The effect of non-normality in the power exponential distributions

Miguel A. Gómez-Villegas,[1] Eusebio Gómez-Sánchez-Manzano,[2] Paloma Maín,[3] and Hilario Navarro[4]

Abstract As an alternative to the multivariate normal distribution we have dealt with a wider class of distributions, including the normal, that considers slightly different tail behavior than the normal tail. This is the multivariate exponential power family of distributions with a kurtosis parameter to give the possible forms of the distributions. To measure distribution deviations the Kullback-Leibler divergence will be used as an asymmetric dissimilarity measure from an information-theoretic basis. Thus, a local quantitative description of the non-normality could be established for joint distributions in this family as well as the impact this perturbation causes in the marginal and conditional distributions.

1 Introduction

The multivariate normal distribution is traditionally used as a model for multivariate data in applications. However, this assumption may be doubtful in many real data analysis and it demands a wider class of distributions than the normal to be handled. Our choice is the multivariate exponential power family of distributions presented in [5] as a generalization of the multivariate normal family in that a new parameter, β, is introduced, as an exponent (see (1) below), which governs the kurtosis, and so the sharpness, of the distribution; for $\beta = 1$ we have the normal distribution, thus this parameter

Fac. de C.C. Matemáticas, Dpto. de Estadística e I.O., Universidad Complutense de Madrid, Pza. de Ciencias 3, 28040-Madrid, Spain ma_gv@mat.ucm.es · Fac. de C.C. Matemáticas, Dpto. de Estadística e I.O., Universidad Complutense de Madrid, Pza. de Ciencias 3, 28040-Madrid, Spain eusebio_gomez@mat.ucm.es · Fac. de C.C. Matemáticas, Dpto. de Estadística e I.O., Universidad Complutense de Madrid, Pza. de Ciencias 3, 28040-Madrid, Spain pmain@mat.ucm.es · Fac. de Ciencias, Dpto. de Estadística, I.O. y Calc. Num., UNED, C/ Senda del Rey 9, 28040-Madrid, Spain hnavarro@ccia.uned.es

represents the disparity of an exponential power distribution from the normal distribution.

The multivariate exponential power family is also a generalization of the univariate one (see [15] and [1, p. 157]) and can be included in the class of Kotz type distributions (see [4, p. 69] and [13]), which, in its turn, is a subset of the more general class of elliptical distributions (see a survey on these in [7]) Also, a matrix generalization of the exponential power distribution can be found in [6].

This distribution can be used to modelize multidimensional random phenomena with distributions having higher or lower tails than those of the normal distribution. Besides, the use of this distribution can robustify many multivariate statistical procedures. The multivariate exponential power distribution has been used to obtain robust models for nonlinear repeated measurements [10], to modeling dependencies among responses, as an alternative to models based upon the multivariate t distribution, and also to obtain robust models for the physiology of breathing. [2] use the multivariate exponential power distribution, as a heavy tailed distribution, in the field of speech recognition.

In this paper we evaluate the effect of this source of non-normality on the joint distributions and the corresponding marginal and conditional distributions for a specific partition. To measure distribution deviations, the Kullback-Leibler (KL) divergence will be used as an asymmetric dissimilarity measure from an information-theoretic basis. Thus, a local quantitative description of the non-normality can be established for joint distributions in this family as well as the impact this perturbation causes in the marginal and conditional distributions. This approach could be useful in problems where, given a model for the joint distribution, the interest is focussed in the distribution of a subset of variables given some values of the remaining ones. Such situations occur, among others, when we deal with Gaussian Bayesian networks for which the output is the conditional distribution of the variables of interest given fixed values of the evidential variables and a sensitivity analysis to non-normality is performed to prove the robustness and accuracy of the inferences.

The paper is organized as follows. In Section 2 the multivariate exponential power family is presented, highlighting some probabilistic characteristics to be handled in later sections. Section 3 is devoted to describe the impact of non-normality on the probabilistic structures of a random vector. The paper ends with conclusions in Section 4.

2 On the multivariate exponential power distributions

Next, we summarize the most important features of this family of distributions. An absolutely continuous random vector $\mathbf{X} = (X_1, \ldots, X_n)'$ is said to have a power exponential distribution if its density has the form

$$f(\mathbf{x}; \mu, \mathbf{\Sigma}, \beta) = k \, |\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \left((\mathbf{x} - \mu)' \, \mathbf{\Sigma}^{-1} \, (\mathbf{x} - \mu) \right)^{\beta} \right\}, \qquad (1)$$

with $k = \frac{n\Gamma\left(\frac{n}{2}\right)}{\pi^{\frac{n}{2}} \Gamma\left(1+\frac{n}{2\beta}\right) 2^{1+\frac{n}{2\beta}}}$, where $(\mu, \mathbf{\Sigma}, \beta) \in (\mathbb{R}^n, S, (0, \infty))$, S being the set of $(n \times n)$ positive definite symmetric matrices, then, we write $\mathbf{X} \sim EP_n (\mu, \mathbf{\Sigma}, \beta)$. The parameters μ and $\mathbf{\Sigma}$ are location and scale parameters. The parameter β is a shape parameter, as the kurtosis depends only on it. Figs. 1-3 show the graphs of the density $EP_2 (\mathbf{0}, \mathbf{I}_2, \beta)$ for the values 6, 1, 1/2 of β.

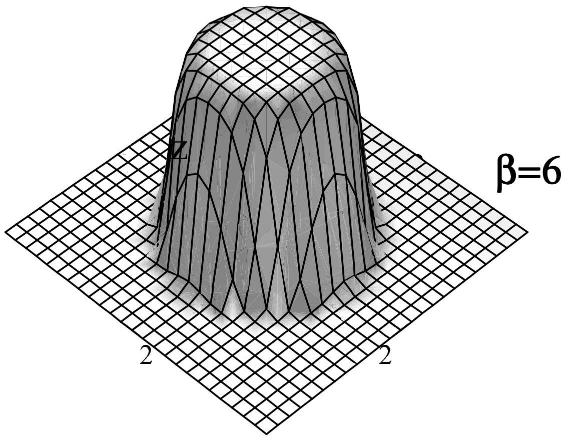

Fig. 1 EP_2 density for $\beta = 6$

It can be pointed that as β increases the sharpness diminishes; for β going to infinity, (1) tends to be uniform in the ellipsoid $(\mathbf{x} - \mu)' \, \mathbf{\Sigma}^{-1} \, (\mathbf{x} - \mu)$ and also when β goes to 0 the pick narrows infinitely and (1) tends to the improper density constant in $\mathbb{R}^n$.

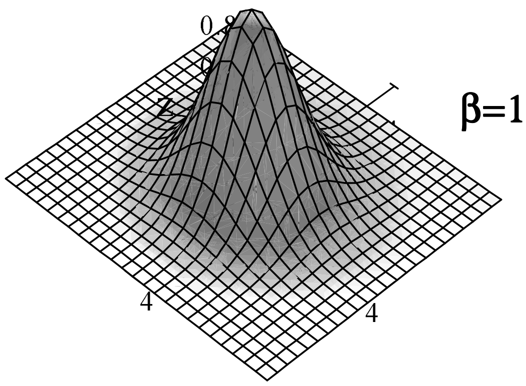

Fig. 2 Multivariate Normal density function, $\beta = 1$

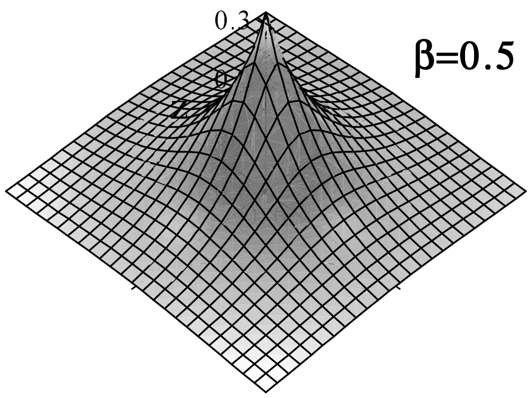

Fig. 3 Multivariate Double Exponential density, $\beta = \frac{1}{2}$

2.1 Some related distributions

Let $\mathbf{X} \sim EP_n(\mu, \Sigma, \beta)$. If $\beta = 1$, then $\mathbf{X}$ has a normal distribution: $\mathbf{X} \sim N_n(\mu, \mathbf{\Sigma})$. In any case, $\mathbf{X}$ has an elliptical distribution: $\mathbf{X} \sim E_n(\mu, \mathbf{\Sigma}, g)$ (in the sense given in [7]) with $g(t) = \exp\left\{-\frac{1}{2}t^{\beta}\right\}$.

An exponential power distribution $EP_n(\mu, \mathbf{\Sigma}, \beta)$ is a scale mixture of normal distributions (see [8]) in the strict sense (namely, with respect to a probability distribution function) if $\beta \in (0, 1]$. If we exclude the normal case, that is, if $\beta \in (0, 1)$, then

$$f(\mathbf{x}; \mu, \mathbf{\Sigma}, \beta) = \int_0^{\infty} N_n\left(\mathbf{x}; \mu, v^2\mathbf{\Sigma}\right) \, dH_{\beta}(v), \tag{2}$$

where $N_n\left(\mathbf{x}; \mu, v^2\mathbf{\Sigma}\right)$ is the normal density with mean μ and covariance matrix $v^2\mathbf{\Sigma}$, and H_{β} is the distribution function having density

$$h_{\beta}(v) = \frac{2^{1+\frac{n}{2}-\frac{n}{2\beta}}\Gamma\left(1+\frac{n}{2}\right)}{\Gamma\left(1+\frac{n}{2\beta}\right)} v^{n-3} S_{\beta}\left(v^{-2}; 2^{1-\frac{1}{\beta}}\right),$$

where $S_{\beta}(\,\cdot\,; \sigma)$ means the density of the (positive) stable distribution having characteristic function (see [14, p. 8])

$$\varphi(t) = \exp\left\{-\sigma^{\beta}\,|t|^{\beta}\, e^{-i\frac{\pi}{2}\beta sign(t)}\right\}.$$

For $\beta = 1$ (the normal case) (2) holds, of course, H_{β} being the distribution function degenerate in 1. For $\beta \in (1, \infty)$, the exponential power distribution $EP_n(\mu, \mathbf{\Sigma}, \beta)$ is a scale mixture of normal distributions too, as all the elliptical distributions are (see [3]), but only in a wider sense, since in this case function H_{β} in (2) is like a distribution function in $(0, \infty)$, but it is not a nondecreasing function.

2.2 Probabilistic characteristics

If $\mathbf{X} \sim EP_n(\mu, \mathbf{\Sigma}, \beta)$, its characteristic function is

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \frac{n}{\Gamma\left(1+\frac{n}{2\beta}\right) 2^{\frac{n}{2\beta}}} \exp\left(i\mathbf{t}'\mu\right) \int_0^{\infty} \Psi_n\left(r\sqrt{\mathbf{t}'\mathbf{\Sigma}\mathbf{t}}\right) r^{n-1} \exp\left\{-\frac{1}{2}r^{2\beta}\right\} dr,$$

where $\Psi_1(x) = \cos x$ and $\Psi_n(x) = \frac{\Gamma\left(\frac{n}{2}\right)}{\pi^{\frac{1}{2}}\Gamma\left(\frac{n-1}{2}\right)} \int_0^{\pi} \exp\{ix\cos\theta\} \sin^{n-2}\theta d\theta$, for $n > 1$. Besides,

$$E[X] = \mu,$$
$$Var[X] = \frac{2^{\frac{1}{\beta}} \Gamma\left(\frac{n+2}{2\beta}\right)}{n\Gamma\left(\frac{n}{2\beta}\right)} \mathbf{\Sigma},$$
$$\gamma_1[X] = 0,$$
$$\gamma_2[X] = n^2 \frac{\Gamma\left(\frac{n+4}{2\beta}\right) \Gamma\left(\frac{n}{2\beta}\right)}{\left(\Gamma\left(\frac{n+2}{2\beta}\right)\right)^2} - n(n-2),$$

where γ_1 and γ_2 are the asymmetry and kurtosis coefficients as shown in [12, p. 31].

Figure 4 shows the kurtosis coefficient as a function of β for $n = 1$ (dotted line), 2, 3, 5 and 7, supporting the previous comments about the monotony relation between kurtosis and the non-normality coefficient in this family.

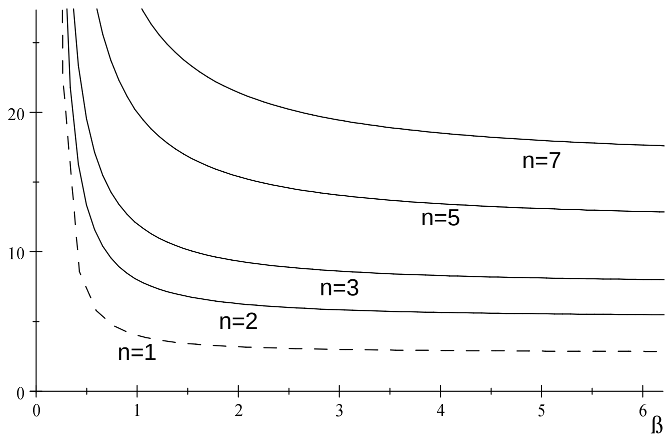

Fig. 4 Kurtosis coefficient as a function of β

2.3 *Marginal and conditional distributions and regression*

The marginal and conditional distributions are elliptical. But the regression function is linear, as in the normal case. Specifically, let $\mathbf{X} \sim EP_n(\mu, \mathbf{\Sigma}, \beta)$ and make $\mathbf{X} = (\mathbf{X}'_{(1)}, \mathbf{X}'_{(2)})'$, with $\mathbf{X}_{(1)} = (X_1, \ldots, X_p)'$ and $\mathbf{X}_{(2)} = (X_{p+1}, \ldots, X_n)'$, with $p < n$; analogously make $\mu = (\mu'_{(1)}, \mu'_{(2)})'$ and $\mathbf{\Sigma} =$

$\begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}$, where $\mathbf{\Sigma}_{11}$ is a $(p \times p)$ matrix. Then $\mathbf{X}_{(1)}$ has an elliptical distribution: $\mathbf{X}_{(1)} \sim E_p(\mu_{(1)}, \mathbf{\Sigma}_{11}, g_{(1)})$, where

$$g_{(1)}(t) = \int_0^\infty w^{\frac{n-p}{2}-1} \exp\left\{-\frac{1}{2}(t+w)^\beta\right\} dw.$$

The distribution of $\mathbf{X}_{(2)}$ conditional to $\mathbf{X}_{(1)} = \mathbf{x}_{(1)}$ is elliptical too. $\left(\mathbf{X}_{(2)} \left| \mathbf{X}_{(1)} = x_{(1)} \right.\right) \sim E_{n-p}\left(\mu_{(2.1)}, \mathbf{\Sigma}_{22.1}, g_{(2.1)}\right)$, with

$$\begin{aligned} \mu_{(2.1)} &= \mu_{(2)} + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\left(\mathbf{x}_{(1)} - \mu_{(1)}\right), \\ \mathbf{\Sigma}_{22.1} &= \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}, \\ g_{(2.1)}(t) &= \exp\left\{-\frac{1}{2}\left(t + \mathbf{q}_{(1)}\right)^\beta\right\}, \end{aligned}$$

where $\mathbf{q}_{(1)} = \left(\mathbf{x}_{(1)} - \mu_{(1)}\right)' \mathbf{\Sigma}_{11}^{-1}\left(\mathbf{x}_{(1)} - \mu_{(1)}\right)$.

3 The effects of deviations from normality

Now, we are interested in the effects of small changes in the parameter β of the $EP_n(\mu, \mathbf{\Sigma}, \beta)$ distribution taking as a reference the one with $\beta_0 = 1$, that, as pointed above, corresponds to a normal distribution with parameters μ and $\mathbf{\Sigma}$. When β is close to $\beta_0 = 1$, that is, $\beta = \beta_0 + \delta$ with δ representing a small deviation from normality, the Taylor expansion leads to the approximation

$$D_{KL}(f, f^{(\delta)}) \approx \frac{1}{2}F_\beta(1)\delta^2 \;, \tag{3}$$

being f the normal density, $f^{(\delta)}$ the perturbed density and $F_\beta(1)$ the Fisher information with respect to β in $\beta_0 = 1$. The same problem can be formulated in terms of the marginal and conditional distributions for a fixed partition of the random vector $\mathbf{X}$. From now on, our goal is both analytical and graphical description of the function (3).

3.1 Joint Distributions

Let $f(\mathbf{x})$ be a density function of the family $EP_n(\mu, \mathbf{\Sigma}, \beta_0 = 1)$, that is a normal density $N_n(\mu, \mathbf{\Sigma})$ and $f^{(\delta)}(\mathbf{x})$ be the perturbed density $EP_n(\mu, \mathbf{\Sigma}, \beta = 1 + \delta)$, then the KL divergence between these densities can be calculated using that, if $\mathbf{X} \sim N_n(\mu, \mathbf{\Sigma})$, the quadratic form $(\mathbf{X} - \mu)' \mathbf{\Sigma}^{-1} (\mathbf{X} - \mu)$ is distributed as a chi-square distribution with n degrees of freedom. Specifically, since

$$D_{KL}(f, f^{(\delta)}) = E_f \left[\log \frac{f(\mathbf{X})}{f^{(\delta)}(\mathbf{X})} \right]$$

it follows

$$D_{KL}(f, f^{(\delta)}) = \log \frac{2^{\frac{n}{2(1+\delta)}} \Gamma\left(\frac{n}{2(1+\delta)}\right)}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)(1+\delta)} - \frac{1}{2}\{E_f\left[(\mathbf{X}-\mu)' \mathbf{\Sigma}^{-1} (\mathbf{X}-\mu)\right] + \\ - E_f\left[\left((\mathbf{X}-\mu)' \mathbf{\Sigma}^{-1} (\mathbf{X}-\mu)\right)^{1+\delta}\right]\}$$

that is

$$D_{KL}(f, f^{(\delta)}) = \log \frac{2^{\frac{n}{2(1+\delta)}} \Gamma\left(\frac{n}{2(1+\delta)}\right)}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)(1+\delta)} - \frac{1}{2}\left(n - \frac{2^{(1+\delta)} \Gamma\left(\frac{n}{2} + (1+\delta)\right)}{\Gamma\left(\frac{n}{2}\right)}\right) \quad (4)$$

According to this result, the divergence between joint densities depends on the dimension of the random vector n and the perturbation δ applied to the reference normal distribution. Figure 5 illustrates the relation (4) when δ is small and, consequently, the approximation (3) holds. Observe that there is a monotone behavior with respect to the dimension n with a faster growth for high dimensions. From a local point of view, Figure 5 confirms the approximation /3).

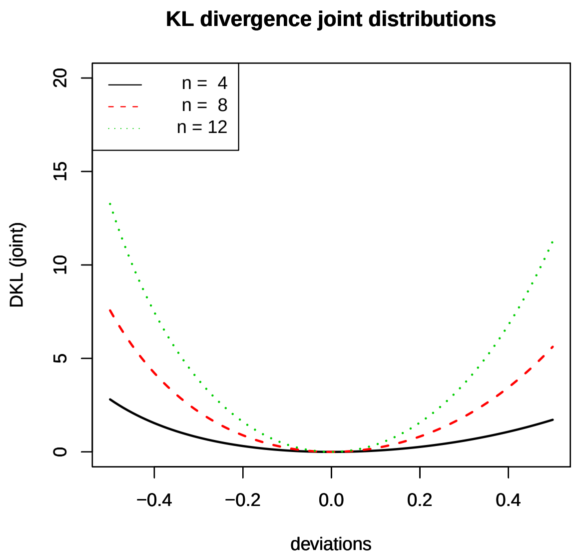

Fig. 5 KL divergence of the joint distributions for $n = 4, 8, 12$

3.2 Conditional Distributions

Now we focus on the analysis of conditional distributions sensitivity to small perturbations of the parameter β. Using previous notation it follows

$$f_{2.1}^{(\delta)}\left(\mathbf{x}_{(2)} \mid \mathbf{x}_{(1)}\right) =$$
$$= k_1 \left|\mathbf{\Sigma}_{22.1}\right|^{-\frac{1}{2}} \exp -\frac{1}{2}\left\{\left[\left(\mathbf{x}_{(2)}-\mu_{(2.1)}\right)' \mathbf{\Sigma}_{22.1}^{-1}\left(\mathbf{x}_{(2)}-\mu_{(2.1)}\right)+\mathbf{q}_1\right]^{(1+\delta)}\right\},$$

being

$$k_1 = \frac{\Gamma\left(\frac{n-p}{2}\right)}{\pi^{\frac{n-p}{2}} \int_0^\infty t^{\frac{n-p}{2}-1} \exp\left\{-\frac{1}{2}\left(t+\mathbf{q}_1\right)^{(1+\delta)}\right\} dt}$$

and consequently the KL divergence is [11]

$$D_{KL}\left(f_{2.1}, f_{2.1}^{(\delta)}\right) = \log \frac{\int_0^\infty t^{\frac{n-p}{2}-1} \exp\left\{-\frac{1}{2}\left(t+\mathbf{q}_1\right)^{(1+\delta)}\right\} dt}{2^{\frac{n-p}{2}} \Gamma\left(\frac{n-p}{2}\right)}$$
$$-\frac{1}{2}\left[n-p-\frac{\mathbf{q}_1^{(1+\delta)+\frac{n-p}{2}}}{2^{\frac{n-p}{2}}} U(a,b,x)\right],$$

where $U(a,b,x)$ is the *Confluent Hypergeometric Function* calculated in

$$a = \frac{n-p}{2}\ ,\ b = 2+\delta+\frac{n-p}{2}\ ,\ x = \frac{\mathbf{q}_1}{2}.$$

Figure 6 shows the KL divergence, as a function of δ, for the conditional distributions corresponding to selected values of the $\mathbf{q}_1$ distribution: the mean and the $10th$ and $90th$ quantiles. In this setting the divergence is affected by the dimension of X, the dimension of the conditioning random vector $X_{(1)}$ and the particular value of the conditioning variables through the Mahalanobis distance to its mean. As it was expected, the KL divergence has a quadratic appearance compatible with Equation (3) for δ close to zero. From a statistical point of view, a larger variability is found for distributions with lighter tails than the normal. Also, for the chosen values (mean and quantiles) of the $\mathbf{q}_1$ distribution, the KL divergence functions are monotone and their relative position are directly related to the ratio p/n..

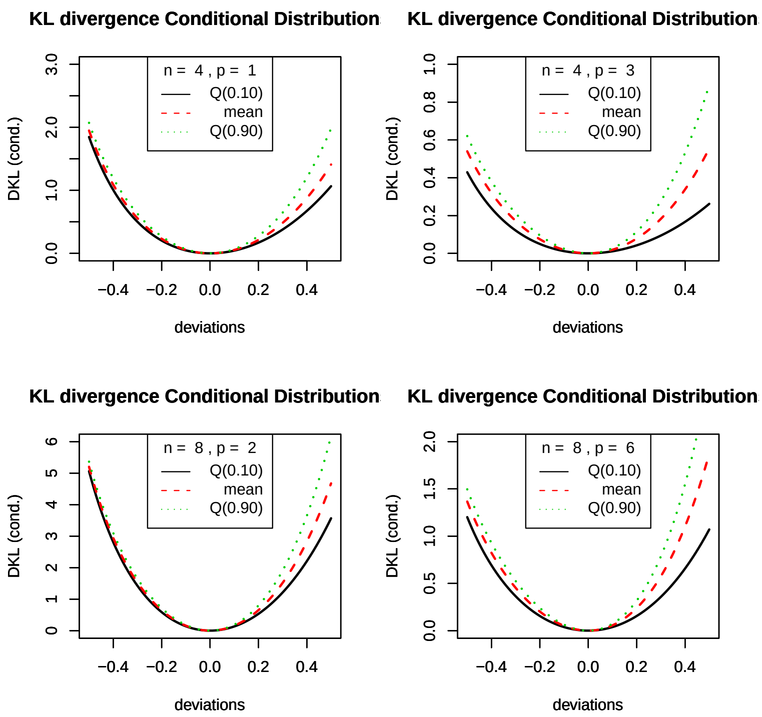

Fig. 6 KL divergence of the conditional distributions: $n = 4, 8$; $p/n = 0.25, 0.75$

3.3 Marginal Distributions

A similar approach holds for the case of marginal distributions. However, deriving an exact expression for the KL divergence using analytical methods appears to be a complicated task. Here, Monte Carlo simulation data were used to approximate the value of this measure, under a variety of conditions.

$$D_{KL}\left(f_1, f_1^{(\delta)}\right) = \log \frac{2^{\frac{n}{2(1+\delta)} - \frac{p}{2}} \Gamma\left(\frac{n}{2(1+\delta)}\right) \Gamma\left(\frac{n-p}{2}\right)}{\Gamma\left(\frac{n}{2}\right)(1+\delta)\exp\left(\frac{p}{2}\right)} - \log E_{\chi^2_p}\left[\int_0^\infty w^{\frac{n-p}{2}-1} e^{-\frac{w}{2}} e^{-\frac{1}{2}\left(\chi^2_p + w\right)^\delta} dw\right]$$

Figure 7 shows the simulation results obtained for different values of n, p and δ, using 50,000 replications of the random variable χ^2_p for each case. In general the behavior observed is similar to that of the previous sections.

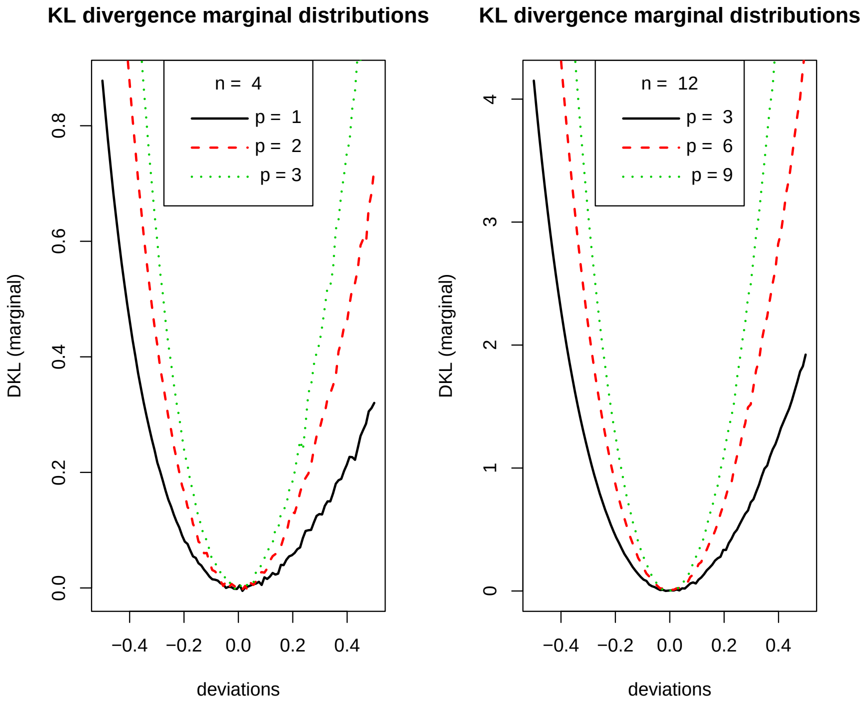

Fig. 7 KL divergence of the marginal distributions: $n = 4, 8$; $p/n = 0.25, 0.50, 0.75$

On the other hand, it is well known that the divergences between the different distributions we have considered are related as follows

$$D_{KL}(f, f^{(\delta)}) = E_{f_1}\left[D_{KL}\left(f_{2.1}, f_{2.1}^{(\delta)}\right)\right] + D_{KL}\left(f_1, f_1^{(\delta)}\right) \tag{5}$$

and therefore the divergence between marginal densities would also be approximated from the previous identity with Monte Carlo simulations to estimate the conditional KL divergence mean.

Finally, Equation (5) suggests the definition of a relative divergence measure for the conditional and marginal distributions in terms of the ratios

$$\frac{E_{f_1}\left[D_{KL}\left(f_{2.1}, f_{2.1}^{(\delta)}\right)\right]}{D_{KL}(f, f^{(\delta)})}, \frac{D_{KL}\left(f_1, f_1^{(\delta)}\right)}{D_{KL}(f, f^{(\delta)})}.$$

A recent approach to this problem is presented in [9].

4 Conclusions

In this paper we considered the multivariate exponential power family of distributions as an alternative model when normality assumption was doubtful. The Kullback-Leibler divergence measure is used as a tool for exploring the influence of deviations from multivariate normal in joint, conditional and marginal distributions. The obtained expressions for divergence measures provide quadratic sensitivity functions both globally and locally. Moreover, it results that this effect depends on the dimension of the vectors involved as well as the values of the conditioning variables through the Mahalanobis distance to its mean, for the case of conditionals.

References

1. Box G, Tiao G (1973) Bayesian Inference in Statistical Analysis. Reading: Addison-Wesley
2. Basu S, Miccheli CA, Olsen P (2001) J Comput Graph Statist 10(1): 158–184
3. Chu KC (1973) IEEE Trans Automat Control 18: 499–505
4. Fang K T, Kotz S, Ng KW (1990) Symmetric Multivariate and Related Distributions. London: Chapman and Hall
5. Gómez E, Gómez-Villegas MA, Marín, JM (1998) Comm. Statist.-Theory Methods 27: 589–600
6. Gómez E, Gómez-Villegas MA, Marín, JM (2002) Comm. Statist.-Theory Methods 31(12): 2167–2182
7. Gómez E, Gómez-Villegas MA, Marín, JM (2003) Rev Mat Complut 16: 345–361
8. Gómez E, Gómez-Villegas MA, Marín, JM (2008) Comm. Statist.-Theory Methods 37(6): 972–985
9. Gómez-Villegas MA, Main P, Navarro H, Susi R (2010) Procedia Social and Behavioral Sciences 2: 7664–7665
10. Lindsey JK (1999) Biometrics 56: 1277–1280
11. Main P, Navarro H (2009) Reliab Eng Syst Safety 94: 92–926
12. Mardia KV, Kent JT, Bibby JM (1979) Multivariate Analysis. Academic Press, London
13. Nadarajah S (2003) Statistics 37(4): 341–358
14. Samorodnitsky G, Taqqu MS (2000) Stable Non-gaussian Random Processes. Chapman and Hall, Boca Raton
15. Subbotin M (1923) Mathematicheskii Sbornik 31: 296–301