**Concentration function for skew-symmetric models with application to Bayesian robustness**

Luciana Graziela de Godoi
Universidade Federal do Espirito Santo, Vitoria, Brasil - lgraziela25@gmail.com

Marcia D'Elia Branco*
Universidade de Sao Paulo, Sao Paulo, Brasil - mbranco@ime.usp.br

Fabrizio Ruggeri
IMATI-CNR, Milano, Italia - fabrizio@mi.imati.cnr.it

**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, $J-$distance 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.
**Keywords**: Bayesian robustness; skew-symmetric models; $L_1$ distance; concentration function.

## 1. Introduction

An approach to model departure from symmetry 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, as given in Azzalini and Capitanio (2003). The probability density function, up to location and scale parameters, is of the form

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

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 $w(-y) = -w(y)$ for all $y \in \mathbb{R}$. We also add a parameter $\alpha \in \mathbb{R}$ in $w$ which controls the shape of the distribution; moreover, we consider that when $\alpha = 0$ then $w(y) = 0, \forall y$. Therefore, when $\alpha = 0$ the symmetric base density $f_0(\cdot)$ is retrieved. 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 $\nu$ 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). 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. To answer this we study some measures of distance between the skewed distribution and their symmetric baseline. 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 close form for the Student-$t$ case. On the other hand, Contreras-Reyes and Arellano-Valle

(2012) obtain the Kullback-Leibler (KL) divergence and the $J$-distance 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 considered 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 neighborhood of baseline prior distribution and to measure ranges spanned, as the prior varies in a class, by the probability of measurable subset 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 in the case the two measures concentrates all the 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 $f_0$. We review some measures already given in the literature and present new results under some special cases.

## 2. $L_1$ and $J$ distances

**Definition 2.1.** *The $L_1$ distance between two densities functions $f$ and $g$ is given by*

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

*where $P_f$ and $P_g$ are measures of probability, associated with the density functions $f$ and $g$, in the same measurable space $(\mathbb{R}, \mathcal{B})$ and $\mathcal{B}$ is the Borel $\sigma-$ algebra.*
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 $f(x) = g(x)$ for all $x$ and $L_1(f,g) = 1$ indicates that the supports of the two densities are disjoint, indicating maximal discrepancy.
For the skew-normal case Vidal et al. (2006) obtained

$$L_1(f_0, f_1) = \frac{1}{\pi}\arcsin\left(\frac{|\alpha|}{\sqrt{1+\alpha^2}}\right) = \frac{1}{2} - \frac{1}{\pi}\arccos\left(\frac{|\alpha|}{\sqrt{1+\alpha^2}}\right). \tag{3}$$

Moreover, it is not difficult to see that this distance does not depend on the location and scale parameters. Then the result given in (3) is also true when comparing a $SN(\xi, \tau^2, \alpha)$ with a $N(\xi, \tau^2)$.
For the case where $Z$ has a skew-t density we consider the fact that this distribution is a scale mixture of the skew-normal, i.e, $Z = V^{-1/2}X$, where $X \sim SN(0, 1, \alpha)$ and $V$ is random variable independent of $X$ with Gamma distribution with both parameters equal to $\nu/2$; then, following Proposition 7 from Vidal et al.(2006), the $L_1$ distance is also given by (3). 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 in the $t$ distribution. In fact, there are several definition of skew-t in the literature and it is important to be aware which one is being considered.

**Definition 2.2.** *The Kullback-Liebler (KL) divergence between two densities $f$ and $g$ is given by*

$$KL(f,g) = \int f(x) log\left\{\frac{f(x)}{g(x)}\right\} dx. \tag{4}$$

*This is not a measure of distance because $KL(f,g) \neq KL(g,f)$. The usual way to make a symmetric measure based in KL divergence is consider $J(f,g) = KL(f,g) + KL(g,f)$: this is known as $J$-distance.*
Contreras-Reyes and Arellano-Valle (2012) present the $J$-distance between two multivariate skew-normal distributions. For the unidimensional case and using their results for our interest, we obtain the following Lemma.

**Lemma 2.1.** *The J-distance between $SN(\xi, \tau^2, \alpha)$ and $N(\xi, \tau^2)$ is*

$$E[log\,(2\Phi(X_1))] - E[log\,(2\Phi(X_2))], \tag{5}$$

*where $X_1 \sim SN(\xi, \alpha^2\tau^2, |\alpha|\tau)$ and $X_2 \sim N(\xi, \alpha^2\tau^2)$.*

Note that, unlike $L_1$, the J-distance depends on location and scale parameters.
In this paper we obtain a similar result for the $J$-distance for the Student-$t$ case.

**Proposition 2.1.** *The J-distance between the $Y \sim ST(\xi, \tau^2, \alpha, \nu)$ and the $T(\xi, \tau^2, \nu)$ is*

$$CH(ST,T) - \frac{1}{\tau}E_S\left[log2T\left(\alpha S\sqrt{\frac{\nu+1}{\nu+S^2}}; \nu+1\right)\right] - \frac{1}{\tau}H^{ST} \tag{6}$$

*where $S \sim T(0,1,\nu)$, $H^{ST}$ is the entropy associated with a $ST(0,1,\alpha,\nu)$ and $CH(ST,T)$ is the cross-entropy between $ST$ and $T$ densities.*

## 3. Concentration function

**Definition 3.1.** *Consider two probability measures $P_f$ and $P_{f_0}$ with density functions $f(\theta)$ and $f_0(\theta)$, respectively. Let $L_y = \{\theta \in \Theta : h(\theta) \leq y\}$ , $h(\theta) = \frac{f(\theta)}{f_0(\theta)}$, $\forall\, \theta$ and $z \in (0,1)$. The concentration function of $P_f$ with respect to $P_{f_0}$ is given by $\varphi_{P_f} : [0,1] \to [0,1]$ such that $\varphi_{P_f}(0) = 0$, $\varphi_{P_f}(1) = P_f(\Theta) = 1$ and*

$$\varphi_{P_f}(z) = P_f(L_y), \quad if \quad z = P_{f_0}(L_y), \tag{7}$$

*where $P_f(L_y) \equiv \int_{L_y} f(\theta)d\theta$ and $P_{f_0}(L_y) \equiv \int_{L_y} f_0(\theta)d\theta$.*
The next lemma presents an important propriety that will be helpful to make graphical interpretation about the distance between two densities.

**Lemma 3.1.** *Let $P_f$ and $P_{f_0}$ be two probabilities measures in the same measurable space $(\mathcal{X}, \mathcal{B})$, for $z \in [0,1]$ and $P_{f_0}(A) = z$ then*

$$\varphi(z) \leq P_f(A) \leq 1 - \varphi(1-z) \quad \forall A \in \mathcal{B}. \tag{8}$$

Another interesting property about the concentration function is that the concentration function does not depend on location and scale parameters, like the $L_1$ distance. We also obtain in this paper a relation between the $L_1$ distance and the concentration function for the normal case, given by

$$L_1(f_1, f_0) = \frac{1}{2} - \varphi^{SN}(0.5),\ \forall\ \alpha. \tag{9}$$

where $\varphi^{SN}$ is the concentration function between a $SN(\xi, \tau^2, \alpha)$ and a $N(\xi, \tau^2)$. Closed expression for $\varphi^{SN}$ was obtained and presented in the next proposition.

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

$$\varphi^{SN}(z) = 2\Phi_2\left[\begin{pmatrix} \Phi^{-1}(z) \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega\right], \quad for\ \alpha > 0 \tag{10}$$

*and*

$$\varphi^{SN}(z) = 1 - 2\,\Phi_2\left[\begin{pmatrix} \Phi^{-1}(1-z) \\ 0 \end{pmatrix}; \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega\right], \quad for\ \alpha < 0, \tag{11}$$

*where $\Phi_2$ is the distribution function of the bivariate normal with mean vector zero and covariance matrix*

$$\Omega = \begin{bmatrix} 1 & -\delta \\ -\delta & 1 \end{bmatrix}, \tag{12}$$

*with* $\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}$, *and* $\Phi^{-1}$ *is the quantile function of the standard normal distribution.*

Figure 1 (left) shows the functions $\varphi_{P_f}(z)$ and $1-\varphi_{P_f}(z)$ for the distributions $SN(0,1,2)$ and $N(0,1)$. We draw a dashed line crossing the curves $\varphi_{P_f}(z)$ and $1-\varphi_{P_f}(z)$ for $z=0.5$. This segment is varying from 0.148 to 0.852 and, according to the Lemma 3.1., $0.148 \leq P_f(A) \leq 0.852$, for all $A \in \mathcal{B}$ such that $P_{f_0}(A) = 0.5$. Moreover, in the second graph on the Figure 1 (right) we show the two probabilities densities functions.
We also obtained closed expression for the concentration function in the Student-$t$ case and showed that the general shape of the concentration curve remains the same for all values of degree of freedom $\nu$. This is not present here for lack of space.

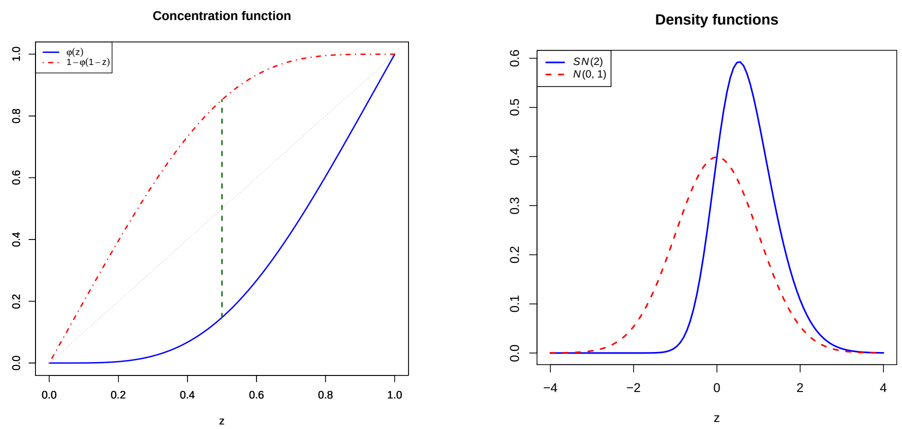


Figure 1: Concentration function of $SN(0,1,2)$ with respect to $N(0,1)$ and respective density functions.

## 4. Bayesian robustness

Bayesian robustness 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, the robustness analysis supposes the likelihood $f(x \mid \theta)$ is fixed and consider a class $\Gamma$ of prior distributions to deal with the uncertainty in specifying one prior distribution. Robustness of a given statistical procedure is measured by the size of the range of posterior measures obtained when the prior distribution varies over $\Gamma$.
Godoi and Branco (2014) studied a multiplicative class of contaminated priors for the location parameter, under a normal likelihood. The authors analyzed the behavior of the posterior mean and posterior variance under changes in the prior distribution.
Suppose that $X_1,\ X_2,\ ...,\ X_n$ is a random sample from a normal distribution with location parameter $\theta$ and scale parameter $\sigma^2$. Usually, the prior distribution specified for $\theta$ 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

$$\Gamma = \left\{ f_\alpha(\theta) = \frac{2}{\tau}\,\phi\left(\frac{\theta-\xi}{\tau}\right)\ \Phi\left(\alpha\frac{\theta-\xi}{\tau}\right) : \alpha \in \mathbb{R} \right\} \tag{13}$$

Note that, the location and scale parameters are considered fixed and the class $\Gamma$ contains an infinite number of density functions arising from the variation of the skewness parameter $\alpha$.
Considering $\sigma^2$ known, Godoi and Branco (2014) showed that under a SN prior distribution for $\theta$ the posterior distribution is in a more general class of skew distribution known as SUN (Skew Unified Normal); for details

about SUN see Arellano-Valle and Azzalini (2006). Therefore, in order to compare the posterior distributions, it is necessary to adapt the results from section 3 for this more general context. Using these results, which are not shown here for lack of space, we made a small simulation study of prior robustness analysis under the class $\Gamma$. In fact, we split $\Gamma$ in two subclass depending on the sign of $\alpha$: one for $\alpha > 0$ and other for $\alpha < 0$. We consider $\tau^2 = 1$, $\xi = 0$ and a random sample of size 20 of normal distribution with $\sigma^2 = 1$.
According to Fortini and Ruggeri (2000), we have to obtain the class of concentration functions between the class of posterior distributions and the baseline posterior and evaluate it considering the infimum of those functions. This measure will be noted by $\hat{\varphi}(z)$ for any $z \in [0, 1]$. In our context, we obtained the infimum for fixed a range of values for $\alpha$. In the subclass where $\alpha > 0$, we considered $\alpha \in (0, 25]$ such that $\alpha_i = 0.1 \times i$, with $i = 1, ..., 250$. Similar results apply to the case $\alpha < 0$.
In the Figure 2 we present the graphs of $\hat{\varphi}(z)$ and $1 - \hat{\varphi}(1 - z)$, when $\bar{x} = -1$, 0 and 1. We note that when the class of contamination prior is indexed by $\alpha$ with the same sign of the sample mean, $\hat{\varphi}(z)$ and $1 - \hat{\varphi}(1 - z)$ are very close and this proximity is more evident for large absolute values of $\bar{x}$. This is a strong evidence that the class of posterior distributions is robust with respect to the skew-normal contamination class considered for the prior. On the other hand, the robustness does not occur when $\alpha$ and $\bar{x}$ have different signs. This lack of robustness in this context can be explained because there is a conflict between the information suggested by the class of prior distribution and the sample mean.

## 5. Conclusions

In this paper we have investigated the use of different measures to compare skew-symmetric distributions with respect to a baseline symmetric distribution. We have been able to provide novel computations, mostly for the Student $t$- distribution. The concentration function has been also proposed. Once the measures used to compare distributions are available, then their different properties could provide insight on the departure from the symmetric baseline. 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 an the implications in Bayesian robustness is currently being pursued.

## 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.

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

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

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

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

Fortini, S. and Ruggeri, F. (2000). On the use of the concentration function in Bayesian robustness, in: D.Rios Insua, F.Ruggeri (Eds), Robust Bayesian Analysis. New York, Springer-Verlag, pp. 109-126.

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.

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.

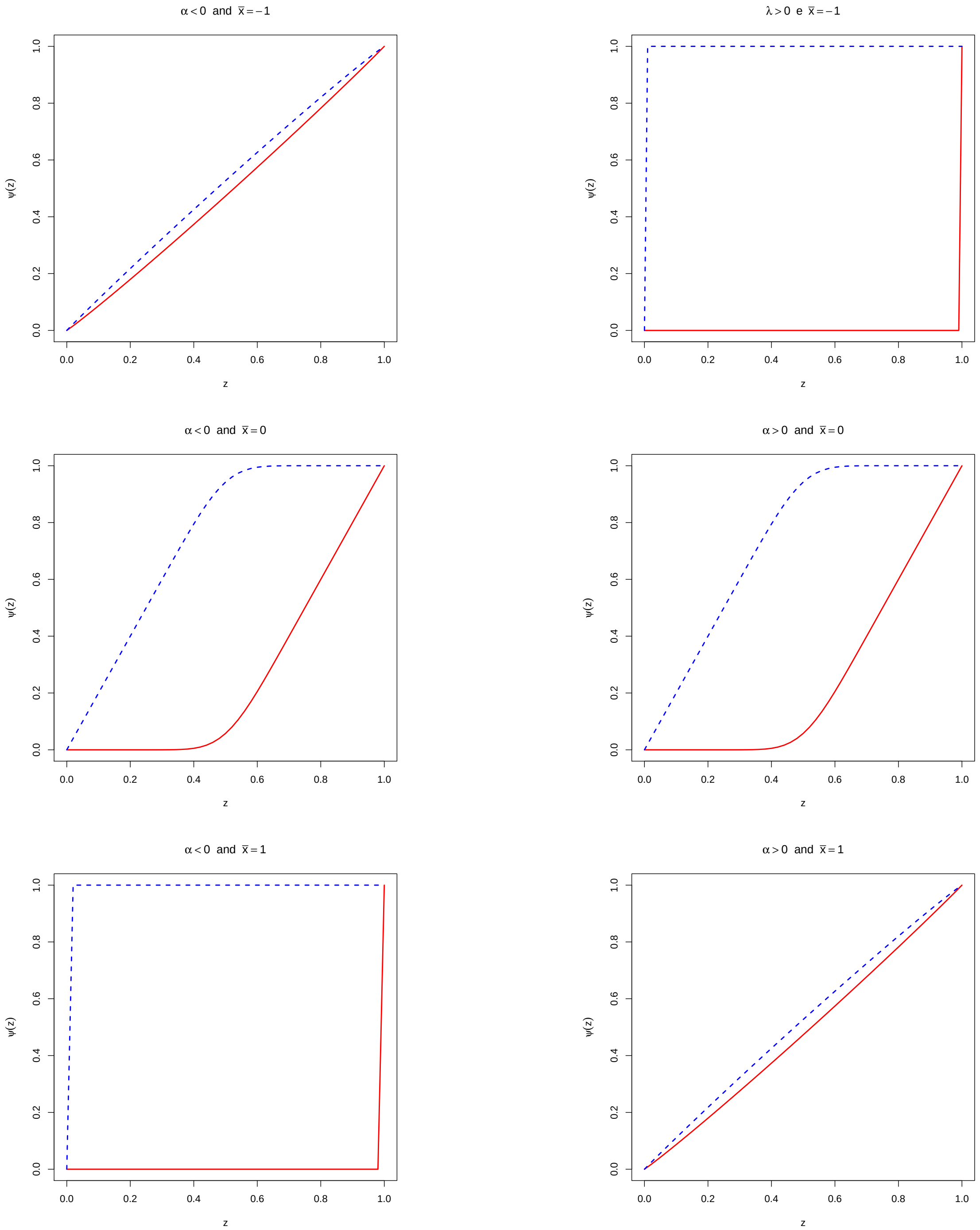


Figure 2: Concentration function for the class of posterior distributions fixed $\bar{x} = -1$, 0 and 1. The lines (–) and (- -) corresponds, respectively, to $\hat{\varphi}(z)$ and $1 - \hat{\varphi}(1 - z)$.