Local Influence in Gaussian Bayesian networks

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

Métodos Bayesianos 2011. Madrid

Noviembre 2011

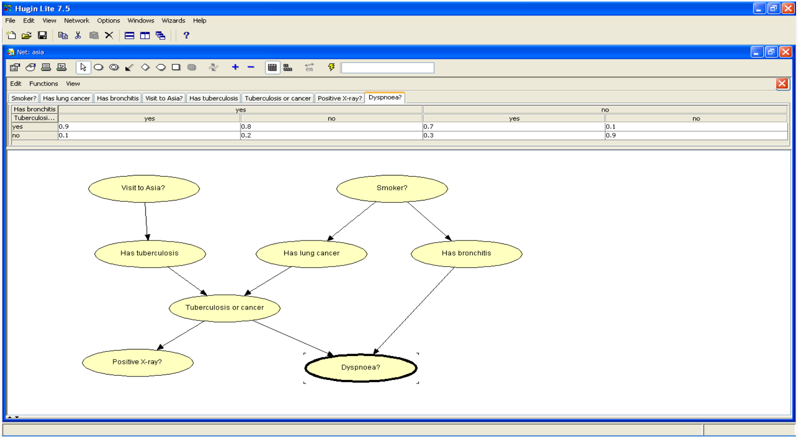
Bayesian networks

Definition

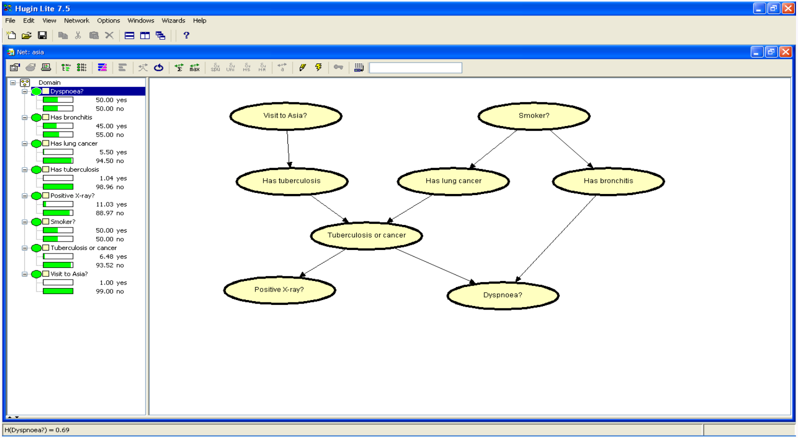
A Bayesian network is a couple $(\mathcal{G}, \mathcal{P})$*, where* $\mathcal{G}$ *is a DAG in which the nodes represent ordered random variables* $\mathbf{X} = \{X_1, \ldots, X_p\}$ *and the edges represent probabilistic dependencies.* $\mathcal{P} = \{P(X_1|pa(X_1)), \ldots, P(X_p|pa(X_p))\}$ *is a set of conditional probability distributions, where* $pa(X_i)$ *is the set of parents of node* X_i *in* $\mathcal{G}$*,* $pa(X_i) \subseteq \{X_1, ..., X_{i-1}\}$*.*
The set $\mathcal{P}$ *defines the associated joint probability distribution:*

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

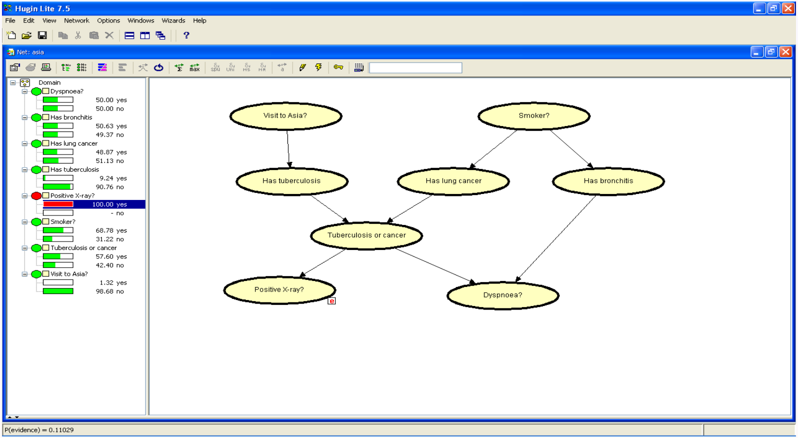
Bayesian network: Asia



Evidence propagation: Asia



Evidence propagation: Asia



Gaussian Bayesian networks

Definition

A Gaussian Bayesian network (GBN) is a BN where the joint probability density of $\mathbf{X} = (X_1, ..., X_p)'$ *is a multivariate normal distribution* $N_p(\mu, \mathbf{\Sigma})$ *with* μ *the* $p-$*dimensional mean vector,* $\mathbf{\Sigma}$ *the* $p \times p$ *positive definite covariance matrix and the dependence structure is fixed in a DAG.*

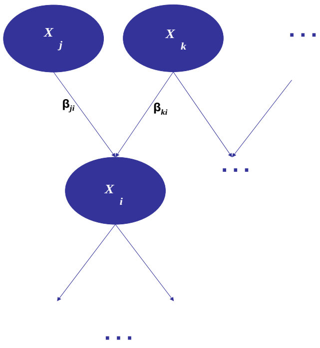
Gaussian Bayesian networks

Then, for every X_i $(i = 1, ..., p)$ given $pa(X_i) \subset \{X_1, ..., X_{i-1}\}$

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

with μ_i the X_i mean, β_{ji} the regression coefficients of X_i with respect to $X_j \in pa(X_i)$ and v_i the conditional variance of $X_i | pa(X_i)$

$\beta_{ji} = 0$ if X_j is not a parent of X_i



Conditional specification of the GBN

From the conditional specification it is possible to determine the parameters of the joint distribution

Conditional specification of the GBN

From the conditional specification it is possible to determine the parameters of the joint distribution

- μ_i are the elements of the $p-$dimensional mean vector μ

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}$$

Conditional specification of the GBN

From the conditional specification it is possible to determine the parameters of the joint distribution

- μ_i are the elements of the $p-$dimensional mean vector μ

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}$$

- v_i and β_{ji} are used to compute the covariance matrix $\boldsymbol{\Sigma}$

Conditional specification of the GBN

Let **D** be a diagonal matrix $\mathbf{D} = diag(\mathbf{v})$ with the conditional variances $\mathbf{v}' = (v_1, ..., v_p)$

$$\mathbf{D} = \begin{pmatrix} v_1 & & 0 \\ & \ddots & \\ 0 & & v_p \end{pmatrix}$$

Let **B** be a strictly upper triangular matrix with the regression coefficients β_{ji} where $j = 1, ..., i-1$.

$$\mathbf{B} = \begin{pmatrix} 0 & \beta_{12} & \dots & \beta_{1p} \\ & \ddots & & \\ & & \ddots & \beta_{p-1p} \\ 0 & & & 0 \end{pmatrix}$$

Conditional specification of the GBN

Let $\mathbf{D}$ be a diagonal matrix $\mathbf{D} = diag(\mathbf{v})$ with the conditional variances $\mathbf{v}' = (v_1, ..., v_p)$

$$\mathbf{D} = \begin{pmatrix} v_1 & & 0 \\ & \ddots & \\ 0 & & v_p \end{pmatrix}$$

Let $\mathbf{B}$ be a strictly upper triangular matrix with the regression coefficients β_{ji} where $j = 1, ..., i-1$.

$$\mathbf{B} = \begin{pmatrix} 0 & \beta_{12} & \dots & \beta_{1p} \\ & \ddots & & \\ & & \ddots & \beta_{p-1p} \\ 0 & & & 0 \end{pmatrix}$$

Then,

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

Objective

Study uncertainty about the parameters of the conditional specification of a GBN evaluating the effect of unknown prior hyperparameters.

Objective

Study uncertainty about the parameters of the conditional specification of a GBN evaluating the effect of unknown prior hyperparameters.

The Kullback–Leibler divergence is used to determine deviations of perturbed models from the original ones and to define a local sensitivity measure to compare prior and posterior deviations.

Objective

Study uncertainty about the parameters of the conditional specification of a GBN evaluating the effect of unknown prior hyperparameters.

The Kullback–Leibler divergence is used to determine deviations of perturbed models from the original ones and to define a local sensitivity measure to compare prior and posterior deviations.

With the obtained results it is possible to decide the values to be chosen for the hyperparameters considered.

Interest model

The interest model is given by the conditional specification of a GBN

Interest model

The interest model is given by the conditional specification of a GBN

Let us suppose that $\mu = \mathbf{0}$

Interest model

The interest model is given by the conditional specification of a GBN

Let us suppose that $\mu = \mathbf{0}$

Then, the parameters to be considered are

$$\beta_{ji} \text{ with } j < i \text{ and } v_i$$

Interest model

The interest model is given by the conditional specification of a GBN

Let us suppose that $\mu = \mathbf{0}$

Then, the parameters to be considered are

$$\beta_{ji} \text{ with } j < i \text{ and } v_i$$

$$\mathbf{B} = \begin{pmatrix} 0 & \beta_{12} & \cdots & \beta_{1p} \\ & \ddots & & \\ & & \ddots & \beta_{p-1p} \\ 0 & & & 0 \end{pmatrix} \Rightarrow \qquad \beta_i = \begin{pmatrix} \beta_{1i} \\ \beta_{2i} \\ \vdots \\ \beta_{i-1,i} \end{pmatrix}, \text{ for } i > 1$$

Interest model

The interest model is given by the conditional specification of a GBN

Let us suppose that $\mu = \mathbf{0}$

Then, the parameters to be considered are

$$\beta_{ji} \text{ with } j < i \text{ and } v_i$$

$$\mathbf{B} = \begin{pmatrix} 0 & \beta_{12} & \cdots & \beta_{1p} \\ & \ddots & & \\ & & \ddots & \beta_{p-1p} \\ 0 & & & 0 \end{pmatrix} \Rightarrow \quad \beta_i = \begin{pmatrix} \beta_{1i} \\ \beta_{2i} \\ \vdots \\ \beta_{i-1,i} \end{pmatrix}, \text{ for } i > 1$$

The parameters to be considered now are

$$\{v_1, \beta_i, v_i\}_{i>1}$$

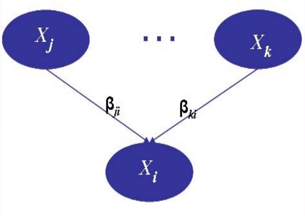
Next, we compute prior distributions, likelihood functions and posterior distributions for the parameters $\{v_1, \beta_i, v_i\}_{i>1}$.

Next, we compute prior distributions, likelihood functions and posterior distributions for the parameters $\{v_1, \beta_i, v_i\}_{i>1}$.

Orphan nodes (node/variable without parents in the DAG) are considered different from nodes with parents in the DAG.

Nodes with parents

Let be X_i a node with a nonempty set of parents $pa(X_i) \subset \{X_1, \dots, X_{i-1}\}$, that is, a node with parents in the DAG.



Prior distributions

$$\Sigma^{-1} \sim W_p(\lambda, \tau^{-1} I_p)$$

Prior distributions

$$\Sigma^{-1} \sim W_p(\lambda, \tau^{-1} I_p)$$

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

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

Prior distributions

$$\Sigma^{-1} \sim W_p(\lambda, \tau^{-1} I_p)$$

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

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

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

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

Prior distributions

$$\Sigma^{-1} \sim W_p(\lambda, \tau^{-1} I_p)$$

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

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

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

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

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

Likelihood function

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

For the variable X_i we consider the observations of its parents $pa(X_i)$

$$\left(\begin{array}{ccc|c|cc} x_{11} & x_{12} & \cdots & x_{1i} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2i} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{ni} & \cdots & x_{np} \end{array}\right) \qquad X_{pa_i} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1i-1} \\ x_{21} & x_{22} & \cdots & x_{2i-1} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{ni-1} \end{pmatrix}$$

$$x_i = (x_{1i}, x_{2i}, ..., x_{ni})^T \text{ observations of } X_i$$

Regression model:

$$x_i = X_{pa_i}\beta_i + \varepsilon_i \text{ for } i = 1, \ldots, p \text{ and } \varepsilon_i \sim N_n(0, v_i I_n)$$

Likelihood function

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

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

for

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

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

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

Posterior distribution

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

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

with

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

$$M_i = \tau I_{i-1} + X_{pa_i}^T X_{pa_i}$$

$$\tilde{\beta}_i = M_i^{-1} X_{pa_i}^T x_i$$

$$q_i = x_i^T x_i - x_i^T X_{pa_i} (M_i)^{-1} X_{pa_i}^T x_i$$

Posterior distributions

It follows the posterior densities of the parameters in the model

Posterior distributions

It follows the posterior densities of the parameters in the model

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

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

Posterior distributions

It follows the posterior densities of the parameters in the model

- $\pi(v_i | x_i X_{pa_i}) \propto \frac{\tau^{\frac{i-1}{2}}}{v_i^{\frac{\lambda+(i-p)+n}{2}+1}} \exp\left\{-\frac{1}{2v_i}(\tau + q_i)\right\}$, $v_i > 0$
  $\Rrightarrow$ an Inverse-Gamma distribution $IG\left(\frac{\lambda+(i-p)+n}{2}, \frac{\tau+q_i}{2}\right)$
- $\pi(\beta_i | v_i x_i X_{pa_i}) \propto \exp\left\{-\frac{1}{2v_i}\left[(\beta_i - \tilde{\beta}_i)^T M_i(\beta_i - \tilde{\beta}_i)\right]\right\}$,
  $\beta_i \in \mathbb{R}^{i-1}$
  $\Rrightarrow$ a normal distribution $N_{i-1}\left(\tilde{\beta}_i, v_i (M_i)^{-1}\right)$

Orphan nodes

X_i is an orphan node if X_i has no parents in the DAG, therefore there is no arc to X_i, then $\beta_{ji} = 0 \ \forall j < i$.

Orphan nodes

X_i is an orphan node if X_i has no parents in the DAG, therefore there is no arc to X_i, then $\beta_{ji} = 0 \; \forall j < i$.

The parameter to be studied is only v_i.

Prior distribution and likelihood function

Prior distribution and likelihood function

- Prior distribution:

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

Prior distribution and likelihood function

- Prior distribution:

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

- Likelihood function:
  The data are the observations of X_i, $x_i = (x_{1i}, x_{2i}, ..., x_{ni})^T$, then

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

Posterior distribution

- Posterior distribution:

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

$\Rightarrow$ an Inverse-Gamma distribution $IG\left(\frac{\lambda+(i-p)+n}{2}, \frac{\tau+x_i^T x_i}{2}\right)$

Divergence measure

We compute the Kullback-Leibler (KL) divergence to evaluate the effect of uncertainty in hyperparameters λ and τ, in terms of additive perturbations $\delta \in \mathbb{R}^+$.

Divergence measure

We compute the Kullback-Leibler (KL) divergence to evaluate the effect of uncertainty in hyperparameters λ and τ, in terms of additive perturbations $\delta \in \mathbb{R}^+$.
Then, we compute

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

and

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

where $\pi^\delta(\cdot)$ is the perturbed model obtained by adding a $\delta \in \mathbb{R}^+$ perturbation to hyperparameters ($\lambda > p$, $\tau > 0$).

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

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

Nodes with parents and orphan nodes

Nodes with parents

- $v_i \sim IG\left(\frac{\lambda+i-p}{2}, \frac{\tau}{2}\right)$ with $\lambda > p$ and $\tau > 0$
- $\beta_i | v_i \sim N_{i-1}\left(0, \tau^{-1} v_i I_{i-1}\right)$ with the hyperparameter $\tau > 0$

Orphan nodes

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

Hyperparameter perturbation $\lambda \Rightarrow \lambda + \delta$

KL divergence to compare prior distributions

Hyperparameter perturbation $\lambda \Rightarrow \lambda + \delta$

KL divergence to compare prior distributions

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

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

Hyperparameter perturbation $\lambda \Rightarrow \lambda + \delta$

KL divergence to compare prior distributions

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

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

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

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

with $\Psi(x)$ the digamma function

Hyperparameter perturbation $\lambda \Rightarrow \lambda + \delta$

KL divergence to compare posterior distributions

Hyperparameter perturbation $\lambda \Rightarrow \lambda + \delta$

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

Hyperparameter perturbation $\lambda \Rightarrow \lambda + \delta$

KL divergence to compare posterior distributions

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

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

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

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

Hyperparameter perturbation $\lambda \Rightarrow \lambda + \delta$

Sensitivity measure: $Sens_i(\lambda)$

Hyperparameter perturbation $\lambda \Rightarrow \lambda + \delta$

Sensitivity measure: $Sens_i(\lambda)$

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

with Ψ' the trigamma function

Hyperparameter perturbation $\tau \Rightarrow \tau + \delta$

KL divergence to compare prior distributions

Hyperparameter perturbation $\tau \Rightarrow \tau + \delta$

KL divergence to compare prior distributions

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

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

and

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

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

Hyperparameter perturbation $\tau \Rightarrow \tau + \delta$

KL divergence to compare prior distributions

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

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

and

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

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

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

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

Hyperparameter perturbation $\tau \Rightarrow \tau + \delta$

KL divergence to compare posterior distributions

Hyperparameter perturbation $\tau \Rightarrow \tau + \delta$

KL divergence to compare posterior distributions

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

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

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

and

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

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

Hyperparameter perturbation $\tau \Rightarrow \tau + \delta$

KL divergence to compare posterior distributions

Hyperparameter perturbation $\tau \Rightarrow \tau + \delta$

KL divergence to compare posterior distributions

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

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

Hyperparameter perturbation $\tau \Rightarrow \tau + \delta$

Sensitivity measure: $Sens_i(\tau)$

Hyperparameter perturbation $\tau \Rightarrow \tau + \delta$

Sensitivity measure: $Sens_i(\tau)$

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

Hyperparameter perturbation $\lambda \Rightarrow \lambda + \delta$

Hyperparameter perturbation $\lambda \Rightarrow \lambda + \delta$

Same results than for nodes with parents

Hyperparameter perturbation $\tau \Rightarrow \tau + \delta$

Hyperparameter perturbation $\tau \Rightarrow \tau + \delta$

KL divergence to compare prior distributions

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

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

Hyperparameter perturbation $\tau \Rightarrow \tau + \delta$

KL divergence to compare prior distributions

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

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

KL divergence to compare posterior distributions

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

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

Hyperparameter perturbation $\tau \Rightarrow \tau + \delta$

Sensitivity measure: $Sens_i(\tau)$

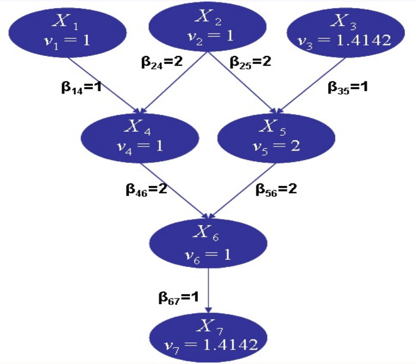
Hyperparameter perturbation $\tau \Rightarrow \tau + \delta$

Sensitivity measure: $Sens_i(\tau)$

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

GBN

Let us consider next GBN

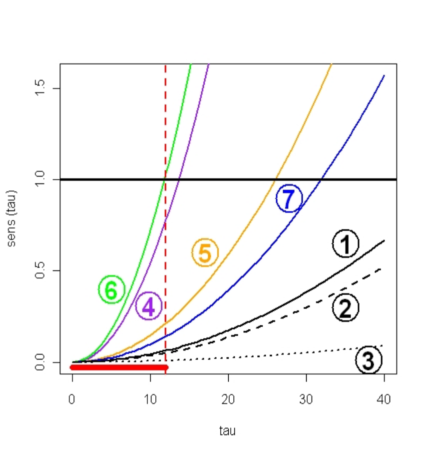


Hyperparameter perturbation $\lambda \rightarrow \lambda + \delta$

Sensitivity measures for different values of perturbed λ are computed, simulating an artificial sample of size $n = 1000$.

$\lambda \backslash X_i$	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

Hyperparameter perturbation $\tau \longrightarrow \tau + \delta$



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 $\tau < 12{,}130363$.

Conclusions

In this work a sensitivity analysis to evaluate the effect of unknown prior hyperparameters in GBN is developed.

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.

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.

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.

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.

This condition is always satisfied for the hyperparameter λ, whereas the hyperparameter τ needs a particular analysis for each network.

Bibliografía

- Lauritzen, S.L., Spiegelhalter, D.J. Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society, Series B* (1988); 50:157–224
- Shachter, R., Kenley, C. Gaussian influence diagrams. *Management Science* (1989); 35:527–550.
- Dobra, A., Hans, C., Jones, B., Nevins, J.R., Yao, G., West, M. Sparse graphical models for exploring gene expression data. *Journal of Multivariate Analysis* (2004); 91:196-212.
- Geiger, D., 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.

Bibliografía

- 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.
- Gómez-Villegas, M.A., Main, P., Susi, R. Sensitivity of Gaussian Bayesian Networks to Inaccuracies in Their Parameters. *Proceedings of the 4th European Workshop on Probabilistic Graphical Models*. Hirtshals, Denmark: (2008): 145-152.
- Susi, R., Navarro, H., Main, P., Gómez-Villegas, M.A. Perturbing the structure in Gaussian Bayesian networks. *Cuadernos de Trabajo - E.U. de Estadística* (2009): 1–21.

Thank you!