Brazilian Journal of Probability and Statistics
2017, Vol. 31, No. 2, 373–393
DOI: 10.1214/16-BJPS318

Concentration function for the skew-normal and skew-*t* distributions, with application in robust Bayesian analysis

Luciana G. Godoi[a], **Márcia D. Branco**[b] and **Fabrizio Ruggeri**[c]

[a]*Universidade Federal do Espírito Santo*
[b]*Universidade de São Paulo*
[c]*Consiglio Nazionale delle Ricerche*

Abstract. Data from many applied fields exhibit both heavy tail and skewness behavior. For this reason, in the last few decades, there has been a growing interest in exploring parametric classes of skew-symmetrical distributions. A popular approach to model departure from normality consists of modifying a symmetric probability density function in a multiplicative fashion, introducing skewness. An important issue, addressed in this paper, is the introduction of some measures of distance between skewed versions of probability densities and their symmetric baseline. Different measures provide different insights on the departure from symmetric density functions: we analyze and discuss L_1 distance, J-divergence and the concentration function in the normal and Student-t cases. Multiplicative contaminations of distributions can be also considered in a Bayesian framework as a class of priors and the notion of distance is here strongly connected with Bayesian robustness analysis: we use the concentration function to analyze departure from a symmetric baseline prior through multiplicative contamination prior distributions for the location parameter in a Gaussian model.

1 Introduction

In the last few decades, there has been a growing interest in exploring parametric classes of non-normal distributions, see Genton (2004) and Azzalini (2005). A popular approach to model departure from normality consists of modifying a symmetric probability density function in a multiplicative fashion, introducing skewness. Following this approach, we consider a class of skew-symmetric distributions given in Azzalini and Capitanio (2003). The probability density function, up to location and scale parameters, is of the form

$$f_1(z|\alpha) = 2 f_0(z) G\bigl(w(z,\alpha)\bigr), \qquad z \in \mathbb{R}, \tag{1}$$

where $f_0(\cdot)$ is a symmetric density in $\mathbb{R}$, that is $f_0(-z) = f_0(z)$ for all $z \in \mathbb{R}$, $G(\cdot)$ is a symmetric absolutely continuous cumulative distribution function, that is $G(-z) = 1 - G(z)$ for all $z \in \mathbb{R}$, with density $g(\cdot)$ and $w(\cdot)$ is a function such that

Key words and phrases. Bayesian robustness, skew-symmetric models, L_1 distance, concentration function.

Received May 2015; accepted April 2016.

$w(-z,\alpha) = -w(z,\alpha)$ for all $z \in \mathbb{R}$. We also add a parameter $\alpha \in \mathbb{R}$ in w which controls the shape of the distribution; moreover, we consider $w(z,0) = 0, \forall z$. Therefore, the symmetric base density $f_0(\cdot)$ is retrieved when $\alpha = 0$. Parameters of location, $\xi \in \mathbb{R}$, and of scale, $\tau > 0$, can be introduced through $Y = \xi + \tau Z$, where Z is a random variable with density (1).

The class (1) contains many interesting distributions. For instance, the choice $f_0(z) = \phi(z)$ and $G(z) = \Phi(z)$, the standard normal density and cumulative distribution function, respectively, with $w(z,\alpha) = \alpha z$, yields the skew-normal distribution, $SN(\alpha)$, introduced by Azzalini (1985). Picking $f_0(z) = t(z;\nu)$ and $G(z) = T(z;\nu+1)$, the standard Student's t density and cumulative distribution function with ν and $\nu+1$ degrees of freedom, respectively, and $w(\alpha,z) = \alpha z\sqrt{\frac{\nu+1}{\nu+z^2}}$, yields the skew-$t$ distribution of Branco and Dey (2001) and Azzalini and Capitanio (2003). Other choices include setting $f_0(z) = \phi(z)$ while letting G be any symmetric cumulative distribution function that is different from the normal distribution (Nadarajah and Kotz (2003)), or setting $G(z) = \Phi(z)$ while letting f_0 be any symmetric density that is different from the normal density (Gomez, Venegas and Bolfarine (2007)).

We consider the usual notation $SN(\xi,\tau^2,\alpha)$, $N(\xi,\tau^2)$, $ST(\xi,\tau^2,\alpha,\nu)$ and $T(\xi,\tau^2,\nu)$ for the location-scale skew-normal, normal, skew-t and Student-t distributions, respectively.

When we work with skew-symmetric models, a natural question which arises is how far we can go from symmetry using this kind of skewness, that is, how the baseline distribution is affected introducing a skewing function. To answer this question, we study some measures of distance between skewed distributions and their symmetric baselines. The L_1 distance was already explored by Vidal et al. (2006) when $w(z,\alpha) = \alpha z$ in (1). The authors use this measure for model comparison. However, they do not present a closed form for the Student-t case. On the other hand, Contreras-Reyes and Arellano-Valle (2012) obtain the Kullback–Leibler (KL) divergence and the J-divergence between two skew distributions in the multivariate skew-normal context. Again, the Student-t case is not explored by the authors. Moreover, as far as we know, nobody has studied the concentration function as a measure of distance for skew-symmetric distributions.

According to Fortini and Ruggeri (2000), the concentration function of a probability measure P with respect to another one, say P_0, extends the classical notion of the Lorenz–Gini curve and it can be used to define neighborhood of probability measures or compare them. A prime use of such property is in Bayesian robustness where the concentration function can be used to define topological neighborhoods of a baseline prior distribution and to measure ranges spanned, as the prior varies in a class, by the probability of measurable subsets with fixed probability under a baseline prior. The concentration function gives different insights with respect to the usual indices when comparing measures: as an example, it is possible that two probability measures have means differing by a very small amount but the concentration function detects a very different behavior when the two measures concentrate all their mass around two very close values (i.e., the mean) but on disjoint intervals, therefore *concentrating* mass in very different subsets. More details on the properties of the concentration function and its applications can be found in the paper by Fortini and Ruggeri (2000) and the references therein.

The focus of this paper is to measure the distance between the density f_1 given in (1) and the symmetric baseline f_0. In Section 2, we review some measures already given in the literature and present new results under some special cases. In Section 3, we present the concentration function as a measure to compare two probability densities and give results in closed form in some special cases. The concentration function is used in a context of Bayesian robustness in Section 4, whereas some final comments are presented in Section 5.

2 Some measures of divergence

The focus of this section is to measure the divergence between the density f_1 given in (1) and the symmetric baseline f_0. Expressions for the L_1 distance and the J-divergence are available in the general context and for the special cases of skew-normal and skew-t models.

2.1 The L_1 distance

The L_1 distance between two density functions f and g is given by

$$L_1(f, g) = \frac{1}{2}\int \left|f(x) - g(x)\right| dx = \sup_{A\in\mathcal{B}} \left|P_f(A) - P_g(A)\right|, \tag{2}$$

where P_f and P_g are probability measures, associated with the density functions f and g, in the same measurable space $(\mathbb{R}, \mathcal{B})$ where $\mathcal{B}$ is the Borel σ-field. Note that L_1 is an upper bound on the differences $|P_f(A) - P_g(A)|$ for any set $A \in \mathcal{B}$. Also, the L_1 distance is bounded and takes values in $[0, 1]$, where $L_1(f, g) = 0$ implies that $f(x) = g(x)$ a.e. and $L_1(f, g) = 1$ indicates that the supports of the two densities are disjoint, indicating maximal discrepancy.

With a slight change in the proof of Proposition 1 in Vidal et al. (2006), we obtain the next result.

Proposition 2.1. *The L_1 distance between f_0 and f_1, as defined in* (1), *is*

$$L_1(f_0, f_1) = E_{Z^*}\left[G\left(\left|w(Z^*, \alpha)\right|\right)\right] - \frac{1}{2}, \tag{3}$$

where the density of Z^* is obtained folding the symmetric distribution f_0 on itself, that is, $f_{Z^*}(z) = 2f_0(z)Id_{(0,\infty)}$, with Id_A the usual indicator function, taking the value one when $z \in A$ and zero otherwise.

For the skew-normal and skew-t distributions, a closed form for the L_1 distance is available.

Proposition 2.2. *The L_1 distance between $SN(\xi, \tau^2, \alpha)$ and $N(\xi, \tau^2)$ distributions is equal to the L_1 distance between $ST(\xi, \tau^2, \alpha, \nu)$ and $T(\xi, \tau^2, \nu)$ distributions, for any ν, and it is given by*

$$\frac{1}{\pi}\arcsin\left(\frac{|\alpha|}{\sqrt{1+\alpha^2}}\right) = \frac{1}{2} - \frac{1}{\pi}\arccos\left(\frac{|\alpha|}{\sqrt{1+\alpha^2}}\right). \tag{4}$$

The proof of Proposition 2.2 follows from Proposition 7 in Vidal et al. (2006) and the fact that the skew-t distribution is a scale mixture of the skew-normal (see Azzalini and Capitanio, 2003). In fact, the result is also true for any member of the scale mixture of the skew-normal family. Vidal et al. (2006) were not able to obtain a closed form for the skew-t case because they were using another way to add skewness to the t distribution. Note that there are several definitions of skew-t in literature and it is important to be aware which each one is being considered. This is even more important when considering the multivariate context (see Chapter 5 in Kotz and Nadarajah (2004)).

Moreover, from (4) we can see clearly that the L_1 distance does not depend on the location and scale parameters, since it only depends on the shape parameter.

2.2 Kullback–Liebler divergence and J-divergence

The Kullback–Liebler (KL) divergence between two densities f and g is given by

$$KL(f, g) = \int f(x) \log\left\{\frac{f(x)}{g(x)}\right\} dx. \tag{5}$$

This is not a measure of distance since $KL(f, g) \neq KL(g, f)$. The usual way to obtain a symmetric measure based on KL divergence is to consider $J(f, g) = KL(f, g) + KL(g, f)$, known as J-divergence. Contreras-Reyes and Arellano-Valle (2012) present the J-divergence between two multivariate skew-normal distributions. Latter, Arellano-Valle, Contreras-Reyes and Genton (2013) showed some results about entropy in the context of the skew-t distribution; however they did not obtain the J-divergence.

In the following, we deal with the one-dimensional case and the more general skewness structure considered in this paper.

Proposition 2.3. *The J-divergence between f_1 and f_0, as defined in (1), is given by*

$$E_S\{\log[G(w(S, \alpha))]\} - E_Z\{\log[G(w(Z, \alpha))]\}, \tag{6}$$

where $S \sim f_1$, $Z \sim f_0$ and α is fixed.

Corollary 2.1.

1. *The J-divergence between the $SN(\xi, \tau^2, \alpha)$ and $N(\xi, \tau^2)$ is given by*

$$E_{X_1}\{\log[\Phi(\alpha X_1)]\} - E_{X_0}\{\log[\Phi(\alpha X_0)]\}, \tag{7}$$

where $X_1 \sim SN(0, 1, \alpha)$ and $X_0 \sim N(0, 1)$; and

2. *The J-divergence between the $ST(\xi, \tau^2, \alpha, \nu)$ and $T(\xi, \tau^2, \nu)$ is given by*

$$\begin{aligned} &E_{Y_1}\left\{\log\left[T\left(\alpha Y_1\sqrt{\frac{\nu+1}{\nu+Y_1^2}}; \nu+1\right)\right]\right\} \\ &\quad - E_{Y_0}\left\{\log\left[T\left(\alpha Y_0\sqrt{\frac{\nu+1}{\nu+Y_0^2}}; \nu+1\right)\right]\right\}, \end{aligned} \tag{8}$$

where $Y_1 \sim ST(0, 1, \alpha, \nu)$ and $Y_0 \sim T(0, 1, \nu)$.

The result for the skew-normal is a particular case of Contreras-Reyes and Arellano-Valle (2012).

The proof of Proposition 2.3 is given in the Appendix. Corollary 2.1 follows from Proposition 2.3 and the fact that the J-divergence does not depend on the location and scale parameters.

Evaluation of the J-divergences presented in the Corollary 2.1 is possible considering numerical procedures like quadrature methods or ordinary Monte Carlo. All the results for the simulation of J-divergences in this work were obtained using the function *NExpectation* in the *Mathematica*® 10.1 software.

Recently, Contreras-Reyes (2014) showed a way to approximate the Kullback–Liebler divergence between two skew-t distributions. Although this result is only an asymptotic approximation, it can be helpful to calculate the J-divergence approximately.

3 The concentration function

According to Cifarelli and Regazzini (1987), the concentration function of a probability measure P with respect to another one, say P_0, is extending the classical notion of the Lorenz–Gini curve. The concentration function studies the discrepancy between two measures defined on the same probability space, comparing the different concentrations of probability determined by the measures. As discussed in Cifarelli and Regazzini (1987), the concentration function is also related to the total variation distance and some indexes, like Gini's and Pietra's, which capture different aspects of the difference between probability measures. In Fortini and Ruggeri (2000), a thorough review of the properties of the concentration function is presented, along with its applications in Bayesian robustness and as a tool to build topological neighborhoods of probability measures.

In the next, we present the formal definition and an important lemma for the interpretation of the concentration function.

We denote by f and f_0 the density functions associated, respectively, to the probability measures P_f and P_{f_0}, absolutely continuous with respect to Lebesgue measure on the same measurable space $(\mathbb{R}, \mathcal{B})$, where $\mathbb{R}$ is the set of the real numbers and $\mathcal{B}$ is the Borel σ-field on $\mathbb{R}$.

Definition 3.1. Let $L_y = \{\theta \in \mathbb{R} : h(\theta) \leq y\}$ for any $y \geq 0$, and $h(\theta) = \frac{f(\theta)}{f_0(\theta)}$ for any $\theta \in \mathbb{R}$. The concentration function of P_f with respect to P_{f_0} is given by φ_{P_f}: $[0, 1] \to [0, 1]$ such that $\varphi_{P_f}(0) = 0$, $\varphi_{P_f}(1) = P_f(\mathbb{R}) = 1$ and

$$\varphi_{P_f}(z) = P_f(L_y), \qquad \text{if } z = P_{f_0}(L_y), \tag{9}$$

where $P_f(L_y) \equiv \int_{L_y} f(\theta)\, d\theta$ and $P_{f_0}(L_y) \equiv \int_{L_y} f_0(\theta)\, d\theta$.

Scarsini (1990) provides a similar definition when he considers two probability measures on the power set of a finite space $\mathcal{X}$. The next lemma (see Cifarelli and Regazzini (1987)) presents an important propriety about the distance between two densities which allows also for a helpful graphical representation.

Lemma 3.1. *Let P_f and P_{f_0} be two probabilities measures on the same measurable space $(\mathbb{R}, \mathcal{B})$. For $z \in [0, 1]$ and any $A \in \mathcal{B}$ such that $P_{f_0}(A) = z$, then*

$$\varphi(z) \leq P_f(A) \leq 1 - \varphi(1 - z). \tag{10}$$

Another interesting property about the concentration function is related to two distributions that have undergone the same location-scale transformation. In this situation, the concentration function is not affected by equal location shifts and/or scale changes applied to both distributions.

3.1 Skew-normal case

We present a closed form expression for the concentration function between skew-normal and normal distributions. Then, we discuss the interpretation of this measure through a special case and complete this section presenting a relationship between the concentration function and the L_1 distance.

Proposition 3.1. *The concentration function between a $SN(\xi, \tau^2, \alpha)$ and a $N(\xi, \tau^2)$ is given by*

$$\varphi^{SN}(z \mid \alpha) = 2\Phi_2\left[\begin{pmatrix} \Phi^{-1}(z) \\ 0 \end{pmatrix} \middle| \begin{pmatrix} 1 & -\delta \\ -\delta & 1 \end{pmatrix}\right], \tag{11}$$

where $\Phi_2(\cdot \mid \Omega)$ is the cumulative distribution function of the bivariate normal with mean vector zero and covariance matrix Ω, $\delta = \frac{|\alpha|}{\sqrt{1+\alpha^2}}$, whereas $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution.

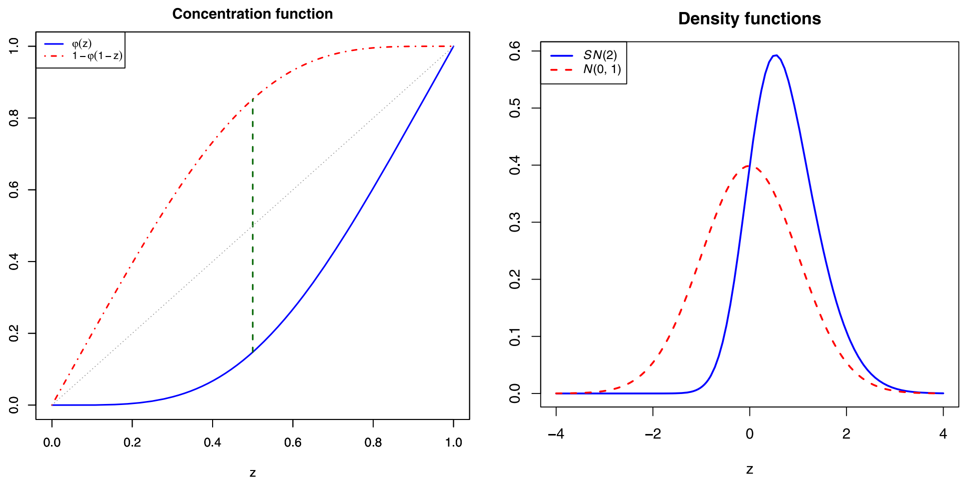

Figure 1 *The concentration function of* $SN(0, 1, 2)$ *with respect to* $N(0, 1)$ *and respective density functions.*

Figure 1 (left) shows the functions $\varphi^{SN}(z)$ and $1 - \varphi^{SN}(1 - z)$ for the distributions $SN(0, 1, 2)$ and $N(0, 1)$. We draw a dashed line crossing the curves $\varphi^{SN}(z)$ and $1 - \varphi^{SN}(1 - z)$ for $z = 0.5$ and it follows from Lemma 3.1 that $0.148 \leq P_{SN}(A) \leq 0.852$, for all $A \in \mathcal{B}$ such that $P_N(A) = 0.5$. Moreover, in the second graph on the Figure 1 (right) we depict the two probabilities densities functions.

The first graph (left) in the Figure 1 shows that a relatively small value of α, that is, 2, implies a quite large difference between the $SN(0, 1, 2)$ and the $N(0, 1)$ distributions, as testified by the vertical line (left plot) which shows how measurable sets of probability 0.5 under the normal distribution can have a probability ranging, approximately, from 0.15 to 0.85 under the skewed normal. The finding is also confirmed by the visual comparison of the densities (right plot) which are concentrating most of their probability on quite different sets. Figure 1 can be considered as a warning when introducing skewness: even small values of α can lead to significantly different concentrations.

Considering expression (4), we obtain that the L_1 distance between the $SN(0, 1, 2)$ and $N(0, 1)$ is 0.352, which corresponds to half distance between $\varphi^{SN}(0.5)$ and $1 - \varphi^{SN}(0.5)$. In Cifarelli and Regazzini (1987) it was proved that the Pietra's index, given by $\sup_{x \in [0,1]}(x - \varphi(x))$, equals the total variation distance between two probability measures, which is also half the L_1 distance. In the next, we will make an explicit computation, showing the relation between the L_1 distance and the concentration function under the normal context.

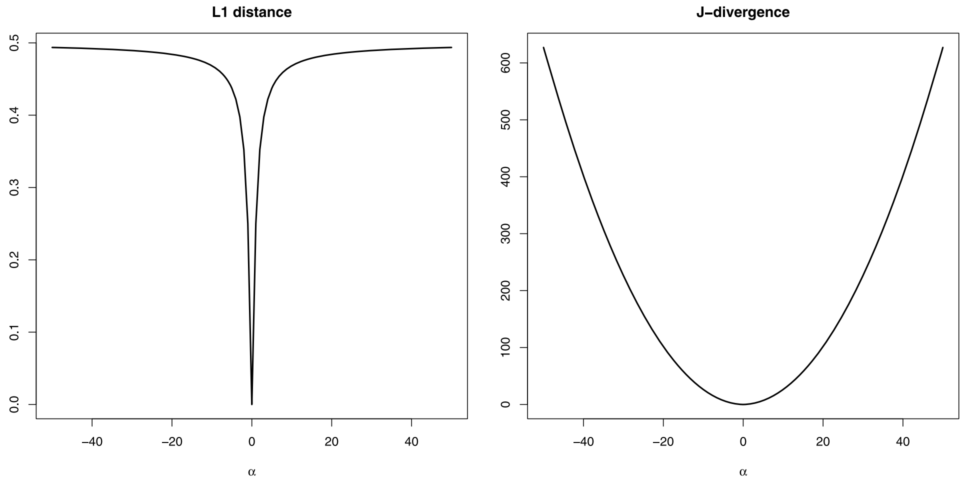

Figure 2 *L_1 and J distances between $SN(0, 1, \alpha)$ and $N(0, 1)$ in function of $\alpha \in [-50, 50]$.*

Proposition 3.2. *The L_1 distance between a $SN(\xi, \tau^2, \alpha)$ and a $N(\xi, \tau^2)$ is given by*

$$L_1(f_1, f_0) = \frac{1}{2} - \varphi^{SN}(0.5), \qquad \forall\alpha. \tag{12}$$

From Proposition 3.2, we see that the L_1 distance can be interpreted as a summary measure of the concentration function.

We also obtained the J-divergence between the $SN(0, 1, 2)$ and $N(0, 1)$ and its value is 1.534. This value is more difficult to interpret and we could not find a relation between it and the concentration function. In Figure 2, we show the behavior of the L_1 and J divergences between $SN(0, 1, \alpha)$ and $N(0, 1)$ for different values of α.

The findings of Figure 1, valid for $\alpha = 2$, can be extended to a larger set of values when looking at the left hand side of Figure 2. The L_1 distance, related to the concentration function by Proposition 3.2, has a dramatic increase for values of $|\alpha| < 10$, providing a warning, as above, on the consequences of the choice of α. The same warning cannot be achieved when looking at the J-divergence on the right-hand side of Figure 2. Note that the L_1 distance increases more smoothly than the J-divergence for $|\alpha| > 20$, when it gets practically constant. Therefore, the L_1 distance is more coherent than the J-divergence since it is known that the density function of a standard skew-normal distribution is minimally affected for $|\alpha| > 20$, when converging to a half normal distribution.

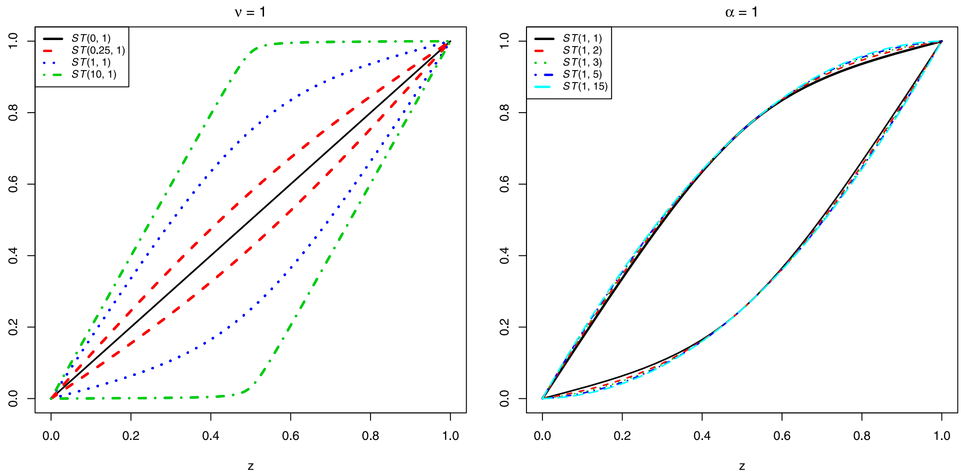

Figure 3 *Graphs of* $\varphi^{ST}(z \mid \alpha, \nu)$ *and* $1 - \varphi^{ST}(1 - z \mid \alpha, \nu)$ *of* $ST(0, 1, \alpha, \nu)$ *with respect to* $T(0, 1, \nu)$.

3.2 Skew-*t* case

We present a closed form expression for the concentration function between the skew-t and the Student-t distributions. Then, using an example, we discuss the influence of the shape parameter α and the degrees of freedom parameter ν.

Proposition 3.3. *The concentration function between* $ST(\xi, \tau^2, \alpha, \nu)$ *and* $T(\xi, \tau^2, \nu)$ *is given by*

$$\varphi^{ST}(z \mid \alpha, \nu) = 2T_2\left[\begin{pmatrix} cT_\nu^{-1}(z) \\ 0 \end{pmatrix} \middle| \begin{pmatrix} 1 & -\delta \\ -\delta & 1 \end{pmatrix}, \nu\right], \tag{13}$$

where $T_\nu^{-1}(\cdot)$ *is the quantile of the standard Student-t distribution with* ν *degrees of freedom,* $\delta = \frac{|\alpha|}{\sqrt{1+\alpha^2}}$ *and* $T_2(\cdot \mid \Omega, \nu)$ *is the distribution function of a bivariate t distribution with mean vector zero, covariance matrix* Ω *and* ν *degrees of freedom.*

Considering that the skew-t distribution can be expressed as a scale mixture of skew-normals, an alternative way to write the concentration function is given by

$$\varphi^{ST}(z \mid \alpha, \nu) = 2E_W\left\{\Phi_2\left[\begin{pmatrix} T_\nu^{-1}(z)W^{1/2} \\ 0 \end{pmatrix} \middle| \begin{pmatrix} 1 & -\delta \\ -\delta & 1 \end{pmatrix}\right]\right\},$$

where $W \sim Gamma(\nu/2, \nu/2)$ and $\delta = \frac{|\alpha|}{\sqrt{1+\alpha^2}}$.

Figure 3 presents $\varphi^{ST}(z \mid \alpha, \nu)$ and $1 - \varphi^{ST}(1 - z \mid \alpha, \nu)$ of the $ST(0, 1, \alpha, \nu)$ with respect to $T(0, 1, \nu)$ in several situations. The graphs were drawn using the expression (13) in Proposition 3.3. The first graph (left) shows these functions when $\nu = 1$ and α is equal to 0, 0.25, 1 and 10. Note that, for fixed $\nu = 1$, the distance between the curves increases with α, as expected. The same behavior is observed when we fix other values for ν. Changes of α (see left-hand side) induce a great variation in the probability of the sets when moving from the $T(0, 1, \nu)$ distribution to the $ST(0, 1, \alpha, \nu)$, as a consequence of Lemma 3.1. In particular, the bottom line, corresponding to $\alpha = 10$, is almost flat up to $z = 0.5$: it means there is a subset of measure 0.5 under the Student-t distribution which has negligible probability under the skew-t distribution. In the second graph (right), we fix $\alpha = 1$ and change the values of ν (1, 2, 3, 5 and 15). In this case, we cannot note many differences between curves indicating that there are no relevant changes in the distributions when considering different degrees of freedom. There are just tiny differences on subsets of small probability, as shown by the nonoverlapping curves for values of z between 0 and 0.3. It is important to remember that, for α fixed, the L_1 distance is the same for all ν.

4 Application in Bayesian robustness

4.1 Prior robustness

Robust Bayesian analysis is concerned with the impact of different specifications for the prior distribution or the likelihood function on the posterior distribution. If, for instance, a specific posterior inference is not much affected by these choices, then we will say that this inference is robust. In general, robustness analysis assumes the likelihood $f(x \mid \theta)$ is fixed and considers a class Γ of prior distributions to deal with the uncertainty in specifying a unique distribution. Robustness of a given statistical procedure is measured by the size of the range of posterior measures or quantities of interest (e.g., mean) obtained when the prior distribution varies over Γ.

A possible way to construct Γ is to consider an elicited prior as baseline prior and contaminate it using a class of functions. In the literature, the most common and well-studied class of priors is the ε-contaminated class when the perturbation of the baseline prior is additive (see O'Hagan (1994)). Considering multiplicative ways of perturbation, Godoi and Branco (2014) studied a multiplicative class of contaminated priors for the location parameter, under a normal likelihood. These authors analyzed the behavior of the posterior mean and posterior variance under changes in the prior distribution. In this section, we explore the use of the concentration function for a robustness study under multiplicative contamination.

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with location parameter θ and scale parameter σ^2. Usually, the prior distribution specified for θ is the normal, that is conjugate with respect to the statistical model considered. The idea here is to propose a class of prior distributions that contains the normal distribution, but allows the inclusion of the assumption of asymmetry. One possibility is to use the skew-normal class of distributions, given by

$$\Gamma = \left\{ f_\alpha(\theta) = \frac{2}{\tau}\phi\left(\frac{\theta - \xi}{\tau}\right)\Phi\left(\alpha\frac{\theta - \xi}{\tau}\right) : \alpha \in \mathbb{R} \right\}. \tag{14}$$

Note that the location and scale parameters are considered fixed and the class Γ contains an infinite number of density functions arising from the variation of the skewness parameter α.

Considering σ^2 known, Godoi and Branco (2014) show that, under a SN prior distribution for θ, the posterior distribution is in a more general class of skew distribution known as SUN (Skew Unified Normal). A continuous p-dimensional random vector $\mathbf{X}$ has a multivariate SUN distribution, denoted by $\mathbf{X} \sim SUN_{p,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \boldsymbol{\Delta}, \boldsymbol{\xi})$, if its density function at $\mathbf{x} \in \mathbb{R}^p$ is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \phi_p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \frac{\Phi_q(\boldsymbol{\xi} + \boldsymbol{\Delta}^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}); \mathbf{0}, \boldsymbol{\Gamma} - \boldsymbol{\Delta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Delta})}{\Phi_q(\boldsymbol{\xi}; \mathbf{0}, \boldsymbol{\Gamma})}, \tag{15}$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\boldsymbol{\xi} \in \mathbb{R}^q$ are location vectors, $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, $\boldsymbol{\Gamma} \in \mathbb{R}^{q\times q}$ and $\boldsymbol{\Delta} \in \mathbb{R}^{p\times q}$ are dispersion matrices. For details about SUN, see Arellano-Valle and Azzalini (2006).

The $SN(\xi, \tau^2, \alpha)$ is recognized as a particular case of a SUN distribution denoted by $SUN_{1,1}(\xi, \tau^2, \alpha\tau^2, 0, \tau^2(1 + \alpha^2))$. Therefore, to compare the posterior distributions under normal prior and SN prior it is necessary to have a distance measure between a normal and a SUN distributions, in a similar fashion as in Section 2. In the next, we will present the results about conjugation under SUN context, given by Godoi and Branco (2014), and the concentration function between posteriors. The proof follows similar steps of the Proposition 3.1, so we will not present it here.

Proposition 4.1. *Let* $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ *be a random sample of size n from* $X \sim N(\theta, \sigma^2)$, *with σ^2 known. If* $\theta \sim SUN_{1,1}(\xi, \tau^2, \alpha\tau^2, 0, \tau^2(1 + \alpha^2))$, *then the posterior distribution is given by*

$$\theta \mid \mathbf{x} \sim SUN_{1,1}\big(\zeta, \omega^2, \alpha\omega^2, -\alpha(\xi - \zeta), \tau^2 + \alpha^2\omega^2\big), \tag{16}$$

where

$$\zeta = \frac{n\bar{x}\tau^2 + \xi\sigma^2}{n\tau^2 + \sigma^2} \quad \textit{and} \quad \omega^2 = \frac{\sigma^2\tau^2}{n\tau^2 + \sigma^2}. \tag{17}$$

Proposition 4.2. *The concentration function between the densities* $N(\zeta, \omega^2)$ *and* $SUN_{1,1}(\zeta, \omega^2, \alpha\omega^2, -\alpha(\xi - \zeta), \tau^2 + \alpha^2\omega^2)$ *is given by*

$$\varphi^{SUN}(z) = \frac{\Phi_2\left[\begin{pmatrix} \frac{\alpha}{\sqrt{n\tau^2+\sigma^2(1+\alpha^2)}} \frac{\tau n(\bar{x}-\xi)}{\sqrt{n\tau^2+\sigma^2}} \\ \Phi^{-1}(z) \end{pmatrix} \middle| \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}\right]}{\Phi\left[\frac{\alpha}{\sqrt{n\tau^2+\sigma^2(1+\alpha^2)}} \frac{\tau n(\bar{x}-\xi)}{\sqrt{n\tau^2+\sigma^2}}\right]}, \tag{18}$$

where $\Phi_2(\cdot \mid \Omega)$ is the cumulative distribution function of the bivariate normal with mean vector zero and covariance matrix Ω, $\rho = \frac{|\alpha|\sigma}{\sqrt{n\tau^2+\sigma^2(1+\alpha^2)}}$ *and $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal distribution.*

In the next proposition, we present an expression for the L_1 distance between a SUN and the normal distribution, which generalizes the result given in Section 2.

Proposition 4.3. *The L_1 distance between the densities $N(\zeta,\omega^2)$ and $SUN_{1,1}(\zeta, \omega^2, \alpha\omega^2, -\alpha(\xi-\zeta), \tau^2+\alpha^2\omega^2)$ is given by*

$$L_1(f_{SUN}, f_N) = \begin{cases} \frac{1}{2} - \Phi[-K(\alpha)] + g(\alpha), & \text{if } \alpha > 0, \\ -\frac{1}{2} + \Phi[-K(\alpha)] - g(\alpha), & \text{if } \alpha < 0, \end{cases} \tag{19}$$

where

$$g(\alpha) = \frac{1}{2\Phi[\alpha Q(\alpha)]}\left\{-\Phi_2\left[\begin{pmatrix} K(\alpha) \\ \alpha Q(\alpha)\end{pmatrix}; -\rho\right] + \Phi_2\left[\begin{pmatrix} -K(\alpha) \\ \alpha Q(\alpha)\end{pmatrix}; \rho\right]\right\}, \tag{20}$$

$$Q(\alpha) = \frac{(\zeta-\xi)}{\sqrt{\tau^2+\alpha^2\omega^2}}, \qquad K(\alpha) = \frac{1}{\sqrt{\omega^2}}\big(Q(\alpha)\tau + \xi - \zeta\big), \tag{21}$$

ζ and ω^2 are as specified in (17) and $\Phi_2(\cdot \mid \rho)$ is the distribution function of the bivariate normal with mean vector zero and $\rho = \frac{\alpha\sigma}{\sqrt{n\tau^2+\sigma^2(1+\alpha^2)}}$ being the correlation coefficient such that the covariance matrix is given by $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$.

4.2 A numerical case study

We develop here a robustness analysis under the skew-normal class of distributions as presented in (14) when a random sample of size n of $X \sim N(\theta, \sigma^2)$, with σ^2 known, is considered. For this, let $\xi = 0$, $\tau^2 = 1$, $\sigma^2 = 1$ and $n = 20$.

Note that this class of prior distributions (Γ) contains an infinite number of standard skew-normal distributions ($SN(0, 1, \alpha)$) arising from the variation of $\alpha \in \mathbb{R}$ and includes the standard normal distribution specified when $\alpha = 0$. In this context, our goal is to observe the influence of this class on the posterior distributions, when α varies.

According to Proposition 4.1, the posterior class of distributions is still a SUN, determined by the variation of α. The posterior distribution associated to the baseline prior is obtained when $\alpha = 0$.

Fortini and Ruggeri (2000) suggest to obtain the class of the concentration functions between the class of posterior distributions and the baseline posterior and analyze the robustness considering the pointwise infimum of the concentration functions in this class. The resulting function will be denoted by $\hat{\varphi}(z)$ for any $z \in [0, 1]$. In our context, for each z fixed in a grid constructed through $z_j = 0.02 \times j$, for $j = 1, \dots, 50$, we consider a grid of values for $\alpha \in [-25, 25]$, namely $\alpha_i = 0.1 \times i$, with $\{i \in \mathbb{Z} \mid i = -250, \dots, 250\}$; using the expression in (18) with known $\bar{x}$, we obtain

$$\hat{\varphi}(z_j) = \inf_{\alpha} \varphi_\alpha(z_j) \approx \min\{\varphi_{\alpha_1}(z_j), \dots, \varphi_{\alpha_{501}}(z_j)\}, \qquad j = 1, \dots, 50. \tag{22}$$

If we suppose, for example, that $\bar{x} = 0.25$ and $\alpha \in [-25, 25]$, we obtain a graph identical to the one presented in Figure 4, bottom row, left-hand side column. The analysis should be performed based on it, showing an evident lack of robustness, as discussed earlier.

In a similar context, Godoi and Branco (2014) showed that the robustness of the posterior mean of θ with respect to the variation of the prior distributions in Γ occurs when both α and $\bar{x} - \xi$ vary in intervals in $\mathbb{R}^+$ (or are, similarly, both in $\mathbb{R}^-$). With this in mind, we split Γ in two subclasses depending on the sign of α: one for $\alpha \geq 0$ and the other for $\alpha \leq 0$. The value $\alpha = 0$ is in both classes to ensure that the baseline prior is inside each of them. In each subclass, the infimum of the concentration function is computed as described previously and the behavior of the posterior distributions is studied to check if the results presented in Godoi and Branco (2014) occur even when using the approach based on the concentration function.

In Figure 4, we present the graphs of $\hat{\varphi}(z)$ and $1 - \hat{\varphi}(1 - z)$ for $\bar{x} = -0.25$, 0 and 0.25. We note that, when α has the same sign of the sample mean, then $\hat{\varphi}(z)$ and $1 - \hat{\varphi}(1 - z)$ are close. This proximity is more evident for large absolute values of $\bar{x}$. In this case, the class of posterior distributions is robust with respect to skew-normal contamination class and there is no significant difference in performing a Bayesian analysis using the prior baseline ($N(0, 1)$) or any other skew prior in Γ. On the other hand, the robustness does not occur when α and $\bar{x}$ have different signs. The lack of robustness in this context can be explained by the conflict between the information suggested by the class of prior distribution (skewness on one side) and the sample mean (obtained on the other side).

It could be possible to consider classes with α in subsets of $\mathbb{R}^+$ (or $\mathbb{R}^-$) and consider the infimum of the concentration function and compute, say, the Pietra's index, given here by $\sup_{x \in [0,1]}(x - \widehat{\varphi(x)})$, to check about robustness. Different classes could give different values of the index, with smaller values denoting improvement in robustness. An issue, common to all robust Bayesian analysis, is about the threshold about the entertained measure (here Pietra's index) which determines if there is robustness or not. This is a critical question whose answer depends on the problem at hand and who is involved: it is a subjective choice. We just provide a tool which can be applied in many practical cases where departure from symmetry is deemed necessary; the consequences of its use (lack or not of robustness) are to be evaluated case by case by the decision makers.

We could also consider classes with both negative and positive values of α. It is evident from the previous discussion and Figure 4 that the infimum of the concentration function will be obtained for those value of α which have the opposite sign of $\bar{x}$.

Similar conclusions are obtained when we consider the L_1 distance. Figure 5 shows the behavior of the L_1 distance as a function of α. Moreover, for $\alpha > 0$ we obtain numerically that the $\max_\alpha L_1(\alpha)$ corresponds to 0.861, 0.499 and 0.137 when $\bar{x} = -0.25$, 0 and 0.25, respectively. For $\alpha < 0$, the $\max_\alpha L_1(\alpha)$ corresponds to 0.137, 0.499 and 0.861 when $\bar{x} = -0.25$, 0 and 0.25, respectively.

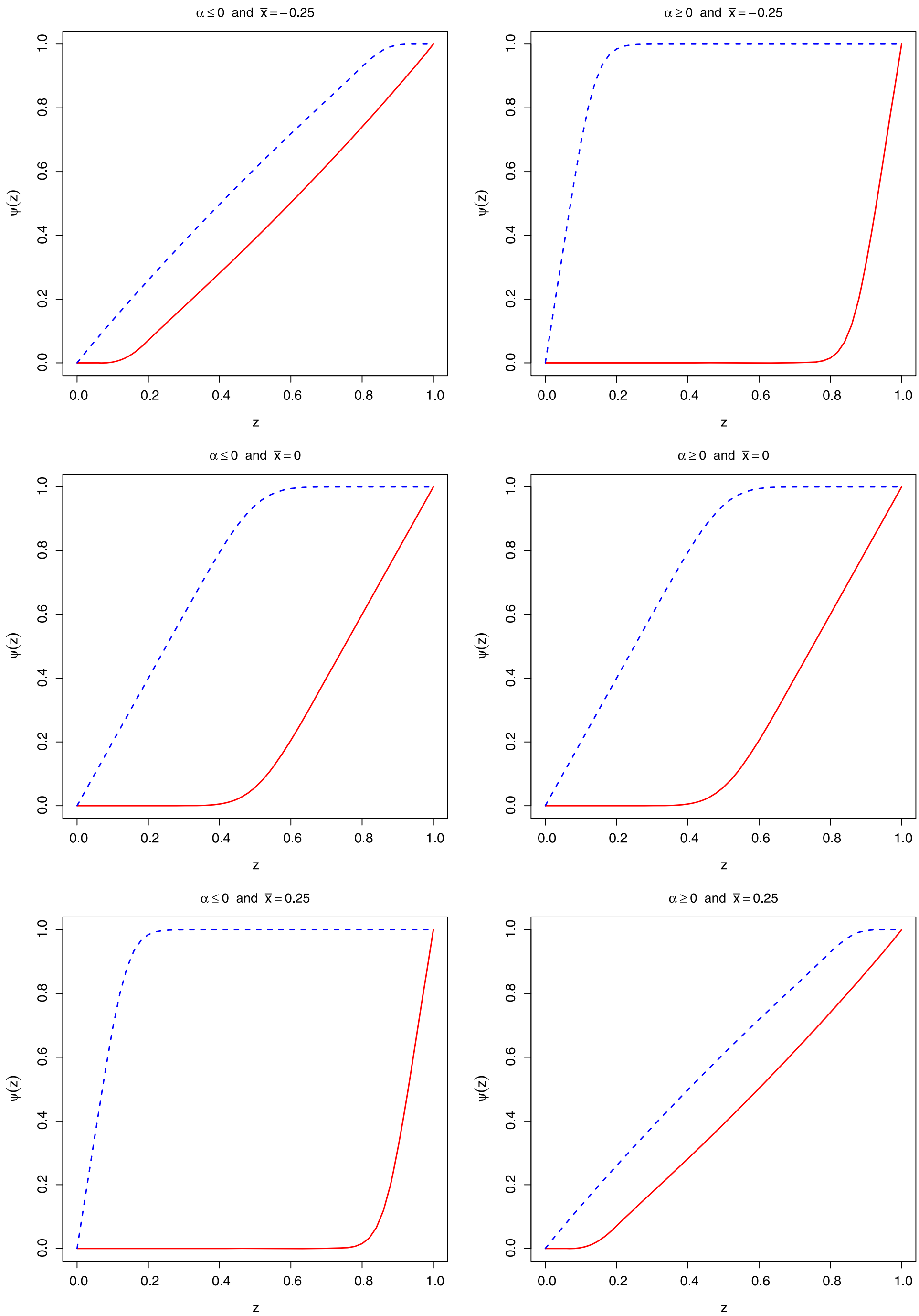

Figure 4 *Concentration function for the class of posterior distributions, with fixed* $\bar{x} = -0.25$, 0 *and* 0.25. *The lines* (–) *and* (- -) *correspond, respectively, to* $\hat{\varphi}(z)$ *and* $1 - \hat{\varphi}(1 - z)$.

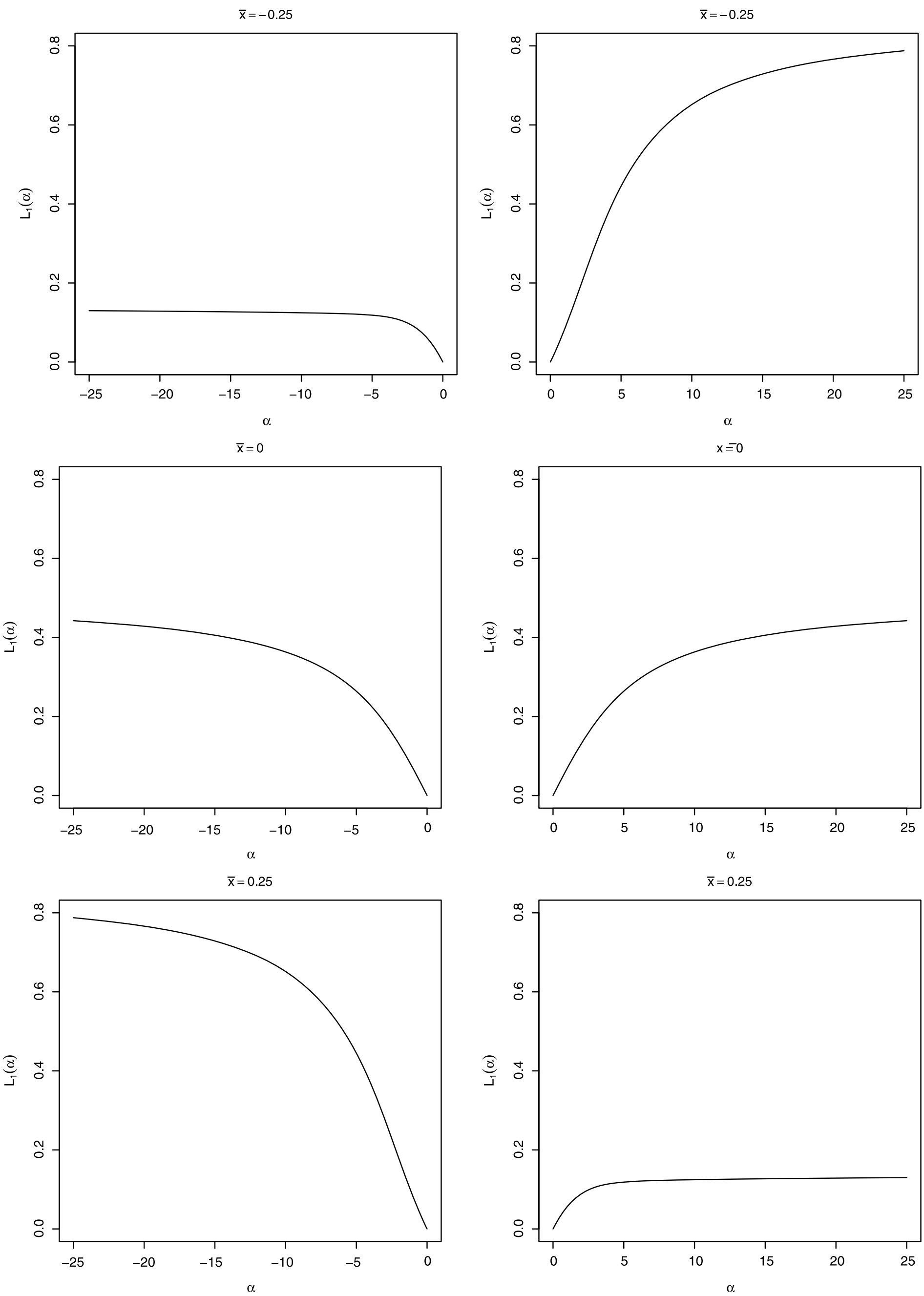

Figure 5 *L_1 distances for the class of posterior distributions, with fixed* $\bar{x} = -0.25$, 0 *and* 0.25.

5 Final comments

In the paper, we have investigated the use of different measures to compare skew-symmetric distributions with respect to a baseline symmetric distribution. In particular, we have proposed the use of the concentration function and discussed findings relate also to L_1 distance and J-divergence. We have been able to provide novel computations, mostly for the Student$-t$ distribution. Once the measures used to compare distributions are available, then their properties could provide different insights on the departure from the symmetric baseline. We have also presented an expression that relates the concentration function and the L_1 distance in the normal case. In this case, we can interpret the L_1 distance as a summary measure of a concentration function. A relevant application of the approach is Bayesian robustness where the class of distributions could be considered a neighborhood, not necessarily in topological sense, of a baseline symmetric prior. A more extensive study on the properties of the measures and the implications in Bayesian robustness is currently being pursued.

Appendix: Proofs

Proof of Proposition 2.3. The J-divergence (see Contreras-Reyes and Arellano-Valle, 2012) between two densities f and g is given by

$$J(f,g) = CH(f,g) + CH(g,f) - H(f) - H(g), \tag{23}$$

where $CH(f,g)$ is the cross-entropy and it is given by

$$CH(f,g) = -\int f(x)\log\big(g(x)\big)\,dx \tag{24}$$

and $H(f) = -\int f(x)\log(f(x))\,dx$ and $H(g) = -\int g(x)\log(g(x))\,dx$ are the usual measures of entropy.

Now we obtain the cross-entropy between f_0 and f_1, as given in (1).

Consider Z and S random variables with density function f_0 and f_1, respectively. Then,

$$\begin{aligned} CH(f_0,f_1) &= -\int f_0(z)\log\big(f_1(z)\big)\,dz \\ &= -\int f_0(z)\big[\log\big(f_0(z)2G\big(w(z,\alpha)\big)\big)\big]\,dz \\ &= H(f_0) - E_Z\big[\log\big(2G\big(w(Z,\alpha)\big)\big)\big]. \end{aligned} \tag{25}$$

On the other hand, the cross-entropy between f_1 and f_0 is

$$CH(f_1,f_0) = -\int f_0(s)2G\big(w(s,\alpha)\big)\log\big(f_0(s)\big)\,ds = -E_S\big[\log\big(f_0(S)\big)\big].$$

Therefore,

$$\begin{aligned} J(f_0, f_1) &= -E_Z[\log(2G(w(Z,\alpha)))] - E_S[\log(f_0(S))] + E_S[\log(f_1(S))] \\ &= E_S\left[\log\left(\frac{f_0(S)2G(w(S,\alpha))}{f_0(S)}\right)\right] - E_Z[\log(2G(w(Z,\alpha)))] \qquad (26) \\ &= E_S[\log(G(w(S,\alpha)))] - E_Z[\log(G(w(Z,\alpha)))]. \end{aligned}$$

□

Proof of Proposition 3.1. According to the discussion made after Lemma 3.1, without loss of generality, this proof can be made for $\xi = 0$ and $\tau^2 = 1$. Then,

$$h(\theta) = \frac{f(\theta)}{f_0(\theta)} = 2\Phi(\alpha\theta). \tag{27}$$

For any $z \in [0, 1]$, we compute the concentration function by finding the value y such that $z = P_{f_0}(\{\theta \in \Theta : h(\theta) \leq y\})$. For $\alpha > 0$,

$$\begin{aligned} 0 < h(\theta) \leq y \quad &\Longleftrightarrow \quad 0 < 2\Phi(\alpha\theta) \leq y \\ &\Longleftrightarrow \quad -\infty < \theta \leq \frac{\Phi^{-1}(y/2)}{\alpha}. \end{aligned} \tag{28}$$

Considering $q = \frac{\Phi^{-1}(y/2)}{\alpha}$ then $z = P_{f_0}(\{\theta \in \Theta : h(\theta) \leq y\}) = P_{f_0}(\{\theta \in \Theta : -\infty < \theta \leq q\}) = \Phi(q)$ and thus $z = \Phi(\frac{\Phi^{-1}(y/2)}{\alpha})$.

Therefore, the value y for which $z = \Phi(\frac{\Phi^{-1}(y/2)}{\alpha})$ is given by $2\Phi[\alpha\Phi^{-1}(z)]$ and $\varphi^{SN}(z \mid \alpha) = P_f(\{\theta \in \Theta : h(\theta) \leq y\}) = P_f(\{\theta \in \Theta : h(\theta) \leq 2\Phi[\alpha\Phi^{-1}(z)]\})$. Noting that $0 < h(\theta) \leq 2\Phi[\alpha\Phi^{-1}(z)] \Longleftrightarrow -\infty < \theta \leq \Phi^{-1}(z)$, we have, for $\alpha > 0$,

$$\varphi^{SN}(z \mid \alpha) = \int_{-\infty}^{\Phi^{-1}(z)} 2\phi(\theta)\Phi(\alpha\theta)\,d\theta. \tag{29}$$

According to Rodríguez (2005), we can rewrite $\varphi^{SN}(z \mid \alpha)$ as presented on (11). For $\alpha < 0$, the proof is similar. □

Proof of Proposition 3.2. For $\alpha > 0$, we consider $\varphi^{SN}(0.5 \mid \alpha)$ as expressed in (29) and using the identity 1010.3 presented in Owen (1980), we rewrite $\varphi^{SN}(0.5 \mid \alpha)$ as $\frac{1}{\pi}\arctan\frac{1}{\alpha}$. Then,

$$\begin{aligned} \frac{1}{2} - \varphi^{SN}(0.5 \mid \alpha) &= \frac{1}{2} - \frac{1}{\pi}\arctan\frac{1}{\alpha} \\ &= \frac{1}{2} - \frac{1}{\pi}\arccos\frac{\alpha}{\sqrt{1+\alpha^2}} \\ &= \frac{1}{2} - \frac{1}{\pi}\left[\frac{\pi}{2} - \arcsin\frac{\alpha}{\sqrt{1+\alpha^2}}\right] = L_1(f_{SN}, f_N). \end{aligned} \tag{30}$$

Similar steps can be done for $\alpha < 0$ knowing that in this case $\varphi^{SN}(0.5 \mid \alpha)$ can be written as $\frac{1}{\pi}\arccos(-\frac{\alpha}{\sqrt{1+\alpha^2}})$ according to the identity 1010.4 in Owen (1980). □

Proof of Proposition 3.3. According to the discussion made after Lemma 3.1, without loss of generality, we consider in this proof $\xi = 0$ and $\tau^2 = 1$. Then,

$$h(\theta) = \frac{f(\theta)}{f_0(\theta)} = 2T\left(\alpha\theta\sqrt{\frac{1+\nu}{\nu+\theta^2}}; \nu+1\right). \tag{31}$$

For any $z \in [0, 1]$, we compute the concentration function by finding the value y such that $z = P_{f_0}(\{\theta \in \Theta : h(\theta) \le y\})$.

Note that $0 < y < 2$ and, for $\alpha > 0$, the function $h(\theta)$ is strictly increasing, for all θ. Then $0 < h(\theta) \le y \iff -\infty < \theta < h^{-1}(y)$, where $h^{-1}(y)$ corresponds to the value of θ such that $2T(\alpha\theta\sqrt{\frac{1+\nu}{\nu+\theta^2}}; \nu+1) = y$. After some calculations, we obtain that $\theta < \sqrt{\frac{K^2\nu}{1-K^2}}$, if $1 \le y < 2$, and $\theta < -\sqrt{\frac{K^2\nu}{1-K^2}}$, if $0 < y < 1$, where $K = \frac{T_{\nu+1}^{-1}(y/2)}{\alpha\sqrt{\nu+1}}$.

For $1 \le y < 2$, we have that $z \ge 0.5$ and that $y = 2T(\frac{\alpha\sqrt{\nu+1}T_\nu^{-1}(z)}{\sqrt{\nu+[T_\nu^{-1}(z)]^2}}; \nu+1)$, since $z = P_{f_0}(\{\theta \in \Theta : h(\theta) \le y\}) = \int_{-\infty}^{\sqrt{\frac{K^2\nu}{1-K^2}}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}\frac{1}{\sqrt{(\nu\pi)}}(1+\frac{t^2}{\nu})^{(\frac{\nu+1}{2})}\,dt$, where $K = \frac{T_{\nu+1}^{-1}(y/2)}{\alpha\sqrt{\nu+1}}$.

Finally, we get that

$$\begin{aligned}\varphi^{ST}(z \mid \alpha, \nu) &= P_f\big(\{\theta \in \Theta : h(\theta) \le y\}\big) \\ &= P_f\left(\left\{\theta \in \Theta : h(\theta) \le 2T\left(\frac{\alpha\sqrt{\nu+1}T_\nu^{-1}(z)}{\sqrt{\nu+[T_\nu^{-1}(z)]^2}}; \nu+1\right)\right\}\right)\end{aligned} \tag{32}$$

and noting that $0 < h(\theta) \le 2T(\frac{\alpha\sqrt{\nu+1}T_\nu^{-1}(z)}{\sqrt{\nu+[T_\nu^{-1}(z)]^2}}; \nu+1) \iff -\infty < \theta \le T_\nu^{-1}(z)$, we have, for $\alpha > 0$ and $z \ge 0.5$

$$\varphi^{ST}(z \mid \alpha, \nu) = \int_{-\infty}^{T_\nu^{-1}(z)} 2t(\theta; \nu)T\left(\alpha\theta\sqrt{\frac{\nu+1}{\nu+\theta^2}}; \nu+1\right)d\theta. \tag{33}$$

If we consider $0 < y < 1$, we have that if $z < 0.5$ we also obtain $y = 2T(\frac{\alpha\sqrt{\nu+1}T_\nu^{-1}(z)}{\sqrt{\nu+[T_\nu^{-1}(z)]^2}}; \nu+1)$ and, following similar steps as before, we obtain the concentration function as presented in (33).

According to Jamalizadeh, Mehrali and Balakrishnan (2009), the expression in (33) can be rewritten as presented in (13). For $\alpha < 0$, the proof is similar. □

Proof of Proposition 4.3. The densities associated to the normal and SUN distributions are given, respectively, by

$$\begin{aligned} f_N(\theta) &= \phi(\theta;\zeta,\omega^2), \\ f_{SUN}(\theta) &= A^{-1}\phi(\theta;\zeta,\omega^2)\Phi(\alpha\theta;\alpha\xi,\tau^2), \end{aligned} \tag{34}$$

with $A = \Phi(\alpha\zeta;\alpha\xi,\tau^2+\alpha^2)$. Then

$$\begin{aligned} &L_1(f_N, f_{SUN}) \\ &\quad = \frac{1}{2}\int \left| f_{SUN}(\theta) - f_N(\theta)\right| d\theta \\ &\quad = \frac{1}{2A}\int \underbrace{\left|\Phi\left[\alpha\left(\frac{\theta-\xi}{\tau}\right)\right] - \Phi\left[\alpha\left(\frac{\zeta-\xi}{\sqrt{\tau^2+\alpha^2\omega^2}}\right)\right]\right|}_{|B|}\phi(\theta;\zeta,\omega^2)\,d\theta. \end{aligned} \tag{35}$$

For $\alpha > 0$,

$$|B| = \begin{cases} \Phi\left[\alpha\left(\frac{\theta-\xi}{\tau}\right)\right] - \Phi\left[\alpha\left(\frac{\zeta-\xi}{\sqrt{\tau^2+\alpha^2\omega^2}}\right)\right], \\ \qquad \text{if } \theta > \tau\left(\frac{\zeta-\xi}{\sqrt{\tau^2+\alpha^2\omega^2}}\right)+\xi, \\ -\Phi\left[\alpha\left(\frac{\theta-\xi}{\tau}\right)\right] + \Phi\left[\alpha\left(\frac{\zeta-\xi}{\sqrt{\tau^2+\alpha^2\omega^2}}\right)\right], \\ \qquad \text{if } \theta < \tau\left(\frac{\zeta-\xi}{\sqrt{\tau^2+\alpha^2\omega^2}}\right)+\xi \end{cases}$$

and then

$$\begin{aligned} &L_1(f_N, f_{SUN}) \\ &\quad = \frac{1}{2A}\left\{\int_{-\infty}^{\tau(\frac{\zeta-\xi}{\sqrt{\tau^2+\alpha^2\omega^2}})+\xi}\left(\Phi\left[\alpha\left(\frac{\zeta-\xi}{\sqrt{\tau^2+\alpha^2\omega^2}}\right)\right] - \Phi\left[\alpha\left(\frac{\theta-\xi}{\tau}\right)\right]\right)\right. \\ &\qquad \times \phi(\theta;\zeta,\omega^2)\,d\theta + \int_{\tau(\frac{\zeta-\xi}{\sqrt{\tau^2+\alpha^2\omega^2}})+\xi}^{\infty}\left(\Phi\left[\alpha\left(\frac{\theta-\xi}{\tau}\right)\right]\right. \\ &\qquad \left.\left. - \Phi\left[\alpha\left(\frac{\zeta-\xi}{\sqrt{\tau^2+\alpha^2\omega^2}}\right)\right]\right)\phi(\theta;\zeta,\omega^2)\,d\theta\right\}. \end{aligned} \tag{36}$$

After some calculations in (36), we obtain the expression in (19), for $\alpha > 0$. The proof for $\alpha < 0$ is similar and thus it is omitted. □

Acknowledgments

This research was partially supported by CNPq joint with CNR-Italy grant 490233/2011-2. The second author also thanks FAPESP grant 2012/21788-2.

References

Arellano-Valle, R. B. and Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics* **33**, 561–574. MR2298065

Arellano-Valle, R. B., Contreras-Reyes, J. E. and Genton, M. G. (2013). Shannon entropy and mutual information for multivariate skew-elliptical distributions. *Scandinavian Journal of Statistics* **40**, 42–62. MR3024031

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**, 171–178. MR0808153

Azzalini, A. (2005). The skew-normal distribution and related multivariate families (with discussion). *Scandinavian Journal of Statistics* **32**, 159–188. MR2188669

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B* **65**, 367–389. MR1983753

Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* **79**, 99–113. MR1867257

Cifarelli, D. M. and Regazzini, E. (1987). On a general definition of concentration function. *Sankhya: The Indian Journal of Statistics* **49**, 307–319. MR1056041

Contreras-Reyes, J. E. (2014). Asymptotic form of the Kullback–Leibler divergence for multivariate asymmetric heavy-tailed distributions. *Physica A* **395**, 200–208. MR3133652

Contreras-Reyes, J. E. and Arellano-Valle, R. B. (2012). Kullback–Leibler divergence measure for multivariate skew-normal distributions. *Entropy* **14**, 1606–1626. MR2979804

Fortini, S. and Ruggeri, F. (2000). On the use of the concentration function in Bayesian robustness. In *Robust Bayesian Analysis* (D. Rios Insua and F. Ruggeri, eds.) 109–126. New York: Springer. MR1795212

Genton, M. G., ed. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton: Chapman & Hall/CRC. MR2156754

Godoi, L. G. and Branco, M. D. (2014). Bayesian robustness under a skew-normal class of prior distribution. *International Journal of Approximate Reasoning* **55**, 1235–1251. MR3197805

Gomez, H. W., Venegas, O. and Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics* **18**, 395–407. MR2370731

Jamalizadeh, A., Mehrali, Y. and Balakrishnan, N. (2009). Recurrence relations for bivariate and extended skew-t distributions and an application to order statistics from bivariate t. *Computational Statistics & Data Analysis* **53**, 4018–4027. MR2744302

Kotz, S. and Nadarajah, S. (2004). *Multivariate t-Distributions and Their Applications*. Cambridge: Cambridge University Press. MR2038227

Nadarajah, S. and Kotz, S. (2003). Skewed distributions generated by the normal kernel. *Statistics & Probability Letters* **65**, 269–277. MR2018040

O'Hagan, A. (1994). *Kendall's Advanced Theory of Statistics, Vol. 2B. Bayesian Inference*. London: Edward Arnold. MR1285356

Owen, D. B. (1980). A table of normal integrals. *Communications in Statistics—Simulation and Computation* **9**, 389–419. MR0570844

Rodríguez, C. L. B. (2005). Inferência bayesiana no modelo normal assimétrico. Master's thesis, IME-USP, São Paulo, Brasil.

Scarsini, M. (1990). An ordering of dependence. In *Topics in Statistical Dependence* (H. W. Block, A. R. Sampson and T. H. Savits, eds.) 403–414. Hayward, CA: Institute of Mathematical Statistics. MR1193994

Vidal, I., Iglesias, P., Branco, M. D. and Arellano-Valle, R. B. (2006). Bayesian sensitivity analysis and model comparison for skew elliptical models. *Journal of Statistical Planning and Inference* **136**, 3435–3457. MR2256274

L. G. Godoi
UFES/CCE 29075-910
Vitória, Espírito Santo
Brazil
E-mail: luciana.godoi@ufes.br

M. D. Branco
USP/IME 05508-090
São Paulo, São Paulo
Brazil
E-mail: mbranco@ime.usp.br

F. Ruggeri
CNR/IMATI 20133
Milano
Italy
E-mail: fabrizio@mi.imati.cnr.it