Applied Mathematics and Computation 219 (2013) 10499–10505

Contents lists available at SciVerse ScienceDirect

Applied Mathematics and Computation

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

Assessing the effect of kurtosis deviations from Gaussianity on conditional distributions

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

[a] Departamento de Estadística e Investigación Operativa, 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

Keywords:
Multivariate exponential power distributions
Kurtosis
Kullback–Leibler divergence
Relative sensitivity

ABSTRACT

The multivariate exponential power family is considered for n-dimensional random variables, $\mathbf{Z}$, with a known partition $\mathbf{Z} \equiv (\mathbf{Y}, \mathbf{X})$ of dimensions p and $n - p$, respectively, with interest focusing on the conditional distribution $\mathbf{Y}|\mathbf{X}$. An infinitesimal variation of any parameter of the joint distribution produces perturbations in both the conditional and marginal distributions. The aim of the study was to determine the local effect of kurtosis deviations using the Kullback–Leibler divergence measure between probability distributions. The additive decomposition of this measure in terms of the conditional and marginal distributions, $\mathbf{Y}|\mathbf{X}$ and $\mathbf{X}$, is used to define a relative sensitivity measure of the conditional distribution family $\{\mathbf{Y}|\mathbf{X} = \mathbf{x}\}$. Finally, simulated results suggest that for large dimensions, the measure is approximately equal to the ratio p/n, and then the effect of non-normality with respect to kurtosis depends only on the relative size of the variables considered in the partition of the random vector.

1. Introduction

In modeling a phenomenon characterized by a random n-dimensional vector $\mathbf{Z}$, very often not all the vector components play symmetrical roles. In this case we consider a partition of the components of $\mathbf{Z}$ in two blocks, $\mathbf{Y}$ and $\mathbf{X}$, with dimensions p and $n - p$, respectively. Examples include regression, classification and Bayesian networks.

The multivariate normal distribution plays a central role in this setting because marginals and conditionals of multivariate normal random variables are also normal. It has also several simplifying properties one of which is that it is uniquely determined by the mean vector and the covariance matrix of the random variables, so it is characterized in terms of central tendency and precision. With respect to the shape characteristics of skewness and kurtosis, it is a symmetric and unimodal distribution and is considered as a reference for the degree of peakedness or flatness for other distributions. The use of kurtosis to describe departures from Gaussianity, often found for real-world data [6,16,22,12], goes back to Pearson [19], however many statistical procedures rely on Gaussianity assumptions and departures from these models can affect final results. Thus a sensitivity analysis to deviations from Gaussianity should be performed.

In this work we investigate the impact of a change in the kurtosis of $\mathbf{Z}$ in the corresponding marginal and conditional distributions of some partition of the random variable. Thus, we introduce a measure to evaluate how the effect of heavy or light tails (peakedness or flatness relative to the normal distribution) of the joint model is transmitted through the structure. We focus on a family of distributions that is of particular interest in applications when the normality assumption is doubtful in

* Corresponding author.
E-mail address: pmain@mat.ucm.es (P. Main).

0096-3003/$ - see front matter
http://dx.doi.org/10.1016/j.amc.2013.04.031

the sense of kurtosis. This is a generalization of the multivariate normal distribution and introduces a parameter, β, that is a measure of the non-normality of the distribution [7]. Members of the multivariate exponential power (MEP) family exhibit a fairly broad spectrum of shapes for the probability density function, ranging from rectangular (short-tailed) to double exponential (long-tailed), with a normal distribution as the central case.

Similar characteristics to the theoretical framework described here can be found in Gaussian Bayesian networks (GBNs), a subclass of Bayesian networks (BNs) for which the joint distribution of **Z** is a multivariate normal distribution. A BN is a probabilistic graphical model for which the information can be updated given some evidence. Therefore, two subsets of variables are found in BNs: the evidential or known variables, corresponding to **X**, and the interest or unknown variables, corresponding to **Y** in our context. The process for updating the information given evidence is known as evidence propagation.

After evidence propagation, the network output is given by the conditional distribution of the variables of interest for observed values of the evidential variables, that is $\{\mathbf{Y}|\mathbf{X}=\mathbf{x}\}$. A joint model deviation with respect to kurtosis of the distribution of **Z** may entail a different network output. In particular, research in linear Causal models with continuous variables, also known as Structural Equation Models (SEM) [4], is based on the assumption of Gaussian data, however if non-Gaussian data are supposed, much stronger results can be obtained [13].

Summing up, in situations where the Gaussianity assumption is unjustified the evaluation of non-Gaussianity is crucial because it can result in a positive or negative effect on the conclusions.

Thus, our aim is to analyze the sensitivity of the normal ($\beta = 1$) conditional distributions to small deviations in β, using the Kullback–Leibler (KL) divergence to measure the discrepancy between distributions. A known additive property of the KL divergence leads directly to the definition of a relative sensitivity measure for the conditionals, taking values in $[0,1]$.

In Section 2 the concepts to be used are presented. In Section 3 we derive the proposed relative sensitivity measure to evaluate normality deviations, making a conjecture about its particular values based on simulation results and both computational and graphical procedures. Section 4 is devoted to a discussion and interpretation of the simulation results. Finally, some conclusions are given in Section 5 while the calculations of the results observed by simulation are presented in Appendix A.

2. The effect of non-normality

There are many reasons for the predominant role of the multivariate normal distribution in statistics. It results from some of its most desirable properties, as that it represents a natural extension of the univariate normal distribution and provides a suitable model for many real-life problems concerning vector-valued data. However, many studies have addressed non-normality in different contexts [5,21,18]. More recently, [9] study some statistics behavior as the kurtosis in the data varies and [11] derive the asymptotic distribution of some known statistics and show they are robust against departure from normality.

As mentioned previously, we consider here some non-normal distributions with respect to kurtosis in the multivariate case. To evaluate the effect of a kurtosis deviation with respect to a Gaussian model, we use the MEP distribution and KL divergence to measure the difference between distributions.

2.1. MEP family

The density function of an n-dimensional MEP distribution with parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and β, $MEP(\boldsymbol{\mu},\boldsymbol{\Sigma},\beta)$ is given by

$$f(\mathbf{z}|\boldsymbol{\mu},\boldsymbol{\Sigma},\beta) = \frac{n\Gamma(n/2)}{\pi^{n/2}\Gamma\left(1+\frac{n}{2\beta}\right)2^{1+\frac{n}{2\beta}}}|\Sigma|^{-1/2}exp\left\{-\frac{1}{2}[(\mathbf{z}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})]^{\beta}\right\},$$

where the vector $\boldsymbol{\mu} \in \Re^n$, $\boldsymbol{\Sigma}$ is a positive definite symmetric matrix and β is the kurtosis parameter $\beta \in (0,\infty)$. Thus, the non-normality parameter is directly related to the shape of the distribution:

β	Distribution
$\beta = 1$	multivariate normal
$\beta = 1/2$	multivariate double exponential
$\beta \to \infty$	multivariate uniform

The parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ play a similar role to their counterparts in the normal case. It is well known [8] that if $\mathbf{Z} = (\mathbf{Y},\mathbf{X}) \sim MEP_n(\boldsymbol{\mu},\boldsymbol{\Sigma},\beta)$, then $\mathbf{Y}|\mathbf{X}=\mathbf{x} \sim E_n(\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}},\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}},g^{(\beta)}_{\mathbf{Y}|\mathbf{X}}(t))$, an elliptical distribution with parameters:

$$\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}} = \boldsymbol{\mu}_{\mathbf{Y}} + \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}^{-1}_{\mathbf{XX}}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}}),$$

$$\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}} = \boldsymbol{\Sigma}_{\mathbf{YY}} - \boldsymbol{\Sigma}_{\mathbf{YX}}\boldsymbol{\Sigma}^{-1}_{\mathbf{XX}}\boldsymbol{\Sigma}_{\mathbf{XY}}$$

and a density generator

$$g^{(\beta)}_{\mathbf{Y}|\mathbf{X}}(t) = \exp\left\{-\frac{1}{2}(t+q_{\mathbf{x}})^{\beta}\right\},$$

where $q_{\mathbf{x}} = (\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})^T\boldsymbol{\Sigma}^{-1}_{\mathbf{XX}}(\mathbf{x}-\boldsymbol{\mu}_{\mathbf{X}})$ is the squared Mahalanobis distance from $\mathbf{x}$ to the center of the $\mathbf{X}$ distribution. The parameter β is a shape parameter, as the kurtosis depends only on it. Then the density function can be expressed as

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \propto \left|\boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}\right|^{-\frac{1}{2}} \times g^{(\beta)}_{\mathbf{Y}|\mathbf{X}}((\mathbf{y}-\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}})^T\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}-\boldsymbol{\mu}_{\mathbf{Y}|\mathbf{X}})). \tag{1}$$

2.2. KL divergence

The KL divergence measure 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 [14].

The KL divergence between two probability densities f and g, defined over the same domain, is given by:

$$D_{KL}(g|f) = \int_{-\infty}^{\infty} f(w)\ln\frac{f(w)}{g(w)}dw.$$

We consider a directed divergence measure because we have the normal distribution, f, as a reference. The following equation relates the joint conditional and marginal divergences:

$$D_{KL}(g_{\mathbf{Y},\mathbf{X}}|f_{\mathbf{Y},\mathbf{X}}) = E_{\mathbf{X}}\left[D_{KL}\left(g_{\mathbf{Y}|\mathbf{X}}|f_{\mathbf{Y}|\mathbf{X}}\right)\right] + D_{KL}(g_{\mathbf{X}}|f_{\mathbf{X}}), \tag{2}$$

where $f,\ f_{\mathbf{Y}|\mathbf{X}}$ and $f_{\mathbf{X}}$ represent the joint, conditional and marginal densities of the reference distribution and $g,\ g_{\mathbf{Y}|\mathbf{X}}$ and $g_{\mathbf{X}}$ are the corresponding densities of the distribution to be compared.

3. Relative conditional sensitivity

Now we address our objective to measure, for an n-dimensional random vector $\mathbf{Z} = (\mathbf{Y},\mathbf{X})$, the relative sensitivity of the conditional distributions $f_{\mathbf{Y}|\mathbf{X}}$ to infinitesimal changes δ in the parameters of the joint density $f_{(\mathbf{Y},\mathbf{X})}$. Other studies have assessed the local divergence of conditional distributions for different purposes. In particular, Blyth measured the local association of the random variable Y with the covariates at $\mathbf{X} = \mathbf{x}$ using an appropriate limit of the divergence between $f_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$ and $f_{\mathbf{Y}|\mathbf{X}=\mathbf{x}+\delta\mathbf{x}}$ and proposed the use of these local divergences to measure the association between random variables [3]. For GBNs, conditional divergence measures to evaluate the sensitivity of model output to parameter perturbations were used in [10]. Formally, (2) is a decomposition into positive summands analogous to that of sums of squares in variance analysis. Then, with the aim of evaluating relative deviations between conditional distributions, relation (2) suggests considering the corresponding ratio of divergences. Finally, the local aspect of this divergence measure is achieved by taking the limit

$$\lim_{\delta\to 0}\frac{E_X\left[D_{KL}(f_{\mathbf{Y}|\mathbf{X};\delta}|f_{\mathbf{Y}|\mathbf{X}})\right]}{D_{KL}(f_{(\mathbf{Y},\mathbf{X});\delta}|f_{(\mathbf{Y},\mathbf{X})})},$$

where the notation $f_{\cdot;\delta}$ represents the density associated with the perturbed parameter. With this methodology, some local sensitivity measures can be applied to probabilistic structures where the interest is in the distribution of one set of variables given the distribution of another set of variables, as is the case of GBNs, already mentioned in Section 2. This measure takes its simplest form when applied to a Gaussian model with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Specifically, if $\mathbf{Z} = (\mathbf{Y},\mathbf{X})$ is a random vector normally distributed with parameters

$$\boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_{\mathbf{Y}}\\ \boldsymbol{\mu}_{\mathbf{X}}\end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix}\boldsymbol{\Sigma}_{\mathbf{YY}} & \boldsymbol{\Sigma}_{\mathbf{YX}}\\ \boldsymbol{\Sigma}_{\mathbf{XY}} & \boldsymbol{\Sigma}_{\mathbf{XX}}\end{pmatrix}$$

and the location parameter $\boldsymbol{\mu}_{\mathbf{X}}$ is additively perturbed to $\boldsymbol{\mu}_{\mathbf{X}} + \delta$, this yields:

$$\lim_{\delta\to 0}\frac{E_X\left[D_{KL}(f_{\mathbf{Y}|\mathbf{X};\delta}|f_{\mathbf{Y}|\mathbf{X}})\right]}{D_{KL}(f_{(\mathbf{Y},\mathbf{X});\delta}|f_{(\mathbf{Y},\mathbf{X})})} = \rho^2,$$

where ρ is the corresponding correlation coefficient for each of the following cases:

(i) If $n = 2$ and $p = 1$ (Y and X are unidimensional variables), then ρ is the simple correlation coefficient between X and Y.
(ii) If $n > 2$ and $n - p = 1$, then ρ is the multiple correlation coefficient between X and $\mathbf{Y}$.
(iii) If both $\mathbf{Y}$ and $\mathbf{X}$ are multidimensional variables and the limit is taken in a canonical direction, ρ represents the corresponding canonical correlation coefficient.

The situation is more complicated when the perturbation affects $\boldsymbol{\Sigma}$. In this case the results are not as easily interpreted and the correlation plays a predominant role. For example, if $n > 2$ and

$$\mathbf{\Sigma}_\delta = \begin{pmatrix} \mathbf{\Sigma_{YY}} & \mathbf{\Sigma_{YX}} \\ \mathbf{\Sigma_{XY}} & \mathbf{\Sigma_{XX}} + \delta \mathbf{I}_{n-p} \end{pmatrix}$$

we obtain the following:

(i) If $n-p=1$, then

$$\lim_{\delta\to 0}\frac{E_X\left[D_{KL}\left(f_{\mathbf{Y}|X;\delta}|f_{\mathbf{Y}|X}\right)\right]}{D_{KL}\left(f_{(\mathbf{Y},X);\delta}|f_{(\mathbf{Y},X)}\right)} = \rho^2(2-\rho^2);$$

it can be pointed that $\rho^2(2-\rho^2) > \rho^2$ for $-1<\rho<1$ and the difference is $\rho^2(1-\rho^2)$, reaching its maximum in $\rho^2 = 0.5$;

(ii) For $n-p=2$, it follows that

$$\lim_{\delta\to 0}\frac{E_X\left[D_{KL}\left(f_{\mathbf{Y}|\mathbf{X};\delta}|f_{\mathbf{Y}|\mathbf{X}}\right)\right]}{D_{KL}\left(f_{(\mathbf{Y},\mathbf{X});\delta}|f_{(\mathbf{Y},\mathbf{X})}\right)} = 1-\left(\frac{|\mathbf{\Sigma_{XX.Y}}|}{|\mathbf{\Sigma_{XX}}|}\right)^2\frac{|\mathbf{\Sigma_{XX}}|-\frac{1}{2}tr^2(\mathbf{\Sigma_{XX}})}{|\mathbf{\Sigma_{XX.Y}}|-\frac{1}{2}tr^2(\mathbf{\Sigma_{XX.Y}})},$$

being $\mathbf{\Sigma_{XX.Y}} = \mathbf{\Sigma_{X|Y}}$, as usual in literature;

(iii) Finally, for $n-p=3$, the previous limit can be expressed as

$$1-\left(\frac{|\mathbf{\Sigma_{XX.Y}}|}{|\mathbf{\Sigma_{XX}}|}\right)^2\times\frac{|\mathbf{\Sigma_{XX}}|tr(\mathbf{\Sigma_{XX}})-\frac{1}{2}(p(\mathbf{\Sigma_{XX}})-c(\mathbf{\Sigma_{XX}}))^2}{|\mathbf{\Sigma_{XX.Y}}|tr(\mathbf{\Sigma_{XX.Y}})-\frac{1}{2}(p(\mathbf{\Sigma_{XX.Y}})-c(\mathbf{\Sigma_{XX.Y}}))^2},$$

where $p(\cdot)$ is the sum of the products of the binary linear combinations in the argument diagonal matrix and $c(\cdot)$ is the sum of the squared elements in the upper triangle.

If instead of looking at $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ we focus on deviations from Gaussianity via kurtosis, the measure representation (3) is not so direct. Calculations and manipulations to get results in this line, will be at the core of this work. Thus, if our interest is in calculating the relative conditional sensitivity to small changes in kurtosis for the MEP family of distributions (3) has the form

$$\lim_{\beta\to 1}\frac{E_X\left[D_{KL}\left(f^{(\beta)}_{\mathbf{Y}|\mathbf{X}}|f_{\mathbf{Y}|\mathbf{X}}\right)\right]}{D_{KL}(f^{(\beta)}|f)}. \quad (3)$$

where $f,\ f_{\mathbf{Y}|\mathbf{X}}$ and $f_{\mathbf{X}}$ represent the joint, conditional and marginal densities for the normal distribution ($\beta = 1$), respectively and the superscript (β) denotes the density when β in the joint model is perturbed to $\beta = 1+\delta$ and δ is the deviation from normality.

KL divergences for the numerator and denominator are non-negative and are set to zero for $\beta = 1$ (Fig. 1). Moreover, the first derivatives with respect to β of such divergences are zero at $\beta = 1$ and the second derivatives are the Fisher information with respect to β, evaluated at the baseline densities, $f_{\mathbf{Y}|\mathbf{X}}$ and f, respectively. Then the limit in (3) will be the quotient of the curvature values for both divergences. In any case, both the KL measure and the corresponding curvature are often intractable. The numerator in (3) is the mathematical expectation of the random divergence between the conditional densities of $\mathbf{Y}|\mathbf{X}=\mathbf{x}$ with respect to the density of $\mathbf{X}$. One particular expression of such divergence as a function of $\mathbf{x}$ is [17]:

$$\log\frac{\int_0^\infty t^{p/2-1}\exp\left\{-\frac{1}{2}(t+q_{\mathbf{x}})^\beta\right\}dt}{2^{p/2}\Gamma\left(\frac{p}{2}\right)}-\frac{1}{2}\left[p-\frac{q_{\mathbf{x}}^{\beta-p/2}}{2^{p/2}}U(a,b,s)\right], \quad (4)$$

where $U(\cdot)$ is the confluent hypergeometric function [1] for $a=\frac{p}{2},\ b=\beta+\frac{p}{2}+1,\ s=\frac{q_{\mathbf{x}}}{2}$, and $q_{\mathbf{x}}$ is the squared Mahalanobis distance from $\mathbf{x}$ to the center of the $\mathbf{X}$ distribution. Assuming that $\mathbf{X}$ has a multivariate normal distribution, the random variable $q_{\mathbf{X}}$ has a χ^2 distribution with degrees of freedom equal to the number of components of the variable $\mathbf{X}$. In our notation, $q_{\mathbf{X}}$ is χ^2-distributed with $n-p$ degrees of freedom. For the denominator in (3) direct calculations yield:

$$D_{KL}(f^{(\beta)}|f) = \log\frac{2^{\frac{n}{2\beta}}\Gamma\left(\frac{n}{2\beta}\right)}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)\beta}-\frac{1}{2}\left(n-\frac{2^\beta\Gamma\left(\frac{n}{2}+\beta\right)}{\Gamma\left(\frac{n}{2}\right)}\right).$$

4. Computation of local sensitivity

Next we deal with determination of the limit. However, computation of the mathematical expectation in (3) presents some difficulties. With respect to (4) and with some calculations given that $q_{\mathbf{X}}$ is χ^2-distributed with $n-p$ degrees of freedom, we obtain:

$$E_{q_{\mathbf{X}}}\left[\frac{q_{\mathbf{X}}^{\beta-p/2}}{2^{p/2}}U\left(\frac{p}{2},\beta+\frac{p}{2}+1,\frac{q_{\mathbf{X}}}{2}\right)\right] = \frac{2^\beta\Gamma\left(\frac{n}{2}+\beta\right)}{\Gamma\left(\frac{n}{2}\right)}.$$

Then, we have to determine the limit, as $\beta\to 1$, of the quotient

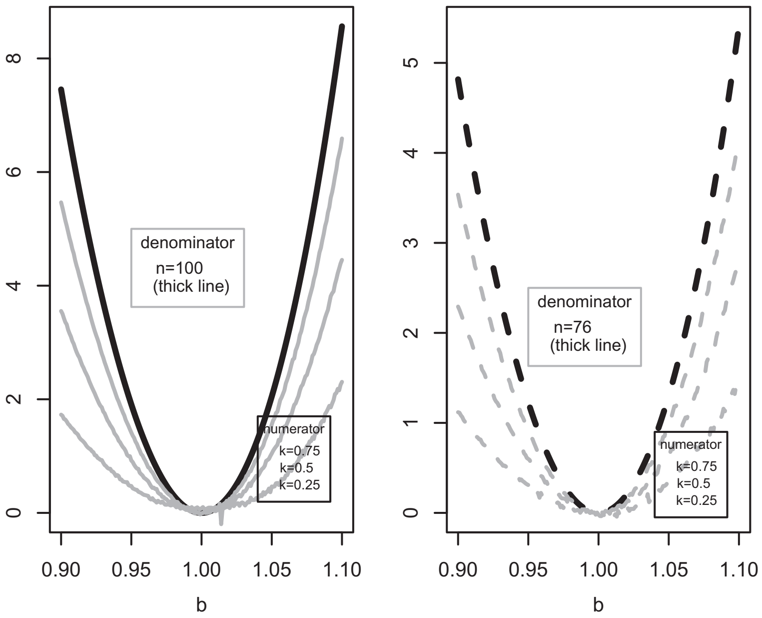

Fig. 1. Quadratic behavior of the simulated conditional KL divergence means (grey lines) and the joint KL divergences (black lines) for joint dimensions of $\mathbf{Z} \equiv (\mathbf{Y}, \mathbf{X}),\ n = 100, n = 76$ and several values of $k = \frac{p}{n}$, the ratio of the dimension of $\mathbf{Y}$ to the joint dimension. The sample size is 100,000.

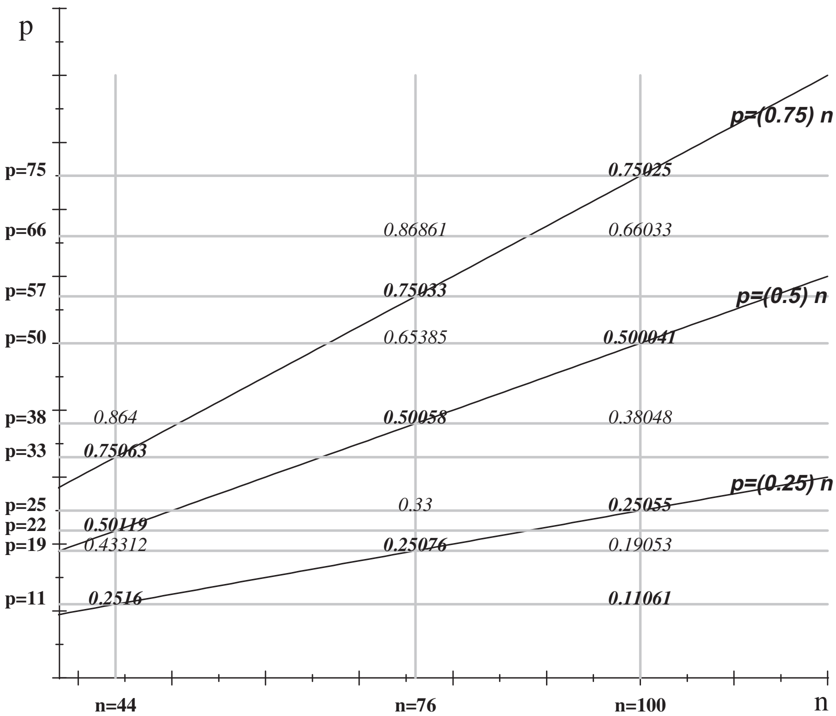

Fig. 2. Results obtained by numerical integration of the limit $\lim_{\beta\to 1} \frac{E_X[D_{KL}(f^{(\beta)}_{\mathbf{Y}|\mathbf{X}}|f_{\mathbf{Y}|\mathbf{X}})]}{D_{KL}(f^{(\beta)}|f)}$ for $n = 44,\ n = 76,\ n = 100$ and different component dimensions p given by some values of $k = \frac{p}{n}$ in the grid, with n and p the dimensions of $\mathbf{Z} \equiv (\mathbf{Y}, \mathbf{X})$ and $\mathbf{Y}$, respectively.

$$\frac{E_{q_{\mathbf{X}}}\left[\log\frac{\int_0^\infty t^{p/2-1}\exp\left\{-\frac{1}{2}(t+q_{\mathbf{X}})^\beta\right\}dt}{2^{p/2}\Gamma\left(\frac{p}{2}\right)}\right]-\frac{p}{2}+\frac{1}{2}\left[\frac{2^\beta\Gamma\left(\frac{p}{2}+\beta\right)}{\Gamma\left(\frac{p}{2}\right)}\right]}{\log\frac{2^{\frac{n}{2\beta}}\Gamma\left(\frac{n}{2\beta}\right)}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)\beta}-\frac{1}{2}\left(n-\frac{2^\beta\Gamma\left(\frac{n}{2}+\beta\right)}{\Gamma\left(\frac{n}{2}\right)}\right)}. \quad (5)$$

It is well known that, in general, the KL divergence between the baseline and a parameter with perturbed density can be approximated by a quadratic function of the deviation [15]:

$$D_{KL}(f^{(1+\delta)}|f) \overset{\delta\to 0}{\approx} \frac{1}{2}F(1)\delta^2,$$

where $F(\beta)$ is the Fisher information for $f^{(\beta)}$ with respect to the parameter β and $F(1)$ is the Fisher information evaluated for $\beta = 1$. Then both the numerator and denominator in (5) can be approximated by a quadratic function of the deviation. As an illustration of this behavior, Fig. 1 shows these functions around $\beta = 1$ for some particular cases of n and $k = p/n$. As mentioned above, the divergence $D_{KL}(f^{(\beta)}_{\mathbf{Y}|\mathbf{X}}|f_{\mathbf{Y}|\mathbf{X}})$ is a random variable given by a function of a χ^2 distribution with $n - p$ degrees of freedom. To evaluate the numerator in (5) for different values of $n,\ k$ and β, 100,000 observations were simulated from a χ^2 distribution with $n - p$ degrees of freedom and further transformed by (4) to obtain an approximation of the mean value. For the calculations, the open source programming language and environment for statistical computing and graphics R [20], is used

Thus, the limit in (3) can be obtained using second derivatives:

$$\lim_{\beta\to 1}\frac{E_X\left[D_{KL}\left(f^{(\beta)}_{\mathbf{Y}|\mathbf{X}}|f_{\mathbf{Y}|\mathbf{X}}\right)\right]}{D_{KL}(f^{(\beta)}|f)} = \frac{\frac{\delta^2}{\delta\beta^2}E_X\left[D_{KL}\left(f^{(\beta)}_{\mathbf{Y}|\mathbf{X}}|f_{\mathbf{Y}|\mathbf{X}}\right)\right]\Big|_{\beta=1}}{\frac{\delta^2}{\delta\beta^2}D_{KL}(f^{(\beta)}|f)|_{\beta=1}}. \tag{6}$$

The limit in (6) gives the ratio of the local curvature of KL divergence between the baseline and perturbed densities. Some details on the calculations and regularity conditions for differentiation under the integral sign are given in Appendix A. This calculations can be computed by numerical integration. Some values are shown in Fig. 2.

The conjecture that measure (3) tends to a constant value of $k = p/n$ as $n, p \to \infty$ is illustrated in Fig. 2. This implies that local deviations from normality in the MEP family affect the conditional distribution to a degree proportional to the dimension of the group of variables considered. In the context of GBNs, this result is intuitively clear, because it implies that network output is greatly affected by normality deviation as the dimension of the evidence variables decreases. Thus, if the set of evidence variables is larger than the set of interest variables, the local effect of normality deviations on network output is minimal.

5. Conclusions

A relative sensitivity measure was proposed for evaluating local effects of kurtosis deviations using the KL divergence measure. We focused on the MEP family, a generalization of the multivariate normal distribution, including distributions with different tails whose variation is controlled by a scalar parameter. Using Monte Carlo simulations of the random quantities involved, an asymptotic result on the relative sensitivity measure was obtained. This limit corresponds to the ratio between the dimensions of the conditional and the joint distributions.

Acknowledgments

This research was supported by the Spanish Ministerio de Ciencia e Innovaci ón, Grant MTM 2008-03282 and in part by GR58/08-A, 910395 – Métodos Bayesianos by BSCH-UCM, Spain.

Appendix A

The second derivative with respect to β of the numerator integrand in $\beta = 1$ for each $q > 0$ is given by

$$\frac{\delta^2}{\delta\beta^2}\left[D_{KL}\left(f^{(\beta)}_{\mathbf{Y}|\mathbf{X}}|f_{\mathbf{Y}|\mathbf{X}}\right)\right]\Big|_{\beta=1} = c(n) + \frac{-\frac{1}{4}\left(\int_0^\infty t^{\frac{1}{2}kn-1}e^{-\frac{1}{2}q-\frac{1}{2}t}(\ln(q+t))(q+t)dt\right)^2}{\left(\int_0^\infty t^{\frac{1}{2}kn-1}e^{-\frac{1}{2}q-\frac{1}{2}t}dt\right)^2} + \frac{\frac{1}{4}\int_0^\infty t^{\frac{1}{2}kn-1}e^{-\frac{1}{2}q-\frac{1}{2}t}\left(\ln^2(q+t)\right)(q+t)(q+t-2)dt}{\left(\int_0^\infty t^{\frac{1}{2}kn-1}e^{-\frac{1}{2}q-\frac{1}{2}t}dt\right)^2}\times\int_0^\infty t^{\frac{1}{2}kn-1}e^{-\frac{1}{2}q-\frac{1}{2}t}dt,$$

where $c(n) = \frac{n}{2}\left(\ln 2 + \psi\left(\frac{n}{2}\right) + \frac{2}{n}\right)^2 + n\psi\left(\frac{n}{2}, 1\right) - \frac{4}{n}$ and $\psi(x)$ is the digamma function, $\psi(x) = \frac{d}{dx}\ln\Gamma(x)$ and $\psi(x, 1)$ is the trigamma function, the first derivative of the digamma function [1].

It is evident that as long as the regularity conditions hold [2], because of the nature of the integrals involved, interchange between derivative and integral signs is allowed and thus

$$\frac{\delta^2}{\delta\beta^2}E_X\left[D_{KL}\left(f^{(\beta)}_{\mathbf{Y}|\mathbf{X}}|f_{\mathbf{Y}|\mathbf{X}}\right)\right]\Big|_{\beta=1} = \int_0^\infty\left(\frac{\delta^2}{\delta\beta^2}D_{KL}\left(f^{(\beta)}_{\mathbf{Y}|\mathbf{X}}|f_{\mathbf{Y}|\mathbf{X}}\right)\Big|_{\beta=1}\right)f_{\chi^2_{(1-k)y}}(q)dq.$$

Furthermore, the second derivative with respect to $\beta = 1$ for the denominator is

$$\frac{\delta^2}{\delta\beta^2} D_{KL}(f^{(\beta)}|f)\bigg|_{\beta=1} = n\left(\ln 2 + \psi\left(\frac{n}{2}\right)\right) + \frac{n^2}{4}\psi\left(\frac{n}{2}, 1\right) + 1 + c(n).$$

References

[1] M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions, Dover, New York, 1970.
[2] P. Billingsley, Probability and Measure, Wiley, New York, 1979.
[3] S. Blyth, Local divergence and association, Biometrika 81 (3) (1994) 579–584.
[4] K.A. Bollen, Structural Equations with Latent Variables, Wiley, New York, 1989.
[5] G.E.P. Box, Non-normality and tests on variances, Biometrika 40 (1953) 318–335.
[6] L.T. DeCarlo, On the meaning and use of kurtosis, Psychol. Methods 2 (1) (1997) 292–307.
[7] E. Gómez, M.A. Gómez-Villegas, J.M. Marín, A multivariate generalization of the power exponential family of distributions, Commun. Stat. Theory Methods 27 (3) (1998) 589–600.
[8] K.T. Fang, Y.T. Zhang, Generalized Multivariate Analysis, Springer-Verlag, Beijing, 1990.
[9] N. Foldnes, U.H. Olsson, T. Foss, The effect of kurtosis on the power of two test statistics in covariance structure analysis, Br. J. Math. Stat. Psychol. (2011), http://dx.doi.org/10.1111/j.2044-8317.2010.02010.x.
[10] M.A. Gómez-Villegas, P. Main, R. Susi, Sensitivity in Gaussian Bayesian networks using a divergence measure, Commun. Stat. Theory Methods 36 (3) (2007) 523–539.
[11] A.K. Gupta, S.W. Harrar, Y. Fujikoshi, Asymptotics for testing hypothesis in some multivariate variance components model under non-normality, J. Multivariate Anal. 97 (2006) 148–178.
[12] M. Haas, Value-at-risk via mixture distributions reconsidered, Appl. Math. Comput. 215 (2009) 2103–2119.
[13] P. Hoyer, S. Shimizu, A.J. Kerminen, M. Palviainen, Estimation of causal effects using linear non-Gaussian causal models with hidden variables, Int. J. Approximate Reasoning 49 (2008) 362–378.
[14] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1951) 79–86.
[15] S. Kullback, Information Theory and Statistics, Wiley, New York, 1959.
[16] T. Lumley, P. Diehr, S. Emerson, L. Chen, The importance of the normality assumption in large public health data sets, Annu. Rev. Public Health 23 (2002) 151–169.
[17] P. Main, H. Navarro, Analyzing the effect of introducing a kurtosis parameter in Gaussian Bayesian networks, Reliab. Eng. Syst. Saf. 94 (5) (2009) 922–926.
[18] K.V. Mardia, Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies, Sankhya Ser. B 36 (2) (1974) 115–128.
[19] K. Pearson, Contributions to the mathematical theory of evolution. II: Skew variation in homogeneous material, Philos. Trans. R. Soc. A 186 (1895) 343–412.
[20] R Development Core Team, R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, 2011, URL <http://www.R-project.org/>.
[21] A.B.L. Srivastava, Effect of non-normality on the power function of t-test, Biometrika 45 (1958) 421–429.
[22] K. Zografos, On Mardia's and Song's measures of kurtosis in elliptical distributions, J. Multivariate Anal. 99 (2008) 858–879.