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