



Conference Proceedings Paper – Entropy

LOCAL EFFECT OF ASYMMETRY DEVIATIONS FROM GAUSSIANITY USING INFORMATION-BASED MEASURES

Paloma Main ^{1,*}, Jorge M Arevalillo ² and Hilario Navarro ²

¹ Departamento de Estadística e Investigación Operativa, Universidad Complutense de Madrid, Plaza de Ciencias 3, 28040 Madrid, Spain

² Departamento de Estadística, Investigación Operativa y Cálculo Numérico, UNED, Paseo Senda del Rey 9, 28040 Madrid, Spain

* Author to whom correspondence should be addressed; E-Mail: pmain@mat.ucm.es,
Tel.: +34-913944431, Fax: +34-913944606.

Published: 13 November 2015

Abstract: In this paper local sensitivity measures are proposed to evaluate deviations from multivariate Gaussianity caused by asymmetry; the model we use to regulate asymmetry is the multivariate skew-normal distribution because it reflects the deviation in a very tractable way. The paper also examines the connection between local sensitivity and Mardia's and Malkovich-Afifi's skewness indices. Once the local sensitivity measures have been introduced, we study the effect of local perturbations in asymmetry on the conditional distributions; this issue has important implications because there are many procedures in statistics and other fields where the output depends on the distribution of some variables for known values of the others. The proposed measures use the Kullback-Leibler divergence to evaluate dissimilarities between probability distributions in order to assess deviation from Gaussianity on the joint distribution and on the marginal and conditional distributions as well. The results are illustrated with some examples.

Keywords: Kullback-Leibler divergence measure; Local sensitivity; Non-Gaussianity

MSC classifications: 62H99

1. Introduction

Because non-Gaussian data are common in many applications, a natural question that arises is whether the usual methods can be applied when the probability models do not satisfy the Gaussianity assumption. We study this issue, present mathematical results, and discuss when and how, in spite of the asymmetry effect, they may be accomplished. The issue we consider here is the effect of deviated joint distributions on the corresponding marginal and conditional ones for a partition multivariate vector of the variables. Many procedures in statistics deal with a Gaussian random vector \mathbf{Z} divided in two blocks with different roles. One of these techniques for modeling continuous multivariate data is Gaussian Bayesian networks (GBN). They represent the dependence structure of a set of random variables with directed acyclic graphs (DAG), where the presence or absence of edges between the nodes with variables represent conditional dependence or independence of the variables in the model [1,2]. For inferential purposes, the network variables are partitioned into two categories: interest and evidence variables, which are also known as non-observable and observable variables and the output is given by the conditional distribution of the interest variables for some known values of the evidential variables. This inference procedure, termed evidence propagation, is computationally tractable under Gaussianity and linearity assumptions and the exact conditional distributions are available. However, Gaussian deviations introduce both computation and interpretation problems. Therefore, a key point is to know if the output of our model is robust to slight perturbations because it would allow the use of the well behaved methods of the Gaussian distribution when the actual model is not exactly as assumed. This problem has been studied for kurtosis perturbations using the multivariate power exponential distribution as the underlying model for handling kurtosis [1]. In this work we focus on asymmetry deviation from Gaussianity so the multivariate skew-normal (SN) distribution [2,3] is used as the alternative model. With this aim, we calculate the relative conditional sensitivity measure to evaluate the effect of the asymmetry on the marginal and conditional distributions. The tool to evaluate deviations from Gaussianity is the Kullback-Leibler (KL) divergence measure. Although it is a non-symmetric measure, it gives a good representation of deviations between Gaussian distributions, which involves tractable calculations; moreover, some recent studies also present fine results for SN distributions [4]. As an example we apply our measures to permutation-symmetric Gaussian variables, illustrated by a complete DAG with a common correlation coefficient; some available results for this case provide revealing findings when the proposed measures are used.

The paper is organized as follows: in the next section we review the fundamentals of the multivariate SN distribution. In Section 3 we propose the new local sensitivity measure and study its connection with Mardia's and Malkovich-Affifi's skewness indices; we also adapt the local measure to derive a local relative sensitivity index. Section 4 is concerned with the evaluation of the effect of asymmetry perturbations on the conditional distributions and a relative conditional sensitivity measure is defined as an essential tool to deal with deviation from Gaussianity characterized by skewness. We finish the paper with a summary of the main findings and some practical implications regarding the proposed measures.

2. The multivariate skew-normal distribution

The multivariate skew-normal distribution was introduced by [2] in order to model departures from multivariate Gaussianity which are related to asymmetry. From the very beginning the SN multivariate distribution has been widely accepted as a modeling tool in scenarios where the Gaussianity assumption on the stochastic mechanism that generates the data fails. Its tractability and appealing properties have enhanced the usefulness of the SN distribution both in theoretical and applied multivariate analysis. Some works discussing its main properties with an emphasis on applications are Azzalini and Capitanio [35], Capitanio *et al.* [6], Azzalini [7], Loperfido [8], just to name a few.

In this paper we adopt the notation of the aforementioned works to define the density function of a p -dimensional SN random vector with location vector $\xi = (\xi_1, \dots, \xi_p)'$ and scale matrix Ω as follows:

$$f(z; \xi, \alpha, \Omega) = 2\phi_p(z - \xi; \Omega)\Phi(\alpha'\omega^{-1}(z - \xi)) \quad : \quad z \in \mathbb{R}^p, \quad (1)$$

where $\phi_p(\cdot; \Omega)$ denotes the p -dimensional Gaussian density with zero mean and full rank covariance matrix Ω , Φ is the distribution function of a standard Gaussian random variable, $\omega = \text{diag}(\omega_1, \dots, \omega_p)$ is a scale diagonal matrix with non negative entries such that $\bar{\Omega} = \omega^{-1}\Omega\omega^{-1}$ is a correlation matrix, and α is a p -dimensional shape vector regulating the skewness of the distribution.

We will write $Z \sim SN_p(\xi, \Omega, \alpha)$ to denote that Z follows a p -dimensional SN distribution with density function (1). When $\alpha = \mathbf{0}$ the multivariate SN becomes a p -dimensional Gaussian variate with mean vector ξ and covariance matrix Ω ; so we can put $Z \sim SN_p(\xi, \Omega, \mathbf{0})$ or equivalently $Z \sim N_p(\xi, \Omega)$.

It is worthwhile noting that ξ and Ω do not correspond to the mean vector and the covariance matrix of the SN variate unless $\alpha = \mathbf{0}$, i.e in the Gaussian case. In fact, $\mathbb{E}(Z) = \xi + \sqrt{2/\pi}\omega\delta$ and $\text{Var}(Z) = \Omega - (2/\pi)\omega\delta\delta'\omega$, where $\delta = (1 + \alpha'\bar{\Omega}\alpha)^{-1/2}\bar{\Omega}\alpha$ [3].

On the other hand, we can also observe that $Z = \xi + \omega Z_0$, where Z_0 is a multivariate skew-normal variable with density function given by

$$f(z; \mathbf{0}, \alpha, \Omega) = 2\phi_p(z; \bar{\Omega})\Phi(\alpha'z) \quad : \quad z \in \mathbb{R}^p. \quad (2)$$

We refer to Z_0 as the “normalized” multivariate SN variate [3].

An appealing feature of the SN distribution with density (1) is that classical non Gaussianity indices like Mardia’s measures of skewness and kurtosis [9] and Malkovich-Afifi’s skewness and kurtosis [10] admit a simple closed form. From [5] we know that Mardia’s measures of skewness and kurtosis are given by

$$\gamma_{1,p}^M = 2(4 - \pi)^2 \left\{ \frac{\alpha'\bar{\Omega}\alpha}{\pi + (\pi - 2)\alpha'\bar{\Omega}\alpha} \right\}^3 \quad (3)$$

$$\gamma_{2,p}^M = p(p + 2) + 8(\pi - 3) \left\{ \frac{\alpha'\bar{\Omega}\alpha}{\pi + (\pi - 2)\alpha'\bar{\Omega}\alpha} \right\}^2. \quad (4)$$

On the other hand, Malkovich-Afifi’s indices of skewness and kurtosis, which will be denoted by $\gamma_{1,p}^D$ and $\gamma_{2,p}^D$, are closely related to Mardia’s measures; they are given by $\gamma_{1,p}^D = \gamma_{1,p}^M$ and $\gamma_{2,p}^D = \gamma_{2,p}^M - p(p + 2) + 3$ [8].

All these measures, along with their omnibus versions, share the common characteristic of being non decreasing functions of the quantity $\alpha_* = \alpha'\bar{\Omega}\alpha$; hence, α_* can be understood as a scalar that

summarizes the departure from Gaussianity in the SN model. This finding has been previously reported by [3] who presented α_* as an indicator that encapsulates the non Gaussianity of the multivariate SN distribution. In the next section we provide a deeper insight about how to interpret this quantity, as well as what it measures intrinsically; we will also examine its meaning in connection with the local deviation index we are going to introduce now.

3. Local deviation and relative local deviation indices

The KL divergence was introduced as a generalization of Shannon's entropy and has been widely used in both statistical inference and information theory. This divergence is a non-symmetric measure that provides global information on the difference between two probability distributions [11].

The KL divergence between two probability densities f and g , over the same domain \mathcal{X} , can be defined as follows

$$D_{KL}(f, g) = \int_{\mathcal{X}} f(w) \ln \frac{f(w)}{g(w)} dw. \quad (5)$$

We can say that (5) is the KL divergence between densities f and g or alternative the KL divergence between their corresponding univariate random variables X and Y , so we can also denote it by $D_{KL}(X, Y)$. The definition can be extended to multivariate density functions in a natural way to deal with the KL divergence of multivariate random vectors.

The main results of this paper rely on the KL divergence between multivariate Gaussian and SN variables. The next proposition provides a result for calculating it. For the sake of simplicity we take $\xi = \mathbf{0}$.

Proposition 1. *Let \mathbf{Z} and \mathbf{Z}^α be p -dimensional random vectors such that $\mathbf{Z} \sim SN_p(\mathbf{0}, \mathbf{\Omega}, \mathbf{0})$ and $\mathbf{Z}^\alpha \sim SN_p(\mathbf{0}, \mathbf{\Omega}, \alpha)$ respectively. Then we obtain that $D_{KL}(\mathbf{Z}, \mathbf{Z}^\alpha) = -E [\log 2\Phi(\sqrt{\alpha_*}Z)]$, where $\alpha_* = \alpha' \bar{\mathbf{\Omega}} \alpha$ and Z is a standard Gaussian random variable.*

Proof. We proof the assertion of the statement by adapting the argument in [4] to the previous parametrization of the SN model. From expression (1) we obtain that the KL divergence between \mathbf{Z} and \mathbf{Z}^α is given by

$$D_{KL}(\mathbf{Z}, \mathbf{Z}^\alpha) = \int_{\mathbb{R}^p} \phi_p(\mathbf{z}; \mathbf{\Omega}) \log \frac{\phi_p(\mathbf{z}; \mathbf{\Omega})}{2\phi_p(\mathbf{z}; \mathbf{\Omega}) \Phi(\alpha' \boldsymbol{\omega}^{-1} \mathbf{z})} d\mathbf{z} = -E [\log 2\Phi(\alpha' \boldsymbol{\omega}^{-1} \mathbf{Z})].$$

Taking into account that \mathbf{Z} follows a multivariate Gaussian distribution, we obtain that $Z = \alpha' \boldsymbol{\omega}^{-1} \mathbf{Z}$ is a univariate Gaussian random variable with variance α_* , which implies the result of the statement ■

3.1. Local deviation index

The result in Proposition 1 provides quantitative insights on how to assess the effect of small perturbations in asymmetry when we deal with multivariate Gaussian variables; it will be used to define our local deviation measure. First of all, we show how we represent perturbations in asymmetry.

In the multivariate SN model, departures from Gaussianity are regulated by the shape vector parameter α . Hence, the effect of infinitesimal perturbations in asymmetry can be quantified by comparing the distribution of a baseline p -dimensional vector \mathbf{Z} such that $\mathbf{Z} \sim N_p(\mathbf{0}, \Omega)$ with the distribution of a perturbed SN vector \mathbf{Z}^ε such that $\mathbf{Z}^\varepsilon \sim SN_p(\mathbf{0}, \Omega, \varepsilon \mathbf{e})$, where \mathbf{e} is a given direction driving the asymmetry and ε is a perturbation factor such that $\varepsilon \downarrow 0$. In this framework, it is well-known that the KL divergence, as a function of ε , can be approximated around 0 by a quadratic function of the form $a\varepsilon^2$ [12]; it therefore stands to reason to define our index as follows.

Definition 2 (Local deviation index). *Let \mathbf{Z} be a Gaussian random vector with $\mathbf{Z} \sim N_p(\mathbf{0}, \Omega)$. For every non-null direction \mathbf{e} such that $\mathbf{e}'\mathbf{e} = 1$, let us consider a vector \mathbf{Z}^ε such that $\mathbf{Z}^\varepsilon \sim SN_p(\mathbf{0}, \Omega, \varepsilon \mathbf{e})$ with $\varepsilon > 0$. Then departures from Gaussianity in asymmetry can be locally quantified by*

$$LD(\mathbf{e}, \Omega) = \lim_{\varepsilon \downarrow 0} \frac{D_{KL}(\mathbf{Z}, \mathbf{Z}^\varepsilon)}{\varepsilon^2}. \tag{6}$$

Note that measure (6) depends on the direction of asymmetry \mathbf{e} and on the covariance matrix of the underlying multivariate Gaussian model. Using the result of Proposition 1, the limit in (6) can be calculated as follows:

$$LD(\mathbf{e}, \Omega) = \lim_{\varepsilon \downarrow 0} \frac{D_{KL}(\mathbf{Z}, \mathbf{Z}^\varepsilon)}{\varepsilon^2} = - \lim_{\varepsilon \downarrow 0} \int_{-\infty}^{\infty} \frac{\log 2\Phi\left(\varepsilon\sqrt{\mathbf{e}'\Omega\mathbf{e}}z\right)}{\varepsilon^2} \phi(z) dz.$$

After some simple calculus with the integral above, we get

$$LD(\mathbf{e}, \Omega) = - \lim_{\varepsilon \downarrow 0} \int_0^{\infty} \frac{\log \left[4 \left(1 - \Phi\left(\varepsilon\sqrt{\mathbf{e}'\Omega\mathbf{e}}z\right) \right) \Phi\left(\varepsilon\sqrt{\mathbf{e}'\Omega\mathbf{e}}z\right) \right]}{\varepsilon^2} \phi(z) dz.$$

Furthermore, after successive applications of L'Hopital's rule we obtain that the limit of the integrand is

$$\lim_{\varepsilon \downarrow 0} \frac{\log \left[4 \left(1 - \Phi\left(\varepsilon\sqrt{\mathbf{e}'\Omega\mathbf{e}}z\right) \right) \Phi\left(\varepsilon\sqrt{\mathbf{e}'\Omega\mathbf{e}}z\right) \right]}{\varepsilon^2} = - \frac{2z^2\mathbf{e}'\bar{\Omega}\mathbf{e}}{\pi}.$$

Hence,

$$LD(\mathbf{e}, \Omega) = \lim_{\varepsilon \downarrow 0} \frac{D_{KL}(\mathbf{Z}, \mathbf{Z}^\varepsilon)}{\varepsilon^2} = \int_0^{\infty} \frac{2z^2\mathbf{e}'\bar{\Omega}\mathbf{e}}{\pi} \phi(z) dz = \frac{\mathbf{e}'\bar{\Omega}\mathbf{e}}{\pi}. \tag{7}$$

As we have already mentioned, [3] noted that the quantity $\mathbf{e}'\bar{\Omega}\mathbf{e}$ is functionally related to Mardia's skewness and kurtosis indices in SN models, so it can be considered as a quantifier for regulating departures from Gaussianity. The relation derived in (7) gives another way to interpret this quantity; it turns out to be proportional to the limit rate of divergence from the underlying Gaussian model when slight SN perturbations through the direction \mathbf{e} are injected. Local perturbations in asymmetry are described by a dissimilarity measure between distributions through KL divergence, so we argue that the new way of interpreting $\mathbf{e}'\bar{\Omega}\mathbf{e}$ provides an overall picture of what it really measures.

Alternatively, we could have used the scalar $\delta^2 = \boldsymbol{\delta}'\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\delta} \in [0, 1)$ in (7), where $\boldsymbol{\delta} = (1 + \mathbf{e}'\bar{\boldsymbol{\Omega}}\mathbf{e})^{-1/2}\bar{\boldsymbol{\Omega}}\mathbf{e}$. Hence, $LD(\mathbf{e}, \boldsymbol{\Omega}) = \frac{1}{\pi} \frac{\delta^2}{1 - \delta^2}$, from which we can obtain that $\delta^2 = \frac{\pi LD(\mathbf{e}, \boldsymbol{\Omega})}{1 + \pi LD(\mathbf{e}, \boldsymbol{\Omega})}$ is a bounded non decreasing quantity of our local deviation index.

Example 1. Maximum LD for permutation-symmetric normal variables

In this example we are going to consider a vector \mathbf{Z} belonging to the family of permutation-symmetric normal variables [13], i.e. random variables having multivariate normal distribution with equal means, equal variances and equal correlation coefficients. Therefore, the covariance matrix is given by

$$\boldsymbol{\Omega} = \begin{pmatrix} \sigma^2 & \rho\sigma^2 & \dots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \dots & \rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \dots & \rho\sigma^2 & \sigma^2 \end{pmatrix}$$

with $\rho > -\frac{1}{p-1}$ in order to have a positive definite matrix.

For the sake of simplicity we can assume that \mathbf{Z} has a zero mean vector, $\sigma^2 = 1$ and $\rho > 0$ since the case $\rho < 0$ can be studied in a similar way using an argument based on the symmetry of the problem. With these assumptions we obtain a correlation matrix $\bar{\boldsymbol{\Omega}} = \boldsymbol{\Omega}$ with appealing properties: its largest eigenvalue is $\lambda_1 = 1 + (p - 1)\rho$, and the remaining eigenvalues are $\lambda_2 = \dots = \lambda_p = 1 - \rho$; therefore, the local deviation index can be bounded as follows

$$LD(\mathbf{e}, \boldsymbol{\Omega}) = \frac{\mathbf{e}'\bar{\boldsymbol{\Omega}}\mathbf{e}}{\pi} \leq \frac{\mathbf{e}_1'\bar{\boldsymbol{\Omega}}\mathbf{e}_1}{\pi} = \frac{1 + (p - 1)\rho}{\pi},$$

where \mathbf{e}_1 is the normalized eigenvector $\mathbf{e}_1 = \frac{1}{\sqrt{p}}\mathbf{1}_{p \times 1}$ corresponding to the largest eigenvalue λ_1 , with $\mathbf{1}_{p \times 1}$ denoting a p-dimensional vector whose components are all equal to one.

Hence, vector \mathbf{e}_1 gives the direction yielding the maximum local deviation from Gaussianity. The maximum is a non decreasing function of the dimension p and the correlation coefficient ■

3.2. Relative local deviation index

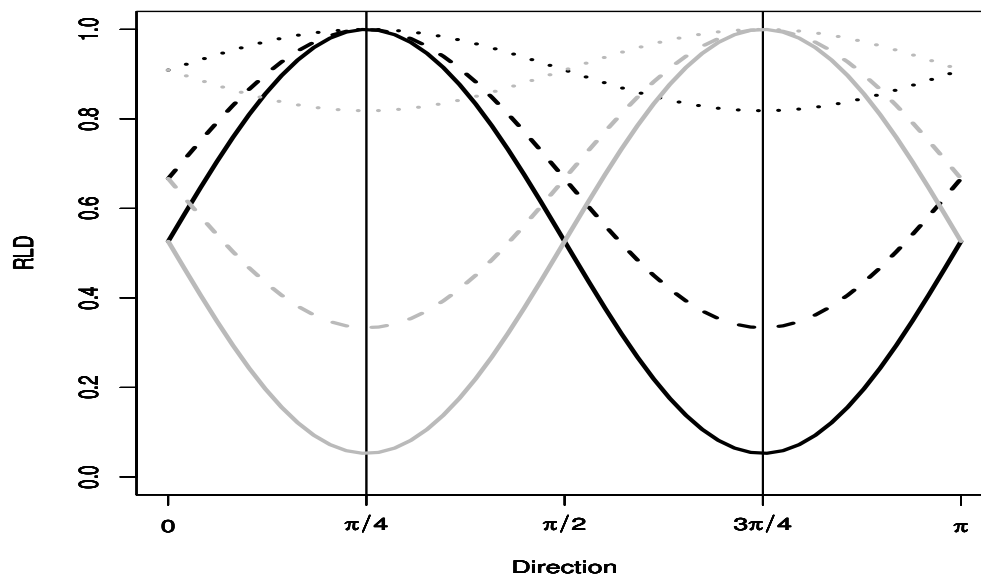
The result derived in (7) gives a valuation of how the multivariate Gaussian distribution is deformed by slight perturbations in the direction \mathbf{e} . The maximum is obtained when moving through the first principal component \mathbf{e}_1 of the correlation matrix $\bar{\boldsymbol{\Omega}}$. This fact leads to the following natural measure for quantifying relative local deviation (RLD) from Gaussianity.

Definition 3 (Relative local deviation index). Let \mathbf{Z} be a Gaussian random vector such that $\mathbf{Z} \sim N_p(\mathbf{0}, \boldsymbol{\Omega})$ with correlation matrix $\bar{\boldsymbol{\Omega}} = \boldsymbol{\omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\omega}^{-1}$. For every non-null normalized direction \mathbf{e} , the relative local deviation from Gaussianity is measured by

$$RLD(\mathbf{e}, \boldsymbol{\Omega}) = \frac{\mathbf{e}'\bar{\boldsymbol{\Omega}}\mathbf{e}}{\mathbf{e}_1'\bar{\boldsymbol{\Omega}}\mathbf{e}_1}, \tag{8}$$

where \mathbf{e}_1 is the normalized eigenvector corresponding to the largest eigenvalue, λ_1 , of the correlation matrix $\bar{\boldsymbol{\Omega}}$.

Figure 1. Plots of RLD against direction —angles $\theta \in [0, \pi]$ — for different correlations: $\rho = 0.9$ (black solid curve), $\rho = 0.5$ (black dashed curve), $\rho = 0.1$ (black dotted curve) and $\rho = -0.9$ (gray solid curve), $\rho = -0.5$ (gray dashed curve), $\rho = -0.1$ (gray dotted curve).



When departures from Gaussianity are modeled in accordance to a SN normalized vector Z_0 , with shape vector α such that e_1 is proportional to α , we know that the principal components of Z_0 are independent and proportional to the components of its canonical transformation [8]. In addition, the direction yielding the maximal skewness is proportional to e_1 , with all the multivariate skewness being absorbed into this direction, and with the remaining canonical variates having null skewness [8]. Thus, the denominator in (8) can be understood as the non Gaussianity index of the only skewed component in the canonical transformation of Z_0 .

Example 2. RLD for bidimensional permutation-symmetric normal vectors

In this example we consider the class of permutation-symmetric normal vectors with $p = 2$. In this case, the denominator in (8) corresponds to the maximum eigenvalue $\bar{\Omega}$, which is given by $\lambda_1 = 1 + \rho$, provided that $\rho > 0$. Hence, the RLD measure can be written as a function, $RLD(\theta)$, of the angle between the direction e and the positive direction of x-axis, where $0 \leq \theta \leq \pi$. The black curves in Figure 1 display this function for different non-negative correlation coefficients. As we have already mentioned, the maximum RLD is attained at $\theta = \frac{\pi}{4}$ with $RLD(\frac{\pi}{4}) = 1$, while the minimum is attained at $\theta = \frac{3\pi}{4}$ with $RLD(\frac{3\pi}{4}) = \frac{1-\rho}{1+\rho}$. Gray plots for the RLD measure at the corresponding negative values of ρ are also included. In this case, by the symmetry of our problem, we obtain that the minimum and maximum are given by $RLD(\frac{\pi}{4}) = \frac{1+\rho}{1-\rho}$ and $RLD(\frac{3\pi}{4}) = 1$ respectively.

In Figure 1 we show several plots of the RLD index for different correlation coefficients. Observe that when $\rho > 0$ the steepest increases of RLD from its minimum are obtained for the larger values of the correlation. On the contrary, when $\rho < 0$ they occur for the smaller values. Thus, we conclude that the largest ρ^2 will yield the steepest increases in local sensitivity ■

4. Application to assess relative conditional sensitivity

In this section we apply the ideas we used to introduce the LD index in order to assess relative sensitivity of conditional distributions. This issue has important implications in modeling continuous multivariate data with GBN. These are graphical displays of the dependence structure of a set of Gaussian variables with DAG, where the presence or absence of edges between the nodes represent conditional dependence or independence of the variables in the model. As the specialized literature establishes, the p -dimensional random vector can be partitioned as (\mathbf{Y}, \mathbf{X}) with \mathbf{X} the set of observable or evidential variables and \mathbf{Y} the set of non-observable or interest variables; when some particular values of the evidential variables are known, the conditional distribution $\mathbf{Y}|\mathbf{X}$ has to be determined in order to update our knowledge about \mathbf{Y} . This inference procedure, termed evidence propagation, is computationally tractable under Gaussianity as the exact conditional distributions are available. However, deviations from Gaussianity may cause computational and interpretation problems. Therefore, an open problem is to know if the output of the network is robust to slight perturbations in our model since this robustness would allow using the well-behaved methods under Gaussianity when the actual model is not exactly Gaussian. We propose to utilize multivariate SN distributions for modeling asymmetry deviations. With this aim, we apply measure (7) from Definition 2 to introduce a Relative Conditional Sensitivity Measure (RCSM) that will be used to evaluate the effect of slight perturbations in skewness on the conditional distributions.

Our study of the relative sensitivity of conditional distributions to infinitesimal changes in the symmetry of the joint density relies on KL divergence measure and the well-known chain rule:

$$D_{KL}(f_{(\mathbf{Y}, \mathbf{X})}, g_{(\mathbf{Y}, \mathbf{X})}) = E_{\mathbf{X}} [D_{KL}(f_{\mathbf{Y}|\mathbf{X}}, g_{\mathbf{Y}|\mathbf{X}})] + D_{KL}(f_{\mathbf{X}}, g_{\mathbf{X}}), \tag{9}$$

where $f_{(\mathbf{Y}, \mathbf{X})}$, $f_{\mathbf{Y}|\mathbf{X}}$ and $f_{\mathbf{X}}$ represent the joint, conditional and marginal densities of the reference distribution and $g_{(\mathbf{Y}, \mathbf{X})}$, $g_{\mathbf{Y}|\mathbf{X}}$ and $g_{\mathbf{X}}$ are the corresponding densities of the distribution to be compared. As we are interested in assessing deviations from Gaussianity we take as reference, f , the density of a Gaussian vector and g the density of the perturbed vector.

Some studies have assessed the local divergence of conditional distributions for different purposes. In particular, [14] measured the local association of the random variable Y with a set of covariates at $\mathbf{X} = \mathbf{x}$ using an appropriate limit of the divergence between $f_{Y|\mathbf{X}=\mathbf{x}}$ and $f_{Y|\mathbf{X}=\mathbf{x}+\varepsilon}$; a proposal to use local divergence as a measure of association between random variables was also provided. Conditional divergence measures were also utilized in the context of GBN to evaluate the sensitivity of model output to parameter perturbations [15] as well as to assess the effect of deviations regulated by a kurtosis parameter [1,16].

In view of relationship (9), an appropriate measure of relative conditional sensitivity can be defined as follows.

Definition 4 (Relative Conditional Sensitivity Measure). *Let (\mathbf{Y}, \mathbf{X}) be a p -dimensional random vector. The relative conditional sensitivity measure of (\mathbf{Y}, \mathbf{X}) with reference density function f is given by*

$$RCSM = \lim_{\varepsilon \downarrow 0} \frac{E_{\mathbf{X}} [D_{KL}(f_{\mathbf{Y}|\mathbf{X}}, f_{\mathbf{Y}|\mathbf{X}}^{\varepsilon})]}{D_{KL}(f_{(\mathbf{Y}, \mathbf{X})}, f_{(\mathbf{Y}, \mathbf{X})}^{\varepsilon})} = 1 - \lim_{\varepsilon \downarrow 0} \frac{D_{KL}(f_{\mathbf{X}}, f_{\mathbf{X}}^{\varepsilon})}{D_{KL}(f_{(\mathbf{Y}, \mathbf{X})}, f_{(\mathbf{Y}, \mathbf{X})}^{\varepsilon})}, \tag{10}$$

with f^ε denoting the corresponding joint, conditional and marginal densities associated with the perturbation.

Under our assumptions it follows that expression (10) has a simple closed form, as we will show in Proposition 5. Before presenting this result, we introduce some notation.

Let (\mathbf{Y}, \mathbf{X}) be a p -dimensional random vector following a $N_p(\mathbf{0}, \Omega)$ distribution and let $(\mathbf{Y}^\varepsilon, \mathbf{X}^\varepsilon)$ be the perturbed random vector following a multivariate $SN_p(\mathbf{0}, \Omega, \varepsilon \mathbf{e})$ with \mathbf{e} a fixed asymmetry direction, and \mathbf{X} and \mathbf{X}^ε ($p - k$)-dimensional vectors with $k < p$. Without loss of generality we can assume \mathbf{e} is a normalized vector. The scale matrix Ω , the correlation matrix $\bar{\Omega}$ and the diagonal matrix ω can be partitioned as follows:

$$\Omega = \begin{pmatrix} \Omega_{\mathbf{Y}\mathbf{Y}} & \Omega_{\mathbf{Y}\mathbf{X}} \\ \Omega_{\mathbf{X}\mathbf{Y}} & \Omega_{\mathbf{X}\mathbf{X}} \end{pmatrix}, \bar{\Omega} = \begin{pmatrix} \bar{\Omega}_{\mathbf{Y}\mathbf{Y}} & \bar{\Omega}_{\mathbf{Y}\mathbf{X}} \\ \bar{\Omega}_{\mathbf{X}\mathbf{Y}} & \bar{\Omega}_{\mathbf{X}\mathbf{X}} \end{pmatrix} \text{ and } \omega = \begin{pmatrix} \omega_{\mathbf{Y}} & \mathbf{0} \\ \mathbf{0} & \omega_{\mathbf{X}} \end{pmatrix}.$$

Similarly, we have the partition $\mathbf{e} = \begin{pmatrix} \mathbf{e}_{\mathbf{Y}} \\ \mathbf{e}_{\mathbf{X}} \end{pmatrix}$ for the asymmetry direction with $\mathbf{e}^\varepsilon = \varepsilon \mathbf{e}$ being the perturbed direction and $\varepsilon > 0$ an infinitesimal perturbation factor.

Proposition 5. Let (\mathbf{Y}, \mathbf{X}) be a p -dimensional random vector following a $N_p(\mathbf{0}, \Omega)$ distribution, with \mathbf{Y} a k -dimensional random vector, $k < p$. Let us denote by $(\mathbf{Y}^\varepsilon, \mathbf{X}^\varepsilon)$ the perturbed random vector which follows a multivariate $SN_p(\mathbf{0}, \Omega, \varepsilon \mathbf{e})$. Then the RCSM measure is given by

$$RCSM = \frac{\mathbf{e}'_{\mathbf{Y}} \bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}} \mathbf{e}_{\mathbf{Y}}}{\mathbf{e}' \bar{\Omega} \mathbf{e}}, \tag{11}$$

where $\bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}} = \bar{\Omega}_{\mathbf{Y}\mathbf{Y}} - \bar{\Omega}_{\mathbf{Y}\mathbf{X}} \bar{\Omega}_{\mathbf{X}\mathbf{X}}^{-1} \bar{\Omega}_{\mathbf{X}\mathbf{Y}}$ is the Schur complement of the submatrix $\bar{\Omega}_{\mathbf{X}\mathbf{X}}$ in $\bar{\Omega}$.

Proof. Since $(\mathbf{Y}^\varepsilon, \mathbf{X}^\varepsilon)$ follows a multivariate $SN_p(\mathbf{0}, \Omega, \varepsilon \mathbf{e})$ distribution, it can be shown that $\mathbf{X}^\varepsilon \sim SN_k(\mathbf{0}, \Omega_{\mathbf{X}\mathbf{X}}, \mathbf{e}_{\mathbf{X}(\mathbf{Y})}^\varepsilon)$, where the asymmetry direction $\mathbf{e}_{\mathbf{X}(\mathbf{Y})}^\varepsilon$ is given by $\mathbf{e}_{\mathbf{X}(\mathbf{Y})}^\varepsilon = \frac{\mathbf{e}_{\mathbf{X}} + \bar{\Omega}_{\mathbf{X}\mathbf{X}}^{-1} \bar{\Omega}_{\mathbf{X}\mathbf{Y}} \mathbf{e}_{\mathbf{Y}}^\varepsilon}{\sqrt{1 + \mathbf{e}_{\mathbf{Y}}^{\varepsilon'} \bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}} \mathbf{e}_{\mathbf{Y}}^\varepsilon}}$ [3]. Therefore, using a similar argument as in Proposition 1, we have

$$\begin{aligned} D_{KL}(\mathbf{X}, \mathbf{X}^\varepsilon) &= \int_{\mathbb{R}^{p-k}} \phi_{p-k}(\mathbf{x}; \Omega_{\mathbf{X}\mathbf{X}}) \log \left[\frac{\phi_{p-k}(\mathbf{x}; \Omega_{\mathbf{X}\mathbf{X}})}{2\phi_{p-k}(\mathbf{x}; \Omega_{\mathbf{X}\mathbf{X}}) \Phi(\mathbf{e}_{\mathbf{X}(\mathbf{Y})}^{\varepsilon'} \omega_{\mathbf{X}}^{-1} \mathbf{x})} \right] d\mathbf{x} \\ &= -E [\log 2\Phi(\mathbf{e}_{\mathbf{X}(\mathbf{Y})}^{\varepsilon'} \omega_{\mathbf{X}}^{-1} \mathbf{X})] \end{aligned}$$

with

$$\mathbf{e}_{\mathbf{X}(\mathbf{Y})}^{\varepsilon'} \omega_{\mathbf{X}}^{-1} = \frac{\mathbf{e}_{\mathbf{X}}^{\varepsilon'} \omega_{\mathbf{X}}^{-1} + \mathbf{e}_{\mathbf{Y}}^{\varepsilon'} \bar{\Omega}_{\mathbf{Y}\mathbf{X}} \bar{\Omega}_{\mathbf{X}\mathbf{X}}^{-1} \omega_{\mathbf{X}}^{-1}}{\sqrt{1 + \mathbf{e}_{\mathbf{Y}}^{\varepsilon'} \bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}} \mathbf{e}_{\mathbf{Y}}^\varepsilon}} = \frac{\mathbf{e}_{\mathbf{X}}^{\varepsilon'} \omega_{\mathbf{X}}^{-1} + \mathbf{e}_{\mathbf{Y}}^{\varepsilon'} \omega_{\mathbf{Y}}^{-1} \Omega_{\mathbf{Y}\mathbf{X}} \Omega_{\mathbf{X}\mathbf{X}}^{-1}}{\sqrt{1 + \mathbf{e}_{\mathbf{Y}}^{\varepsilon'} \omega_{\mathbf{Y}}^{-1} \Omega_{\mathbf{Y}\mathbf{Y}.\mathbf{X}} \omega_{\mathbf{Y}}^{-1} \mathbf{e}_{\mathbf{Y}}^\varepsilon}}$$

and \mathbf{X} a $(p - k)$ -dimensional Gaussian variable $N_{p-k}(\mathbf{0}, \Omega_{\mathbf{X}\mathbf{X}})$.

Now, if we consider the scalar variable $T^\varepsilon = \mathbf{e}_{\mathbf{X}(\mathbf{Y})}^{\varepsilon'} \omega_{\mathbf{X}}^{-1} \mathbf{X}$ then we know that T^ε is a scalar Gaussian variable whose variance is given by $V[T^\varepsilon] = \mathbf{e}_{\mathbf{X}(\mathbf{Y})}^{\varepsilon'} \omega_{\mathbf{X}}^{-1} \Omega_{\mathbf{X}\mathbf{X}} \mathbf{e}_{\mathbf{X}(\mathbf{Y})}^\varepsilon \omega_{\mathbf{X}}^{-1} = \varepsilon^2 \frac{\mathbf{e}' \bar{\Omega} \mathbf{e} - \mathbf{e}'_{\mathbf{Y}} \bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}} \mathbf{e}_{\mathbf{Y}}}{1 + \varepsilon^2 \mathbf{e}'_{\mathbf{Y}} \bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}} \mathbf{e}_{\mathbf{Y}}}$.

Therefore, $D_{KL}(\mathbf{X}, \mathbf{X}^\varepsilon)$ can be written as

$$D_{KL}(\mathbf{X}, \mathbf{X}^\varepsilon) = -E \left[\log 2\Phi \left(\varepsilon \sqrt{\frac{\mathbf{e}'\bar{\Omega}\mathbf{e} - \mathbf{e}'_{\mathbf{Y}}\bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}\mathbf{e}_{\mathbf{Y}}}{1 + \varepsilon^2\mathbf{e}'_{\mathbf{Y}}\bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}\mathbf{e}_{\mathbf{Y}}}} \right) Z \right],$$

where Z is a standard Gaussian univariate variable.

Then, with similar calculations as in the limit of Proposition 1 we obtain

$$\lim_{\varepsilon \downarrow 0} \frac{D_{KL}(\mathbf{X}, \mathbf{X}^\varepsilon)}{\varepsilon^2} = \frac{\mathbf{e}'\bar{\Omega}\mathbf{e} - \mathbf{e}'_{\mathbf{Y}}\bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}\mathbf{e}_{\mathbf{Y}}}{\pi},$$

from which we get

$$\begin{aligned} RCSM &= \lim_{\varepsilon \downarrow 0} \frac{E_{\mathbf{X}} \left[D_{KL}(\phi_{\mathbf{Y}|\mathbf{X}}, \phi_{\mathbf{Y}|\mathbf{X}}^\varepsilon) \right]}{D_{KL}(\phi_{(\mathbf{Y},\mathbf{X})}, \phi_{(\mathbf{Y},\mathbf{X})}^\varepsilon)} = 1 - \lim_{\varepsilon \downarrow 0} \frac{D_{KL}(\phi_{\mathbf{X}}, \phi_{\mathbf{X}}^\varepsilon)}{D_{KL}(\phi_{(\mathbf{Y},\mathbf{X})}, \phi_{(\mathbf{Y},\mathbf{X})}^\varepsilon)} \\ &= 1 - \frac{\mathbf{e}'\bar{\Omega}\mathbf{e} - \mathbf{e}'_{\mathbf{Y}}\bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}\mathbf{e}_{\mathbf{Y}}}{\mathbf{e}'\bar{\Omega}\mathbf{e}} = \frac{\mathbf{e}'_{\mathbf{Y}}\bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}\mathbf{e}_{\mathbf{Y}}}{\mathbf{e}'\bar{\Omega}\mathbf{e}}, \end{aligned}$$

where ϕ and ϕ^ε are obviously denoting the density functions of the Gaussian and perturbed vectors in the different ways they are used. Thus, the proposition is proved ■

In this setting it may be worthwhile finding the directions leading to the optimum of the RCSM index, since they would be directions through which the impact of non Gaussianity attains a maximum or a minimum. We can solve this issue by rewriting (11) as a generalized Rayleigh’s quotient as follows

$$RCSM = \frac{\mathbf{e}'\bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}^*\mathbf{e}}{\mathbf{e}'\bar{\Omega}\mathbf{e}}, \tag{12}$$

where $\bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}^* = \begin{pmatrix} \bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}} & \mathbf{0}_{k \times (p-k)} \\ \mathbf{0}_{(p-k) \times k} & \mathbf{0}_{(p-k) \times (p-k)} \end{pmatrix}$ is an extended version of $\bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}$ with dimension $p \times p$.

It is well known that the maximum and minimum of the quotient in expression (12) is given by the maximum and minimum eigenvalues of the matrix $\bar{\Omega}^{-1}\bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}^*$. Taking into account the form of the inverse of a matrix partitioned by blocks we obtain that

$$\begin{aligned} \bar{\Omega}^{-1}\bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}^* &= \begin{pmatrix} \bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}^{-1} & -\bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}^{-1}\bar{\Omega}_{\mathbf{Y}\mathbf{X}}\bar{\Omega}_{\mathbf{X}\mathbf{X}}^{-1} \\ -\bar{\Omega}_{\mathbf{X}\mathbf{X}}^{-1}\bar{\Omega}_{\mathbf{X}\mathbf{Y}}\bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}^{-1} & \bar{\Omega}_{\mathbf{X}\mathbf{X}}^{-1}\bar{\Omega}_{\mathbf{X}\mathbf{Y}}\bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}^{-1}\bar{\Omega}_{\mathbf{Y}\mathbf{X}}\bar{\Omega}_{\mathbf{X}\mathbf{X}}^{-1} + \bar{\Omega}_{\mathbf{X}\mathbf{X}}^{-1} \end{pmatrix} \bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}^* \\ &= \begin{pmatrix} \mathbf{I}_{k \times k} & \mathbf{0}_{k \times (p-k)} \\ -\bar{\Omega}_{\mathbf{X}\mathbf{X}}^{-1}\bar{\Omega}_{\mathbf{X}\mathbf{Y}} & \mathbf{0}_{(p-k) \times (p-k)} \end{pmatrix}, \end{aligned}$$

whose eigenvalues are given by $\lambda_1 = 1$ and $\lambda_2 = 0$ with multiplicities k and $p - k$ respectively.

The eigenspace for the maximum eigenvalue $\lambda_1 = 1$ is spanned by the columns of the matrix

$$\begin{pmatrix} \mathbf{I}_{k \times k} \\ -\bar{\Omega}_{\mathbf{X}\mathbf{X}}^{-1}\bar{\Omega}_{\mathbf{X}\mathbf{Y}} \end{pmatrix},$$

which in turn provide directions for which the effect of slight asymmetry perturbations on the conditional distribution attains its maximum. On the other hand, the eigenspace associated with $\lambda_2 = 0$ is the $(p - k)$ -dimensional kernel of the transform $\bar{\Omega}^{-1}\bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}^*$, spanned by the columns of the matrix

$$\begin{pmatrix} \mathbf{0}_{k \times (p-k)} \\ \mathbf{I}_{(p-k) \times (p-k)} \end{pmatrix},$$

which corresponds to the directions of the conditioning variables.

We apply our results to the distributions of Example 1.

Example 3. Maximum RCSM for permutation-symmetric normal vectors

In this example we study the RCSM index for the class of permutation-symmetric normal variables so we use the same notation and restrictions of Example 1. The multivariate vector is partitioned in subvectors with dimensions k and $p - k$, so we obtain that

$$\begin{aligned} \bar{\Omega}_{\mathbf{X}\mathbf{X}}^{-1} &= \begin{pmatrix} 1 & \cdots & \rho \\ \vdots & \ddots & \vdots \\ \rho & \cdots & 1 \end{pmatrix}_{(p-k) \times (p-k)}^{-1} \\ &= \frac{1}{(1 - \rho)[1 + \rho(p - k - 1)]} \left[\begin{pmatrix} 1 & -\rho & \cdots & -\rho \\ \vdots & \vdots & \ddots & \vdots \\ -\rho & \cdots & -\rho & 1 \end{pmatrix} + \rho(p - k - 2)\mathbf{I}_{p-k} \right] \end{aligned}$$

which multiplied by

$$\bar{\Omega}_{\mathbf{X}\mathbf{Y}} = \begin{pmatrix} \rho & \rho & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \cdots & \rho & \rho \end{pmatrix}_{(p-k) \times k}$$

gives $-\bar{\Omega}_{\mathbf{X}\mathbf{X}}^{-1}\bar{\Omega}_{\mathbf{X}\mathbf{Y}} = -\frac{\rho}{1 + \rho(p - k - 1)}\mathbf{1}_{(p-k) \times k}$, where $\mathbf{1}_{(p-k) \times k}$ denotes a $(p - k) \times k$ matrix with all the entries equal to one. Consequently,

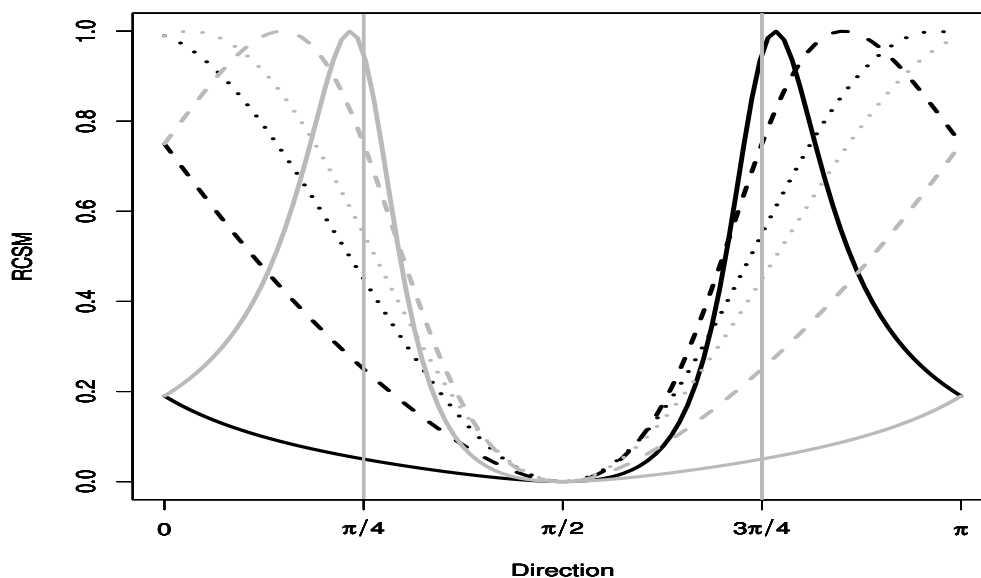
$$\bar{\Omega}^{-1}\bar{\Omega}_{\mathbf{Y}\mathbf{Y}.\mathbf{X}}^* = \begin{pmatrix} \mathbf{I}_k & \mathbf{0}_{k \times (p-k)} \\ -\frac{\rho}{1 + \rho(p - k - 1)}\mathbf{1}_{(p-k) \times k} & \mathbf{0}_{(p-k) \times (p-k)} \end{pmatrix}.$$

Therefore, a direction where a perturbation in asymmetry produces the maximum relative local effect on the conditional distribution is $\mathbf{e}_{\max} = \begin{pmatrix} \mathbf{e}_Y \\ \mathbf{e}_X \end{pmatrix}$ with

$$\mathbf{e}_Y = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{k \times 1}, \quad \mathbf{e}_X = -\frac{\rho}{1 + \rho(p - k - 1)}\mathbf{1}_{(p-k) \times 1}.$$

Example 4. Maximum RCSM for bidimensional permutation-symmetric normal vectors

Figure 2. Plots of RCSM against direction —angles $\theta \in [0, \pi]$ — for different correlations: $\rho = 0.9$ (black solid curve), $\rho = 0.5$ (black dashed curve), $\rho = 0.1$ (black dotted curve) and $\rho = -0.9$ (gray solid curve), $\rho = -0.5$ (gray dashed curve), $\rho = -0.1$ (gray dotted curve).



A particular simple but enlightening case is obtained when $p = 2$ with vector (Y, X) and conditioning variable X . Now, the normalized eigenvector associated with the highest eigenvalue $\lambda_1 = 1$ is $\mathbf{e}_{\max} = \begin{pmatrix} 1/\sqrt{1+\rho^2} \\ -\rho/\sqrt{1+\rho^2} \end{pmatrix}$, which gives the direction where perturbations have the highest impact on the conditional distribution. On the other hand, the minimum impact, or equivalently the maximum impact on the marginal univariate conditioning distribution, is obtained through the direction $\mathbf{e}_{\min} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

In Figure 2 we have plots of RCSM for different values of the correlation coefficient. Since the minimum RCSM is attained in the direction \mathbf{e}_{\min} all the curves intersect at angle $\theta = \frac{\pi}{2}$. Furthermore, the angle that provides the direction with maximum RCSM is functionally related to the correlation coefficient by $\theta = \pi - \arctan(\rho)$ when $0 < \rho \leq 1$ and by $\theta = -\arctan(\rho)$ when $-1 \leq \rho < 0$. Note that when both components of the vectors are linearly dependent, with $\rho = -1$ or $\rho = 1$, the directions yielding the maximum RCSM are given by the angles $\theta = \frac{\pi}{4}$ and $\theta = \frac{3\pi}{4}$ respectively —see their positions at the vertical lines of Figure 2. They are also the directions with the minimum RLD, as was shown in Example 2. This fact is explained by the perfect linear correlation between both variables



5. Conclusions

In this paper we use the Kullback-Leibler (KL) divergence to introduce local deviation (LD) and relative local deviation (RLD) indices designed to assess the effect slight departures from Gaussianity, due to asymmetry, have on the joint distribution of a multivariate vector; the asymmetry is regulated by the family of SN distributions. We have studied the relationship between Mardia and Malkovich-Afifi's

measures of skewness and the LD index, and have also proved the existence of a monotone relation between these measures and the latter; their connection is described by the slant parameter $\alpha_* = \alpha' \bar{\Omega} \alpha$ [3]. This fact has also provided new insights about how α_* can be interpreted as an indicator that encapsulates the non Gaussianity in the SN model.

We have also studied the local effect of asymmetry in the conditional distribution of a block of variables conditioned on the others, as well as the local effect on the marginal distribution of the conditioning variables. With this aim, we have proposed a relative conditional sensitivity measure, RCSM, that evaluates such effects. Its simple closed form as a Rayleigh's quotient allows us to determine the directions of asymmetry for which slight perturbations lead to the highest impact on the conditional distribution; this finding has important practical implications in statistical learning and data analysis procedures that assume the multivariate Gaussianity as a starting hypothesis. In particular, for GBN we can select the partition of the network variables leading to outputs with minimal impact for perturbations of the joint distribution. Then, we can use direct results to propagate evidence in GBN even when the symmetry assumption is not realistic.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Main, P.; Navarro, H. Analyzing the effect of introducing a kurtosis parameter in Gaussian Bayesian networks. *Rel. Eng. & Sys. Safety* **2009**, *94*, 922–926.
2. Azzalini, A.; Dalla Valle, A. The multivariate skew-normal distribution. *Biometrika* **1996**, *83*, 715–726.
3. Azzalini, A.; Capitanio, A. *The Skew-Normal and Related Families*; IMS monographs. Cambridge University Press, 2014.
4. Contreras-Reyes, J.E.; Arellano-Valle, R.B. Kullback-Leibler Divergence Measure for Multivariate Skew-Normal Distributions. *Entropy* **2012**, *14*, 1606–1626.
5. Azzalini, A.; Capitanio, A. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society Series B* **1999**, *61*, 579–602.
6. Capitanio, A.; Azzalini, A.; Stanghellini, E. Graphical Models for Skew-Normal Variates. *Scandinavian Journal of Statistics* **2003**, *30*, 129–144.
7. Azzalini, A. The Skew-normal Distribution and Related Multivariate Families. *Scandinavian Journal of Statistics* **2005**, *32*, 159–188.
8. Loperfido, N. Canonical transformations of skew-normal variates. *TEST* **2010**, *19*, 146–165.
9. Mardia, K.V. Measures of multivariate skewness and kurtosis with applications. *Biometrika* **1970**, *57*, 519–530.
10. Malkovich, J.F.; Afifi, A.A. On Tests for Multivariate Normality. *Journal of the American Statistical Association* **1973**, *68*, 176–179.
11. Kullback, S.; Leibler, R.A. On information and sufficiency. *Annals of Mathematical Statistics* **1951**, *22*, 49–86.

12. Cover, T.M.; Thomas, J.A. *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*; Wiley-Interscience, 2006.
13. Tong, Y. *The Multivariate Normal Distribution*; Springer-Verlag, 1990.
14. Blyth, S. Local Divergence and Association. *Biometrika* **1994**, *81*, 579–584.
15. Gómez-Villegas, M.A.; Maín, P.; Susi, R. Sensitivity Analysis in Gaussian Bayesian Networks Using a Divergence Measure. *Communications in Statistics Theory and Methods* **2007**, *36*, 523–539.
16. Gómez-Villegas, M.A.; Maín, P.; Navarro, H.; Susi, R. Assessing the effect of kurtosis deviations from Gaussianity on conditional distributions. *Applied Mathematics and Computation* **2013**, *219*, 10499–10505.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).