gaussian Bayesian network of interest.

An artificial sample of size $n = 1000$ was simulated using R, an open source programming language and environment for statistical computing and graphics [16]. With the sensitivity measure introduced in Section 4, the following results were obtained for the two types of perturbation.

Sensitivity measure for perturbed first hyperparameter of the inverse gamma distribution

As observed in Table 3, the sensitivity measure for each variable is very similar for all nodes. Moreover, the measure increases with values of λ but is always less than 1. Thus, an effect size requires a value of $\lambda = 1000$.

Sensitivity measure for perturbed second hyperparameter of the inverse gamma distribution

Fig. 2 shows the sensitivity measure obtained for $\tau > 0$, in which different colored lines represent each variable and node numbers are indicated by circles.

When $Sens(\tau) < 1$, the posterior KL divergence is less than the prior KL divergence for infinitely small perturbations. Therefore, τ values for which $Sens(\tau) < 1$ are recommended. In Fig. 2 it is evident that X_6 is the most sensitive node for all values of τ; thus, if its sensitivity measure is restricted to values less than one, the rest of the nodes will be controlled. The red zone for recommended values corresponds to $\tau < 12.1304$.

6. Conclusions

In this work we performed a sensitivity analysis to evaluate the effect of unknown prior hyperparameters in GBNs. We used KL divergence to determine deviations of perturbed models from the original ones, for both prior and posterior distributions. With those deviations, a local sensitivity measure to compare posterior and prior behavior for small hyperparameter perturbations is proposed.

Determining the sensitivity to small changes in λ and τ hyperparameters is useful to study the robustness from a practical standpoint because from a robust Bayesian perspective, a range of values for the hyperparameters satisfying our sensitivity measure of less than one is desirable to obtain a posterior effect for hyperparameter perturbations smaller than the prior. We showed that this condition is always satisfied for the hyperparameter λ, whereas the hyperparameter τ needs a particular analysis for each network.

With the sensitivity analysis proposed it is possible to determine the hyperparameters values when describing a GBN with the conditional specification, to get a posterior effect for uncertain hyperparameters smaller than prior. Therefore, this methodology introduces a new method for determining how to select the tuning parameter appropriately in Bayesian learning of GBNs. We propose select the tuning parameter by introducing a measure of sensitivity for each node. Then, Bayesian robustness is applied by requiring smaller posterior deviations than priors for each node. Finally an admissible range of values for the tuning parameter is obtained and the maximum is selected.

Acknowledgments

This research was supported by the Spanish Ministerio de Ciencia e Innovación, Grant MTM 2008 · 03282, and in part by GR58/08-A, 910395—Métodos Bayesianos, BSCH-UCM, Spain. The very constructive comments of the Editor and reviewers are gratefully acknowledge.

Appendix A

KL divergence for uncertainty in τ

Prior distributions:

$$\begin{aligned} D_{KLprior} &= D_{KL}(\pi^{\delta}(\beta_i, v_i) \mid \pi(\beta_i, v_i)) \\ &= D_{KL}(\pi^{\delta}(v_i) \mid \pi(v_i)) + \int \pi(v_i) D_{KL}(\pi^{\delta}(\beta_i \mid v_i) \mid \pi(\beta_i \mid v_i)) dv_i \\ &= \frac{(i-1)}{2}\left[\left(\frac{\delta}{\tau}\right) - \log\left(1+\frac{\delta}{\tau}\right)\right] + \frac{\lambda+(i-p)}{2}\left[\left(\frac{\delta}{\tau}\right) - \log\left(1+\frac{\delta}{\tau}\right)\right] \\ &= \left(i + \frac{\lambda-(p+1)}{2}\right)\left[\left(\frac{\delta}{\tau}\right) - \log\left(1+\frac{\delta}{\tau}\right)\right]. \end{aligned}$$

Posterior distributions:

$$\begin{aligned} D_{KLposterior} &= D_{KL}(\pi^{\delta}(\beta_i, v_i \mid x_i, X_{pa_i}) \mid \pi(\beta_i, v_i \mid x_i, X_{pa_i})) \\ &= \int \pi(v_i \mid x_i, X_{pa_i}) D_{KL}\left(\pi^{\delta}\left(\beta_i \mid v_i, x_i, X_{pa_i}\right) \mid \pi\left(\beta_i \mid v_i, x_i, X_{pa_i}\right)\right) dv_i \\ &\quad + D_{KL}(\pi^{\delta}(v_i \mid x_i, X_{pa_i}) \mid \pi(v_i \mid x_i, X_{pa_i})) = (1) + (2). \end{aligned}$$

The first summand is

$$\begin{aligned} (1) = \frac{1}{2}\Bigg[&\log\frac{|M_i|}{\left|M_i^{\delta}\right|} + tr\left(M_i^{\delta} M_i^{-1}\right) - (i-1) + (\tilde{\beta}_i - \tilde{\beta}_i^{\delta})' M_i^{\delta} (\tilde{\beta}_i - \tilde{\beta}_i^{\delta}) \\ &\times \int \frac{1}{v_i} \frac{\left(\frac{\tau+q_i}{2}\right)^{\frac{\lambda+(i-p)+n}{2}}}{\Gamma\left(\frac{\lambda+(i-p)+n}{2}\right)} v_i^{-\left(\frac{\lambda+(i-p)+n}{2}+1\right)} \exp\left\{-\frac{1}{2v_i}(\tau+q_i)\right\} dv_i \Bigg]. \end{aligned}$$

Then with some calculations and using

- $$\begin{aligned} M_i^{\delta} &= M_i + \delta I_{i-1} \to \begin{cases} M_i^{\delta} M_i^{-1} = I_{i-1} + \delta M_i^{-1} \\ M_i^{-1} = \left(M_i^{\delta}\right)^{-1}\left(I_{i-1} + \delta M_i^{-1}\right) \end{cases} \\ &\to \left\{ M_i^{-1} - \left(M_i^{\delta}\right)^{-1} = \delta\left(M_i^{\delta}\right)^{-1} M_i^{-1} \right. \end{aligned}$$
- $tr\left(M_i^{\delta} M_i^{-1}\right) = (i-1) + \delta tr\left(M_i^{-1}\right)$
- $(\tilde{\beta}_i - \tilde{\beta}_i^{\delta})' M_i^{\delta} (\tilde{\beta}_i - \tilde{\beta}_i^{\delta}) = x_i' X_{pa_i}\left(M_i^{-1} - \left(M_i^{\delta}\right)^{-1}\right)' M_i^{\delta}\left(M_i^{-1} - \left(M_i^{\delta}\right)^{-1}\right) X_{pa_i}' x_i = \delta^2 \tilde{\beta}_i'\left(M_i^{\delta}\right)^{-1} \tilde{\beta}_i$
- $\int \frac{1}{v_i} \frac{\left(\frac{\tau+q_i}{2}\right)^{\frac{\lambda+(i-p)+n}{2}}}{\Gamma\left(\frac{\lambda+(i-p)+n}{2}\right)} v_i^{-\left(\frac{\lambda+(i-p)+n}{2}+1\right)} \exp\left\{-\frac{1}{2v_i}(\tau+q_i)\right\} dv_i = \frac{\lambda+(i-p)+n}{\tau+q_i}$,

we obtain

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

The last summand is

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

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

Appendix B

Sensitivity measures for nodes with parents

Uncertainty about λ:
In this case,

$$\begin{aligned} Sens\,(\lambda) &= \lim_{\delta\to 0} \frac{D_{KL}(\pi^{\delta}(\beta_i, v_i \mid x_i, X_{pa_i}) \mid \pi(\beta_i, v_i \mid x_i, X_{pa_i}))}{D_{KL}(\pi^{\delta}(\beta_i, v_i) \mid \pi(\beta_i, v_i))} = \lim_{\delta\to 0} \frac{D_{KL}(\pi^{\delta}(v_i \mid x_i, X_{pa_i}) \mid \pi(v_i \mid x_i, X_{pa_i}))}{D_{KL}(\pi^{\delta}(v_i) \mid \pi(v_i))} \\ &= \lim_{\delta\to 0} \frac{\log \frac{\Gamma\left(\frac{\lambda+\delta+(i-p)+n}{2}\right)}{\Gamma\left(\frac{\lambda+(i-p)+n}{2}\right)} - \left(\frac{\delta}{2}\right)\Psi\left(\frac{\lambda+(i-p)+n}{2}\right)}{\log \frac{\Gamma\left(\frac{\lambda+\delta+(i-p)}{2}\right)}{\Gamma\left(\frac{\lambda+(i-p)}{2}\right)} - \left(\frac{\delta}{2}\right)\Psi\left(\frac{\lambda+(i-p)}{2}\right)} = \lim_{\delta\to 0} \frac{\frac{d}{d\delta}\Psi\left(\frac{\lambda+(i-p)+n+\delta}{2}\right)}{\frac{d}{d\delta}\Psi\left(\frac{\lambda+(i-p)+\delta}{2}\right)} \\ &= \frac{\Psi'\left(\frac{\lambda+(i-p)+n}{2}\right)}{\Psi'\left(\frac{\lambda+(i-p)}{2}\right)} < 1, \quad \text{where } \Psi' \text{ is the trigamma function.} \end{aligned}$$

Uncertainty about τ:
First we can consider

$$\begin{aligned} Sens\,(\tau) &= \lim_{\delta\to 0} \frac{D_{KL}(\pi^{\delta}(\beta_i, v_i \mid x_i, X_{pa_i}) \mid \pi(\beta_i, v_i \mid x_i, X_{pa_i}))}{D_{KL}(\pi^{\delta}(\beta_i, v_i) \mid \pi(\beta_i, v_i))} \\ &= \lim_{\delta\to 0} \frac{(1)+(2)}{D_{KL}(\pi^{\delta}(\beta_i, v_i) \mid \pi(\beta_i, v_i))} = \left(1^*\right) + \left(2^*\right). \end{aligned}$$

By calculating the two summands separately we obtain the limit.

$$\begin{aligned} \left(1^*\right) &= \lim_{\delta\to 0} \frac{\frac{1}{2}\left[\log \frac{|M_i|}{\left|M_i^{\delta}\right|} + \delta tr\left(M_i^{-1}\right) + \frac{\lambda+(i-p)+n}{\tau+q_i}\delta^2 \tilde{\beta}_i' \left(M_i^{\delta}\right)^{-1} \tilde{\beta}_i\right]}{\frac{\lambda+(i-p)+(i-1)}{2}\left[\left(\frac{\delta}{\tau}\right) - \log\left(1+\frac{\delta}{\tau}\right)\right]} \\ &= \lim_{\delta\to 0} \frac{-\frac{d}{d\delta}\log\left|M_i^{\delta}\right| + tr\left(M_i^{-1}\right) + \frac{\lambda+(i-p)+n}{\tau+q_i}\frac{d}{d\delta}\left(\delta^2\tilde{\beta}_i'\left(M_i^{\delta}\right)^{-1}\tilde{\beta}_i\right)}{(\lambda+(i-p)+(i-1))\,\frac{\delta}{\tau(\tau+\delta).}} \end{aligned}$$

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

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

where $P = \left(e_1 \vdots \cdots \vdots e_{i-1}\right)$ are the eigenvectors of the orthogonal matrix. Then

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

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

$$= \frac{d}{d\delta}\sum_{k=1}^{i-1}\frac{z_{ik}^2\delta^2}{\lambda_k+\tau+\delta} = \sum_{k=1}^{i-1} z_{ik}^2 \frac{\delta^2+2\delta(\lambda_k+\tau)}{\lambda_k+\tau+\delta}, \quad \text{with } z_i = P'\tilde{\beta}_i = \begin{pmatrix} z_{i1} \\ \vdots \\ z_{ii-1}\end{pmatrix}.$$

Therefore,

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

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

Conversely,

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

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

$$\lim_{x\to 0}\frac{-\log\left(1+\frac{x+h(x)}{c_2}\right)+\frac{x+h(x)}{c_2}}{-\log\left(1+\frac{x}{c_1}\right)+\left(\frac{x}{c_1}\right)} = \frac{c_1^2}{c_2^2}\lim_{x\to 0}\left(1+\frac{d}{dx}h(x)\right)^2$$

and 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 $\frac{d}{d\delta}q_i^\delta$:

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

and with an eigenanalysis of the $X_{pa_i}'X_{pa_i}$ matrix and P as above, it follows that

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

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

with $w_i = P'X_{pa_i}'x_i = \begin{pmatrix} w_{i1} \\ \vdots \\ w_{ii-1}\end{pmatrix}$.

Thus,

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

and

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

yielding

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

As a final result, we obtain $\lim_{\delta\to 0}\frac{D_{KL}(\pi^{\delta}(\beta_i,v_i|x_i,X_{pa_i})|\pi(\beta_i,v_i|x_i,X_{pa_i}))}{D_{KL}(\pi^{\delta}(\beta_i,v_i)|\pi(\beta_i,v_i))} = (1^*)+(2^*)$.

Appendix C

Sensitivity measure for orphan nodes

$$\begin{aligned} Sens\,(\tau) &= \lim_{\delta\to 0}\frac{D_{KL}(\pi^{\delta}(v_i\mid x_i)\mid \pi(v_i\mid x_i))}{D_{KL}(\pi^{\delta}(v_i)\mid \pi(v_i))} = \lim_{\delta\to 0}\frac{\lambda+(i-p)+n}{\lambda+(i-p)}\frac{-\log\left(1+\frac{\delta}{\tau+x_i'x_i}\right)+\frac{\delta}{\tau+x_i'x_i}}{-\log\left(1+\frac{\delta}{\tau}\right)+\left(\frac{\delta}{\tau}\right)} \\ &= \frac{\lambda+(i-p)+n}{\lambda+(i-p)}\frac{\tau^2}{\left(\tau+x_i'x_i\right)^2}. \end{aligned}$$

References

[1] E. Castillo, J.M. Gutierrez, A.S. Hadi, Expert Systems and Probabilistic Network Models, Springer-Verlag, New York, 1997.
[2] E. Castillo, U. Kjærulff, Sensitivity analysis in Gaussian Bayesian networks using a symbolic-numerical technique, Reliab. Eng. Syst. Saf. 79 (2003) 139–148.
[3] R.G. Cowell, A.P. Dawid, S.L. Lauritzen, D.J. Spiegelhalter, Probabilistic Networks and Expert Systems, Springer, Barcelona, 1999.
[4] M.H. DeGroot, Optimal Statistical Decisions, McGraw-Hill, New York, 1970.
[5] D. Dey, S. Ghosh, K. Lou, On local sensitivity measures in Bayesian analysis, in: J.O. Berger, B. Betro, E. Moreno, L.R. Pericchi, F. Ruggeri, G. Salinetti, L. Wasserman (Eds.), Bayesian Robustness, in: IMS Lecture Notes Monograph Series, vol. 29, 1996, pp. 21–39 (with discussion).
[6] A. Dobra, C. Hans, B. Jones, J.R. Nevins, G. Yao, M. West, Sparse graphical models for exploring gene expression data, J. Multivariate Anal. 91 (2004) 196–212.
[7] D. Geiger, D. Heckerman, Parameters priors for directed acyclic graphical models and the characterization of several probability distributions, Ann. Statist. 30 (2002) 1412–1440.
[8] M.A. Gómez-Villegas, P. Main, R. Susi, Sensitivity analysis in Gaussian Bayesian networks using a divergence measure, Comm. Statist. Theory Methods 36 (3) (2007) 523–539.
[9] M.A. Gómez-Villegas, P. Main, R. Susi, Extreme inaccuracies in Gaussian Bayesian networks, J. Multivariate Anal. 99 (2008) 1929–1940.
[10] M.A. Gómez-Villegas, P. Main, R. Susi, The effect of block parameter perturbations in Gaussian Bayesian networks: sensitivity and robustness, Inform. Sci. 222 (2013) 439–458.
[11] P. Gustafson, C. Srinivasan, L.A. Wasserman, Local sensitivity analysis, in: J.M. Bernardo, J.O. Berger, A.P. Dawid, A.F.M. Smith (Eds.), Bayesian Statistics, vol. 5, Oxford University Press, 1996, pp. 197–210 (with discussion).
[12] F.V. Jensen, T. Nielsen, Bayesian Networks and Decision Graphs, sixth ed., Springer, New York, 2007.
[13] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Statist. 22 (1951) 79–86.
[14] S.L. Lauritzen, Graphical Models, Clarendon Press, Oxford, 1996.
[15] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, CA, 1988.
[16] R Development Core Team 2011, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN: 3-900051-07-0. http://www.R-project.org.
[17] R. Shachter, C. Kenley, Gaussian influence diagrams, Manag. Sci. 35 (1989) 527–550.
[18] A.F.M. Smith, Bayesian approaches to outliers and robustness, in: P. Florens, et al. (Eds.), Specifying Statistical Models, Springer-Verlag, New York, 1983.